[Course Overview](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live)

[Course Overview](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live)

[Hi. My name is Janani Ravi, and welcome to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=0) [this course on Architecting Serverless Big](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=7.755" \t "psplayer) [Data Solutions using Google Dataflow. A](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=10.665) [little about myself, I have a masters](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=14.20075) [degree in electrical engineering from](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=16.005) [Stanford and have worked at companies such](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=17.738) [as Microsoft, Google, and FlipKart. At](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=20.54975) [Google, I was one of the first engineers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=23.204) [working on real time collaborative editing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=24.794) [in Google Docs, and I hold four patents](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=27.207600000000003) [for its underlying technologies. I](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=30.501333333333342)[currently work on my own startup,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=32.38242857142857) [LoonyCorn, a studio for high-quality video](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=34.149) [content. In this course, you will be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=37.5887142857143) [exposed to full potential of Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=39.88066666666665)[Dataflow and its radically innovative](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=41.80550000000001) [programming model. We start this course](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=44.336999999999996) [off with a basic understanding of how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=46.80250000000001) [Dataflow works for serverless compute.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=49.180800000000005) [We'll study the Apache Beam APIs used to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=52.094) [build pipelines and understand what data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=54.508) [sources, syncs, and transformations are.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=57.074)[We'll study the stages in a Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=60.477) [pipeline and visualize it as a directed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=62.56) [async link graph. We'll then use Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=65.476) [Beam APIs to build pipelines for data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=68.219) [transformations in both Java, as well as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=70.97333333333333) [Python, and execute these pipelines](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=73.74999999999997) [locally and on the cloud. We'll integrate](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=75.88699999999999) [our pipelines with other GCP services such](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=78.4375) [as BigQuery and see how we can monitor and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=80.90828571428571) [debug slow pipeline stages. We'll then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=83.957) [study different pipeline architecture such](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=87.2698)[a branching and pipelines using side](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=89.05433333333335) [inputs. We'll also see how we can apply](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=91.62) [windowing operations to perform](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=94.44166666666668) [aggregations on our data. Finally, we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=95.903) [work with Dataflow without writing any](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=98.741) [code using prebuilt Dataflow templates](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=101.39333333333332) [that Google offers for common operations.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=104.17442857142858) [At the end of this course, you should be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=107.201) [comfortable using Dataflow pipelines to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=108.811) [transform and process your data and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=111.398) [integrate your pipelines with other Google services.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=c3e69c87-1c4b-486a-9052-0b71772a45e5&clip=0&mode=live&start=113.866)

[Introducing Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live)

[Module Overview](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live)

[Hi, and welcome to this course on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=0) [Architecting Serverless Big Data Solutions](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=3.737) [Using Google's Dataflow. It's pretty](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=6.571) [common nowadays for all organizations to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=9.523333333333333) [collect huge amounts of data. Now this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=11.446571428571428) [data is typically in a raw format or](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=14.126000000000001) [unstructured format and might have to be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=17.126333333333335) [cleaned up and transformed before it's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=19.639285714285716) [really useful and that's where Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=21.226999999999997) [comes in. Dataflow is the GCP's serverless](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=23.360999999999997) [and fully managed data transformation](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=26.87375) [service. Serverless because you provision](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=29.88066666666667) [no machines to get up and running with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=31.853199999999994) [Dataflow. Data transformation actions on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=34.927374999999984) [your raw data are specified in the form of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=36.958) [pipelines and you can set up pipelines for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=39.838374999999985) [batch processing, as well as for streaming](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=42.20237500000001) [data. Dataflow uses the open source Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=45.04885714285715) [Beam API to specify transformations and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=48.03085714285714) [other operations. Apache Beam is available](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=51.00985714285716) [in both Python, as well as Java. If you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=53.4852857142857)[work with the very popular Spark data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=56.40866666666668) [analytics engine, you'll find that the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=59.23833333333333) [constructs and operations in Dataflow seem](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=61.40033333333333) [very, very familiar. However, if you want](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=64.027) [to use Spark with distributed computing,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=66.312) [you need to set up a distributed Hadoop](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=67.71842857142856) [cluster. Not so with Dataflow. There are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=71.13874999999999)[important architectural differences between Spark and Dataflow.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=0&mode=live&start=74.06549999999999)

[Prerequisites and Course Outline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live)

[Before we get started with actual course](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=1.944) [content, here are some pre-reqs that you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=3.88) [need to have in order to make the most of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=5.659) [your learning. This course assumes that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=8.031666666666666) [you have a basic understanding of cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=10.041999999999998) [computing and that you worked with some](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=12.1715) [cloud computing platform. It may not be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=14.66114285714286)[the GCP. This course also assumes that you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=16.983999999999995) [have a basic understanding of how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=18.90475) [distributed computing and processing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=21.202999999999996) [works, that you have a basic understanding](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=23.323749999999997) [of how compute and storage work on a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=25.543499999999998) [distributed platform such as Hadoop. All](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=28.29685714285715) [of the code in this course will be written](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=30.907444444444444) [using the Python programming language so](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=32.823) [you should have a good understanding of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=34.85557142857142) [basic Python constructs. There's exactly](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=37.2702) [one coding demo that we'll do in the Java](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=39.803428571428576) [programming language so you get a feel for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=42.50450000000001) [how pipelines are written and executed in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=45.138999999999996) [Java. All of the Python code in this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=48.08975) [course will be written using Python 2. 7](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=50.033888888888875) [because that's the version of Python](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=52.695499999999996) [currently supported by Dataflow. If you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=55.01579999999999) [have prior experience working with the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=57.60685714285714) [Spark data analytics engine, that would be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=59.550000000000004) [very useful. This is not really required](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=62.3028) [though. If you feel that you lack some](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=64.94283333333334) [cloud computing pre-reqs, especially on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=66.92862500000003) [the GCP, here are some courses that you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=68.85266666666665) [can watch on Pluralsight. Architecting big](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=71.55220000000001) [data Solutions Using Google Dataproc will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=73.59899999999999)[show you how Hadoop and Spark work on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=75.84214285714286) [Google Cloud. If you're working with big](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=79.15257142857143) [data on an on-premise Hadoop cluster and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=80.93124999999999) [you want to move to the GCP, this is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=82.63771428571427) [probably where you'll start. If you want](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=85.49414285714288) [an introductory course to the Google Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=87.90371428571424) [Platform and understand how storage works,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=90.205) [Architecting Google Cloud Storage](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=92.521) [Configurations is a good course for you.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=94.395) [Here is a quick overview of what we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=97.344) [cover in this course. We start off by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=99.191) [introducing Dataflow, we'll understand the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=102.19633333333334) [basic components of a Dataflow pipeline,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=103.86) [and execute a pipeline locally and on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=105.9627142857143) [GCP. We'll then move onto building more](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=109.69140000000002) [complex pipelines using Apache Beam APIs.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=112.16200000000002) [We'll write Dataflow jobs in both Java and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=115.034) [Python. Exactly one example is in Java,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=117.81444444444448) [all other examples are in Python. We'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=120.22) [also see how we can write code in Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=122.73655555555555) [to integrate with other GCP services such](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=124.92099999999999) [as cloud storage buckets and BigQuery.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=127.70075) [We'll then move onto some advanced](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=130.7) [pipeline constructs such as windowing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=132.53942857142857) [multiple outputs from a pipeline,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=134.82360000000006) [branching, side inputs, and so on. And](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=137.3004) [finally, we'll see how you don't always](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=139.745125) [need to write code on the GPC to use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=142.161) [Dataflow. We'll use Dataflow templates to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=144.59190000000012) [create pipelines. We'll try out some](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=147.34771428571423) [predefined templates that Google makes](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=149.59466666666665) [available to us. All of the demos in this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=151.42100000000002) [course will assume that we are engineers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=153.85033333333334) [working at this hypothetical online](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=155.73071428571424) [retailer, SpikeySales. com. They](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=158.16160000000002) [specialize in flash sales of trending](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=160.6872857142857) [products, which means they have very large](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=162.507) [spikes in user traffic. Engineers at](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=165.348) [SpikeySales want to move away from their](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=167.99599999999998) [on-premise computing options. Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=170.8084) [computing fits their bill perfectly. They](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=172.87616666666668) [want the pay-as-you-go model and don't](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=175.0812857142857) [want to hold idle capacity during off-sale](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=177.237) [periods. They're in the process of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=180.84899999999993) [evaluating the GCP. They're moving data to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=183.30749999999998) [the cloud and compute as well. They want](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=185.6695714285715) [elastic, pay-as-you-go global access.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=188.19783333333334)[SpikeySales collects a lot of data and all](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=190.894) [of this data is currently processed on a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=193.61742857142858) [large on-premise Hadoop cluster, which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=196.4646666666667) [means there our utilization concerns and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=198.78416666666666) [administration of this cluster is a huge](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=201.2895) [overhead. The on-premise Hadoop cluster](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=203.8793333333333) [has also proved expensive to scale. There](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=206.1342857142857)[is a huge overhead in ordering and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=208.91171428571428) [provisioning machines. This also results](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=211.12733333333333) [in an overprovisioning problem. On sale](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=213.65914285714288) [days, the cluster is fully loaded. On](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=216.046) [off-sale days, the cluster is mostly idle.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=219.006) [This is why they're excited about Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=222.338) [on the GCP. They're interested in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=224.295) [serverless compute, which means they don't](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=227.413) [have to provision machines to handle huge](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=228.653) [loads. Developers are also excited about](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=231.99549999999994) [focusing on actual development rather than cluster administration.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=1&mode=live&start=234.26049999999995)

[Introducing Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live)

[In this clip, we'll see how Google Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=2.002) [Dataflow can help the engineers at](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=4.1) [SpikeySales. com. Because they specialize](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=6.783500000000002) [in flash sales and have huge sale days,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=8.62042857142857) [the user traffic to SpikeySales. com tends](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=11.386999999999999) [to have a huge variation, which means](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=14.167285714285715) [they're big data processing needs also](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=16.912999999999997) [vary. They're in the process of moving](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=19.82425) [away from their large on-premise Hadoop](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=21.788750000000004) [cluster, but they don't want to rewrite](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=23.6754) [all of their jobs. The Day-1 solution on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=25.8554) [the GCP is to use Dataproc. Cloud Dataproc](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=28.671666666666663) [on the GCP offers you Hadoop on the cloud.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=31.673499999999997) [Hadoop, Spark, Pig code can simply be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=35.407) [migrated as-is. Dataproc, however, still](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=38.755) [requires you to spin up virtual machines](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=41.69157142857143) [that make up your cluster. As SpikeySales](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=43.8145) [engineers become more familiar with GCP's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=46.5036) [offering, they're moving towards the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=49.06800000000001) [Day-365 solution where they use Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=51.369333333333344) [where there is no reason to provision](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=53.480500000000006) [resources at all. Dataflow is entirely](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=55.97250000000001) [serverless and will scale workers to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=58.465999999999994) [handle additional workloads automatically.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=60.8815) [However, there is a tradeoff here. All of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=63.898) [their data processing code needs to be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=65.72475) [recoded to use Apache Beam. Let's compare](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=67.81066666666666) [a traditional on-premise Hadoop cluster](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=71.48679999999999) [and Cloud Dataproc in a little more detail](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=73.21514285714285) [before we move onto Dataflow. Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=75.78812500000001) [Dataproc is a fully-managed service for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=78.21166666666667) [running MapReduce, Spark, Pig, and Hive on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=81.1302) [the GCP. Even though it's Hadoop on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=83.62319999999998) [cloud, the way you work with traditional](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=85.67571428571428) [on-premise Hadoop and Cloud Dataproc is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=87.302) [completely different. Traditional Hadoop](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=90.49333333333333) [clusters are always provisioned, they stay](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=92.4196) [in existence forever. When you use managed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=95.12249999999999) [Hadoop with Dataproc though, you can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=97.66171428571431)[create clusters dynamically as in when you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=99.82449999999999) [need them to run your compute jobs.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=101.94849999999998) [Clusters can be spun up very, very easily](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=103.89437500000003) [and can be thought of as temporary, use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=106.59366666666668) [only when needed. The distributed file](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=109.2926) [system that is HDFS in traditional](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=112.02457142857142) [on-premise Hadoop runs on the actual](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=114.566) [cluster nodes. Since managed Hadoop with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=117.44800000000001) [Dataproc runs on virtual machines on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=119.89057142857145) [Google Cloud, HDFS is configured and runs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=122.36600000000001) [on the persistent disks of these VMs. In](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=124.91628571428573) [traditional on-premise Hadoop, storage is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=128.59675) [tightly coupled with compute. The data for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=131.061) [the jobs are stored in HDFS on the actual](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=133.74130000000002) [cluster nodes where the processes are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=136.404) [executed. Cloud Dataproc on the other hand](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=138.826) [allows you to separate storage from](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=141.31125000000003) [compute. The data will ideally be stored](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=143.79959999999994) [in Google Cloud storage buckets rather](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=145.88200000000006) [than in the cluster itself. Tight coupling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=147.74699999999999) [of storage and compute means that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=151.3968333333334) [traditional on-premise Hadoop clusters](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=153.4306666666666) [encapsulates state, state as represented](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=155.98825000000002) [by the data stored on the cluster nodes.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=158.05014285714287) [On Dataproc, you can decouple storage and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=160.714)[compute and store data in cloud storage](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=163.171) [buckets, which means the cluster will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=165.594) [ideally be stateless. Dataflow, however,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=167.432) [takes the advantages of distributed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=170.47433333333333) [processing on the cloud one step further.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=172.742) [It's a completely serverless approach for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=175.42199999999997) [data transformation, which works for batch](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=178.05474999999998)[processing, as well as with streaming](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=181.005) [data. We'll now compare and contrast](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=184.18328571428566) [Dataflow with Cloud Dataproc and see what](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=186.49716666666663) [the advantages are. Cloud Dataproc is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=188.525) [essentially managed Hadoop on the cloud,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=191.011) [which means you're still required to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=193.214) [provision clusters in order to run data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=194.881)[processing. Dataflow is entirely](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=198.02224999999993) [serverless. You don't have to know about](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=200.25225) [the clusters running your jobs at all. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=202.42099999999994) [GCP completely abstracts you away from any](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=204.86285714285714) [machines that are provisioned for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=207.43914285714285) [processing. Scaling a Cloud Dataproc](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=209.9748571428572) [cluster is easy. You simply add VMs to it,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=211.9757142857143) [but it's still a manual process. Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=214.53600000000006) [is completely autoscale. If the data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=217.89428571428576) [processed by our job grows, Dataflow will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=220.68224999999995) [simply autoscale the number of workers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=222.82933333333335)[assigned to perform processing. Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=225.42933333333332) [Dataproc is still part of the Hadoop](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=227.781) [ecosystem. It's managed Hadoop on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=230.196) [Cloud. Dataflow on the other hand uses the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=231.97516666666667) [open source Apache Beam API. The jobs that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=234.40137499999994) [you run on Dataflow are Dataflow jobs, not](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=238.18925000000002) [MapReduce. The Apache Beam API is simply](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=241.47724999999997) [the programming model that you'll use to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=244.09442857142855) [define your pipelines. The actual](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=246.376) [execution is on Dataflow. Cloud Dataproc](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=249.21766666666667) [is inherently batch oriented and it's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=252.234) [typically used for batch processing, even](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=254.393) [though you can work with streaming data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=256.2247142857143) [when you use Spark on Cloud Dataproc. With](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=259.0632222222222)[Dataflow, however, batch and streaming is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=262.26683333333335) [completely integrated. You can use pub/sub](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=265.2) [to feed in streaming data and perform](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=267.5230000000001) [windowing operations within your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=269.9275999999999) [pipelines. If you want to build and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=271.9213333333333) [execute machine learning models on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=274.0032499999999) [Dataproc, you'll use the Spark machine](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=275.4545714285714) [learning libraries. With Dataflow, you'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=277.66150000000005) [use TensorFlow with Cloud ML engine. There](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=279.65299999999996) [is one abstraction that is common between](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=282.53014285714283) [Cloud Dataproc and Dataflow and that is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=285.05785714285724) [the complete separation of compute and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=287.29533333333336) [storage. Just like you can work on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=290.0439999999999) [datastore in cloud storage buckets when](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=292.2757499999999) [you use Dataproc, Dataflow inherently](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=293.91799999999995) [works with data that is stored separately from where processing occurs.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=2&mode=live&start=296.23439999999994)

[Enabling APIs and Installing Client Libraries](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live)

[Let's get a little hands-on before we dive](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=2.183) [deeper into Dataflow. We'll enable APIs on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=3.88) [the GCP to work with Dataflow, and we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=6.468777777777777) [install client libraries to write our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=9.145) [Dataflow code in Python. The URL console.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=11.806500000000003) [cloud. google. com will allow us to log](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=14.941500000000001) [into our GCP account. Now GCP can be used](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=18.159285714285712) [with any Gmail account or with the G-suite](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=21.317999999999998) [account associated with your organization.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=24.305333333333337) [The account I'm using here is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=26.344) [spikeyengineer@loonycorn. com, specify the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=27.661) [password and log into the GCP web console.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=30.69142857142857) [You need to have billing enabled on your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=33.596625) [current GCP project in order to execute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=36.79685714285714) [the code that you'll write in this course.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=39.132625000000004) [However, the total cost of running all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=41.345375000000004) [the code in this course should be less](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=43.549375) [than $10. All of GCP's resources are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=46.156) [associated with the project, which is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=48.95214285714286) [billing unit on the GCP. Spikey-dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=51.31271428571428) [is my current project. You can click on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=53.881) [this drop-down and change your project or](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=56.888000000000005) [create new projects if you want to. If you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=58.79757142857143) [click on the hamburger icon on the top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=61.60811111111109) [left, that will open up the Navigation](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=63.839400000000005)[menu that will allow you to access all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=65.43171428571428) [GCP's services. We'll enable APIs to use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=68.991) [Dataflow by going to APIs and Services and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=72.04900000000002) [clicking on the Dashboard option. Certain](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=74.68485714285711) [APIs are enabled by default for a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=77.869) [particular project. You can see those](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=79.904) [here. If you want to enable APIs, simply](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=81.7252) [click on the link at the top. You can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=83.58144444444449) [simply type in the name of the API that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=86.27480000000001) [you're interested in the search box here.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=88.092375) [I'll say Dataflow, that'll give me the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=89.901)[Dataflow API. Go ahead and click on this,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=92.1565) [that'll take you to a page where you'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=94.42462499999999) [have a blue button to enable this API.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=97.02955555555555) [Once we have this API enabled, the next](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=100.418) [step is to set up the credentials for a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=102.72477777777777) [service account that our program will use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=105.36475) [to run. You can create a service account](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=107.573875) [by clicking on the Create Credentials](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=110.1654) [button here or you can use the Navigation](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=111.82414285714285) [menu to go to IAM and administration and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=114.60866666666666) [choose the Service accounts option. A](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=117.834) [service account on the GCP can be thought](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=121.65566666666666) [of as a robot account which allows program](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=123.341) [to program authentication. Click on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=125.995) [Create Service Account link on top and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=129.43299999999996) [we'll call this the spikey-dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=131.33900000000003) [service account. This is the service](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=133.29366666666667) [account that they'll use for Dataflow to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=135.65842857142854) [run jobs. This account should allow access](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=137.69733333333335) [to other GCP services. Click on the Create](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=139.8211666666666) [button to create this account and let's go](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=143.462) [ahead and select a role associated with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=145.756) [this account. In order to keep things](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=148.4395) [simple, I'm going to associate the project](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=150.606) [owner role with the service account in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=153.297) [order to have our jobs have access to all](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=155.681) [of the client libraries and other GCP](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=158.725) [services. In the real world though, you'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=161.78400000000008) [explicitly add only those roles for the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=163.829) [services that your Dataflow job is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=166.21424999999994) [planning to access. Click Continue to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=168.7194285714286) [associate this role with our service](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=170.82457142857143) [account. We're now ready to create a key.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=172.83999999999997) [Our Dataflow jobs will use the service](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=176.425) [account JSON key in order to authenticate](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=179.128) [themselves to other GCP services. Download](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=181.754) [this private key JSON file onto your local](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=184.71171428571427) [machine and store it safely somewhere.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=187.451) [This is the spikey-dataflow. json file.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=189.464) [Let's go to the Navigation menu and take a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=193.168) [look at the Cloud Dataflow page. Notice](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=195.81442857142858) [the pin icon right next to Dataflow. You](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=198.397) [can click on this and pin Dataflow to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=201.07955555555554) [very top right below your Home button.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=204.09199999999998) [Click on Dataflow and that will take you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=206.668) [to the Dataflow page. If you've never used](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=207.9897777777777) [Dataflow before, this is what the page](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=210.233) [looks like. You can click on the Try](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=212.37611111111113) [Dataflow button and walk through a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=213.937) [tutorial that the GCP offers. I'm not](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=216.58642857142854) [really interested in the tutorial at this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=219.45571428571427) [point in time. I'm going to cancel this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=221.42180000000002) [tutorial and explore it further. Let's now](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=223.64116666666666) [take a look at CloudShell. The CloudShell](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=226.52642857142854) [on the GCP is an extremely useful service.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=229.0117142857143) [It's an ephemeral VM on the Google Cloud,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=231.712) [which comes preinstalled with all of GCP's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=234.312) [command-line tools and utilities. If](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=237.286) [you're working from an insecure local](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=240.45742857142858) [machine where you don't want to install](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=242.128) [Cloud SDKs, you can simply use this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=244.207) [CloudShell VM. If you click on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=247.0205) [three-dot menu at the top right of your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=249.957) [CloudShell terminal, this will bring up](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=251.976) [menu options that allow you to upload a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=254.386) [file to your CloudShell. Let's upload the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=256.754) [file with our service account credential](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=259.8747142857142) [that is present on our local machine. If](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=261.69499999999994) [you now run an LS command on your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=264.5132222222223)[CloudShell terminal VM, you'll see that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=267.69900000000007) [this file is now present in your home](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=269.64766666666674) [directory. In order to have the Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=272.57166666666666) [jobs that we'll launch from our CloudShell](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=274.68975000000006) [terminal window, use this service account](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=276.96783333333343) [to connect to other GCP services. We need](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=279.30628571428576) [to set up the Google application](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=282.65475000000004) [credentials environment variable. Once you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=284.478) [set this environment variable, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=287.90600000000006) [Dataflow jobs will automatically pick up](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=289.82616666666667) [these service account credentials. You can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=292.3692) [click on the full screen icon on the top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=294.93288888888895) [right of your CloudShell terminal window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=297.20385714285715) [in order to utilize the entire screen.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=299.170625) [We'll now install the Python client](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=302.283) [libraries that we need. First run sudo](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=303.958) [apt-get update to update all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=306.506625) [packages in apt-get. The Apache Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=309.38800000000003) [Python library for the GCP can be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=311.8656249999999) [installed using a simple pip install. When](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=314.86175000000003) [you perform this install, you might see a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=318.3884) [few errors. This is because certain Google](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=320.31866666666673)[technologies are only compatible with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=322.97599999999994) [specific OAuth client versions. Even](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=325.6514000000001) [though we don't use OAuth in this course,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=328.351875) [I'm going to update the version of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=330.25712500000003) [oauth2client to 1 that is compatible with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=332.232) [these GCP services. The OAuth client was](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=335.59) [installed successfully. There is still](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=338.813) [another error here for a compatible HTTP](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=340.57899999999995) [lib2client, so I'm going to change that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=343.4903333333333) [version as well. All of the Python](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=346.41733333333326) [libraries have been installed, this time](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=348.32599999999996) [with no errors. We can now move onto](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=349.9915) [initializing the gcloud SDK. This is not](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=353.2085) [really needed when you're working on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=356.01449999999994) [CloudShell, but if you're working with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=358.1175999999999) [gcloud on your local machine, make sure](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=359.50912500000004) [you run gcloud in it and follow through](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=361.7932857142858) [with the steps that it'll lead you to. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=364.3898750000001) [steps are self-explanatory and pretty](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=367.4675) [straightforward, so I'm going to move on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=369.085) [to gcloud auth login, which we'll use to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=370.65474999999986) [authenticate your terminal window with the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=373.533) [Google Cloud. Once again, this is needed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=376.5724) [only if you're running the gcloud SDK on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=378.861625) [your local machine. It's not needed if](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=381.25300000000004) [you're working on CloudShell. Once you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=383.8990000000001)[finished initializing and authenticating](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=386.2414) [the gcloud SDK, you're ready to get](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=387.95725) [moving. Run gcloud components update in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=390.6998) [order to get the latest components and you're now good to go.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=3&mode=live&start=392.8901428571428)

[Using Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live)

[Let's now get a big picture understanding](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=2.118) [of how you would use Dataflow on the GCP.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=3.858) [The first thing that you need to do is to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=7.46) [write code either in Python or in Java to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=9.072) [build up your pipeline. A pipeline can be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=12.066) [thought of as a series of transformations](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=14.4478) [that you perform on your data. Once your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=16.141285714285715)[pipeline is ready, you're now ready to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=19.0462) [submit the Dataflow job for execution.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=21.201999999999995) [Dataflow on the GCP will pick up your job,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=23.714285714285715) [which has been defined using Apache Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=27.190750000000005) [APIs and then assign workers to execute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=29.886000000000006) [the processing tasks. You can think of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=33.57714285714286) [every processing pipeline as a directed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=37.04828571428571) [acyclic graph. Dataflow will take care of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=39.4986) [paralyzing this pipeline and executing it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=41.97542857142858) [in the most performant and efficient](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=44.20485714285716) [manner. You define the exact processing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=47.10274999999999) [steps in your pipeline using Apache Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=49.256714285714295) [APIs. Apache Beam is a unified programming](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=52.22866666666667) [model which you'll use to define the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=55.57357142857143)[transformations and operations on your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=57.684) [data. Dataflow on the GCP supports Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=60.54733333333331) [APIs in both Python, as well as Java so](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=63.33850000000002) [you can choose the programming language](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=65.83111111111108) [that is the most familiar to you. If you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=67.46628571428573) [worked with Spark before, you'll find that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=70.49033333333334) [the Beam programming APIs will be very](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=72.49457142857145)[familiar to you. The code looks very](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=74.50599999999999) [similar to Spark. We'll, of course, see a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=76.93742857142855) [lot of Apache Beam in the demos as we move](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=79.2617777777778) [forward, but here is a link to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=81.78485714285715) [complete programming guide for Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=83.85549999999998) [Beam. This is a handy reference link for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=86.06059999999998) [you to have around when you write your own](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=88.01488888888888) [code. What you should know about Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=90.57075) [Beam is that it's a programming model that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=92.39849999999998) [you use to define how you want to process](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=94.745) [your data. The actual execution of every](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=98.00322222222223) [processing step is performed by something](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=101.01399999999998) [called the Dataflow Runner. And Google's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=103.05499999999999) [Dataflow on the GCP is only a backend for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=106.453) [the Apache Beam APIs and you can specify](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=110.195) [other runners as well, which run on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=112.86166666666666) [different distributed computing platforms.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=115.41300000000001) [Let's say you have a program where Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=117.867) [Beam APIs are used to specify your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=120.049) [processing steps and you use the direct](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=122.54542857142857) [runner with this program. This will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=125.2294) [execute your pipeline on your local](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=127.65233333333333) [machine, it'll not use the cloud at all.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=129.4012857142857) [Interestingly enough, Spark clusters also](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=131.9532) [support Apache Beam. If you specify a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=134.4295)[Spark Runner, this will execute your job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=137.2003333333333) [on a Spark cluster. And if you want to run](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=139.88022222222222) [Apache Beam programs on the Google Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=142.82733333333334) [using Dataflow, you'll specify the Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=144.94800000000004) [Dataflow runner. Dataflow on the GCP can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=147.8983333333333) [also work with streaming data. Streaming](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=150.39671428571427) [jobs run indefinitely forever until you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=152.93959999999998)[explicitly stop them. All jobs on Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=156.3652) [whether they're streaming or batch jobs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=159.35466666666662) [have monitoring and visualization tools](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=161.45057142857144) [which you can use to monitor the current](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=163.50799999999998) [state of the pipeline. Let's visualize how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=167.0907142857143) [Dataflow executes the jobs that you submit](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=169.85700000000003) [to the Google Cloud. Here is your code](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=172.3935)[that performs a bunch of transformations](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=175.01577777777777) [on your data. You then submit this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=176.8675) [processing pipeline to Dataflow, which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=179.4708571428572) [then behind the scenes provisions machines](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=181.35466666666667) [and spins off workers to perform all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=184.41500000000002) [the data processing that you've specified.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=187.92385714285717) [The provisioning of machines, the spinning](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=190.411)[up of workers, all of this is completely](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=192.35966666666667) [abstracted away from you, which is why](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=194.5172) [Dataflow is set to be serverless. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=197.35600000000002) [workers used by Dataflow to execute your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=200.05871428571427) [pipelines are said to be serverless and no](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=202.36785714285713) [ops. GCP automatically assigns resources](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=205.69699999999997) [needed to complete the job. Let's say the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=208.34516666666667) [amount of data that your job has to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=211.28266666666664) [process is very large, it has increased](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=213.32266666666666) [when GCP takes care of autoscaling as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=215.822875) [well. GCP automatically adds or removes](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=217.96566666666666)[capacity based on your current workload.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=221.37300000000005) [Before executing code, Dataflow will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=224.177) [analyze your pipeline and see which of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=226.45214285714286) [these operations can be performed in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=229.2646666666667) [parallel. It'll then assign parallel](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=231.58533333333332) [processing tasks to different workers so](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=233.56133333333332) [that they're completed in the most](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=235.3113333333333) [efficient manner. All of the data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=237.83933333333331) [processing operations that you perform on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=239.91742857142862) [input data can be set to make up a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=241.75766666666664) [pipeline. A pipeline is essentially a DAG,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=244.75542857142855) [or a directed acyclic graph. All of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=247.73624999999998) [edges specified in this directed acyclic](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=251.44849999999997) [graph are essentially unbounded](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=253.41483333333335) [collections. Unbounded because they can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=255.69016666666673) [also represent streaming data which run to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=257.74766666666665) [infinity. These are called PCollections in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=261.17283333333324) [Apache Beam. And all of the operations](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=264.24614285714296) [that you perform on the input data so that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=266.53849999999994) [you get it in the final form in which you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=268.6342727272726) [needed are called transforms. These are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=271.7936666666666) [Apache Beam operations to modify the input data.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=4&mode=live&start=274.1321428571428)

[Introducing Apache Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live)

[So what exactly is Apache Beam and how is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=1.768) [it used with Dataflow. Apache Beam is an](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=4.259) [open source, unified programming model](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=7.7625) [that can work for both batch, as well as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=9.783428571428571) [streaming data. The primary purpose of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=13.545285714285713) [Apache Beam is to build pipelines which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=15.687999999999999) [define a series of transformations on data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=18.182000000000006)[so that raw data is converted to its final](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=20.923874999999995) [form. It's important to realize that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=24.333400000000005) [Apache Beam is not a runtime or an](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=26.24642857142857) [execution environment. Pipelines defined](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=29.741249999999997) [using Apache Beam has to be executed on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=31.389142857142858) [different backends. This backend can be a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=34.50119999999999) [Dataflow backend, a Spark backend, or any](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=36.94185714285715) [platform that supports Apache beam. These](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=39.4482857142857) [are the basic components of any Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=42.62871428571428) [Beam pipeline. The first is the data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=44.690999999999995) [source, this is the source of batch or](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=46.96544444444444) [streaming data that needs to be processed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=49.40442857142856) [and transformed. A pipeline then comprises](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=51.97166666666667) [of a series of transformations. There can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=54.181) [be one or more transforms that modify the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=57.100249999999996) [data to get it into the right final form.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=59.51325) [And finally, a beam pipeline will write](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=62.6466) [the data out to a data sink. This is where](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=65.19511111111109) [the data will be stored in some kind of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=67.59133333333334) [persistent storage. In order to get data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=70.88600000000002) [from its raw to its final form, a number](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=72.91099999999999) [of different kinds of transformations may](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=75.29749999999999) [be applied on this data. All of these](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=77.46885714285713) [transformations and the entire pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=80.97357142857145) [can be represented as a directed acyclic](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=83.029) [graph, a graph with no cycles. This entire](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=86.26485714285714) [set of complications make up our Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=90.04828571428571) [Beam pipeline. A pipeline in Apache Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=93.03450000000001) [is a single potentially repeatable job. A](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=95.89874999999999) [pipeline is what you execute over and over](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=98.873125) [again as new data flows in. A pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=101.30757142857144) [comprises of a number of transformations](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=105.159) [that you apply to the input data to get it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=106.74228571428571) [in its final form and a pipeline may](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=109.74919999999999) [perform operations in parallel. Certain](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=112.74966666666667) [transformations may have dependencies on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=115.86925) [transformations which come before it.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=118.1225) [Others may not. All of the edges that make](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=120.42599999999999) [up this directed acyclic graph form data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=123.53433333333331) [which are stored in PCollections. A](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=126.67399999999999) [PCollection in Apache Beam is a special](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=129.9185) [container class that represents an](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=132.134) [unbounded dataset, datasets of unlimited](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=134.005)[size. If you're performing transformations](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=137.43) [on batch data, the PCollection will be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=139.563) [bounded on streaming data such as from a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=142.3772857142858) [\_\_\_\_\_ subscription, the dataset will be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=145.032125) [unbounded. There are two terms that you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=148.3846666666667) [should be clear about when you're using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=150.455) [Apache Beam with Dataflow, and those are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=152.48014285714288)[driver and runner. The driver program is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=155.30562500000002) [what you define using Apache Beam APIs.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=157.79749999999996) [This defines the computation directed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=160.32000000000002) [acyclic graph or the pipeline. The runner](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=162.91271428571426) [is what executes this directed acyclic](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=166.64342857142856) [graph or the pipeline on a backend. Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=168.69487500000002) [Beam works with a number of different](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=172.05257142857144)[runners such as Apache Spark, Flink, Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=173.81366666666668) [Dataflow. Apache Beam programs can also be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=177.23066666666665) [run locally on your machine using the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=179.68785714285713) [direct runner. Here are the basic steps](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=182.66074999999995) [that you'll follow when you write code for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=184.71724999999998) [any Beam pipeline. You'll first create a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=186.99887499999997) [pipeline object. You can use Python or](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=189.2513333333333)[Java. This is done using Apache Beam APIs.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=191.6846666666667) [You'll then specify a data source for your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=194.695) [input data. Remember, all data is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=196.84628571428573) [represented in the form of a PCollection.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=199.086) [Your data source on the Google Cloud can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=201.786) [be Big Query, Cloud Storage, or even Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=203.5557777777778) [pub/sub. Cloud pub/sub is Google's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=206.69074999999992)[streaming messaging service. You'll then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=208.362) [transform your data to get it in its final](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=211.38375000000002) [form. You'll use transform such as ParDo,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=214.1447142857143) [or ParallelDo, GroupByKey, Join, any other](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=217.58450000000002) [transform. This is analogous to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=220.89100000000002) [transformations in Apache Spark. When you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=222.96440000000004) [get the final form of your data, that'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=225.03737500000003) [be a PCollection as well, you'll write it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=227.26900000000003) [out to a data sink and this can be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=229.32749999999993) [BigQuery, Cloud Storage, or even Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=231.8577142857143) [Pub/Sub. Once you've defined your entire](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=234.06474999999995) [pipeline, you'll then submit it to a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=236.82549999999995) [pipeline runner which can be Dataflow on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=239.13149999999996) [the GCP, Spark, or your local machine.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=241.54500000000007) [This is responsible for executing your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=243.754) [pipeline. The first four steps make up the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=246.48571428571424) [driver that you'll define using Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=249.12500000000009) [Beam APIs. And the final step here is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=251.50057142857145) [runner. The runner for this course, of course, is Cloud Dataflow.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=5&mode=live&start=254.85912499999998)

[Exploring the Wineries Dataset](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live)

[Let's get started with Dataflow. We'll see](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=1.813) [how we can run a simple Dataflow job on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=3.9572222222222226) [our local machine using the direct runner](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=6.3717500000000005) [for Apache Beam. In order to not have to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=9.2545) [worry about Platform differences, we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=12.148749999999998) [activate CloudShell and write all of our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=14.246142857142857) [code here. Create a directory called](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=16.73314285714286)[spikey\_dataflow and cd into this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=19.158) [directory. This directory is currently](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=22.537999999999993) [empty as you can see by running the ls-l](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=24.43171428571428) [command. I'm going to use the three-dot](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=28.332142857142856) [menu here on the top right in order to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=30.311999999999998) [upload the dataset that I'm going to use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=32.582249999999995) [for this particular job. Click on upload](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=34.444571428571436) [file, and in the Datasets folder that is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=36.831374999999994) [available within your downloadable](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=39.3025) [resources, you'll find a number of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=40.5808) [different datasets. These are the sales](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=43.626) [and offers related datasets used for all](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=46.26457142857144) [for the Dataflow jobs that we'll write in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=48.68250000000001) [this course. We'll pick a simple dataset](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=51.66266666666668) [for this very first program that is a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=53.667375000000014)[spikey\_whinery\_list. csv. This dataset has](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=56.559333333333335) [been modified from its original source,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=59.55371428571429) [which is available on Kaggle. com. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=61.814499999999995) [dataset will now be uploaded to the home](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=64.68775000000001) [directory of your CloudShell VM. Copy this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=66.98466666666667) [dataset over to your spikey\_dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=70.51033333333332) [directory and run an ls command to ensure](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=72.57316666666667)[that this wine list is now present within](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=76.06075) [spikey\_dataflow. When you're working on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=79.82725) [the GCP and you need modify source files,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=81.40544444444447) [a config file, you can use code editor.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=83.71400000000003) [The beta version is now available and it's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=87.385) [this nifty IDE that you can use directly](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=89.339) [from within your browser. This has now](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=92.315) [opened up an Explorer window to your home](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=95.277375) [directory on CloudShell. There you see it,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=97.90350000000001) [your spikey\_dataflow folder, and within](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=100.49300000000002) [that, we have the spikey\_winery\_list. csv.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=102.59275000000002) [Let's take a look at the actual data that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=106.43) [you're planning to process using Dataflow.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=107.84) [This contains winery information, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=110.58)[variety of wine, the province, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=112.99083333333334) [country, a rating score, and the price of one bottle of this particular wine.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=6&mode=live&start=115.01775)

[Running a Dataflow Job Locally](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live)

[We are now ready to write the code for our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=1.923) [very first dataflow program using Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=3.665) [Beam. Right-click on the spikey\_dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=6.596428571428571) [folder and click on New, File. Create a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=9.0362) [file named spikey\_wineries. py that'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=12.733666666666668) [contain our Python code. And here is a new](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=15.2894) [blank file created using code editor on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=18.55277777777777) [the GCP. I'm going to paste in all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=20.8326) [Python code here. We'll walk through and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=22.915) [see exactly what it does. The Beam APIs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=25.218) [are available in the Apache Beam package](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=28.199625000000005) [typically aliased as Beam. Instantiate a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=30.525) [new Beam pipeline in the main function by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=33.50150000000001) [calling beam. Pipeline. Options to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=36.748200000000004) [configure the pipeline can be passed in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=39.17400000000001) [using the command line, but we won't](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=41.147000000000006) [really use that in this program. We'll get](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=43.25566666666666) [to that a little later on in this course.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=45.641363636363636)[Specify where the input file is available](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=48.343) [and where the output of this pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=50.711) [should be stored. Both of these are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=52.983) [folders on the local machine because we're](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=55.006) [going to execute this pipeline locally.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=56.813) [This job is going to perform a very simple](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=60.143) [grep operation. The search term is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=62.66766666666667) [California. We're going to find all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=65.16779999999999) [wineries that are based in California.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=66.94125000000003) [With the pipeline instantiated, we can use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=69.467) [the pipe operator in Python to apply all](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=72.15633333333335) [of our transforms. You can associate a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=75.6026) [string or a message with every transform](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=78.17999999999999) [that tells you what exactly that transform](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=80.23857142857142) [is about. We first read in all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=82.89039999999999) [winery data from our input file using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=85.69499999999998) [beam. io. ReadFromText. Every line in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=88.46625) [input csv file becomes an entity in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=92.41949999999997) [PCollection that is then passed on to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=94.77699999999999) [beam. FlatMap operation. A flatmap is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=97.99257142857144) [typically used to process a single entity](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=100.409125) [and the output of a flatmap can be one or](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=102.54125)[more entities for each input entity. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=106.64860000000002) [want to apply the find\_wineries method to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=110.02671428571429) [all of our input entities. If the search](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=113.60833333333333) [term is present in a particular input](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=116.580125) [record, then we return that line,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=118.78266666666666) [otherwise, we do nothing. Once we've](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=121.315) [grepped all of the wineries that are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=123.62950000000001) [present in California, we write the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=125.44485714285715) [results out to file using beam. io.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=127.472) [WriteToText. This pipeline can now be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=131.5775) [executed by calling p. run.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=133.58828571428566) [wait\_until\_finish. Let's create the folder](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=137.55325000000002)[under spikey-dataflow where our final](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=139.82260000000002) [results will be stored. Right-click, click](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=141.7782) [on New, Folder, and create a folder called](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=143.90242857142852) [output. The final results of this pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=147.84042857142853) [will be written out in the output folder](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=149.84999999999994) [with the result prefix. Use the icon on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=152.06014285714286) [the top right to switch to CloudShell.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=154.80088888888892) [We'll be doing this really often and we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=157.567) [use CloudShell to execute this code.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=159.653) [Running ls-l will show you the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=162.262) [spikey\_wineries. py, that is the code file](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=165.173) [that we just created, the csv file that is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=167.520625) [going to process, and the output directory](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=170.09500000000003) [where the results will be stored. We can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=172.17342857142856) [simply run Python spikey\_wineries. py to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=175.9085714285714) [execute this code on our local machine. No](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=179.06537500000002) [additional arguments are needed to use the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=182.08942857142858) [direct runner. The code ran through to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=184.75028571428572) [completion successfully. Let's confirm](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=188.26666666666668) [from the Dataflow page that no Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=190.20685714285713) [job was created and you can see that this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=192.86488888888888) [is indeed the case. There is no Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=195.54000000000002) [job to monitor as the code ran locally on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=197.7226666666667) [our CloudShell machine. We'll now switch](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=201.4348571428571) [tabs and close our CloudShell window and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=203.9612857142857) [go back to the code editor. Within the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=206.27866666666668) [output folder, we should find the results](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=209.18200000000002) [that we are looking for, and here you can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=210.95) [see it, the output file with the result](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=213.00911111111105)[prefix. And if you take a look at the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=215.92149999999998) [results here, you'll find all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=218.07459999999995) [wineries that are present in California.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=220.0055) [We filter out only those records that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=222.968)[are interested in from the original dataset.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=7&mode=live&start=224.922)

[Running a Dataflow Job on the Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live)

[In this demo, we'll work with the same](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=1.67) [Dataflow job that we saw earlier, but this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=3.28) [time, we'll see how we can run it on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=5.828) [Google Cloud. When you work with Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=8.326) [jobs in the Google Cloud, the best](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=10.543125000000003) [practice is to store your data on cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=12.3305) [storage buckets. Gsutil is a tool on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=15.646666666666667) [GCP that allows you to work with Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=18.560000000000006) [storage buckets from your command line.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=20.83428571428571) [Gsutil config will authorize your session](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=23.127999999999997) [to work with cloud storage buckets. Click](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=25.593) [on the URL that you see here on screen,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=28.535222222222224) [this will take you to a web page where you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=30.34) [log in with your GCP account in order to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=32.96155555555555)[authorize gsutil to access cloud storage](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=36.04233333333334) [resources. Click on Allow, this will give](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=38.91833333333333) [you a code that you need to paste into](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=41.343) [your CloudShell window. You're now](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=44.637666666666675) [authorized to create and manage cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=46.49042857142857) [storage buckets from your CloudShell](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=48.686) [terminal. We'll now run the gsutil mb](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=51.096) [command in order to create a new cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=53.440250000000006) [storage bucket called spikey-df-store.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=56.211250000000014) [This bucket will exist within your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=59.911) [spikey-dataflow project. Once this bucket](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=61.316) [has been created successfully, you can use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=64.245) [the gsutil cp command to copy over the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=66.92457142857143) [spikey\_winery\_list. csv from your current](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=69.57) [working directory to this cloud storage](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=72.534) [bucket. The csv file will now be present](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=75.28885714285711) [within the data folder on this bucket.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=77.42644444444443) [Let's switch over to our web console and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=81.036) [use the Navigation menu to view our cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=83.188) [storage buckets. I'm going to pin it to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=86.23533333333333) [the very top so that it's easily](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=87.99581818181817) [accessible. Click on Storage, and here is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=89.822) [the bucket that we just created, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=92.47449999999999) [spikey-df-store. This is a multi-regional](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=95.69983333333332) [bucket based in the United States. Click](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=97.43028571428572) [through, and within the data folder in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=100.42839999999998) [this bucket, you should find the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=102.48549999999997) [spikey\_winery\_list. csv. Let's now switch](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=105.51200000000001) [to the spikey\_dataflow directory on our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=107.75700000000002) [code editor and create a new file within](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=109.879375) [this. We'll call this the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=112.69475) [spikey\_wineries\_dataflow. py. We'll run](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=116.659) [the exact same bit of code where we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=119.03077777777779) [grep or search for wineries based in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=121.41488888888888)[California. There are a few differences](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=124.78111111111107) [though because this Dataflow code is going](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=126.48271428571431) [to run on the Google Cloud, we need to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=128.38266666666664) [specify the project for this particular](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=131.12985714285713) [Dataflow job and the bucket where we have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=133.78833333333336) [our data stored. We'll also specify a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=136.58466666666666) [number of command-line arguments that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=138.87999999999997) [we'll use to instantiate our pipeline.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=140.91450000000003) [Instead of using the command line, I'm](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=143.868) [going to specify these arguments here](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=145.69385714285713) [directly within code. We'll see how we can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=147.763) [use the arg parser to parse command-line](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=150.23950000000002) [arguments in later examples. Your Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=152.68866666666668) [job needs to know the project that it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=155.83125) [belongs to and also needs to know the job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=158.3816666666667) [name. The Save\_main\_session is an](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=161.2502857142857) [important command-line argument that you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=163.201) [need to pass into Dataflow. This causes](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=165.219) [the state of the global namespace within](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=167.70585714285713) [your Python code to be pickled and loaded](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=170.2882857142857) [into your Dataflow workers. Let's say you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=173.503) [import a few libraries in the global](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=176.15900000000002) [namespace, you'll find that you won't be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=177.97383333333335) [able to use these libraries within a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=180.60375000000002) [Dataflow worker if you don't specify](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=182.85233333333332)[save\_main\_session. All Dataflow jobs that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=186.385) [run on the GCP require a bucket to store](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=188.94100000000003) [staging and temporary data. This is a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=192.46933333333328) [bucket in your project that you need to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=195.12225) [pass in using command-line arguments. And](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=197.41800000000006) [here is the most important bit. In order](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=200.76357142857142) [to indicate that this job should be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=202.92049999999998) [executed on the GCP using Dataflow, you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=205.2752857142857) [need to specify that the runner program is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=208.292) [the Dataflow runner. Instantiate the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=211.09625) [pipeline and pass in the command-line](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=213.25542857142858) [arguments that we've hardcoded in the form](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=215.2412) [of a list here to this pipeline. Specify a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=218.12759999999997) [complete path to where our input data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=221.38275) [lives within the bucket that we just](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=223.87875) [created and specify an output path where](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=225.795) [the results will be stored. Observe how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=228.723) [the URL of cloud storage buckets are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=231.12814285714285)[specified using the gs:// prefix. The rest](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=234.14985714285712) [of the Apache Beam pipeline code is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=238.8585) [exactly the same. We search for all](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=240.82575) [wineries that are based in California and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=242.87671428571434) [here are all the transformations that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=244.918) [apply in our pipeline. We call p. run to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=247.55542857142854) [start the execution of our program, and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=250.70899999999995) [since all of this code is defined within](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=252.6965) [the run function, simply invoke the run](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=255.13359999999997) [function from within main. We'll now](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=257.7704999999999) [switch over to CloudShell and execute this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=260.01185714285714)[Dataflow job on the cloud. We simply call](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=262.09328571428574) [Python spikey\_wineries\_dataflow. py.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=265.64500000000004) [Remember, all of the command-line](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=267.311) [arguments are already hardcoded within](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=269.35100000000006) [this program. We get a bunch of log](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=271.78166666666675) [messages on screen indicating that our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=274.0763333333334) [Dataflow job has started execution. Use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=276.17300000000006) [the Navigation menu to head over to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=279.54044444444446) [Dataflow page and there you see it, our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=281.488) [first Dataflow job. When you click](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=284.72825) [through, you'll find this very nice visual](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=287.76114285714283) [representation of all of the stages in our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=290.0598571428571) [Beam pipeline. The first stage in our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=293.5356666666666) [pipeline is where we get wineries and then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=295.40966666666657) [we grep for wineries in California and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=297.5449999999999) [write out to file. On the right pane,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=300.10928571428565) [you'll find a summary for this particular](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=302.9980000000001) [job. You can see the name of the job, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=305.12275) [id, the region where it's being executed,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=307.43779999999987) [the current status of the job which is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=309.821) [running, and so on. You'll also see how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=311.588) [long this job has been running. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=314.57) [elapsed time here is just 15 seconds. At](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=316.48575) [this point in time, the machines for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=319.251) [executing our worker are being](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=321.372) [provisioned. Notice autoscaling doesn't](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=323.37533333333334) [have any information yet. The resource](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=325.21299999999997) [metrics for this job are currently empty.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=327.8906666666667) [No machine or worker has been allocated](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=330.82) [for this job yet. You can click on every](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=333.084) [stage in the pipeline in order to get](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=336.038) [additional information for this particular](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=337.922) [stage. You can see how many input elements](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=340.26283333333333)[have been processed and how many elements](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=341.90612499999986) [have been added to the output of this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=344.15257142857155) [stage. Since the job isn't completely up](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=347.2552500000001) [and running yet, all of these stages do](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=349.49957142857147) [not have any elements. If you wait for a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=351.77675) [little while, you'll observe that a worker](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=354.805) [will be provisioned in order to execute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=357.471) [this job. You can see that autoscaling is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=359.383) [in the process of assigning one worker,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=361.72357142857146) [which is the default for this job. If a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=364.44255555555554) [particular stage has multiple operations](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=367.54966666666667)[within it, you can click on this little](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=368.82975) [drop-down in order to view more details.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=370.93825) [As we wait a little longer, you'll see](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=374.489) [that our read operations slowly starts](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=376.344111111111) [processing the input file. Observe from](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=378.749) [the right-hand pane that it's reading in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=381.9901428571428) [from spikey\_winery\_list. csv stored in our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=384.86580000000004) [cloud storage bucket. The grep code that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=387.9525) [we specified using a lambda transform has](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=390.39750000000004) [also started executing here and you can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=392.89) [see that the input collections and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=395.35) [output collections now have some elements.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=397.46849999999995) [The input to our grep transform had](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=399.41849999999994) [roughly 50, 000 elements. We filter out](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=401.86733333333336) [many of these and the output is just](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=404.3933333333334) [around 16, 000 elements. I wait for a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=407.1283333333333) [little longer until our code runs through.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=410.4030000000001) [Grep has processed 100, 000 records and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=412.63)[extracted around 30, 000 wineries in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=415.8670000000001) [California. Click anywhere outside the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=419.489) [execution graph and you'll see details for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=421.68466666666654) [the entire pipeline. We have currently one](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=424.2846249999999) [worker, which is assigned one vCPU that is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=427.17140000000006) [running. If you scroll down to the bottom](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=430.2737500000002) [on the right pane, you'll find all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=432.1810000000001) [options which we use to configure our beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=434.74412500000005) [pipeline. And finally, some more time](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=438.186625) [passes and the green checkmark show that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=439.94812500000006) [your pipeline has executed successfully.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=442.20866666666655) [The read operation has processed a total](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=445.219) [of 150, 000 elements. The wall time here](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=447.046) [gives you an idea of how much time was](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=451.2535999999999) [spent per entity on this particular step.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=453.43957142857136) [This includes all of the time taken to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=456.583) [initialize the data, shuffling data,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=458.686) [processing the threads, and so on. Very](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=461.066) [slow pipeline stages perform a lot of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=464.0291666666667) [processing or initialization will have a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=466.75933333333336) [high value for wall time. If you wait for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=469.16228571428576) [a little longer, you'll find that Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=471.6307142857143)[will gradually wind down the workers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=473.73800000000006) [allocated to this job because this job is now complete.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=8&mode=live&start=475.8455000000001)

[Staging and Output on Cloud Storage](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live)

[You can also monitor the current status of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=1.83) [your Dataflow jobs using the command line](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=3.81) [and the gcloud utility. Gcloud dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=6.131) [jobs list will list out all of your jobs.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=8.5285) [We have just the one job that we've run so](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=11.758333333333333) [far that is currently in state running. It](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=14.496500000000001) [hasn't fully wound down yet. If you want](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=16.619333333333334) [to view the details of one particular job,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=18.725666666666665) [you can simply copy over the job id. I'm](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=21.253) [going to assign the job id environment](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=23.60825) [variable to this unique id and then I'm](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=25.598) [going to view the logs for this particular](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=28.261) [job by running gcloud beta dataflow logs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=30.717) [list and then specify the job id. And here](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=34.107) [are all of the logs that are associated](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=37.39944444444445) [with this current job. There are none so](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=39.1822) [far. You can also view all of the metrics](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=42.33819999999999) [related to this job by calling gcloud beta](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=44.611)[dataflow metrics list. This will show you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=47.394) [on the command line some of the metrics](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=51.22611111111111) [that you might view using a monitoring](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=52.92833333333334) [tool such as Stackdriver. Now that our job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=55.3222) [has executed successfully, let's navigate](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=59.537285714285716) [to cloud storage buckets and see whether](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=61.13842857142857) [we have the final csv file that is our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=63.37344444444445) [result. Go to the spikey-df-store bucket,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=66.35828571428571) [and within here, you'll observe some of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=69.0735) [the folders that have been created by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=72.02114285714286) [Dataflow itself. The staging folder will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=74.74366666666667)[contain the binaries, the Python packages,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=76.78833333333334) [and other details that Dataflow needs to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=78.915) [execute the Dataflow job within the Spikey](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=81.681) [wineries job. Here are the wineries, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=84.44266666666667) [Python packages, the SDKs, and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=87.11857142857146) [protocol buffers that Dataflow has used](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=89.0625714285714) [internally. The output that is the final](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=91.76075) [result of this Dataflow job will be within](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=93.8703333333333) [the output folder in the spikey-df-store](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=96.66399999999999) [bucket. Go to output, the go to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=99.3864) [wineries folder here, and within this,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=102.3453333333333)[you'll find all of the wineries that are located in California.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=9&mode=live&start=104.50950000000003)

[Dataflow Pricing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live)

[Now that we have a basic understanding of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=1.599) [how to execute jobs on Dataflow using the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=3.3) [GCP, let's see how GCP compute resources](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=7.083166666666667) [are prized for running these jobs. We know](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=9.839285714285715) [that Dataflow assigns workers to execute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=12.61257142857143) [the processes on the Google Cloud. When](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=14.707142857142857) [you use Cloud Dataflow, there are three](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=17.5468) [components to the price that you have to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=19.65814285714286) [pay to execute your jobs, the compute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=22.171222222222216) [cost, the storage cost, and the data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=24.263333333333332) [processing cost. The compute cost relate](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=26.9042) [to the vCPUs that I use to process your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=28.91633333333333) [jobs. This involve per-second billing of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=31.799600000000005) [vCPUs and memory used in both batch, as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=34.65499999999999)[well as streaming workers and let's just](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=36.84311111111113) [specify a different machine type. GCP will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=39.24528571428573) [use the standard worker configuration,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=41.63133333333333) [which involves 1 vCPU, 3. 75 GB of memory,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=43.6694) [and 250 GB persistent disk. This is for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=48.650500000000015) [batch jobs. For streaming jobs, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=52.21040000000001) [standard machine type is a little](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=54.16133333333333) [different. This involves 4 vCPUs, 15 GB](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=55.99480000000001) [memory, and 420 GB persistent disk. These](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=59.315142857142845) [compute costs are generally calculated in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=63.366) [terms of per vCPU per hour. You can see](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=65.32674999999999) [that streaming is more expensive than](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=68.69757142857144) [batch. Memory can also be calculated in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=70.89766666666665) [terms of per gigabyte per hour. Here,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=73.05299999999998) [there's not much difference between batch](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=75.85033333333334) [and streaming data. Storage costs will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=77.918) [vary based on whether you choose to use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=80.53742857142858) [standard persistent disks or SSD disks.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=83.40714285714284)[SSD persistent disks configured with your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=86.61449999999999) [machine offer lower latency, and hence,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=89.9535) [they're more expensive. And finally, data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=92.77700000000002) [processing costs for workers depend on how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=95.46599999999998) [much data you process, whether it's batch](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=97.85571428571431) [or streaming for batch data, it's about 1](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=100.21) [cent per GB, and for streaming data it's roughly 2 cents per GB of data.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=10&mode=live&start=103.23366666666666)

[Monitor Dataflow Jobs Using Stackdriver](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live)

[Before we go further into more complex](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=1.77) [processing tasks, let's see how we can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=3.619) [monitor Dataflow jobs using Stackdriver.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=5.567) [Here is the one job that we've executed on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=9.125) [Dataflow. Use the Navigation menu and go](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=12.336199999999998) [to the monitoring option under](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=14.209500000000004) [Stackdriver. If this is the first time](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=16.6425) [that you're using Stackdriver on this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=18.48099999999999) [project, you'll need to set up a new](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=19.90528571428571) [workspace for your project. You can add](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=22.904499999999995) [multiple projects to a workspace if you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=25.459428571428568) [want to. I'll go with just one project for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=27.183799999999994) [now. Stackdriver is primarily used on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=30.078250000000004) [GCP, however, it can also be used to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=32.873571428571424) [monitor your AWS resources in case you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=35.26062500000001)[have a hybrid cloud setup. We'll just skip](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=37.571000000000005) [this step for now and then move on. In the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=40.74557142857141) [next page, you're asked to install](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=44.04225) [monitoring agents in order to get more](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=45.324) [information about your resources. Once](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=47.729) [again, we can skip this, configure no](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=49.6975) [reports, wait for your workspace to be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=52.01933333333333) [configured, and then click on the Launch](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=54.326857142857136) [monitoring button, which will then take](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=56.873000000000005) [you to Stackdriver monitoring for this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=58.640714285714274) [project. This is the page that you'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=61.426750000000006) [start off with when you want to launch any](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=63.30624999999999) [kind of monitoring set up. Here, the first](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=65.25844444444442) [thing that we'll do is to create a new](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=67.7822857142857) [dashboard to monitor our Dataflow jobs.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=70.27477777777777) [Click on the Dashboards link to the left](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=72.904) [Navigation pane and then click on Create](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=75.07) [Dashboard. This will bring up a new](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=78.73449999999998) [dashboard where you can group your charts](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=80.67624999999997) [together, give it a meaningful name and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=82.48733333333335) [we'll call it Spikey Dataflow jobs and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=84.219) [click on Add Chart. This will open up a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=87.01899999999998) [Dialog pane where you can configure what](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=90.17755555555557) [is it that you want to monitor. You first](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=91.8282) [need to specify the target resource type.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=94.118) [I'm going to go with Dataflow job. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=96.356) [next thing you need to specify is what](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=99.32288888888888) [metric you want to monitor. I'll go with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=101.618) [element count here to track how many](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=104.70187499999999) [elements flow through our Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=106.83850000000001) [pipeline. For the one Dataflow job that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=108.87225000000002) [executed, you can see that we started with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=111.36149999999999) [over 100, 000 elements, which was filtered](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=113.33725000000001) [down to below 50, 000. Click on Save and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=115.69242857142854) [this chart will now be added to your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=119.23344444444449) [monitoring dashboard. You can customize](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=122.63344444444448) [your dashboard with multiple charts](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=124.5435) [showing the different details that you're](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=126.34200000000001) [interested in. Once again, I'm going to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=128.241) [monitor my Dataflow jobs, and this time,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=129.83671428571432) [I'll track the total memory usage time of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=132.475) [my job. You can play around with different](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=135.355875) [metrics and view the visualization that is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=137.7976) [displayed to the very right of your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=139.93671428571426) [screen. If you're interested in one](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=142.74175) [specific job, you can also add filters to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=144.65371428571433) [your visualization. I'll leave this as-is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=148.58049999999992) [and add this visualization to my](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=150.79342857142854) [dashboard. Another important metric for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=153.112) [any Dataflow job is the vCPU time. I'm](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=155.02266666666665) [going to add in another chart that tracks](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=158.147875) [the vCPU time for my Dataflow jobs. Click](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=160.60575) [on Save and add this chart to your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=164.291) [dashboard as well. Now I've not gone into](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=165.843) [the details of the different ways in which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=168.214) [you can have your visualization displayed.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=170.946) [Those are for you to explore. Instead,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=173.343) [we'll move onto creating an alerting](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=175.43) [policy for our Dataflow jobs. Click on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=177.56916666666666) [Alerting on the left Nav pane and choose](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=180.1431111111111) [Create a Policy. You can opt in to the new](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=182.635) [alerting user interface that's currently](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=186.02266666666665)[in beta. This will give you many more](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=187.66899999999998) [options for alerting. The very first step](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=190.85699999999994) [is to specify the condition under which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=193.0855) [you want your alert to be sent. The alerts](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=195.03671428571425) [are going to be configured on Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=197.47955555555552) [jobs and we'll configure total memory](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=199.81250000000003) [usage time and set an alert based on this.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=202.2782857142857) [Scroll down to the bottom here and you can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=205.798) [specify the threshold condition for your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=208.28) [alert. If the total memory usage time is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=211.4428571428571) [about 200, you want to be notified.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=214.08662500000008)[Observe that you can see the threshold](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=216.989) [that you just configured off to the right](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=218.68124999999998) [on your graph. You want to be alerted only](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=220.543) [if this threshold has been reached for a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=223.521) [continuous 3 minutes. Click Save to set up](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=226.31) [our alerting condition and we can go ahead](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=230.16524999999996) [and complete configuring our alert. You](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=232.48442857142857)[can add in a notification channel to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=235.657375) [indicate how you want to be alerted. Click](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=237.361) [on Add notification channel, we'll go](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=240.50060000000002) [ahead and choose to be notified by email.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=243.15179999999998) [The email address that I'll put in here is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=246.163) [the address associated with this GCP](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=248.293) [account, spikey. engineer@loonycorn. com.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=250.323) [I can also add in additional](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=254.565) [documentation, which will make this alert](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=256.3418571428571) [more useful when I'll receive an email. My](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=257.88237499999997) [documentation says high memory usage for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=260.5276666666666) [Dataflow jobs. Now I'm going to go ahead](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=262.81975) [and name this alerting policy so I can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=265.131) [identify it. It's the Dataflow high memory](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=267.73685714285716) [usage policy, click on Save, and you will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=270.2398333333333) [have configured your alert. If this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=274.09671428571426) [threshold condition is ever breached,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=276.40333333333336) [you'll find that you'll receive an email.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=277.91380000000004) [And on this note, we come to the very end](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=281.313) [of this introductory module on Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=283.27) [Dataflow. Cloud Dataflow is the serverless](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=285.741142857143) [and fully managed data transformation](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=288.05728571428585)[service offered on the GCP. Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=290.53739999999993) [allows you to construct pipelines to apply](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=293.37328571428566) [transformations to both batch, as well as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=296.538) [streaming data. Dataflow uses the Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=299.519) [Beam unified programming model in order to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=302.14700000000005) [define your pipelines. You can write code](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=305.52887499999997) [for Dataflow using either the Python or](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=307.9782857142857)[the Java SDKs. You'll find that the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=310.83600000000007) [transforms that you apply on data using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=313.42742857142855) [Apache Beam are very similar to those that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=315.60999999999996) [you'd find on Apache Spark. However, there](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=318.4004444444444) [are major differences in the architecture](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=320.6781428571429) [for Dataflow and Spark, specifically](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=322.757) [because Dataflow is completely serverless.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=325.673)[Dataflow offers complete separation of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=329.22) [storage and compute. You store your data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=331.461) [in cloud storage buckets and operate on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=333.7195714285715) [them using autoscale workers. In the next](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=336.9004999999999) [module, we'll see how we can transform our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=340.132) [data using Apache Beam APIs in both Java, as well as Python.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=00f5e4d5-0fdc-47ef-9193-b55d93af030f&clip=11&mode=live&start=342.4334285714285)

[Understanding and Using the Apache Beam APIs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live)

[Module Overview](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live)

[Hi, and welcome to this module where we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=0) [study the Apache Beam APIs used to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=4.262) [transform our data in a lot more detail.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=6.429) [This module will be very, very hands on.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=9.47) [We'll see how we can build pipelines for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=12.088) [data transformation actions and execute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=14.139) [them on Dataflow. We'll build and execute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=16.884)[pipelines in both Java, as well as Python.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=19.075) [We'll use these pipelines to process some](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=22.743) [real world ecommerce data, and we'll see](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=25.297) [how we can visualize and debug pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=28.271) [stages using Dataflow monitoring. We'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=30.774) [see how we can configure a pipeline to run](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=33.560874999999996) [with multiple workers to deal with higher](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=36.08285714285714)[workloads. We'll also see how it's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=38.78633333333334) [possible to identify slow pipeline stages.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=40.86414285714285) [In addition, we'll also see how we can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=44.122) [monitor logs in Dataflow and use them as a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=46) [debugging aid. And finally, we'll also](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=49.192) [integrate Dataflow with other GCP](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=51.838) [services, specifically BigQuery. For all](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=54.47633333333334) [of the demos in this module, we'll use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=58.086) [this ecommerce dataset. This ecommerce](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=60.5046) [dataset was made available by Flipkart,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=62.861333333333334) [which is a very popular ecommerce site in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=65.21771428571428)[India. This data was originally uploaded](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=68.91749999999999) [to the data. world site. We've taken the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=71.65099999999998) [original dataset and modified it a little](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=75.7405) [bit to fit our use case. The columns in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=78.03539999999998) [this dataset are the time a particular](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=80.84733333333334) [product was sold, the name of the product,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=82.91214285714285) [the quantity of sales, and the price of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=85.33266666666667) [the original product. You can imagine that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=88.43028571428572) [this is information that SpikeySales. com](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=90.536) [gets every 5 minutes when it has its sale](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=93.06671428571427) [days. We'll perform a bunch of operations](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=98.23800000000001) [on this data using Dataflow in order to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=99.97100000000003) [extract insights. Let's upload this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=103.67074999999998) [dataset to our cloud storage bucket where](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=105.591) [we can then process it using Dataflow. Go](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=107.49042857142857) [to spikey-df-store, and within the data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=110.67033333333335) [folder here, we'll click on the Upload](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=113.69399999999999) [button to upload the csv file from our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=116.10125000000001)[local machine to cloud storage. This file is called the spikey\_sales\_weekly. csv.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=0&mode=live&start=118.56185714285715)

[Create a Java Project Using Maven](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live)

[In this demo, we'll see how we can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=1.918) [implement an Apache Beam pipeline in Java.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=3.579) [If you're not familiar with Java and are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=6.329) [only interested in Python, you can skip](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=7.986) [this demo and move onto the next one. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=10.935) [is the only demo in this course that's in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=13.540222222222223) [Java, the rest all is in Python. We'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=15.247) [then execute this pipeline on the local](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=18.503285714285717) [machine, as well as the Google Cloud.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=20.535) [We'll write our code within a directory](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=22.943) [called spikey\_revenue, cd into the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=24.449)[spikey\_revenue folder. If you're working](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=26.996) [with Java to write an Apache Beam program,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=29.909375) [you can use the Maven built-in Project](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=32.281625) [Management tool to write your code and to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=34.967) [manage your dependencies. Maven is an open](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=38.443) [source tool that can be used with any](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=40.8038888888889) [Java-based project. If you're interested](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=43.78920000000001) [in learning more about Maven, maven.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=45.59285714285714) [apache. org is a very useful site. Maven](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=47.59022222222223) [has a repository of Java archives or JAR](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=51.8005) [files and you can specify dependencies](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=55.341857142857144) [from within this repository using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=57.963) [XML-based templates. You can generate the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=61.12942857142856) [build, the dependencies, and the examples](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=63.42559999999999)[from an original archetype that is a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=65.59300000000002) [template that Maven already has for Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=68.13557142857142) [Beam. We'll start with version 2. 6 of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=71.74699999999999) [Apache Beam archetype that Maven has](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=74.60877777777773) [within its repository. You can use this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=77.69024999999996) [Apache Beam template to generate our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=80.50400000000002) [spikey-product-revenue artifact. An](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=83.16000000000001) [artifact is generally a JAR file in Java.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=85.14588888888889) [Specifying the Maven generate command with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=88.873) [these options will simply create the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=90.957) [Apache Beam project for you with all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=93.77957142857142) [the dependencies set up correctly. With](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=96.10324999999995) [our project successfully generated, let's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=99.247) [close our CloudShell window and switch to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=101.08933333333334) [the code editor IDE. If you hit File,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=103.96466666666667) [Refresh, you can view your Maven project](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=107.02844444444442) [within this Explorer window under](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=109.10700000000001) [spikey\_revenue. You'll find the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=111.458) [spikey-product-revenue folder, this was](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=114.15974999999999) [the name of our artifact ID, and under](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=115.65687499999999) [that, you'll find the pom. xml. Every](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=118.384) [Maven project comes with a pom. xml. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=121.344) [is Maven's project object model file,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=124.67228571428572) [which is used to specify your JAR version](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=126.918) [and all of its dependencies. Here is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=129.762) [artifact ID or the Java archive or JAR](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=134.31799999999996) [that this particular project will build.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=137.441) [If you scroll down this file, you'll find](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=139.928) [all of the dependencies for Apache Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=141.7097142857143) [are also specified using XML. These](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=144.5465714285714) [dependencies will be pulled from the Maven](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=147.49133333333333) [repository. Under the source folder will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=150.389) [be a folder hierarchy that represents our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=152.37387499999997)[namespace, and deep inside that, you'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=155.00240000000002) [find a bunch of Java examples. These are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=156.70100000000005) [all Dataflow examples that you can use as your template to start off with.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=1&mode=live&start=159.59828571428577)

[Writing a Dataflow Job in Java](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live)

[I'm just going to create a new file here](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=2) [under this examples folder where we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=4.216499999999999) [write our Dataflow code. This is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=6.433285714285714) [SpikeyProductRevenue. java. I'm going to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=9.638) [paste in the code and walk through it step](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=11.801000000000002) [by step showing you exactly what it does.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=13.815749999999998) [Here are all the import statements that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=16.060125) [you'll need to write your Dataflow code](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=18.163714285714285) [using Apache Beam APIs. We first import](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=20.703999999999997) [all of the standard Java libraries that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=23.738125000000004) [might use and below these are the Apache](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=26.205125000000002) [Beam imports. Let's start off at the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=29.8522) [bottom here. This is our main method. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=32.274) [is where we parse the command-line](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=35.51757142857143) [arguments to set up the options for our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=37.537) [pipeline. The options for this particular](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=40.96544444444444) [program are specified in a class called](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=42.81983333333334) [revenue count options. The main function](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=45.078) [then calls the runProductRevenue function.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=48.47839999999999) [We'll set up this Dataflow job to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=51.302) [calculate the total revenue for each of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=53.701250000000016) [the products sold on our ecommerce](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=56.28349999999999)[platform across all of the sale days. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=58.76714285714286) [first step is to instantiate a Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=61.65000000000001) [pipeline class using all of the options](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=64.31028571428571) [that were passed in on the command line.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=66.86537500000001) [The options are used to configure our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=69.002) [pipeline by specifying where the input](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=71.357) [data might be found, where the output](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=73.51833333333335) [results should be stored, and so on. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=75.91842857142855) [apply function is called on the pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=78.26744444444444) [and on PCollections to transform our input](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=80.47433333333333) [data. The first step is to read all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=83.97474999999997)[records from the input file which we do](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=86.404) [using TextIO. read. From every record, we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=88.642) [extract the name of the product and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=92.99933333333334) [revenue that that product generated.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=95.506) [Remember, every product might be present](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=97.586) [in our data more than once. It depends on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=99.58742857142859) [how many times it's sold. Observe that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=102.68275) [use ParDo or ParallelDo to perform this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=105.80262499999998) [transformation. This means we want this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=108.885) [operation to be executed in parallel with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=110.73837499999999) [multiple workers. The extract product](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=113.84466666666665) [revenue function is a class that we've](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=115.87414285714286) [written. We'll see the details of this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=118.16125000000002) [code in just a little bit. Let's first get](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=119.80266666666665) [the big picture of the various transforms](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=122.18199999999999) [applied. Every product and its associated](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=125.1648333333333) [revenue is then converted to key value](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=127.57428571428574) [pairs where the product name is the key](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=130.17537499999995)[and the revenue is the value. Once we have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=132.62800000000004) [product name as keys and the revenue as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=135.86566666666673) [corresponding values, we can perform a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=139.19383333333337) [grouping operation. We'll group all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=141.7248571428572) [revenue numbers on a per product basis,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=144.09925000000004) [and once we do that, we can add up all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=146.67844444444447) [the revenue numbers for each product to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=150.30999999999997)[find the total revenue per product. Notice](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=152.59485714285714) [the combine operation says we perform it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=156.2155) [per key and we perform a summing operation](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=158.7804444444444) [of doubles. These are all built-in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=161.43566666666666) [aggregations available in Apache Beam.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=164.08059999999995) [Once we have the total revenue per product](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=166.935) [present as a key value pair, we can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=169.477)[convert the keys and the values to a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=172.16322222222212) [string format and write it out. Our map](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=174.43411111111115) [operation processes every key value pair](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=177.81300000000002) [and converts it to a string format. And](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=180.32828571428564) [finally, we'll write out the total revenue](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=183.43757142857143) [per product to an output file stored on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=185.52185714285716) [cloud storage. We start execution on this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=189.197) [pipeline by calling p. run and wait until](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=191.84071428571426) [it's complete. Let's now see some of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=195.09874999999994) [more granular details of this code.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=197.35624999999996) [Revenue count options are what we use to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=199.721)[specify the input arguments that we pass](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=201.822) [in on the command line. This extends the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=204.63512499999993) [pipeline options base class and every](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=208.02050000000003) [argument needs to be annotated. Here is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=210.88) [our input file with an @Description](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=214.27675000000002) [annotation and a default value. And here](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=216.532) [is the option for where the output will be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=220.25844444444445) [written, the output file with an](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=222.6235) [@Description annotation and a default](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=224.065) [value. For every getter, you need to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=227.39833333333328) [specify setters as well. Set input file](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=229.99528571428567) [and set output. Let's take a look at how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=232.81699999999995) [each of our transforms are implemented.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=235.28166666666664) [The extract product revenue function](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=238.503) [operates on every record that is fed in in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=240.539) [the form of text and extracts the product](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=243.4215) [field, as well as the revenue field.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=245.66575) [Transformations that can be applied in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=248.121) [parallel to your input data typically](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=250.166) [extend the DoFunction base class. This is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=252.234) [a templated function. The input is in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=255.562) [form of a string, the output is also a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=257.7717777777777) [string. The actual transformation is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=260.85150000000004) [specified within the process element](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=262.84779999999995) [method of the DoFunction. You can process](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=265.2070000000001) [the input string that's passed in in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=268.164)[element variable and pass the process](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=270.8430000000001) [output onto the receiver. Every string](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=273.1813333333333) [record that we read in has comma-separated](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=275.77250000000004) [values. We'll split on the comma and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=278.4705) [extract the product name and the revenue.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=281.1565000000001) [The name of the product is available at](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=284.665) [index1 and the revenue is obtained by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=286.547) [multiplying the price of the product by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=289.582) [the quantity. We'll then write the product](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=291.8657142857143) [and the revenue out in a string format.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=294.1898888888889) [The PCollection with the product and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=297.869) [revenue in the string format is then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=299.663) [passed on to the extract product revenue](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=301.8671250000002) [pair function. This once again extends the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=305.227) [DoFunction class. The input to this is a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=308.03842857142854) [string and the output is a key value pair.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=310.21233333333333) [Key value pairs in Apache Beam are stored](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=313.21) [in the KV class. In the process element](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=315.796)[method, we'll split on the comma once](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=319.384) [again, extract the name of the product and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=321.462) [the corresponding revenue as a double,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=324.17285714285714) [instantiate a new key value pair, and pass](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=326.53) [this onto the output. In our key value](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=329.60100000000006) [pair, the key is the product name which is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=332.47255555555563) [of type string and the value is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=334.8185000000001) [revenue, which is of type double. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=337.54980000000006) [GenerateProductRevenuePairs that generates](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=340.86733333333336) [these key value pairs extends the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=342.9937142857143) [PTransform base class, and within the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=345.368)[PTransform, we apply a DoFunction.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=348.055) [PTransforms basically operate on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=352.023) [PCollections that they can then input](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=354.193) [PCollection and produce a modified output](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=356.473) [PCollection. This PTransform simply calls](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=360.44157142857154) [the extract product revenue pair function](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=363.1763333333333) [that we saw earlier. We also saw that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=366.19699999999995) [had applied the FormatAsTextFn as a part](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=369.34287499999994) [of our pipeline. This takes in a simple](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=372.2377142857143) [key value pair and converts it to a string](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=375.23988888888886) [so that it can be written out to a file.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=377.3344999999999)[The input to this FormatAsTextFn is a key](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=380.009) [value pair where the key is a string and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=383.068) [the value is a double and the output is of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=385.155) [type string. The actual code for this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=388.15700000000004) [transformation is very simple. We simply](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=390.607) [construct a string using key, value. Now](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=393.58099999999996) [that you've understood the actual](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=397.1451666666667) [transformation that we apply as a part of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=398.436) [our Beam pipeline, let's switch over to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=401.033) [CloudShell and run this code. In order to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=402.937) [build and execute this Java project using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=406.413375)[Maven, you need to be in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=408.3801111111111) [spikey-product-revenue folder, which is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=410.484) [where your pom. xml lives. Let's first run](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=414.1391428571428) [this dataflow job locally using mvn](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=417.9385714285715)[compile exec:Java and specify that the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=420.65066666666667) [main class that we want to execute is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=423.6006666666667) [SpikeyProductRevenue. Wait for this job to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=427.307) [run through successfully. This might take](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=429.7297500000001) [a minute or so. If you take a look at the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=431.62242857142866) [Dataflow page on your web console, you'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=434.3851999999999) [find that this job was executed on your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=436.1694999999999) [local CloudShell machine and cannot be monitored using Dataflow.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=2&mode=live&start=439.15266666666673)

[Examining Output Files on Cloud Storage](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live)

[Our Dataflow job in Java executed locally,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=1.931) [but it was pointing to data that lived in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=4.937) [cloud storage buckets. Let's head over to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=7.327) [cloud storage and see whether our spikey](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=10.302499999999997) [weekly sales data was processed correctly](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=12.577714285714285) [by our job. Within our spikey-df-store](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=15.234000000000002) [bucket, head over to the output folder](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=18.42) [where the output would have been produced.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=20.471750000000007) [If you remember the output file format, we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=23.419) [stored it within the output folder with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=25.89075) [the product prefix. If you click on one of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=28.556) [these files and open it up, you'll see the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=31.066) [total revenue per product across all](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=33.493) [products that sold during sale days. If](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=36.915) [you perform a search based on product](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=40.25875) [name, you'll find that there's exactly one](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=42.063) [entry for each product with the total](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=44.162) [revenue number associated with that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=46.715)[product. You can see here that there are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=49.758333333333326) [two entries for Funcart. The first is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=51.40037499999998) [Funcart Frolic Birthday and the second](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=54.71885714285715) [product is the Funcart Heart Goggles.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=57.16133333333333) [Let's search for another product here, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=59.882) [Shoetopia Girls Sandals. There's exactly](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=61.98025) [one entry and you can see that the total](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=63.97025) [revenue that it generated across all sale days was 29, 441.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=3&mode=live&start=66.45285714285716)

[Find the Top Selling Products](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live)

[Now that you got a taste for what a Java](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=2.275) [Dataflow job looks like, from this demo](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=4.609) [onwards, we'll go back to working in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=6.738) [Python. We'll build and execute a Python](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=9.853874999999997) [pipeline to find the top selling products](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=12.320142857142859) [across all of our sale days. We'll also](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=14.649000000000004) [see how we can monitor the logs generated](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=17.179) [by this pipeline. We start off in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=20.305125000000004) [spikey\_revenue folder of our code editor](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=23.356000000000005) [IDE, and within this, we are going to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=25.542999999999996) [create a new Python file where we'll write](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=28.058222222222227) [our code. We'll call this Python program](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=30.5384) [spikey\_top\_selling\_products. py. I'm going](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=33.286) [to paste in all of the code and work](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=37.079800000000006) [through it step by step starting with the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=38.979375000000005) [import statements. We'll use the RParse](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=41.498666666666665) [library in order to parse command-line](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=44.4365) [arguments and we'll import the logging](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=46.7682)[library in order to generate our own](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=48.98957142857144) [custom logs. All of the other imports are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=51.939400000000006) [Apache Beam specific and you'll understand](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=54.38949999999999) [them when you see how they're used. We'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=56.24700000000001) [work with the same ecommerce data which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=59.579) [represents the sales of various products](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=61.675000000000004) [during sale week on SpikeySales. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=64.01566666666666)[objective of this Dataflow program is to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=67.27714285714285) [find the top 10 selling products during](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=69.50042857142857) [sale week. The first transformation that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=72.23566666666667) [we apply to the CSV data is to extract the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=74.9708) [name of the product and the revenue that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=77.90124999999999) [it generated. In Python, we need to write](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=80.47425000000001) [a class that inherits from the beam. DoFn](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=83.89416666666666) [class. The actual transformation logic](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=87.418) [will be written within the process](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=89.44266666666667) [function of this class. The input to this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=92.2045) [function is one record from our csv file](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=95.48949999999998)[and we extract all of the columns from](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=98.0512) [this data. This includes the timestamp,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=101.0845) [the name of the product, the number of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=103.35114285714286) [units sold, and the retail price. We then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=105.79185714285713) [multiply units sold by retail price to get](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=109.26849999999999) [the revenue that this product generated](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=111.70942857142856) [and return a tuple which contains the name](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=113.4312857142857) [of the product and the corresponding](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=116.599) [revenue. We also generate a custom log](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=119.143) [within this function using the logging](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=121.93800000000003) [Python module. Once you've extracted the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=124.84833333333333) [product and the corresponding revenue, we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=127.10757142857145) [calculate the total revenue on a per](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=129.1846) [product basis. The input to this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=132.97371428571427) [particular stage in the pipeline is all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=135.0035) [the revenues associated with the same](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=137.4901428571429) [product that was sold at different points](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=139.538125) [in time. We extract the name of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=141.924) [product and the list of revenue values](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=144.56233333333333) [from the input and find the total revenue](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=146.87050000000002) [by performing a sum operation. We then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=150.40644444444442) [return the name of the product and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=153.42566666666664) [total revenue across all of the sales of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=155.4375) [the product. This is then passed into a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=158.14077777777777) [function that calculates the top 10](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=160.923) [products by revenue. We generate a custom](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=163.691) [info log here as well. Notice that all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=167.19122222222217) [our log messages are prefixed by the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=169.677) [asterisks. This is just for demo purposes](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=172.30075) [so that it's easy to search for these](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=174.29999999999998) [particular messages within the logs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=176.4195) [generated. A list of all of the products](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=179.004) [that were sold on SpikeySales are passed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=181.20577777777785) [into this function. We perform a sort](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=183.70937499999994) [operation and sort in the descending](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=186.29499999999996) [order. We performed the sort under revenue](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=188.99166666666665) [field of the tuple and this sort is in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=192.24242857142855) [descending order as you can see from the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=195) [reverse equal to True. Once we have all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=196.797) [the products sorted in the descending](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=200.52612499999995) [order of revenue, we extract the top 10](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=202.55399999999997) [selling products. We write the final top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=206.08628571428568) [10 products out to a file, we first need](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=208.34699999999995) [to convert it to a string format and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=210.57866666666666)[that's what format output does. For each](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=212.90650000000002) [product, we extract the product name and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=215.61185714285716) [the total revenue and write it out in a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=218.03344444444443) [comma-separated fashion. Let's take a look](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=221.49699999999999) [at the main run method that executes our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=223.84466666666668) [pipeline. We use the argument parser in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=226.278) [order to parse arguments that we pass in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=228.64324999999994) [from the command line. The input file that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=230.91471428571435) [this job processes can be passed in as a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=233.87362500000003) [command line argument, this has a default](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=236.41266666666658) [value though, so we can choose to leave it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=238.54371428571434) [empty. The output file where the final](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=241.52244444444452) [result will be stored is a required](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=243.55625000000003) [argument. You do need to specify this. Use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=245.86383333333328) [the argument parser to parse the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=248.58314285714286) [command-line arguments. The arguments that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=250.69) [were added, the input and the output will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=253.43) [be stored in the known args. All of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=255.93066666666664) [remaining arguments will be the pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=258.7695714285714) [args. Instantiate the pipeline options](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=261.6396666666667) [that we'll use to configure our Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=263.716) [pipeline using the pipeline arguments](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=266.50828571428565) [passed in. Even if you haven't specified](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=269.05199999999996) [this using the command line, make sure](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=271.6377142857144) [that you set save\_main\_session to True.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=273.7792500000001)[Remember that this is what ensures that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=277.216) [the global namespace of your Python](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=278.9362499999999) [program is pickled and passed into the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=281.12700000000007) [Dataflow workers. This is what gives the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=284.2614285714286) [Dataflow workers access to all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=286.68971428571416) [libraries and functions defined in your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=288.6461428571429) [global namespace. And here are the various](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=291.3183333333333) [stages in our Beam pipeline. We first read](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=294.0399999999999) [in the input file using the ReadFromText](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=296.70399999999995) [function. For each record in our csv file,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=299.6735) [we apply a ParDo or ParallelDo operation](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=303.28890000000007) [which extracts the product name and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=307.04319999999996) [total revenue. This is the processing that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=310.3684999999999) [we specified in the revenue per product](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=312.62212500000004) [class. Since the same product might have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=315.6975999999999) [been sold at different points in time over](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=317.72750000000013) [different days, we then perform a grouping](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=320.076) [operation on a per product basis. We then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=322.84857142857135) [apply another ParDo transformation to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=326.7316666666667) [calculate the total revenue per product.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=329.4635714285714) [We finally combine all of the product](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=333.211) [revenue pairs that we generated to one](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=335.615) [list so that we can sort and find the top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=338.76062499999983) [end products. The output of this stage is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=342.3804545454547) [a list of product revenue tuples. We then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=345.0456666666668) [pass this list into the top revenue](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=349.0425555555556) [products function, which you'll find the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=351.449) [top 10 products. We format these top 10](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=353.801) [products as a string and write this output out to file.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=4&mode=live&start=357.41)

[Executing a Java Pipeline on Cloud Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live)

[We have our Java code already written.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=2) [Let's see how we can use the CloudShell](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=4.077) [terminal window in order to execute this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=5.928) [job on the Google Cloud. The most](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=8.254) [important argument that you need to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=10.999571428571427) [specify is the runner for your job.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=12.607999999999999) [Instead of the local machine, we are going](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=14.792) [to use the DataflowRunner. The main class](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=16.258) [that is the entry point for our job is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=19.299) [SpikeyProductRevenue and here are the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=23.102333333333334) [other parameters that our Dataflow job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=25.048285714285715) [needs. The project in which this job is to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=27.038285714285717) [be executed, the bucket which forms its](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=30.616333333333333) [staging location, and where the output](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=32.94083333333334) [should be placed. In addition, we also](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=35.077625000000005) [need to specify the DataflowRunner as a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=37.230999999999995) [pipeline option. Hit Enter to submit this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=40.312200000000004) [job to Dataflow and head over to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=43.26424999999999)[Dataflow web console using your Navigation](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=45.61475) [menu. This time around, we did not specify](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=48.4025) [a job name using the command line and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=51.145666666666685) [Dataflow automatically assigned one for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=53.96642857142859) [us. You can click through to this job and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=56.5466) [visualize the various stages of execution](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=59.165) [in our graph. If you wait for the little](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=61.594) [bit, you'll find that Dataflow spins up](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=63.943) [workers to execute our pipeline. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=66.501) [number of workers move from 0 to 1. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=69.10000000000001) [interesting stage here is the total](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=72.42071428571428) [revenue per product. If you expand this,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=74.297) [you'll find that it performs two](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=77.32250000000002) [operation, a grouping by key and then a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=79.14985714285714) [combining of grouped values. It's still](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=82.7577142857143) [too early in our pipeline execution. No](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=85.22875) [elements have been added and processed by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=87.46328571428572) [this stage. Wait a little longer and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=89.790875) [resources will soon be allocated to our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=92.464) [worker. Our worker uses 1 vCPU, that is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=95.36987500000002) [the default. If you wait a little longer,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=98.85333333333332) [you'll slowly see that the various stages](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=100.828) [of execution of your pipeline will be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=103.202) [completed. The very first stage that read](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=106.27150000000003) [in our input file has processed a total of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=108.12733333333334) [around 20, 000 elements. That's how many](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=111.65779999999997) [records were present in our input dataset.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=114.25214285714287) [The next stage is the extract product](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=116.531625) [revenue. It operated on all of its input](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=119.77066666666666) [elements and produced corresponding](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=122.34362500000002) [outputs. These text inputs are then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=124.80924999999999) [converted to key value pairs in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=126.94585714285715) [GenerateProductRevenue pairs stage. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=129.144)[process the roughly 20, 000 product](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=132.88642857142855) [entries and generate product revenue pairs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=135.145) [for each entry. The next stage is where](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=137.982) [things get interesting. You can click on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=141.415) [the internal stages within the overall](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=144.24199999999996) [combine operation. You can see from the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=147.0835) [elements processed that out of roughly 20,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=150.06200000000004) [000 entries, there are just about 12, 000](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=152.5978571428571) [entries for unique products. The Combine.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=155.4704285714286) [GroupedValues operation will find the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=160.1168) [total revenue for each product. And](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=163.28614285714283) [finally, we convert the key value pairs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=166.414625) [for these 12, 000 products to strings and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=168.729) [write it out to a file. The job has run](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=171.069) [through successfully, even though Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=174.64033333333333) [hasn't completely wound down. Let's head](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=176.3625) [over to our cloud storage buckets and see](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=179.30271428571427) [the result of this job. When you execute a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=181.74024999999997) [pipeline written in Java using Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=184.89650000000006) [Dataflow, you'll find that Dataflow sets](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=186.79633333333334) [up a bucket as its staging area. This is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=189.375) [where it'll upload all of the Java SDKs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=193.45666666666668) [and other binaries and jobs that it needs.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=195.81777777777776) [Let's head over into spikey-df-store and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=199.792) [see the results of executing this Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=202.303) [job on the cloud. Within the outputs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=204.751) [folder, we'll have a product revenue](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=206.483) [folder where our results will be stored.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=208.82471428571426) [If you open up one of these files, you'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=211.641) [find the total revenue calculated on a per product basis across all sale days.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=5&mode=live&start=213.73785714285714)

[Execute Jobs and Monitor Logs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live)

[Now that we've defined our processing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=1.677) [pipeline, let's switch over to our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=3.219) [CloudShell terminal window and execute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=5.239) [this on Cloud Dataflow. Recall the Python](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=7.374833333333332) [executable on our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=11.86) [spikey\_top\_selling\_products. py file and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=13.381) [we pass in a bunch of command-line](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=15.8354) [arguments. The input file to this Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=18.99042857142857) [job is the spikey\_sales\_weekly. csv file](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=21.345999999999997) [stored in the spikey-df-store bucket. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=24.0415) [output file will be generated in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=28.0185) [output folder in the same bucket in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=29.562) [top selling subfolder. Specify the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=32.028) [DataflowRunner in order to run this on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=35.461571428571425) [Cloud Dataflow. Specify the name of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=37.99250000000001) [project, the temporary location that this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=40.37449999999998) [Dataflow job can use, and the number of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=42.42633333333333) [workers. We'll set this to one, which is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=45.200624999999995) [anywhere the default value. Let's switch](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=47.03374999999998) [over to the Dataflow page on our web](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=49.0552857142857) [console and you can see our Dataflow job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=51.51450000000002) [running. And if you click through, you'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=55.08185714285716) [see all of the stages being executed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=57.173333333333346) [within this Beam pipeline. We've written](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=59.922) [out a few custom logs in our Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=62.447250000000004) [program. We can view these using the Logs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=64.95500000000001)[link on top. The Logs view here only shows](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=68.16509999999998) [you the most recent logs. It's not a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=71.00100000000002) [Detail view. It's useful for quick](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=74.25599999999999) [debugging though. You can click on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=76.88199999999998) [Logs link once again to close this view.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=79.038) [Wait for a little bit and you'll see that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=81.37933333333335) [Dataflow brings up the resources for this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=83.93100000000001) [job. The number of workers have been](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=86.1275) [scaled to one. As usual, you can click on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=88.52966666666669) [a particular stage in this pipeline and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=90.82300000000001) [view how many elements it has processed.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=92.654875) [If you click on the Logs link once again,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=96.188) [you can see information about what each](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=98.47) [worker executes. For a complete list of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=101.07866666666666) [all of the logs generated by our Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=103.91900000000003) [job and the ability to filter and search](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=106.44585714285716) [for specific log messages, you need to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=109.451) [click on the Stackdriver link here. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=111.41587499999997) [will take you to the Stackdriver Login](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=114.33914285714286) [page where you'll have a complete list of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=116.252) [all logs, and here you can see the custom](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=118.5114) [logs that we wrote out from within our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=120.904125)[Dataflow program. You can search for the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=123.989) [term spikey to filter and view only our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=126.27812499999997) [custom logs. If you scroll to the very](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=129.23785714285714) [bottom here, you'll see the one log](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=131.70950000000008) [message that was written out when we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=133.67374999999996) [calculated the top 10 products. Our job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=136.29477777777774) [seems to have completed successfully. You](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=139.1) [can switch back to the main Dataflow page.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=140.92055555555555) [You'll find that one worker was](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=143.325) [provisioned and one worker has now been](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=145.17557142857146) [deallocated. Let's now view the results of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=148.80800000000002) [this execution in our spikey-df-store](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=150.8937142857143) [cloud storage bucket. Within the output](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=153.031) [directory, you'll find a directory named](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=155.6422857142857) [top\_selling. There's just a single file](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=158.89000000000001) [written out here with the top 10 selling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=161.30850000000004) [products and here they are. You can see](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=163.92571428571432) [that the GAGA Canvas Painting is the one that has generated the most revenue.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=6&mode=live&start=166.76733333333328)

[Scaling Number of Workers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live)

[In this demo, we'll see how we can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=1.519) [provision more workers to execute our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=3.263) [Dataflow job and we'll also see how we can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=5.727) [monitor slow stages using graph](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=8.095) [visualization. We are currently within the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=11.498333333333337) [spikey\_revenue folder on our CloudShell](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=13.696499999999999) [terminal window. We'll run the same](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=16.04866666666667) [Dataflow job once again, the one which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=18.487500000000004) [finds the top selling products. All of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=21.041624999999996) [input arguments remain exactly the same as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=23.954571428571427) [in the last example. The only difference](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=26.369285714285713)[here is that we specify num\_workers equal](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=28.611250000000002) [to 10. This will allow Cloud Dataflow to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=31.205) [autoscale up to 10 workers if needed.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=34.71957142857141) [Submit this job to Dataflow and we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=38.307) [switch over to the web console to monitor](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=40.311) [it. As you can see here, our batch](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=43.952777777777804) [processing job is currently executing. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=46.28750000000002) [can click through and view the execution](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=48.523875000000004) [graph and all of the other information.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=50.721) [Our job has just spun up, no resources](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=53.94) [have been provisioned so far. Wait for a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=56.16399999999999) [little bit of time, and after about a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=59.51371428571429) [minute or so, you'll see that Dataflow has](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=61.2542857142857) [now allocated 1 vCPU to this particular](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=63.876857142857155) [job. If you scroll down in the right](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=68.14433333333334) [Navigation pane, you'll see the options](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=70.50849999999998) [that our pipeline was configured with.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=72.24457142857143) [Num\_workers was set to 10. Wait for a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=74.54) [little bit longer and you'll slowly see](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=77.90424999999999) [that the number of workers allocated to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=79.52642857142857) [process this job is slowly moving up. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=82.17666666666666) [have six workers allocated so far and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=86.01342857142858) [Dataflow is trying to bring this up to 10](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=88.77000000000002) [workers. Dataflow determines how much it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=92.50983333333336) [needs to autoscale the number of workers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=94.87128571428572) [based on the current progress of your job.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=97.55914285714289) [Wait for a little bit longer and your job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=101.397) [will complete processing. Observe one](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=103.648) [thing here though, we had eight workers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=106.61819999999999) [provisioned. Dataflow tried to scale it up](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=109.98099999999998) [to 10, but it could not. If you happen to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=112.07000000000002) [be working on a GCP project that you just](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=114.88233333333329) [created, you'll find that this is true for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=117.813) [you as well. If you take a look at the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=119.916) [indication right next to the logs, you'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=122.692) [see that there was a warning generated in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=124.82685714285715) [the logs. Let's see if this gives us a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=126.7985) [clue as to what happened. If you take a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=128.94811111111113) [look at the log messages, you'll find that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=131.24833333333325) [Dataflow encountered some kind of errors](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=133.67271428571436) [while it tried to scale the number of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=135.52971428571428) [workers from 8 to 10. This is because of a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=137.83499999999995) [constraint that Google has just set up.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=141.2868888888889) [New GCP projects have an IP address limit](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=144.006)[per region of 8. Each Dataflow worker is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=147.638) [assigned to a separate virtual machine and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=151.885) [has a separate IP address. Because of this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=154.54) [regional limit of 8, Dataflow was unable](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=158.09871428571427) [to create 10 workers, it stopped at 8 and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=160.6094285714286) [our job ran through with 8 workers. If](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=163.92499999999995) [you're using a GCP project that has been](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=166.79399999999995) [in existence for a while, you might find](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=169.40500000000003) [that you don't encounter this error. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=171.81614285714284) [is interesting to observe and gives you an](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=174.916) [idea of how you can use log messages to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=177.1988) [debug your pipeline. However, this should](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=179.872) [not impact the completion of your job.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=182.09099999999998) [Dataflow scaled to 8 workers, and after](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=185.144) [the job was complete, it scaled down to 0.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=187.19149999999993) [Wait for a little bit longer until all](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=191.011) [resources are freed up and the job will finally be marked as succeeded.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=7&mode=live&start=193.066)

[Identifying Slow Pipeline Stages](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live)

[We worked with the same Python code as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=1.939) [before, which finds the top 10 selling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=3.843) [products. This time, we'll see how we can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=6.340249999999999) [use the execution graph visualization to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=8.248666666666669) [identify slow stages. For this to work, we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=11.691) [need to artificially slow down one of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=15.369499999999999) [execution stages of this pipeline. We'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=17.96699999999999) [add a sleep statement to the process](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=20.266714285714283) [function within the calculate total](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=23.12866666666667) [product revenue class. This stage before](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=25.551499999999997) [it processes any record will sleep for one](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=28.255714285714284) [second. We'll now switch back to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=31.807600000000004) [CloudShell and execute this Python once](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=33.68185714285713) [again with 10 workers. Head over to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=36.666000000000004) [Dataflow page and click through to the job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=40.164500000000004) [that's currently running and let's see](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=42.48942857142858) [what happens. At this point in time, this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=44.858399999999996) [job has been running for about a minute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=47.02599999999998) [and Dataflow is trying to scale up to 10](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=49.790124999999996) [workers. We're still in the process of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=53.73742857142856) [provisioning. None of the pipeline stages](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=55.556571428571424) [have really started yet. We wait a little](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=57.226833333333346) [longer and the first stage seems to have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=60.721) [succeeded. This is the read from our input](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=64.11566666666668) [file and the wall time for this read was 1](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=66.18733333333331) [second. Roughly 20, 000 records were read](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=69.57977777777772) [from the input dataset and added to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=71.8845) [output. The next stage where we extract](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=75.35185714285717) [the product revenue for each line item](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=77.697) [seems to have succeeded as well. It read](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=79.95824999999996) [in roughly 20, 000 records and output 20,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=83.16357142857142) [000 records. The group per product has not](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=87.31175000000002) [completed execution yet, but it seems to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=90.74999999999997) [be running just fine, but the total](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=92.59924999999998) [revenue per product seems to have a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=94.89857142857144) [problem. The wall time for the various](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=97.65566666666668) [stages that we saw so far were under 10](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=99.36174999999997) [seconds, but for this particular stage,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=102.34124999999997) [the wall time is currently at 33 seconds.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=104.128) [The wall time is an indication of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=107.858)[total processing time for this stage. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=110.05088888888889) [is the stage where we added in 1 second of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=113.07900000000001) [sleep for every element that it processed.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=116.2754) [As the graph executes, you'll find that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=119.242) [this wall time constantly increases. It's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=121.3) [now at 4 minutes, 31 seconds and this is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=124.00944444444444) [how you'll identify slow pipeline stages](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=127.56885714285714)[using the wall time. It's pretty clear](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=130.568) [that this job is going to take really long](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=133.24733333333336) [to run through to completion. I'm going to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=135.57175000000004) [forcibly end this job by clicking on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=138.07475000000002) [Stop job button. This will bring up a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=140.75899999999996) [dialog warning you that this job will be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=143.32387500000002) [forcibly stopped. Go ahead and hit Stop](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=146.17614285714288) [Job and stop execution. Your job will be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=149.5381111111111) [cancelled and the number of workers will be scaled down to 0.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=8&mode=live&start=152.52499999999998)

[Integrating Dataflow with BigQuery](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live)

[Writing the results of a Dataflow job out](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=1.787) [to file is useful, but kind of limited.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=4.239) [What you really want is to store these](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=7.011) [results in a data warehouse such as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=9.573) [BigQuery. In this example, we'll see how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=12.462249999999994) [you can integrate Dataflow with BigQuery](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=14.449999999999996) [to store results. BigQuery is the GCP's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=17.217285714285715) [serverless data warehouse. Within big](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=21.180714285714288) [data, choose the BigQuery option, you can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=23.61133333333333) [pin it to the very top if you want to. Big](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=25.551600000000004) [query allows you to store data in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=28.285111111111114)[petabytes scale on the Google Cloud and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=31.222249999999995) [you can query this data in a matter of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=33.33299999999999) [seconds. BigQuery is highly performant.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=36.09842857142857) [Here is the BigQuery web UI and on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=38.882) [left Navigation pane, you can see your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=40.915100000000024) [current project. All of the data in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=43.789999999999985) [BigQuery is stored within datasets which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=45.86674999999998) [are containers which hold tables which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=48.004) [store the actual data. I'm going to create](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=50.94257142857143) [a new dataset within my Spikey Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=54.106444444444456) [project, I'll call it the sales dataset.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=56.367000000000004)[Click on the Create dataset button and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=59.313285714285705) [left Navigation pane will show you that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=62.71385714285715) [your dataset has been created. Select this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=65.37733333333335) [newly created dataset, and within this,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=67.98160000000001) [we'll now create a table that will hold](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=70.09750000000001) [our data. This will be a new empty table](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=72.55119999999998) [within the Spikey Dataflow project. And](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=75.67755555555553) [the dataset which contains this table is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=78.77085714285714) [the sales dataset. I'm going to call this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=81.505) [table the product\_revenue table. BigQuery](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=84.42514285714286) [offers a number of advanced options that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=87.52833333333334) [you can view within the Advanced Options](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=89.9242857142857) [menu. You can partition this table based](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=92.971) [on a timestamp, specify a customer managed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=95.41750000000005)[key, and so on. We'll go with the default](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=98.1803333333333) [options here. Click on Create table and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=101.278) [wait for this table to be created. If you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=103.40812499999997) [select the table on the left Navigation](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=106.40825) [pane, you can view a number of details](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=107.95524999999999) [about this table. Here in the table info,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=110.42375) [observe the table ID. The table ID shows](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=114.28922222222222) [you how you can reference this table](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=117.11549999999998) [outside of BigQuery. Specify the name of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=120.03549999999998) [the project followed by a colon and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=123.07700000000001) [name of the dataset. the table\_name. Let's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=125.67549999999999) [head back to our code editor UI, and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=128.3777142857143) [within the spikey\_revenue folder, create a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=131.4926666666667) [new file. We'll call this file the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=134.27366666666666)[spikey\_top\_selling\_products\_bigquery. py.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=136.677) [The code for this Dataflow job is very](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=140.391) [similar to the code that we've seen before](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=142.905) [when we calculated the top selling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=145.042) [products. There are a few differences that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=146.9277142857143) [I'll highlight here. The first change here](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=148.949) [is the schema that we specify for the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=151.82000000000005) [BigQuery table into which we'll write our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=154.41828571428576) [results. Remember that we did not specify](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=157.36275) [a schema when we created the BigQuery](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=159.6117500000001) [table. We can specify the schema when we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=161.88900000000004) [write data into it. Our table contains two](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=164.34587499999995) [columns, the product name of type string](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=167.901) [and the product revenue of type float. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=170.28)[other change here is that we don't format](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=173.46050000000002) [the final result as a string. Instead, we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=175.59042857142856) [format it as JSON. Extract the product](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=178.99533333333332) [name and the product revenue and code the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=182.25099999999998) [product name to the ASCII format and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=184.92900000000003) [convert the product\_revenue to a float.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=187.70199999999997) [The keys in our JSON format correspond to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=191.787)[the names of our columns in our BigQuery](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=194.035) [table, product\_name and product\_revenue.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=197.011) [We'll use the argument parser to parse](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=202.07) [command-line arguments specifying the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=204.51) [input and the output files. This is as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=206.20680000000004) [before. Instantiate the Beam pipeline with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=209.46328571428575) [the pipeline options specified and apply](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=211.8538)[all of the transforms that we did earlier.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=214.43628571428576) [The one important difference here is that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=218.092) [we don't find the top 10 selling products.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=220.037) [Instead, we calculate the total revenue](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=222.643) [per product and write this out to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=225.12528571428575) [BigQuery. Format the total revenue per](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=228.05300000000005) [product in the JSON format and write to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=230.23512499999995) [BigQuery using beam. io. WriteToBigQuery.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=233.93716666666663) [Because our BigQuery table does not have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=238.569) [its own schema, make sure you specify the schema when you write data to it.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=9&mode=live&start=240.597)

[Writing Results to BigQuery](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live)

[Open up the CloudShell terminal window in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=1.961) [order to execute our job from the command](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=4.189) [line. The one interesting thing here is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=7.24411111111111) [how we specify the output. The output is a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=9.262500000000001) [BigQuery table. Make sure you specify the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=12.344857142857142) [project name: the dataset name. the name](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=14.855833333333333) [of the table. We'll also specify that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=18.942999999999998)[dataflow should use up to four workers to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=21.184142857142852) [execute this job. Let's head over to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=24.174124999999997) [Dataflow page and observe the progress of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=26.900125) [this job. Observing your job and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=30.180666666666674) [monitoring the various pipeline stages](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=32.601333333333336) [should be very familiar to you now. You](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=34.486999999999995) [can see that Dataflow is trying to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=37.14925) [autoscale your job up to four workers.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=39.258) [Wait for a little bit for your job to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=42.485) [complete execution. You can see that the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=44.464) [final stage here is Write to BigQuery and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=46.57700000000001) [it has processed around 12, 000 elements.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=49.26099999999999) [Our job seems to have run through](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=53.344) [successfully. Let's take a look at our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=55.14299999999999) [BigQuery table to see whether results are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=57.428000000000004) [available there. Select the table on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=60.22599999999999) [left Navigation pane and you can see that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=62.522999999999996) [this table is now associated with a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=64.6108) [schema. This is the schema that we have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=67.68975) [specified from within our Dataflow job. If](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=70.13574999999999) [you take a look at the Details tab, you'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=73.34255555555556) [find that this table now has 12, 000](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=75.32277777777777) [records written to it. That sounds about](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=77.881) [right. The Preview pane will give you a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=80.61099999999999) [sampling of some of the products and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=83.13087499999996) [total revenue information available in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=85.39762499999998)[this table. You can use the query editor](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=87.77900000000002) [to execute a query. BigQuery gives you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=90.75380000000003) [some nifty validation of your queries.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=93.1605) [Let's say you misspelled product revenue,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=95.41928571428572) [you'll find that there is a query](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=97.729) [validator that runs and tells you there is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=100.36875) [an error. The error is pretty specific in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=103.002) [this particular case and recognized name](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=105.74199999999999) [product\_revenu. Let's fix our spelling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=109.588) [mistake and run this query to find our top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=111.90099999999998) [selling products in the descending order.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=115.24471428571425) [Hit the Run query button and here is our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=118.211) [GAGA Canvas Painting at the very top of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=121.102) [our list. And on this note, we come to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=124.47240000000002) [very end of this module where we got](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=126.95739999999995) [really hands on with Dataflow. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=129.12675000000002) [constructed pipelines for data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=131.73433333333335) [transformation and executed them on Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=133.68333333333334) [Dataflow. We built and executed this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=136.36285714285714) [pipeline both in Java, as well as Python.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=138.8328888888889) [We performed a number of transformations](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=142.536) [that we executed in parallel using ParDo.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=144.344)[We also studied how to run grouped by,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=148.173) [combine, and other operations. You saw how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=150.795) [we could use the Dataflow web console to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=154.19199999999998) [visualize and debug the various stages](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=156.44285714285715) [within our pipeline. We saw how we could](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=159.31825) [monitor logs in Dataflow and write results](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=161.78924999999995) [out to BigQuery. In the next module, we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=165.08385714285717) [see some variations in the pipelines that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=168.2421428571428) [we execute on Dataflow. We'll study, as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=170.3681666666666) [well as implement windowing operations,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=173.71100000000004) [branching operations, and work with side inputs.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=8cdd5e62-6dec-4b00-84fb-a9495f600845&clip=10&mode=live&start=176.012)

[Creating and Using PCollections and Side Inputs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live)

[Module Overview](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live)

[Hi, and welcome to this module on creating](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=0) [and using PCollections and side inputs in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=4.311) [your pipelines. We've already seen that a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=7.455857142857143) [pipeline is essentially just a directed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=9.808666666666667) [acyclic graph. A pipeline comprises of a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=12.265666666666666) [data source and sink and a series of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=15.344249999999997) [transformations on your data. PCollections](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=17.8765)[form the edges of your pipeline and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=20.065999999999995) [transformations that you apply to modify](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=23.447000000000003) [your data are the nodes of the DAG. In](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=25.55855555555555) [this module, we'll study and implement](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=28.848857142857145) [some advanced operations such as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=31.11333333333333) [windowing, which allows you to perform](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=33.485749999999996) [aggregations on subsets of your data.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=35.53200000000001) [We'll also see how we can perform](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=37.891) [branching operations to split a single](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=39.82825) [pipeline into two or more pipelines, which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=42.09928571428573) [are then processed separately. And](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=44.9908) [finally, we'll see how we can merge](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=47.2825) [together multiple pipelines into one by using side inputs.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=0&mode=live&start=49.05)

[Window Operations and Types of Windows](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live)

[So far, the aggregations that we've](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=2.182) [performed on the input data of our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=3.794) [pipeline have applied to all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=6.289499999999998) [entities in that pipeline. However, data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=8.605999999999998) [is often associated with time and you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=10.757714285714286) [might want to work on subsets of data.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=13.2308) [Let's assume that the X axis is time and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=16.066) [the Y axis represents data streaming into](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=18.832) [our system or just time associated with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=21.61) [each entity in our system. Often the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=24.243) [records that we process using Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=26.82685714285714) [have timestamps associated with them. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=28.96) [might be the case, even if this is a batch](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=32.3196) [processing operation. In such situations,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=35.15166666666667) [you might want more granular information](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=37.78571428571429) [from your data, in which case, you might](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=40.0856) [want to define a window or a subset of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=42.533333333333346) [data on which you perform aggregations. A](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=45.359125)[window can be thought of as a subset of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=49.1039) [streaming data or any data associated with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=51.389) [a timestamp and this window can be based](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=54.246) [on a time interval, the count of entities](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=56.592666666666695) [within that particular window, or a window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=59.852) [can be based on the time interval between](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=62.353) [two entities. This is a Session window.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=64.79866666666666) [Once you've defined the subset of data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=67.873) [using a window, transformations can then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=70.296) [be applied on all entities within a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=72.73885714285716) [window. These transformations typically](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=75.47742857142855) [take the form of grouping and aggregation](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=77.1308) [operations. You might want to find the sum](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=80.87816666666666) [of all entities, the minimum, maximum, or](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=82.99511111111114) [even the average. Windows can be divided](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=86.30774999999997) [into four broad categories. The first of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=89.50899999999999) [these and the simplest to understand is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=91.55900000000001) [the Tumbling window where a time interval](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=93.44657142857143) [slides over your input data in a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=96.5437142857143) [non-overlapping manner. A Sliding window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=99.60749999999999) [is also one with a fixed time interval,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=101.55533333333335) [except that when it moves over your input](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=104.03714285714285) [entities, there is some overlap between](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=107.51985714285713) [consecutive windows. A Count window is not](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=110.80699999999999) [based on a fixed time interval. Once](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=113.98736363636361) [you've received say 10 entities, those 10](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=116.208) [entities make up a window and they may be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=118.39611111111111) [received at any point in time. A Session](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=121.1685) [window is defined by the time interval](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=123.67075000000001) [between two entities. A session is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=125.73225000000001) [considered to be ended. A Session window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=128.30357142857144) [opens up and you start receiving entities](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=130.38262499999996) [or elements, it's considered closed once](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=132.84449999999998) [there's a certain interval between two](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=135.773) [entities. That interval could be anything](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=138.3568)[that's something that you define. Maybe](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=140.221) [the interval is one minute, then the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=141.84216666666666) [session is closed. And finally, we have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=144.4128) [the Global window which includes all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=146.8678571428571) [the entities in the batch or streaming](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=148.8495) [data. In order to understand windowing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=151.85875000000001) [better, let's quickly understand how the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=154.07866666666663) [Tumbling, Sliding, and Global windows](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=156.00616666666664) [work. You have a number of entities that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=158.8312) [you're operating on using Dataflow. These](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=160.85262500000002) [entities are basically elements in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=163.9467142857143)[PCollection. These entities can be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=166.46528571428578) [streaming data, in which case, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=168.61966666666666) [PCollection is unbounded or batch data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=170.89528571428576) [associated with timestamps. When you use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=173.86350000000002) [tumbling windows, imagine that you're](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=176.6428) [dividing the input data into fixed window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=178.5037142857143) [sizes based on a time interval. An](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=181.84566666666666) [important characteristic of the tumbling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=185.50983333333332) [window is the fact that the time intervals](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=187.234) [are non-overlapping, which means a single](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=189.878) [entity can belong to just one window. In](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=192.344) [addition, the number of entities might be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=196.11314285714286) [different within a window. This, of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=198.6356) [course, depends on the timestamp](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=200.34600000000003) [associated with that entity. A Tumbling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=201.666) [window is so called because the window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=204.368) [tumbles over the input data in a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=206.339125) [non-overlapping manner. The data available](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=208.965) [to us is ordered by timestamp. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=212.23066666666668) [Tumbling window simply extracts subsets of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=214.74416666666667) [this data in a non-overlapping fashion.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=217.6237142857143) [Once you have a subset of data defined by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=221.216) [a window, you'd perform some kind of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=223.3237999999999) [aggregation operation on these entities.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=225.129) [Let's assume that all of the entities that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=227.982) [we are working with are integers. You](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=229.648) [might want to apply the sum operation on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=231.821125) [each window. So we have the first window,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=233.999875) [the sum of these numbers is 14. Then we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=236.482) [have the second window, the sum of these](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=239.096) [numbers is 7. We have the third window,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=241.11257142857147) [the sum of these numbers is 19. And](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=243.3395555555556) [finally, the very last window here where](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=245.59350000000006) [the sum of these numbers is 21. Sliding](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=247.6884285714286) [windows are very similar to tumbling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=251.20683333333332) [windows with one crucial difference. They](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=253.6445) [do have a fixed time interval which forms](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=256.1945) [the size of the window. However, sliding](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=258.6565) [windows slide over the input data with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=260.8776666666667) [overlapping time. That is the sliding](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=263.8066) [interval. The number of entities in each](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=266.3954) [window might be different. Since time is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=268.41357142857146) [overlapping the sliding interval, an](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=270.959) [entity might belong to more than one](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=273.86811111111115) [window. Imagine this data associated with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=276.535888888889) [timestamps and we have a window moving](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=278.795) [over this data and notice that there are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=281.42787500000003)[some entities that are overlapping between](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=283.913625) [consecutive windows. We'll use the same](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=287.44) [integer entities example as before and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=290.64314285714283) [apply the sum operation on each window,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=292.86212500000005) [this time it's a sliding window. For the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=295.416) [first window, the entity sum up to 14, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=298.0433333333333) [window then slides over and the sum of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=300.5516666666667) [these entities is equal to 8. Observe that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=303.65500000000003) [the integer 3 was part of the previous](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=306.2718888888889) [window and is part of the current window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=308.47475000000003) [as well. The window then slides over once](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=310.62714285714287) [again and we get a sum of 13. And this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=313.94179999999994) [sliding continues over all of the entities](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=317.567) [that we have within our PCollection.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=319.6424285714286) [Observe that as the window slides over,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=321.835) [there is some overlapping time. That's the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=324.593) [main difference between sliding and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=327.5023333333333) [tumbling windows. And finally, the Global](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=329.43699999999995) [window is one that's very easy to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=332.25185714285715) [understand. All of the data in this stream](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=334.9544000000001) [or our timestamped data is simply in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=337.06311111111097) [one Global window. Aggregations apply to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=340.041) [all of the data, and in some manner, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=343.544) [Global window is what we've been working with so far.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=1&mode=live&start=346.1957142857143)

[Windowing Operations](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live)

[Now that we've seen how windowing works,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=2.008) [let's see how we can perform](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=3.956) [transformations and aggregations using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=5.186) [window operations in Dataflow. We start](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=7.732) [off in the GCP web console's cloud storage](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=10.158666666666665) [buckets page. Spikey-df-store is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=13.183) [bucket that we had created earlier. Within](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=15.093000000000002) [this, we have the data folder, and within](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=17.407857142857143) [here, we're going to upload some new files](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=19.9765) [to process in this module. I've modified](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=22.095857142857145) [the format of this data to be slightly](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=24.518555555555555) [different based on our current use case.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=26.505875) [The first column represents an ID for that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=29.011) [particular transaction, then we have the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=31.388) [timestamp, product, price, and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=33.246) [quantity sold. We'll now switch over to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=36.11171428571428) [our CloudShell terminal window and create](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=38.60442857142858) [a directory under our home directory](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=40.655)[called spikey\_sales. Cd into this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=42.512) [directory. All of the code that we write](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=45.847500000000004) [in this module will be within this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=47.50888888888888) [directory. Switch over to code editor on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=50.315)[your browser window, right-click on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=52.491999999999976) [spikey\_sales folder, and create a new](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=54.45280000000001) [file. We'll name this file](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=57.26866666666666) [spikey\_hourly\_sales. py. We have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=60.45599999999999) [timestamps associated with each of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=62.30328571428571) [transactions in our input data. We want to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=63.97866666666667) [track the number of products sold for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=67.15462500000001) [every hour of the sale. This will allow us](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=69.77900000000001) [to track whether our sales perform better](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=72.8815) [at certain time intervals or when certain](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=74.72299999999998) [offers are applied, for instance. I'm](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=76.81433333333331) [going to import the print\_function so that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=79.74657142857143) [my Python 2 print statements are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=82.55699999999999) [compatible with the Python 3 syntax. Set](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=84.639) [up the other imports for the Apache Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=87.43962499999999)[modules and other Python modules as you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=89.49785714285714) [see on screen. This Dataflow job works on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=92.026375) [batch data, however, each of our records](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=95.546125) [is associated with a timestamp, which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=97.969) [we're going to extract, then use this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=100.2435) [timestamp to perform windowing operations.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=102.466) [We know the exact structure of each line](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=105.582) [of the text records that we read in. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=108.38555555555558) [use a regular expression to parse out the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=111.542375) [portion which contains the timestamp](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=114.08000000000001) [information. The timestamp data will be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=116.67500000000004) [available in the form of a string. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=118.41774999999998) [simply parse it to the datetime format and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=120.65866666666666) [convert it to an integer. We'll have a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=123.57155555555553) [processing step in our pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=126.5914285714286) [implemented using a do function which will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=128.28266666666667) [extract timestamps from the input records](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=131.59333333333336) [and then emit a timestamp value. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=134.15500000000003)[processing step will return a beam.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=137.73883333333333) [window. TimestampedValue element. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=141.03075) [contains the original text record that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=144.63042857142858) [processed in the element field and we've](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=147.193125) [extracted the timestamp and associated it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=149.742) [with that element. Once we've performed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=151.882) [the aggregations that we are interested in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=154.71)[within a window of data, the PrintWindowFn](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=156.606) [function will simply print out the result](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=159.17280000000002) [of this window aggregation to a string](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=161.73471428571435) [format. Observe that the process function](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=164.4606) [that acts on the windowed elements takes](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=166.8248571428572) [in a window param. We'll extract the start](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=169.90744444444448) [time and the end time of the window that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=172.63149999999996) [we just operated on, we'll print it out to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=175.36862500000007) [screen, this is only for debugging](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=177.7184285714286) [purposes, you can eliminate this line if](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=179.36066666666662) [you want to. We'll return the range of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=181.19133333333335) [window defined by the start time and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=183.6755714285715) [end time in some kind of readable format](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=185.54377777777788) [and output the actual element itself.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=188.06450000000004) [We'll get the input file to process and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=191.11) [the output file to write data into using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=192.77755555555552) [command-line arguments. Set up the options](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=195.63) [to configure our pipeline, and then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=199.3927142857143)[instantiate the Beam pipeline using a with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=201.41500000000002) [statement. This is a much better and more](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=204.85279999999995) [readable way to write your pipeline code.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=207.50375000000003) [The very first step is to read all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=210.212) [text records from our input text file. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=212.228) [then perform a windowing operation which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=216.10185714285714) [slices all of our input data into hourly](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=218.561)[intervals. We want to sum up the total](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=222.39655555555564) [number of products sold for every hour of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=224.19044444444438) [sales and the overlapping interval between](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=226.85499999999993) [2 windows is 30 minutes. Windowing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=229.12039999999996) [operations only work with data that is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=231.95466666666667) [associated with a timestamp. So we first](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=234.37933333333334) [extract the timestamp and associate it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=236.82285714285712) [with each element. The timestamp is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=239.202) [embedded into each record, we simply](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=241.71128571428574) [extract it. We then apply a window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=244.13950000000003) [operation into this timestamp data by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=246.65174999999994)[calling beam. WindowInto. This window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=249.62699999999995) [operation groups our data into hourly](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=252.4265714285714) [intervals and the sliding interval is 30](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=254.72533333333334) [minutes. We haven't applied the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=257.4882857142857) [aggregation to find the total count of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=259.304) [products yet, but we will in the very next](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=261.51) [step. Notice that the sliding window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=264.4250999999998) [requires two intervals, the window](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=266.6154285714286) [interval and the sliding interval, and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=268.4175) [they are specified in seconds. Now that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=270.641875) [have divided our data into windows, we can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=273.33662499999997) [now apply an aggregation within a window.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=275.75442857142866) [The aggregation that we choose here is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=278.1032857142857) [CountCombineFn that simply counts the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=281.27166666666665) [number of entities in a particular window.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=283.437) [This is one of the built-in default](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=286.134) [combiners available in Beam. Once we have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=288.527) [the count of the number of products](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=291.4126666666667) [associated with each hourly window, we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=293.4508) [format this information to a text format](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=296.2562857142857) [using the PrintWindowFn that we defined earlier and write it out to a file.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=2&mode=live&start=299.0037142857143)

[Executing and Monitoring Windowed Pipelines](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live)

[With our code fully defined, we can switch](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=1.688) [over to CloudShell to execute this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=3.562) [Dataflow job. We need a few libraries](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=5.898250000000001) [installed on our machine though before we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=7.9257500000000025) [can run this code. We'll install the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=9.715) [python-dateutil package first. Once this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=12.5595) [package has been installed successfully,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=14.894000000000002) [we are ready to run our Dataflow code,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=16.25542857142857) [which we do in a manner that should be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=18.678888888888892) [very familiar to us. Make sure you specify](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=21.625666666666664) [the input and the output command-line](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=24.212500000000002) [arguments, the text file from where we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=26.427999999999997) [read in our data, and the text file where](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=28.173000000000002) [we want to write out our results. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=30.21500000000001) [remaining arguments are all of those that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=33.46828571428571) [we've seen earlier. Make sure you specify](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=35.55425) [that the runner is the Dataflow runners](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=37.45700000000001) [that we run it on Cloud Dataflow. After](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=39.24757142857142)[submitting this job, we'll use the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=42.53433333333334) [Navigation menu to go to the Cloud](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=44.664444444444435) [Dataflow page in order to monitor the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=46.753499999999995) [progress of this job. And here is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=49.3494)[execution graph displaying all of our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=52.302428571428564) [pipeline stages. The new interesting](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=54.5592) [operation that we've performed here is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=56.3075) [windowing. Expand the WindowCount](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=59.147) [operation and you'll find that it's made](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=60.68150000000001) [up of many substages. You can see](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=63.486500000000014) [internally that windowing involves](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=66.4685) [associating some kind of window key with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=68.10566666666666) [all of the entities or elements that are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=70.75525) [present within a window. Our job has been](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=73.27333333333333) [running for a minute or so now and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=75.65520000000001) [Dataflow is slowly scaling up to four](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=77.6525) [workers to process our data. Wait a couple](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=79.954) [of minutes for the job to run through](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=82.8032) [successfully and let's take a look at each](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=84.297) [pipeline stage. The first is the create](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=86.76650000000001) [staqe, which has processed roughly 2300](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=89.19699999999999) [elements. This new dataset that we used is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=91.87000000000002) [a little smaller. We then extracted an](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=94.71280000000004) [associated timestamps with each element.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=97.70999999999998) [The window in operations apply to all 2300](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=100.02450000000002) [elements. Observe the very last stage of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=104.46133333333336)[the windowing operation though. The output](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=106.70828571428575) [here is just 34 records. The output that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=109.42585714285715) [we have finally is for 34 overlapping](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=113.23266666666667) [hourly ranges. Because the 60-minute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=116.31366666666668) [ranges overlap by 30 minutes, we finally](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=119.49100000000001) [get 68 records at our output. Sixty-eight](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=122.61528571428573) [records were written out to our text file.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=125.87966666666665) [Use the Navigation menu to head over to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=129.195) [cloud storage buckets where we can view](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=131.397) [the final result of this Dataflow job.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=133.58187500000005) [Within spikey-df-store, we have the output](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=136.128) [folder, and within this, the hourly\_sales](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=139.055) [folder contains the information that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=142.152) [are looking for. Click to open this file,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=144.038) [and here is the number of products that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=147.336) [were sold for every hour over the week.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=149.19) [You can see from the patterns here that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=152.868) [every hour SpikeySales sells between](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=154.584)[120-160 products, except in a few cases.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=157.617) [These particular results bear further](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=161.859) [investigation. Maybe this hour is at night](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=164.67650000000003) [where not many sales occur. Maybe there is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=167.15175) [some problem with the reporting of this data. It could be anything.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=3&mode=live&start=169.79814285714286)

[Branching Operations](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live)

[In this demo, we'll see how we can perform](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=1.941) [branching operations with Dataflow where](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=4.044) [we split a single pipeline into multiple](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=6.402333333333333) [pipelines with different data which are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=9.428) [then processed separately. Here is a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=11.205) [visualization of your Apache Beam pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=14.426624999999998) [that we are familiar with. You're using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=16.22) [this to process data. It's often the case](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=18.47733333333333) [that one particular stage of the pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=21.07885714285715) [will have multiple outputs, which are then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=23.78114285714286) [processed by different pipelines. This is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=26.6804) [a branching operation. The data has been](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=30.123399999999997) [split after this particular pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=32.45357142857143) [stage. Two separate pipelines have now](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=34.43333333333333) [been created to process this split data.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=36.804833333333335) [Different transformations will apply after](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=40.298) [the data has been split. We'll now use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=42.298) [branching operations to operate on our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=45.073499999999996)[weekly sales data to find the top selling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=47.206999999999994) [products and the bottom selling products.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=49.453) [We start off in GCP's code editor, and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=52.388) [under the spikey\_sales folder, we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=55.278999999999996) [create a new file in order to write our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=57.19244444444444) [Python code, this file will be called the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=59.888875) [spikey\_top\_bottom\_selling\_products. py.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=63.214000000000006) [I'm going to paste in all of the code and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=65.876) [walk through what exactly this pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=67.647) [does. Here are the imports as before. All](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=69.63985714285714) [of this should be pretty self-explanatory.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=72.24871428571429) [Here is the pipeline stage that performs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=74.809) [the actual splitting of the records that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=77.164) [it reads in. I've called this class the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=79.554) [split items based on sales function and it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=82.181) [extends the do function. When we split the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=85.692) [input data, we can think of this as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=88.56255555555558) [creating two separate pipelines. We're](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=90.91971428571428) [going to associate different tags with the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=92.93866666666666) [data as we split them. The top sellers and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=95.37057142857144) [the poor sellers are both different tags.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=97.92449999999998) [After we read in every record, we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=100.774) [extract the name of the product and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=102.835125) [quantity of the product that was sold. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=105.1385) [are working with a tab separated text file](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=107.906) [here so each record has to be split on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=110.06277777777778) [tab. We'll keep our logic for splitting](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=112.74624999999999) [data very simple. We have a quantity](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=115.17975) [associated with each product that was](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=118.01757142857144)[sold. If the quantity is greater than or](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=120.56800000000001) [equal to 90, that's a top selling product.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=122.70111111111112) [If the quantity is less than five, that's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=125.322) [a poor seller. We process each of these](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=127.8205) [elements and produce a tag output. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=130.7272857142857) [output is either tagged as a top seller or](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=133.898) [a poor seller. Pvalue. TaggedOutput allows](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=136.68949999999998) [us to split and mark our output so that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=140.39366666666666) [they can then be processed separately. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=143.25357142857143) [rest of the code to parse the pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=145.92000000000002) [options from command-line arguments and to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=147.736) [instantiate the pipeline is the same as](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=149.938) [before. Let's go directly to how we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=152.24100000000004) [process the data in our pipeline stages.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=154.67375000000004) [We read the lines of the input dataset in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=158.141) [the string format using ReadFromText and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=160.191) [then we pass it into our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=163.147) [SplitItemsBasedOnSalesFn. When a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=166.8498571428571) [particular pipeline stage performs a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=168.08985714285714) [branching operation, you need to specify](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=169.647) [the two different outputs of this pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=172.9) [stage. We've clearly indicated here that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=175.83912499999994) [this stage has two tagged outputs, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=178.07483333333334) [top\_sellers and the poor\_sellers. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=181.6576) [PCollections containing the top selling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=184.322) [items and the poor selling items can now](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=186.052) [be accessed as though from a dictionary.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=189.09555555555548) [You can use the tags to extract the top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=191.979) [sellers PCollection, as well as the poor](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=194.388) [sellers PCollection from](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=197.106) [split\_items\_result. You could choose to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=199.86599999999996) [process each individual PCollection.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=202.07800000000003) [Further, we are going to keep things](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=203.64) [simple here and simply write them out to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=205.0155000000001) [separate files. The top selling products](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=208.35577777777775) [will be in a file prefixed by top\_ sellers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=210.19311111111105) [and the poor selling products will be in a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=212.68757142857137)[file prefixed by poor\_sellers. Our code is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=215.57314285714284) [now complete. We can now switch over to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=218.6648) [CloudShell to execute this pipeline. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=220.83575) [actual execution is very straightforward,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=223.17600000000002) [exactly the same as before. Once we've](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=225.48333333333335) [submitted this job to Beam, let's head](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=228.18942857142855) [over to the Dataflow page and monitor the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=229.99175)[status of this job. And here is our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=232.45266666666666) [pipeline. Observe how our execution graph](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=235.84680000000003) [clearly indicates that we split our data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=238.68216666666672) [into two pipelines. The top selling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=242.86499999999998) [products form a different PCollection and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=245.67871428571425) [can be processed differently. We write it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=247.917) [out to a different file and the same is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=250.61933333333334) [true for our bottom selling products. Our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=253.3146666666667) [workers are still in the process of being](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=255.73457142857143) [provisioned. Dataflow is slowly scaling up](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=257.87399999999997) [to around 5 workers. Wait for a little](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=260.39933333333335) [longer and let's examine the individual](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=263.26949999999994) [stages. You can see that we read about](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=265.19166666666666) [2300 elements from the input file, that's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=267.746625) [about right. Select the stage which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=270.494) [performs the actual branching operation.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=272.30266666666665) [You can see that it has one input](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=274.667) [PCollection and two output PCollections.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=276.786) [At the inputted process, roughly 2300](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=280.181) [records or entities, around 91 of these](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=282.9918) [elements went into poor selling products,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=286.609) [and about 229 elements went into top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=289.37800000000004)[selling products. The remaining elements](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=293.5262000000001) [were all in between and were not included](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=295.5151428571428) [in any output. Click on the stage which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=298.5717142857142) [writes out the top selling products. You](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=301.2813749999999) [can see that at its input, it receives 229](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=303.1345714285714) [entities and the stage which writes out](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=306.12942857142866) [the poor sellers received 91 entities.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=309.32849999999996) [We'll now head over to cloud storage to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=312.538) [view the results written out to file. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=314.597) [go to the spikey-df-store bucket, within](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=317.2778333333333) [that, we go to the output folder, and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=319.418)[within that, we go to the top. bottom](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=321.9856) [folder. All of the top selling products](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=324.8303999999998) [have been written out to files prefixed by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=326.66275) [top\_sellers. Let's open one of these and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=329.86014285714293) [let's take a look at some of our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=331.6703333333332) [best-selling products. We can also observe](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=334.5842857142857) [some of the products that have been doing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=336.50750000000005) [poorly. This information can be used](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=339.41540000000003) [further by the analysts at SpikeySales in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=341.3237500000001) [order to maybe offer some discounts on these poor sellers.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=4&mode=live&start=343.80871428571425)

[Side Inputs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live)

[In this demo, we'll see how we can build](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=2.044) [and execute Apache Beam pipelines using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=3.851) [side inputs. Side inputs are used when we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=6.375857142857143) [want multiple streams of data to be fed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=9.63722222222222) [into the same processing stage. The offer](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=11.997) [related data is fairly simple. It has just](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=14.734142857142858) [three columns of information. The first is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=17.246142857142857)[the product ID to which this offer](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=19.543) [applies, the name of the offer, and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=21.552142857142858) [discount offered on the price of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=24.080428571428573) [product. We'll now switch over to code](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=25.991) [editor beta on the GCP. We are working](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=28.095) [within the SpikeySales folder. We'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=30.692428571428575) [create a new file here. This is a Python](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=32.233666666666664) [file called spikey\_top\_sellers\_offers. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=35.153800000000004) [code that we are going to look at which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=38.9662) [uses side inputs will check how many of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=40.426) [our top selling products have offers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=43.4428888888889) [applied on them. Here is the code pasted](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=46.3575) [in. I've also made code editor full screen](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=49.37750000000001) [so that it's easy for us to view all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=51.588875) [the code. Let's now walk through what's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=54.222454545454575) [happening here step by step. We have a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=55.91866666666666) [class named top\_selling\_products\_fn, which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=58.655888888888896) [extends Beam's do function. This simply](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=61.7614)[processes every record and extracts only](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=64.49) [the top selling products. Split every](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=67.37433333333333) [record on the tab, extract the product ID,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=70.039) [the name of the product, and the quantity](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=72.6305) [of the product that was sold. If the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=74.97742857142858) [quantity of product sold was greater than](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=77.55687499999999) [90, we'll categorize this as a top seller](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=80.26939999999998) [and we'll output the product ID and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=83.02674999999998) [product name. The next do function](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=86.4378888888889) [processes offer related data. It gets all](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=88.60419999999998) [of the product IDs which currently have an](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=91.04925000000001) [offer applied. This extracts all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=94.85524999999998) [product IDs from the offers text file. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=97.16050000000004) [now have a function which checks whether a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=100.642875) [particular top selling product has an](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=103.12657142857144) [offer applied to it. This is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=105.434) [match\_ids\_fn that we'll use within our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=108.52114285714286) [transform and its input arguments are a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=110.72466666666666) [single top selling product and the list of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=113.54100000000001) [products that have offers applied. These](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=116.76588888888887) [are the discounted items. Top\_seller is a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=119.1704) [tuple which contain the product ID and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=122.20057142857142) [product name. If the product ID of a top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=125.0573333333333) [selling product is present in the list of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=128.32755555555553) [discounted product IDs, that means there](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=131.2757142857143) [is an offer applied to this product. If a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=133.48899999999998) [particular product has an offer on it,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=136.9512222222222) [we'll return the name of the product. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=138.674) [one difference to the input files that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=142.13637500000002) [pass to this job is the fact that there](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=144.3685) [are two input files, the actual products](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=146.227) [that were sold and the offers applied to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=149.303) [products. Input1 refers to the sales data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=152.47188888888888)[and input2 refers to the offers data.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=155.34000000000003) [Instantiate the pipeline and configure it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=158.282) [using our command-line arguments. All of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=160.311) [the interesting stuff comes in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=162.92975) [transforms that we apply in the pipeline.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=164.564) [We'll read from the two input text files](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=168.002) [that we specified. The sales related data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=170.596) [which is the sold\_items PCollection, and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=173.883) [offer related data, which is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=176.69385714285713) [offer\_applied\_items PCollection. We then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=178.445) [applied the filter operation to find the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=181.53057142857142) [top selling products, products, which have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=183.6464) [a quantity sold of over 90. We then apply](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=185.778) [our transformation on the second data set](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=188.89557142857151) [on the offer\_applied\_items and we extract](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=191.45414285714284) [the product IDs of all of the products](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=194.55050000000003) [that have some kind of discount applied.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=197.32724999999996) [This is now stored in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=200.046)[offer\_applied\_ids PCollection. And this is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=201.39) [the stage in the pipeline where we use the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=204.73833333333334) [side input. We want to find those top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=207.34200000000004) [selling products which have some kind of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=210.42700000000002) [offer applied or some kind of discount.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=212.74699999999999) [The main pipeline operation is on the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=215.13) [top\_selling\_products PCollection. In](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=217.225) [addition, we pass in data from the other](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=220.28255555555555) [offer related pipeline in our flatmap](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=222.783) [operation. The match IDs function takes in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=226.09328571428577) [every top selling product, it also takes](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=229.00150000000005) [in the offer\_applied\_ids. The product ids](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=231.381) [on which discounts or offers are available](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=234.72000000000003) [are passed in as a side input to the main](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=237.364625) [pipeline which processes the top selling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=240.71828571428566) [products. We need to convert the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=243.30919999999998) [offer\_applied\_ids, which is a PCollection,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=245.54199999999994) [to a list before we pass this in. This is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=246.9685555555555) [why it's important to ensure that when you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=249.93237499999995) [use side inputs all of the data in your](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=252.57449999999994) [side input can fit in the memory of one](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=255.19259999999997) [worker. We now have a collection of all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=258.8543333333333)[the top selling products on which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=261.23260000000016) [discounts were applied. We can now simply](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=263.3962857142857) [write it out to a text file with the sold-on-discount prefix.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=5&mode=live&start=265.50999999999993)

[Executing and Monitoring Pipelines with Side Inputs](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live)

[With our code complete, we can now turn](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=1.045) [our attention to submitting this job to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=3.1641111111111107) [Dataflow. We are working in our CloudShell](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=6.5867142857142875) [terminal window and all of the arguments](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=8.535) [on the command line should be familiar to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=10.493250000000002) [us. The one interesting thing here is that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=12.923571428571424) [we have two input files. Input1 refers to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=14.851749999999996)[the sales of products and input2 are the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=18.74675) [offers on those products. Submit this job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=22.004428571428566) [to Dataflow and let's head over to our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=24.646) [Dataflow web page in order to monitor the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=26.807999999999996) [progress of this job. And here is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=29.330000000000002) [visualization for the execution graph for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=32.333999999999996) [this Dataflow job. Observe how two](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=34.97449999999999) [pipelines merge into one because of our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=38.04014285714285) [use of side inputs. The first pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=41.019000000000005) [processes the products that were sold. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=43.747) [read in those products and extract the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=46.10199999999999) [product ids and product names. The second](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=48.585) [pipeline processes products which have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=51.633399999999995) [offers applied. Observe that the data from](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=54.407) [both of these branches are then fed into](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=56.79862499999999) [that stage which calculates those top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=59.740750000000006) [selling products on which discounts were](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=62.205999999999996) [applied. Dataflow auto scales the number](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=64.80324999999999) [of workers assigned to this job to five](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=67.13428571428571) [and slowly our job will run through to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=69.73925) [completion. You can see here that we read](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=72.83414285714287) [in roughly 2300 products from the input](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=74.44611111111112) [file and only 229 of these were top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=76.96524999999998) [selling products. And on this branch of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=80.88814285714284) [the pipeline, we processed roughly offers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=83.17785714285712) [on 1, 000 products. This stage where we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=85.81949999999999) [determine which of our top selling](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=89.09750000000003) [products had discounts applied takes in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=91.113) [two input collections, the products](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=93.7134)[themselves and also the offers. You can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=95.85285714285713) [see from the output at this stage that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=98.54322222222223) [around 97 of our top selling products had](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=100.587375) [discounts applied. Once our Dataflow job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=103.894) [runs through to completion, we can view](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=106.55571428571426) [the results written out to cloud storage](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=108.55528571428573) [buckets. Use the Navigation menu, head](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=111.21299999999998) [over to storage, go to spikey-df-store,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=112.95885714285717) [output, and let's find which of these top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=116.60960000000001) [selling products had offers applied. Open](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=118.87800000000003) [up any of these files and here are the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=121.9055) [products which were discounted and were](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=124.151) [top sellers. And on this note, we come to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=127.33325000000004) [the very end of this module where we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=129.63020000000006) [studied some advanced operations on Beam](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=131.77071428571432) [pipelines. We saw how pipelines are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=134.4326) [essentially directed acyclic graphs where](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=136.9236666666667) [PCollections are edges of our graphs and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=139.3074) [transformations form the nodes of our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=142.1405714285714) [graph. We saw how timestamp data could be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=144.51833333333337) [used in windowing operations where](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=146.83711111111108) [aggregations can be performed on subsets](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=149.0684) [of data. We also saw pipelines which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=152.3) [performed branching operations where a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=155.09583333333327) [particular stage split the data into](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=157.2112)[multiple branches which could then be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=159.59371428571433) [processed separately. We also saw how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=162.274) [multiple branches in a pipeline could be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=164.30700000000002) [merged into a single stage using side](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=166.40714285714284) [inputs. So far, we've seen a number of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=169.668) [different Dataflow examples, all of which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=172.07912499999995) [involved writing code. In the next module,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=174.008) [we'll see how we can work with Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=176.5086666666666) [without writing any code at all by using predefined Dataflow templates.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=06c0d589-1463-4aca-9b68-b4742980e6b5&clip=6&mode=live&start=178.64299999999997)

[Creating Pipelines from Google Templates](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live)

[Module Overview](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live)

[Hi, and welcome to this module where we'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=0) [see how we can create and execute Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=4.479) [pipelines using predefined templates that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=7.239) [Google makes available to us. It's often](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=10.069) [the case that you have data available in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=12.7745) [some kind of source and you want to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=14.4695) [perform some kind of standard processing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=16.310142857142857)[on this data and store it to a different](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=18.070875) [sink. If you have to write code for all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=20.725749999999998) [these operations, that can get pretty](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=23.606333333333335) [tedious, which is why Google offers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=25.5075) [predefined templates for ease of pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=27.807000000000002) [creation. In this module, we'll take a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=31.069000000000003) [look at some of the predefined templates](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=32.70333333333335) [that are available to us and see how we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=34.44900000000001) [can create and deploy pipelines from these](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=37.10966666666667) [templates. We'll use a Dataflow template](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=40.550166666666655) [that reads in streaming data from the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=42.62387500000001)[Pub/Sub service and then writes it out to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=45.754142857142845) [BigQuery. We'll do all of this without writing a single line of code.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=0&mode=live&start=49.454333333333345)

[Introducing Dataflow Templates](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live)

[Dataflow on the GCP is an extremely](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=2.413) [powerful service for big data processing,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=4.922) [but it has one major drawback. Dataflow is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=7.01) [very hard for non-developers to use.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=10.53275) [Unless you program well in a high-level](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=13.016) [programming language such as Java or](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=15.578) [Python, you can't use Dataflow at all. And](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=17.034) [this is where Dataflow templates come in.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=20.2425) [Templates allow non-developers to harness](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=22.983) [the part of Dataflow for common processing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=25.791) [tasks. Even if you're familiar with big](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=28.787000000000006) [data processing, it's pretty obvious that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=30.7826) [Dataflow can be fairly intimidating for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=32.749857142857145) [beginner developers. You need dependencies](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=35.87619999999999) [to be set up, code to be written, code is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=38.11442857142857) [being processed behind the scenes on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=40.91400000000001) [multiple workers. All of this needs to be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=43.180333333333344) [taken into account when you develop and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=45.38100000000001) [execute your pipelines. Well, if you can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=47.867000000000004) [develop in a high-level programming](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=50.483000000000004) [language, you can still use Dataflow using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=52.11533333333334) [Dataflow templates. Google has provided a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=55.623000000000005) [number of predefined templates which are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=58.85942857142857) [available for everyone to use. These](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=60.83119999999999) [templates define common processing tasks](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=62.96133333333332) [and typically read data from a source,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=65.126) [perform simple processing, and write it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=67.855) [out to a sink. The code for these Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=69.51442857142858) [jobs have been written by Google](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=72.845) [engineers. You don't have to worry about](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=74.82671428571425) [recompiling the code, you're completely](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=76.98014285714284) [abstracted away from it. You simply use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=79.09440000000002) [runtime parameters to customize the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=81.65814285714285) [execution of that job. Dataflow splits up](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=83.7324) [the execution of the job from the actual](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=87.07214285714282) [job definition. As a user of Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=90.9177777777778) [templates, you don't have to configure](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=93.58133333333332) [dependencies or even write any code.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=95.22771428571428) [You'll have developers who are responsible](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=97.364) [for writing and testing the code and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=99.759) [defining these templates. Non-developers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=102.013) [can then use these templates without](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=104.10533333333333) [writing any code. When you're writing code](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=107.00075000000001) [to execute your own custom Dataflow job,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=109.61342857142854) [all of the steps need to be performed by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=111.58850000000001) [developers. You configure dependencies](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=114.40499999999997) [done by the developer, you write code for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=115.77600000000001) [your pipeline either in Python or Java,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=118.54666666666667) [that's also developer's job, and finally,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=121.069) [you execute your pipeline, once again,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=123.9) [done by developers. When you use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=126.71079999999999) [Dataflow's templates, the first three](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=129.55300000000003) [steps are still done by developers.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=131.27733333333336) [Developers simply create a template by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=133.054) [executing the pipeline. These templates](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=135.825) [that have been created by developers are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=138.12642857142856) [then typically stored on cloud storage](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=140.32033333333334) [buckets. Once the template has been](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=142.88633333333334) [created, the pipeline can be executed by](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=145.22899999999998) [any non-developer. Currently, all of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=148.58525) [templates available on Google perform data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=150.94233333333332) [transfers from one GCP service to another,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=153.46999999999997) [Bigtable to Cloud Storage, Pub/Sub to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=156.04228571428575) [BigQuery, Cloud Storage to Pub/Sub, and so](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=158.64499999999998) [on. There are many more intra-GCP](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=161.26542857142857) [transfers available for which you can use these Dataflow templates.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=1&mode=live&start=163.74283333333332)

[Introducing Pub/Sub](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live)

[In the Dataflow template that we'll use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=2.608) [later on in this module, we'll work with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=4.285) [pub/sub. Pub/Sub is GPC's serverless](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=7.766777777777779) [streaming message delivery service.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=10.369200000000001) [Pub/Sub, which stands for publisher](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=12.2514) [subscriber can be thought of as messaging](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=14.499500000000001) [middleware. This is responsible for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=17.470666666666666) [guaranteed delivery of messages from](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=19.826666666666668) [publishers who publish messages to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=21.927000000000003) [subscribers who are interested in specific](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=24.13533333333333) [messages on specific topics. Pub/sub is a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=26.866000000000003)[distributed messaging delivery system](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=30.127) [which can operate at scale. It offers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=31.945333333333334) [many-to-many asynchronous messaging and it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=35.362) [completely decouples the sender of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=37.888000000000005) [message from the receiver. This means that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=40.407714285714285) [the rates at which senders send messages](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=42.84900000000001) [need not match the rates at which the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=44.8645)[messages are consumed. Here is the general](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=48.0156) [structure of all of the components that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=50.4441111111111) [make up the pub/sub message delivery](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=52.29966666666667) [system. Publishers are responsible for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=55.092333333333336) [publishing messages and they can be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=57.245) [thousands of publishers on pub/sub. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=59.27628571428573) [publisher creates a topic and sends](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=62.25257142857143) [messages which are specific to that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=64.572) [particular topic. A message contains a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=66.733) [text payload and may also contain](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=69.0945) [additional optional metadata attributes.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=71.76766666666666)[Publishers publish messages to a topic and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=74.551) [these messages are stored within pub/sub](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=77.35) [in message storage. Messages are not](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=80.173) [immediately delivered to subscribers, thus](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=83.17) [effectively decoupling publishers and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=85.67683333333333) [subscribers. Subscribers were interested](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=88.36416666666669) [in receiving messages from publishers will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=90.36560000000001) [subscribe to a particular topic. Once a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=92.97200000000001) [subscription has been created, subscribers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=96.13699999999999) [can choose to pull messages from pub/sub.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=98.11171428571429)[Messages can also be pushed to subscribers](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=102.116) [using webhooks. Pub/sub ensures that these](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=105.03433333333334) [messages are delivered to the right](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=108.10966666666666) [subscriber, and once a subscriber receives](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=110.42259999999999) [a message, the subscriber has to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=112.49185714285711) [acknowledge that the message has been](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=114.5215714285714) [received. Only when subscribers have](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=117.05400000000002)[acknowledged receiving these messages are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=119.32419999999999) [they removed from the pub/sub queue and from the message storage within pub/sub.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=2&mode=live&start=121.06099999999999)

[Execute a Pub/Sub to BigQuery Job Using Dataflow Templates](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live)

[In this demo, we'll see how we can execute](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=1.725) [Dataflow jobs by using predefined Google](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=3.789) [templates. Imagine that in the SpikeySales](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=7.894714285714286) [organization real time sales data is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=10.064) [available using the pub/sub message](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=12.844) [delivery service. Go to the Navigation](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=14.939666666666668) [menu and click on Pub/Sub. We're going to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=17.32922222222222) [create a new topic where we'll publish](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=19.933125) [real time sales data. The full path to any](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=22.291666666666668) [topic on pub/sub includes the name of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=25.668124999999993) [project as well. The name of the topic](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=28.022249999999993) [here is spikey-real-time-sales. You can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=29.904714285714277) [see here that a new topic has been created](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=33.0498) [to which publishers can publish messages.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=35.174) [This topic currently has no subscribers.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=37.82) [Let's click through and view details of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=40.273) [this topic. You can use the web console to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=42.653333333333336) [publish messages to this topic. Click on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=45.53816666666667) [Publish Message and this will bring up a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=47.78557142857143) [dialog which will allow you to publish](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=49.8865) [messages. This topic currently has no](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=52.227000000000004) [subscribers. There is no one to whom the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=54.92850000000001) [published messages will be delivered,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=57.60874999999998) [which is why you see this warning up top.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=59.14014285714286) [Let's cancel out of this page and let's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=62.358) [head over to BigQuery, which is where](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=64.563) [we'll store the data that we read in from](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=66.68924999999996) [pub/sub. I'm going to open up the BigQuery](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=70.64100000000002) [web console in a new tab so that it's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=72.53900000000002) [easier for us to navigate. We'll create a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=75.32988888888895) [new table within the sales dataset that we](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=77.51675) [had created earlier. Click on the Create](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=80.065) [Table link after having selected the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=82.373) [dataset in the left Navigation pane. We'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=84.49725000000001) [call this the product\_sales table and this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=87.65442857142857) [will hold the streaming data that we read](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=89.81524999999999) [in from pub/sub using our Dataflow job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=92.314375) [created from templates. Select the table](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=94.9) [on the left Navigation pane and specify a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=97.78757142857143) [schema for this table. This table has](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=100.553) [exactly two columns. The first column](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=103.70000000000002) [holds the name of the product and is of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=105.96400000000001) [type string, and the second column holds](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=108.60900000000001) [the number of units of that product that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=110.77712499999998) [was sold. And this column will be of type](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=113.69585714285715) [integer. Both of these columns are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=118.31137499999997) [nullable, which means that null values can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=120.58016666666663)[be present for any of these fields. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=123.01450000000003) [new schema has been saved. We now have the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=126.96566666666666) [source of our data that is pub/sub and the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=129.47819999999996) [sink for our data which is BigQuery. Let's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=131.98980000000003) [head over to the Dataflow page, which I'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=135.06671428571428) [once again open in a new tab so that it's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=137.315) [easier to navigate. We won't be writing](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=140.1598)[any code in this demo. We'll simply use](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=142.50833333333335) [the Create Job From Template link up top](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=144.8581428571429) [in order to specify our Dataflow job. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=147.77100000000002) [first step is to uniquely identify this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=150.65099999999998) [job by giving it a name. We'll call it the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=153.08866666666665) [spikey-sales-tracker. The next step is to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=157.67279999999997) [choose one of the predefined templates](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=159.36122222222218) [that we'll use to instantiate our pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=160.92433333333332) [and execute our job. All of the predefined](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=164.17485714285712) [templates for data transfer jobs are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=167.7674285714286) [present in this drop-down list. We'll](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=169.70874999999998)[choose PubSub to BigQuery. You can](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=172.0882) [configure the region where you want this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=175.14575) [job to be executed. Ideally, you'd run](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=176.8118) [this job very close to where your data is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=178.92162500000003) [located. I'm going to simply specify](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=181.61320000000003) [us-east1 as my region. We're transferring](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=184.18277777777783) [data from pub/sub to BigQuery. We need to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=186.82814285714284) [specify the topic to which this Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=189.65799999999996) [job should subscribe. This is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=191.84675000000001) [spikey-real-time-sales topic. Make sure](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=194.77520000000004) [you specify the full path to this topic,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=196.53325)[including the name of the project. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=199.43175) [next step is to give the name of the table](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=201.66209999999998) [where you want the results to be stored.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=203.3011111111111) [Remember, Bigtable tables need to be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=205.462) [prefixed by the name of the project:, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=208.728) [name of the dataset, and then the name of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=211.70600000000002) [the table. Dataflow jobs typically need a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=213.7632857142857)[temporary staging location to store](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=216.48716666666667) [intermediate results and specify the temp](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=218.548) [folder under spikey-df-store. If you click](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=221.024) [on Optional parameters, you can see](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=224.60350000000003) [additional ways in which you can configure](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=226.171875) [this Dataflow job. I'll simply set max](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=228.2662) [workers to 2. I have no additional](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=230.88985714285718) [configuration. We've configured our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=234.0912) [Dataflow job. We are now ready to start](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=235.88520000000003) [execution. Click on Run job and watch this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=238.3668571428572) [pipeline be instantiated and set up.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=241.38599999999997) [Because this job operates on streaming](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=244.35) [data, this job will continuously run until](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=246.79) [it's explicitly stopped. It's always](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=249.813) [watching for messages published to that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=252.13633333333334)[topic to process and transfer those message results into Bigtable.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=3&mode=live&start=254.20475000000002)

[Publishing Messages and Reading from BigQuery](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live)

[Our Dataflow job is currently running.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=1.434) [We're now ready to take it for a test](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=3.228) [drive. We start off in the Cloud Pub/Sub](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=5.675099999999998) [page and here is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=8.051666666666666) [spikey-real-time-sales topic to which](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=9.8) [we'll publish messages. Observe as this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=11.382) [topic now has one subscription, that is](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=13.776857142857143) [our Dataflow job. Click on Publish message](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=16.738) [and we'll publish our very first message](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=19.38828571428572) [to this topic. This message contains two](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=21.726142857142857) [pieces of information, the name of the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=23.728333333333335) [product and the units sold. These exactly](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=25.731857142857148) [match the columns in our BigQuery table.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=28.59142857142858) [Click on the Publish button to publish](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=32.199) [this new message to this topic. Let's](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=33.75300000000001)[publish another message using the same](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=36.82766666666667) [link regarding the sales of another](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=38.690374999999996) [product. Click on Publish. Two messages](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=40.81662499999998) [have now been published to this topic. We](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=42.68457142857144) [can now head over to the Dataflow job to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=45.6013) [see whether it's been processing our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=47.449) [messages. If you click on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=49.455571428571446) [ReadPubSubMessages, the messages haven't](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=51.365199999999994) [been received yet. Just wait for a little](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=52.965999999999994) [bit so that the workers are scaled, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=55.48633333333334) [number of workers is equal to one, and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=57.69642857142858) [then click on ReadPubSubMessages once](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=59.69037499999999) [again. And you see here that the two](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=61.873) [messages that we published were processed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=64.82188888888888) [successfully. Observe that the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=67.77379999999998)[ConvertMessageToTable row performs a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=69.76079999999999) [branching operation. A message that was](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=71.79299999999999) [successfully passed is written out to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=74.30114285714288) [BigQuery. A message that failed to pass,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=76.522) [maybe it was malformed in some way, will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=78.95633333333338) [be written out to a different table. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=81.51312499999999) [two messages that we published were](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=84.23542857142857) [correctly formed, they've been written out](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=85.84) [successfully. Let's switch over to the tab](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=88.27420000000002) [where we have the BigQuery web console](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=89.85233333333332) [open and query this table to see whether](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=91.89042857142854) [the results were written out. Do a simple](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=94.21814285714288) [select\* query on this table, click Run](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=97.47071428571425) [query, the query will run through, and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=100.31271428571429) [you'll find that, yes indeed, our two](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=102.21149999999997) [messages have been written out to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=104.40557142857145) [BigQuery. Let's head back to Pub/Sub and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=107.179) [publish two more messages. The first of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=109.58357142857145) [these messages is a correctly formed one.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=111.6498571428571) [The second message that we'll publish will](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=114.708) [have one of the fields missing. This has](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=116.595) [only the name of the product, no units](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=119.6335) [sold. Let's head back to observing our](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=121.76157142857143) [Dataflow job. You can see that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=124.54) [ConvertMessageToTable row has now](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=127.05266666666665) [processed three messages. We've submitted](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=128.74259999999998) [a total of 4 and they have been added to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=130.58166666666668) [the output. Three records have been](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=133.91940000000002) [written to BigQuery successfully so far.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=135.83585714285715)[Let's see what happens to the fourth one.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=137.855) [Click on ReadPubsubMessages. It's now](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=139.491) [processing the fourth of these messages](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=142.10157142857145) [and the fourth message has also been](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=144.01399999999998) [processed successfully. Back to the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=147.53233333333333) [BigQuery tab where we'll display all of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=149.9502857142857) [the records in our BigQuery table once](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=151.84425000000002) [again. Click on Run query and you'll find](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=153.92424999999997) [that we now have four records in our table](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=156.52100000000002) [and one of these records has a null value](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=159.14300000000003) [for units sold. That's perfectly](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=162.1845) [acceptable because that's we had defined](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=163.9988) [the schema for this table. We'll now go](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=166.29500000000002) [back to the Pub/Sub page for our topic,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=168.81600000000003) [and this time we'll publish a new message](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=170.667) [which contains malformed data. This](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=173.44233333333335) [message contains a list of products, but](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=176.15142857142857) [our Dataflow job can process just one](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=178.154)[product at a time. Hit Publish on this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=180.73) [message and let's head over to Dataflow to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=183.36) [see what happens. The](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=185.574) [ConvertMessageToTable row has processed](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=187.8226) [five messages, the fifth of these is the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=189.853) [malformed one. One of these messages, the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=193.207) [malformed one has gone into the other](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=196.638125) [branch, this branch performs a flatten](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=198.984) [operation on the input message converting](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=201.111) [it to a string format and writes out a](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=203.05) [failed record. These failed records are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=205.95499999999998)[automatically written out to a different](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=208.559) [BigQuery table so that you can track them.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=210.643) [If you head over to the BigQuery UI and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=213.272) [refresh the page, you'll find that under](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=215.52) [your sales dataset, there is now a table](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=217.887) [containing records which were in error.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=220.713) [The schema for this table has been defined](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=224.385) [by the templated Dataflow job and it](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=226.589) [contains a bunch of different information.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=229.111) [And if you run a query on this table,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=231.244) [you'll find that the one failed record was](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=233.492) [written out here. This is the record that](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=236.577) [our Dataflow job wasn't able to parse.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=238.898) [Let's head back to the page where we are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=241.869) [currently monitoring our streaming job.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=243.349) [Now streaming jobs on Dataflow will run](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=245.669) [forever until you explicitly stop them and](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=247.887) [you can do so by clicking on the Stop job](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=250.484)[button. When you stop a streaming job, you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=253.78218181818184) [have to make a choice as to what to do](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=255.72044444444438) [with the data that has already been](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=258.09280000000007) [buffered by Dataflow. You can choose the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=260.52662499999997) [drain option where your job will process](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=262.84100000000007) [all of the data that has been buffered,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=264.73333333333335) [but will not ingest any new data, or if](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=266.83299999999997) [you don't care about the buffered data,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=269.52025) [you can simply cancel the existing job.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=271.8753333333333) [That's the option we'll choose here, click](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=274.645) [on cancel and click on Stop Job. Wait for](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=276.8131428571428)[a little bit. Your job will be cancelled and Dataflow will wind down its workers.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=4&mode=live&start=279.9473333333332)

[Summary and Further Study](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live)

[And with this, we come to the very end of](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=1.66) [this module where we saw how we could](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=3.581) [execute Dataflow jobs using pipeline](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=5.873) [templates. Templates democratizes Dataflow](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=9.014666666666669) [jobs allowing you to perform data transfer](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=11.6875) [operations between different GCP services](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=14.764) [using Dataflow without writing a single](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=17.067) [line of code. In this module, we saw how](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=19.691) [we could specify runtime parameters to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=22.193) [configure our pipeline, and we then](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=24.365) [deployed these pipelines using predefined](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=26.865) [Google templates. We saw a hands-on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=29.441333333333333) [example using templates where we read in](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=32.72571428571428) [streaming data from pub/sub and wrote](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=35.22442857142858) [these messages out to BigQuery. And on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=37.75299999999999) [this note, we come to the very end of this](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=40.8352) [course on Dataflow. Now if you're](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=43.27759999999999) [interested in studying further and you're](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=45.15275) [specifically interested in big data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=47.177666666666674) [processing, here are some courses on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=49.41340000000001) [Pluralsight that you might find useful.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=51.27771428571429) [The building Blocks of Hadoop - HDFS,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=53.589) [MapReduce, and YARN will introduce you to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=55.886) [big data processing using MapReduce in the](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=58.97971428571429) [Hadoop ecosystem. Or you might be](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=61.775333333333336) [interested in the Spark Data Analytics](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=64.1985) [engine, which means Getting Started with](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=66.47033333333333) [Spark 2 is a course that will be useful](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=68.87700000000001) [for you. If you're interested in big data](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=71.303) [processing on Cloud platforms, here are](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=73.57271428571428) [some resources that you can use,](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=75.63885714285716) [Architecting Big Data Solutions Using](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=77.7685714285714) [Google Dataproc will introduce you to](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=80.2195) [managed Hadoop on the GCP. And Big Data on](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=83.05466666666669) [Amazon Web Services will show you how you](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=86.0873333333333) [can work with truckloads of data on AWS.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=88.53900000000003) [It's now time for me to say goodbye.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=92.481) [That's it from me here today. Thank you for listening.](https://app.pluralsight.com/player?course=google-dataflow-architecting-serverless-big-data-solutions&author=janani-ravi&name=caf89f6c-4d6a-4bc9-b635-2ee51f275292&clip=5&mode=live&start=94.356)