[Course Overview](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live)

[Course Overview](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live)

[Hi! My name is Janani Ravi, and welcome to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=0) [this course on Designing Scalable Data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=8.338" \t "psplayer) [Architectures on the Google Cloud. A](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=10.456) [little about myself. I have a master's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=13.1455) [degree in electrical engineering from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=14.901714285714284) [Stanford and have worked at companies such](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=16.378) [as Microsoft, Google, and Flipkart. At](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=19.186999999999998) [Google, I was one of the first engineers](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=21.58822222222222) [working on real-time collaborative editing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=23.398) [in Google Docs, and I hold four patents](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=25.538) [for its underlying technologies. I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=27.8348888888889)[currently work on my own startup,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=30.33342857142857) [Loonycorn, a studio for high-quality video](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=32.004) [content. The GCP offers up a very large](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=35.532000000000004) [number of services for every important](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=38.19199999999999) [aspect of public cloud computing. This](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=40.21933333333334) [array of services and choices can often](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=42.946285714285715) [seem intimidating. Even a practitioner who](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=45.42366666666667) [understands several important services](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=48.48233333333334) [might have trouble connecting the dots, as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=49.9492) [it were, and fitting together those](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=52.333) [services in meaningful ways. At such](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=54.296666666666674)[terms, a need is felt for end-to-end](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=56.34585714285714) [integrated examples which span different](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=59.74533333333333) [services and focus on how they all fit](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=61.976285714285716) [together into the big picture. In this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=64.82555555555555) [course, you will gain the ability to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=67.08699999999999) [design Lambda and Kappa Architectures that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=68.95666666666666) [integrate batch and streaming, plan](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=71.8412) [intelligent migration and disaster](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=74.364) [recovery strategies, and pick the right](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=75.796) [machine learning workflow for your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=78.18) [enterprise. First, you will learn why the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=80.65999999999998) [right choice of stream processing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=82.51649999999998) [architecture is becoming key to the entire](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=84.01999999999998) [design of a cloud-based infrastructure.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=86.2975) [This has to do with the need for real-time](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=89.452) [data processing and a specific issue in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=91.292) [machine learning modeling known as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=93.691) [training-serving skew. Increasingly,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=95.682) [organizations are confronted by ML models](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=97.9655) [that work well in the lab but perform](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=100.7372) [poorly in real-time prediction. This is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=103.065) [often because real-time data is not](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=105.8185) [processed as accurately and cleanly as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=107.431) [batch data. You will learn two specific](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=109.828) [architectural patterns, Lambda and Kappa](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=112.303) [Architectures, that seek to mitigate this.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=115.153) [Next, you will discover how the Transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=117.7) [Service is an invaluable tool in planning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=119.744) [both migration and disaster recovery](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=122.403) [strategies on the GCP. Currently, the GCP](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=124.579)[does not offer an equivalent of the AWS](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=127.05914285714287) [Data Migration Service, so when you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=129.64000000000004) [migrate to the GCP, it is rational to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=131.77299999999997) [first move all your structured and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=134.381)[unstructured data to cloud storage rather](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=136.19899999999998) [than committing upfront to a specific](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=138.32471428571426) [structured storage solution. The Transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=140.959) [Service is great for this specific use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=143.50671428571428) [case. Finally, you will explore how to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=145.44866666666667) [pick the right machine learning technology](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=148.00985714285716) [for your specific use. You can leverage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=149.89766666666668) [pre-trained ML APIs if you don't need to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=152.77700000000002) [build a custom model at all. Or you could](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=155.4896666666667) [leverage auto-ML to implement transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=157.9967142857143) [learning, which retrains a sophisticated](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=159.667)[model to your specific data. You can even](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=162.02166666666665) [build ML models in SQL using BigQuery ML.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=165.29366666666667) [When you're finished with this course, you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=168.533) [will have the skills and knowledge of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=170.039875) [entire cross-section of big data and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=171.89285714285714) [machine learning offerings on the GCP to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=173.818) [build cloud architectures that are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=176.074)[optimized for scalability, real-time](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=178.186) [processing, and the appropriate use of deep learning and AI technologies.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=c985bfcc-00de-4383-8e04-a56593fd950e&clip=0&mode=live&start=180.6316)

[Implementing Integrated Batch and Streaming Architectures on the GCP](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live)

[Module Overview](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live)

[Hi, and welcome to this course on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=0) [Designing Scalable Data Architectures on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=3.5743750000000003) [the Google Cloud. In this module, we'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=6.099285714285716) [focus on discussing the importance of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=8.675714285714285) [stream processing and how it's very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=10.567857142857141) [important to have an integrated pipeline,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=12.618714285714285) [which combines batch, as well as stream](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=14.509857142857143) [processing, so that you have just one code](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=17.21857142857143) [base to maintain. Now back in the day, how](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=19.7508) [you stored data was extremely important,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=22.552999999999997) [whether you stored them in files or in a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=24.271) [structured database. Data architectures](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=27.689) [nowadays are being designed around machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=29.275399999999998) [learning rather than storage. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=31.7696) [importance of having data which you can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=33.757000000000005) [use to extract insights by building ML](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=36.34249999999999) [models has grown, which means that stream](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=38.8405) [processing has gone from being peripheral](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=41.41114285714287) [to being a critical aspect that we have to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=43.698428571428565) [take into consideration from the very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=46.23328571428573) [beginning. Why is that? Because data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=48.57980000000001) [that's available in the real world is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=50.61728571428573) [usually streaming data, and this is what](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=53.01614285714284) [you'd pass in for prediction into your ML](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=55.18422222222222) [pipelines. Now batch processing of huge](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=58.63559999999999)[amounts of data stored in the form of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=60.64628571428571) [files or in a structured database](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=62.69859999999999) [somewhere has been around for a long time,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=64.5986) [which means that many aspects of designing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=67.1345) [data architectures are now standardized](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=69.38080000000002) [especially in the case of batch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=71.87) [processing. But streaming data, well,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=74.34885714285716) [stream processing is a key exception to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=76.51066666666668) [this trend. As developers and architects,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=78.98928571428569) [we are still coming to terms with the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=80.90299999999999) [volume and variety of stream processing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=83.55012500000004) [options and the kind of data that we are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=85.60957142857141) [dealing with in the real world. There is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=88.01488888888892) [one thing that we realize, though. It's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=90.102) [hard to maintain separate architectures](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=92.45750000000001) [and code bases for batch processing and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=94.417) [stream processing. These lead to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=97.96260000000002) [unexpected issues when we are working with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=99.93)[data in the real world such as the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=101.68366666666665) [training-serving skew when we are dealing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=103.6575) [with machine learning models, which is why](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=105.8284) [you should seriously consider integrating your batch and streaming architectures.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=0&mode=live&start=108.12042857142856)

[Prerequisites and Course Outline](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live)

[Before we dive into actual course content,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=2.165) [let's take a look at some of the pre-reqs](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=4.41) [that you need in order to make the most of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=6.211) [your learning. This is an advanced course](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=8.315) [on the GCP. This course assumes that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=11.068) [you're familiar with many of GCP's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=13.372) [products and services. As this is a course](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=15.575714285714284) [on data architectures, this course assumes](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=18.259999999999998) [that you work with big data processing on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=20.136714285714284) [the Google Cloud. Pre-reqs to this course](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=22.8762) [are Architecting Serverless Big Data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=25.364666666666672) [Solutions Using Google Dataflow,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=27.2965) [Architecting Big Data Solutions Using](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=29.2168) [Google Dataproc, that is, managed Hadoop](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=31.964599999999997) [on the Google Cloud, and, finally,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=34.41442857142857) [Architecting Production-Ready ML Models](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=36.909) [Using Google Cloud ML Engine. If you feel](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=39.22866666666667) [that you're not familiar with these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=42.205749999999995)[products and services on the GCP, these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=43.596714285714285) [are courses that you might want to take](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=45.963) [before you launch into this one. I feel](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=47.79042857142857) [like this course is a little different](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=50.3042) [from other technology-based courses that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=52.03866666666667) [you might have watched on Pluralsight.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=53.801399999999994) [This course is not shy about offering](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=56.554) [opinions based on facts as of December](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=58.162) [2018. The point of this course is to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=61.81599999999999) [introduce certain best practices and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=63.898999999999994) [architectures, but it's mostly to get you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=66.10599999999998) [thinking. We'll start this course off by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=68.6865714285714) [discussing the importance of integrating](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=70.46066666666667) [your batch and stream processing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=72.2116) [pipelines. Specifically, we'll talk about](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=74.9945) [the Lambda Architecture and the Kappa](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=76.80033333333333) [Architecture here. We'll then move on to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=79.41366666666669) [discussing migration and disaster recovery](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=82.02271428571427) [for your on-premises data center and on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=84.045) [the GCP. We'll specifically talk about](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=87.17099999999999) [migrating your data to Cloud Storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=89.318) [buckets and see hands-on demos for how it](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=91.162125) [can be done. And, finally, we'll discuss](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=93.90662499999998) [machine learning workflows. We'll discuss](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=96.43220000000002) [the suite of AI products that the GCP has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=98.33779999999997) [to offer and briefly discuss how you would](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=100.79433333333336) [choose the right product that fits your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=103.541) [use case. This is the module where we'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=106.10455555555554) [also introduce and work with Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=107.9478571428571) [Dataprep, a very interesting UI-based tool](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=109.97050000000002) [to clean and prepare your data for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=112.56442857142858) [building machine learning models. As we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=115.0164285714286) [work through this course, we'll assume](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=117.02500000000002) [that we are developers and architects at](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=118.91799999999999) [this hypothetical online retailer called](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=121.186) [SpikeySales.com. SpikeySales specializes](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=124.36066666666667) [in flash sales of trending products, which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=126.17666666666666) [means that they experience huge spikes in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=128.9) [user traffic on sale days. On other days,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=131.6894285714286) [traffic is fairly low. SpikeySales is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=134.57800000000006)[contemplating a move to a cloud provider,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=137.2773333333334) [specifically the GCP. Cloud computing fits](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=140.54887500000004) [their bill perfectly. They want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=143.17849999999999) [pay-as-you-go functionality and don't want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=145.49059999999997) [to hold idle capacity during off-sale periods.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=1&mode=live&start=147.6787142857143)

[Changing Architectural Considerations in the Real World](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live)

[The one thing that we as developers know](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=2.142) [for sure is that technology that is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=4.053) [cutting edge today will be outdated](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=6.3918750000000015) [tomorrow. And as technology evolves, the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=9.001) [architectural considerations that we have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=11.5578) [to keep in mind as we are designing our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=13.267222222222221) [systems also change. Not so very many](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=15.654) [years ago as the amount of data that your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=18.568199999999997) [organization dealt with kept growing, one](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=21.235) [thing that you might have discussed is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=23.043125) [should you be switching to Hadoop.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=25.023249999999997) [Horizontal scaling with a distributed](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=27.103) [cluster was getting to be mainstream, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=28.773) [Hadoop was where everyone was headed.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=31.407285714285713) [MapReduce on Hadoop was very cool. But](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=34.013) [maybe it wasn't so very developer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=36.886428571428574) [friendly, which is why technology kept](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=38.509) [evolving, and people started moving](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=40.261) [towards Spark. Spark gave you a REPL,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=43.23242857142856) [read-evaluation-print-loop, environment,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=45.448) [an interactive environment where you could](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=47.978) [prototype stuff easily. In order to meet](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=49.956) [their compute needs, a vast majority of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=52.77957142857144) [organizations had and still have a large](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=55.22471428571428) [on-premises data center. However, cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=58.30099999999999) [computing started getting more mainstream.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=60.6208) [Companies wanted to get away from the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=63.01866666666667) [administrative overhead of maintaining](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=66.10549999999998) [their own machines, and they wanted the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=67.46366666666667) [convenience of the cloud. They started](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=70.03833333333333) [moving from on-premises to cloud. It was](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=72.48842857142859) [easy to spin up VMs to process huge](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=74.93919999999999) [amounts of data with cloud providers. But](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=77.068) [what do you do when data is real time?](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=79.9725) [That's when stream processing got to be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=82.524) [mainstream. With the growth of IoT and the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=85.38962499999998) [ability to access and process real-time](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=87.33977777777777) [data, stream processing became an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=90.0008) [important compute component in any org. As](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=92.17519999999998)[organizations have grown in their ability](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=95.65016666666668) [to collect, store, and process data,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=97.38933333333333) [analytical processing to extract insights](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=99.859) [is no longer sufficient. Building machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=102.5462) [learning models that help businesses make](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=105.064) [predictions has grown in importance. We](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=107.38285714285715) [are now in a period where machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=110.32849999999999) [and deep learning are slowly getting](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=112.30966666666666) [democratized. It's not just reserved for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=114.7185) [experts to harness the power of these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=116.90828571428568) [models. Deep learning is mainstream, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=119.314) [there are a bunch of technologies such as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=121.25114285714285) [AutoML or BigQuery ML that allow people](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=123.96925000000002) [with no programming experience to harness](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=126.88857142857142)[the power of neural networks. Once we know](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=129.45799999999997) [that real-time machine learning is where](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=133.34000000000003) [we are headed, we have to plan for a world](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=134.92499999999998) [in which everyone uses machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=137.48550000000003) [models. Everyone wants the predictions](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=139.7926666666667) [that ML can offer them, and almost](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=141.7341428571429) [everyone will be building these models.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=144.2151428571428) [The democratization of ML means that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=146.9) [harnessing the power of deep learning is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=149.414) [no longer restricted to data scientists or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=151.32325) [researchers. And what's definitely true](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=154.62685714285715) [today not just in the future is that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=156.29428571428576) [everyone queries data. Everyone has the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=159.31974999999994) [ability to pull the data that we need to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=161.434) [make business decisions. If everyone's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=164.33142857142857) [building and using ML models, an important](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=166.26014285714288) [thing that they have to watch out for is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=168.73549999999997) [training-serving skew. Your model might](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=172.19660000000002) [perform very well during training but](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=174.28842857142857) [perform poorly in the real world. This](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=176.21325) [could be either because you're overfitting](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=178.93366666666668) [on the training data or you're not](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=180.42785714285714) [processing your real-time data for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=183.05116666666666) [prediction the way you processed your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=184.892) [training data. Another trend that can be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=186.921) [observed in the world today is that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=189.651) [organizations no longer want to be tied to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=191.892) [a single cloud provider. They don't want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=194.3) [their on-premises data center, but they](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=197.35819999999998)[don't want to be locked into a single](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=198.96299999999997) [platform as well. And, finally, an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=201.84585714285717) [important consideration for developers](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=204.494) [today as they are building their data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=205.8417142857143)[processing pipelines is the kind of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=207.931) [architecture that they choose. Stream](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=210.3664) [processing is no longer a secondary](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=212.533875) [consideration. It's a core component of an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=215.08912499999994) [architecture. So which do you choose, a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=217.9835714285715) [Lambda Architecture or Kappa Architecture?](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=219.84699999999995) [What you see here onscreen might seem like](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=222.829) [a grab bag of things that you have to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=224.619) [consider, but each is important in its own right.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=2&mode=live&start=226.459)

[Discussion: Using and Building ML Models, Querying Data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live)

[Now that we've introduced a bunch of ideas](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=1.906) [here, let's dive into each of these in a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=3.986) [little more detail. We'll start off with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=6.464) [the first statement here, which says that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=8.987) [we need to plan for a world in which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=11.171) [everyone uses models. Machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=12.976) [today is no longer a province of a few](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=15.997285714285713) [engineers at Google, Facebook, or any of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=18.62514285714286) [those other technology companies. Models](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=21.2705) [are becoming objects of daily use.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=23.29657142857143) [Everyone wants to harness the power of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=25.778) [machine learning, and it's not just for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=27.459) [tech-based use cases. Models are important](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=29.389) [for business teams as well. Sales teams](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=32.873999999999995) [today often consult models before setting](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=35.62433333333333) [out on sales calls. Sales teams often buy](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=38.05866666666667) [huge databases of leads. But which of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=41.81750000000001) [these leads will convert? What is the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=44.63771428571428)[probability that a customer will place a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=46.45814285714286) [huge order with their organization? If](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=48.660999999999994) [they have machine learning models that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=51.16657142857142) [give them this kind of information, this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=53) [will allow sales teams to focus their](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=55.396) [efforts better. Now it's very likely that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=57.671) [people who make up the sales team are good](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=60.942)[salespeople. They're not engineers, which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=63.93089999999998) [means that machine learning models may](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=65.28833333333333) [need to be packaged in such a way that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=66.86042857142856) [anyone can use them. Now, of course, in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=69.78079999999997) [order to get a good machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=72.02777777777779) [model, you have to train on huge dataset,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=73.6272857142857) [which will require a distributed training](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=76.12971428571429)[platform. But the emphasis is shifting](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=78.96266666666666) [from distributed training towards](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=80.92299999999997) [user-friendly prediction. Even the best](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=83.3446) [models will be useless unless they're](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=85.914)[accessible to people who don't know how to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=88.08514285714286) [code. And this takes us to the second](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=90.37044444444444) [point. As more and more people use machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=92.48162500000001) [learning models, we'll find that almost](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=95.13133333333333) [everyone will be building these models.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=97.21014285714286) [Models will be built by those who need](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=99.833) [them, not just those who have expertise.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=101.483) [If building powerful machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=104.974) [models is accessible to only those who](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=106.944) [know how to code and who have ML](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=109.26975000000002) [expertise, that's a severe limitation on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=112.105375) [our ability to harness the power of ML.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=114.18514285714282) [With the advent of technologies such as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=117.426) [AutoML, BigQuery ML, and machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=119.058) [APIs, we are slowly moving towards the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=121.984) [democratization of machine learning and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=124.327) [artificial intelligence. This process was](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=126.643) [begun with the introduction of open-source](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=129.05649999999997) [libraries such as SciKit Learn and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=130.9284) [TensorFlow to build ML models. And as we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=133.56466666666662) [move further along in that direction, we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=136.46887499999997) [should now be able to build ML models with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=137.96475) [a few clicks of the UI. And we've seen](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=140.4246666666667) [this trend in the technologies of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=143.16971428571432) [past. MapReduce, which involved Java, gave](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=145.39625) [way to Spark, which offered an interactive](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=147.5838571428572) [environment in Python. And there is no](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=150.211) [doubt that Python has made programming](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=152.9214444444444) [very, very accessible to even](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=154.595) [non-developers. TensorFlow gave way to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=157.55999999999995) [Keras, which is far easier to use. And in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=159.35800000000003) [this world, AutoML, BigQuery ML and ML](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=162.61487499999993) [APIs are big winners. And with this, we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=165.81449999999995) [can move on to the third talking point. We](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=168.67500000000004) [need to plan for a world in which everyone](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=171.28433333333334) [queries data. One thing that's pretty](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=173.52666666666667) [obvious is that organizations today are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=175.84885714285713) [flatter and more data driven. Everyone](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=178.05599999999998) [uses data to make their decisions. Today](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=180.5957142857143) [everyone's building their own](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=183.7388) [spreadsheets. Maybe they have someone else](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=185.94066666666666) [pull data, but tomorrow everyone will have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=188.1595714285714) [to pull their own data because only they](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=190.4262857142857) [know what's relevant and important to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=193.39971428571437) [them. If more and more people are pulling](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=195.66800000000003) [their own data, it has to get easier to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=197.63037500000007) [query data, and that's why SQL is getting](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=200.097) [to be more, not less, important. SQL is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=202.57966666666658) [the common language used by data analysts,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=205.72375) [scientists, everyone. People are moving](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=208.923)[towards data warehouses rather than](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=211.95250000000004) [RDBMSs, and here BigQuery, the serverless](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=214.35299999999998) [data warehouse on the GCP, is a big](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=217.57642857142858) [winner. Why data warehouses? Well, you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=220.74528571428573) [want the ability to store huge amounts of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=222.85425) [data from disparate sources and still](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=225.51857142857142) [query them all in one place. Data in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=228.11871428571428)[relational databases is too constrained](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=231.39366666666666) [and too structured to support this use case.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=3&mode=live&start=233.41816666666665)

[Discussion: Training Serving Skew, a Multi Cloud World, Lambda, and Kappa Architectures](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live)

[In a world where machine learning is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=1.936) [mainstream, what you have to worry about](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=3.667) [is training-serving skew. So what exactly](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=5.869375000000002) [does this mean? If you've built machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=8.702) [learning models, you know that models are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=10.883) [trained on pre-classified, pre-labeled](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=12.744) [batch data. So you might find that it's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=15.585) [the case that that you train your model](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=17.792) [and test it with your historic data, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=19.831111111111113) [your model performs well during your back](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=21.80712500000001) [test. But this model performs poorly in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=24.82585714285715) [production, in the real world. Now an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=27.096) [important reason that a model might](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=29.590500000000002) [perform poorly in the real world is that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=31.146) [it's not general enough. Maybe your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=33.216) [training data wasn't large or diverse](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=35.954428571428565) [enough. In such cases, your model is set](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=37.69871428571429) [to be overfitted on the training data, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=40.475750000000005) [it doesn't really perform well in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=42.879111111111115) [prediction. However, there is another](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=45.38444444444447) [important reason that the model that you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=47.15666666666667) [have trained is not really very useful](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=49.228444444444435) [when used for prediction, and that is the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=51.513000000000005) [training-serving skew. This is a big, but](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=54.133) [neglected, cause. Now the training-serving](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=58.42614285714287) [skew comes into the picture when there is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=60.20100000000001) [a difference in how you process the data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=62.430375) [that you use for training and the data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=64.5788888888889) [that you pass in for prediction. Training](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=66.93280000000001) [data is typically sourced from batch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=69.18766666666667) [pipelines. This is data that you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=71.14671428571428) [understand which you've processed](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=72.981) [meticulously well. Prediction data,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=75.081)[though, comes from streaming pipelines,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=77.705) [and this data is often processed in an ad](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=80.417) [hoc manner with many shortcuts. You don't](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=83.35233333333333) [really give this prediction data the same](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=86.592) [care or the same processing treatment that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=88.78228571428572) [you do for training. When there is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=91.32) [difference in how you treat training data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=93.68377777777779)[versus prediction data, that thought gives](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=95.8772) [rise to the training-serving skew. Let's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=98.18800000000002) [now move on to the next point that we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=100.93966666666667) [brought up and what we have to consider](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=102.7541111111111) [when we build our data architectures,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=104.86499999999998) [multi-cloud is in. Organizations today are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=106.89375) [no longer one cohesive or monolithic unit,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=110.33999999999999) [which means that different teams and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=112.89499999999998) [stakeholders will have different](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=115.07966666666664) [preferences, and an organization as a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=116.915) [whole fears platform lock-in. They don't](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=119.24185714285716) [want to be tied to one cloud provider or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=121.13957142857143) [one platform. Organizations prefer working](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=124.42724999999996) [with their on-premises data center because](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=126.5256) [they feel that their data is secure and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=129.09733333333332) [their processing is close to where they](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=131.63600000000002) [are. At the same time, organizations](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=134.24928571428572) [realized the advantages of not having to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=136.47299999999996) [deal with the administrative overhead of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=138.64000000000007) [their own machines. This conflicting set](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=141.216) [of priorities within an organization](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=143.79300000000003) [basically means that traditionally Infra](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=145.85240000000002) [as a Service, that is, their on-premises](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=148.0092857142857) [DC, does not work for them, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=150.41642857142855) [traditional Platform as a Service where an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=152.84071428571428) [org is tied into a single cloud provider](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=155.45922222222222) [also does not work, which is why](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=157.568375) [containers and Kubernetes are big winners.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=159.979) [A Kubernetes cluster allows you to tie in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=163.835) [your on-premises data center with the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=166.085) [cloud provider, which means you have the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=168.1232857142857) [best of both worlds. And let's move on to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=170.04625000000007) [the last consideration that we brought up](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=173.092625) [here and the implications of real-time ML,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=174.99328571428572) [Lambda or Kappa Architectures. Now if you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=177.9453333333333) [haven't heard of what these are, don't](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=181.22533333333337) [worry. We'll discuss this in much more](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=182.74171428571427) [detail in the next clip. Now Lambda and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=184.336) [Kappa Architectures have this similarity,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=187.178) [that they give equal importance to stream](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=188.98499999999999) [processing, as well as batch processing.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=191.64642857142857) [Both of these architecture setups try to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=193.47220000000002) [combine batch and stream data, but they do](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=196.08933333333334) [so in different ways. Their approach is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=198.83400000000006) [completely different. When you use the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=201.176) [Lambda Architecture, the coupling between](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=203.87266666666662) [batch and stream processing pipelines is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=205.338) [loose. The loose coupling implies that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=208.48649999999998) [with Lambda Architecture, you might have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=210.70742857142858) [to maintain separate processing pipelines](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=212.56199999999995)[for batch and stream data. But in some](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=214.37750000000003) [ways, this separation of processing makes](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=217.47700000000006) [Lambda Architectures more robust. Kappa](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=219.77583333333328) [architectures, on the other hand, have a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=222.88266666666667) [far tighter coupling between batch and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=224.84171428571432) [stream processing. You have just one](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=227.047) [processing pipeline, but that might be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=229.6396666666667) [more fragile. Lambda architectures are an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=232.23100000000005) [older concept. They have been around for a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=234.95133333333337) [while and are a standard data processing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=236.84899999999993) [design. You combine batch and stream](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=239.3875) [processing with explicitly separate](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=241.93166666666662) [layers, so you have separate layers for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=243.62550000000002) [each. And once the processing of each is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=246.81519999999998) [complete, they are merged into a serving](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=248.821) [layer. This is robust, but this difference](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=251.61662499999994) [in processing can lead to training-serving](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=253.984) [skew that we discussed earlier. The Kappa](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=256.583) [Architecture, on the other hand, is a far](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=259.16614285714286) [newer idea which tightly integrates the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=261.495) [pipeline for batch and stream processing.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=263.5623333333333) [There is no separate code for each.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=265.985) [Instead, the batch code is simply fed](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=268.77425) [through the streaming layer. So you have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=271.5666666666667) [the exact same processing layer for both](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=274.4851249999999)[kinds of data. In theory, this is great](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=276.6862) [because it completely eliminates the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=279.0192857142858) [training-serving skew. However, in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=281.4614) [practice this can be fragile and can be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=283.487)[quite complicated to implement. Here is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=286.6451428571429) [summary of all of the points that we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=289.06714285714287) [discussed over the past few clips. Each](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=290.81449999999995) [point by itself is worthy of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=292.785)[consideration, and each is a factor that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=294.85625) [you need to take into account when you're](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=297.245) [designing your architectural patterns. All](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=298.995) [of these lead up to one significant point](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=302.37666666666667) [though. Make sure you design your data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=304.748) [architecture so that real-time ML prediction is easy, not just possible.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=4&mode=live&start=306.7515714285715)

[Lambda and Kappa Architectures](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live)

[Of all of the points that we brought up,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=2.077) [let's discuss the Lambda and Kappa](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=4.029) [architectures in a little more detail.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=6.079) [These architectures are at the heart of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=8.969) [designing your processing so that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=10.547) [real-time ML prediction is easy, not just](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=12.327) [possible. Now when you're working with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=15.680428571428571) [data processing in the real world, you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=17.371666666666666) [probably have historical data that is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=19.20428571428571) [stored on file somewhere. Maybe it's in a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=21.45166666666667) [Cloud Storage bucket. And at the same](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=24.15871428571429) [time, you have live data coming in in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=25.947999999999993) [form of a stream. If you're working on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=28.277000000000005) [GCP, you're quite possibly using Pub/Sub.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=30.731) [When you use the Lambda Architecture, your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=33.652) [batch data passes through a separate](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=35.934) [processing layer, that is, the batch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=38.01) [layer. A great option that you have on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=40.44599999999999) [GCP for batch processing is running Spark](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=42.507199999999976) [on a Dataproc cluster. Dataproc is managed](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=44.888857142857134) [Hadoop on the GCP. Your streaming data is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=48.068) [processed in a separate layer that is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=51.166) [referred to as the speed layer. A good way](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=53.483) [to process streaming data on the GCP is to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=56.533666666666655) [wire up your Pub/Sub to trigger a cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=58.73259999999999)[function. Once batch processing in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=61.4255) [batch layer and stream processing in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=64.23414285714286) [speed layer is complete, the output of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=66.831375) [both of these will be fed into the serving](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=69.214) [layer. The serving layer is where the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=72.02919999999996) [final output resides and is available for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=74.09400000000001) [further analysis. That might be your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=76.98800000000001) [BigQuery table. And once your data lives](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=79.29866666666666) [in BigQuery, there is no shortage of tools](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=81.63557142857142) [available on the GCP to visualize and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=83.56349999999998) [analyze this data. You can use Datalab,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=86.12762499999997)[Tableau, or the BigQuery UI. This is a big](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=88.7262857142857) [picture overview of how you might design a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=91.9654285714286) [Lambda Architecture on the GCP. The Lambda](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=94.74200000000002) [Architecture takes a hybrid approach to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=98.10199999999999) [batch and near real-time processing. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=100.366) [basic component of any Lambda Architecture](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=103.38766666666668) [is that it contains these three layers,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=105.46966666666667) [the speed layer, the batch layer, and,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=108.30728571428571) [finally, the serving layer. The batch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=110.66666666666669) [layer is often the source of truth. When](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=113.0111111111111) [you use the Lambda Architecture, it's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=115.31816666666667) [possible that you store your streaming](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=116.725) [data in the batch layer as well. Observe](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=118.651) [that the Lambda Architecture at its heart](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=121.59128571428572) [uses different code to process the batch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=123.64514285714286) [layer and the speed layer. That means you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=126.73) [have to maintain separate code bases for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=129.3755714285714) [batch and streaming data. Now if you're](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=131.62314285714288) [working on the GCP, you have other options](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=134.08433333333335) [that you can use to process your batch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=136.0855) [layer. Instead of Spark on Dataproc, you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=138.14385714285714) [might want to use Dataflow templates. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=140.60475000000002) [big advantage of using Dataflow is that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=143.952625) [you don't have to provision clusters for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=146.204) [your processing. Processing is entirely](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=148.231) [serverless. If you're building a big data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=151.506) [processing pipeline using the Lambda](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=153.57614285714294) [Architecture, with your data stored in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=155.7264) [serving layer, that is, BigQuery, you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=158.15125000000003) [use Datalab to analyze this data and maybe](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=160.1194285714286) [pass this data on to be processed on Spark](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=163.03399999999996) [in Dataproc clusters. If you're building a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=167.04766666666663)[machine learning pipeline using the Lambda](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=169.19966666666667) [Architecture, you'll pass on your data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=171.657) [from the serving layer to Datalab, which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=173.65787500000005) [you'll use to visualize and manipulate](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=175.8228571428572) [your data. You'll then use ML Engine for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=178.35519999999997) [distributed training. And, finally, once](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=181.87824999999998) [you have a trained machine learning model,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=183.41485714285716) [you'll expose this model as a RESTful API](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=185.66450000000003) [endpoint. Client applications will then](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=189.47166666666664) [call into your prediction API in order to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=191.29166666666669) [make predictions. The Lambda Architecture](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=194.9) [is a concept that has been around for a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=196.94) [while. The Kappa Architecture changed how](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=199.38350000000003) [people thought about their data processing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=201.30499999999995)[pipelines. The original idea of the Kappa](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=204.08583333333326) [Architecture was proposed by an engineer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=206.6417142857143) [named Jay Kreps when he worked at](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=208.76099999999994) [LinkedIn. The Kappa Architecture builds on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=211.06199999999998) [the concept that the stream is the source](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=212.75433333333336) [of truth, not any batch or storage layer.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=215.56062500000007) [The original form of the Kappa](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=219.252)[Architecture recommended the use of a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=221.30914285714277) [technology such as Apache Kafka, which is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=222.95824999999994) [a message transport system. A message](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=225.6785) [transport system would then retain a log](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=228.22833333333335) [of all of the data that it processes,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=230.23739999999998) [whether it's batch data or streaming data.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=232.46316666666667) [And once you have this transport system,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=235.21800000000002)[you'll always process directly from the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=238.02660000000003) [stream, never from a storage layer.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=240.57) [Because you're always working with the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=242.9252) [logs of your records, you use a single](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=245.30814285714285) [processing framework, whether it's batch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=248.46966666666668) [data or stream data. This was the original](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=250.50814285714284) [Kappa Architecture as proposed by Jay](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=253.33899999999997) [Kreps. But the basic implication of this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=256.0315) [is that batch and stream processing should](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=258.1929999999999) [be completely integrated and should not be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=260.6305000000001) [different processing pipelines. And this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=263.646) [is what the basic structure of a Kappa](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=266.8655555555555) [Architecture looks like. You process batch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=269.09749999999997) [and streaming data using the same code.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=271.6708571428571) [You have batch data fed in from, say,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=274.4813333333334) [Cloud Storage buckets and streaming data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=277.7076666666666) [from Pub/Sub, and both of these pass](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=279.6247142857142) [through the same Dataflow pipeline. And in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=282.3927142857143) [this way by integrating batch and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=285.37) [streaming, you maintain exactly one code](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=286.999) [base for simplicity. In its simplest form,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=289.331) [the Kappa Architecture would simply](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=291.937) [integrate batch and streaming and save the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=293.687) [data out to a data warehouse such as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=296.487) [BigQuery. But if your client applications](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=299.8389999999999) [have different requirements which need](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=301.4879999999999) [different storage services, you might want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=303.5292) [to save your data out to multiple services](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=305.6208571428571) [such as Spanner and BigTable as well. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=308.2355000000001) [key idea behind the Kappa Architecture is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=311.9651428571429) [the single processing pipeline, whether it's batch data or streaming data.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=5&mode=live&start=314.51160000000004)

[Setting up a GCS Bucket, PubSub Topic, and BigQuery Table](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live)

[In this demo, we'll see an example of how](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=1.967) [we can use the same code for batch, as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=3.597) [well as stream processing. We'll be using](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=6.108) [Cloud Dataflow to process batch data from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=9.303285714285712) [Cloud Storage and stream data from cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=12.067428571428573) [Pub/Sub. Engineers at SpikeySales want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=15.07) [analyze their historical order-related](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=17.54) [information, along with live orders that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=20.50475) [they receive on Pub/Sub, and see whether](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=22.809375000000003) [there are any trends or any insights that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=24.984571428571424) [they can extract. They want to batch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=28.10475) [process the historical orders and stream](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=30.30742857142857) [process the live orders and save them both](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=32.03885714285714) [in BigQuery. We are currently logged in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=35.01400000000001) [with our G Suite account to GCP's web](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=36.78477777777779) [console. We are working in the spikey-da](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=39.422999999999995) [project. Whichever project you're using](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=42.60500000000002) [for your demos, make sure that you have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=44.03000000000001) [billing enabled for that project so that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=46.36057142857144) [you can provision resources and use other](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=47.72550000000002) [features on the GCP. When you want to run](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=50.44466666666665) [code on the Google Cloud Platform, you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=53.20600000000001) [should have the code run using service](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=55.26371428571429) [account credentials. Use the Navigation](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=58.031571428571425) [menu, head over to IAM and admin, and go](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=59.846) [to Service accounts. Click on the Create](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=62.91814285714285) [Service Account button up top here in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=65.12957142857141) [order to create a new service account. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=66.77444444444447)[going to call this service account the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=68.65685714285716) [spikey-kappa-arch service account, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=71.79000000000002) [that's my service account ID as well.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=73.14999999999999) [Click on the Create button. That will](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=75.21933333333335) [create a service account for you. The next](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=77.51775) [step is to select a role to associate with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=79.66577777777779) [this service account. For the sake of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=81.95719999999999) [simplicity, I'm going to give the service](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=84.02099999999999) [account the project owner role. I want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=85.91762499999999) [this service account to have full full](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=88.33733333333332) [access to all of the resources on the GCP.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=89.86749999999999) [However, when you're working with service](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=92.383) [accounts in the real world, you'd, of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=93.814) [course, want to give it more granular](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=95.47955555555556) [access. Click on the Continue button here.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=97.6884444444445) [And the next step is to create a key](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=99.578) [associated with the service account. Our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=101.946) [cloud Dataflow program will use this key](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=104.34942857142858) [to get programmatic access to resources.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=106.69257142857143) [This key is going to be a JSON key that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=109.396) [I'll now download to my local machine.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=111.25) [Once I have my service account key, I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=113.867) [now ready to set up my data for batch, as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=115.64388888888884) [well as stream processing. Head over to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=117.90180000000005) [Storage and Browser. I'm going to create a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=120.70700000000001) [Cloud Storage bucket here that will hold](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=123.13037499999997) [my batch processing data. The name of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=125.42249999999999) [bucket on the GCP has to be globally](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=127.67181818181815) [unique. I'm going to call this the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=129.1825714285714) [spikey-historical bucket because this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=131.25) [holds my historical data that I want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=133.074) [integrate with live data that I received](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=135.705) [from Pub/Sub. We'll go with the defaults](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=137.826) [here and have this bucket be a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=140.195) [multi-regional bucket located in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=141.766) [United States. Go ahead and click on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=144.1831428571428) [Create button here to get this bucket all](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=146.794) [set up. Once the bucket has been created,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=149.16) [I'm going to upload a CSV file that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=151.03724999999994) [represents my historical data. This, of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=153.92849999999999) [course, is a very simple file with just](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=156.20871428571425) [about 30 rows of historical data. It's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=158.86200000000005) [called inputs\_orders\_small.csv. The schema](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=162.22766666666666) [for the data that we've uploaded is very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=165.01674999999997)[simple and straightforward. It's simply an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=167.464) [order\_id, which is a string, and an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=169.699375) [order\_size. The order\_id is simply a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=171.952) [unique identifier for an order placed on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=174.19225) [the e-commerce site, and the order\_size is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=176.539) [the price that the customer paid for that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=179.337) [order. The batch, as well as the streaming](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=182.44366666666656) [data, will have the same schema. Now let's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=184.6029999999999) [head over to Pub/Sub and go to Topics.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=187.5642) [This is where we'll create the Pub/Sub](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=189.757) [topic to which live orders are posted.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=191.292) [Right here on this page, click on Create a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=194.347) [topic. If you've never used Pub/Sub in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=196.52950000000007) [this project before, you'll have to enable](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=198.26685714285713) [the Pub/Sub API first. I'm going to call](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=200.0014285714286) [this topic live\_sales because it gives us](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=202.71300000000002) [live streaming information about sales](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=205.40414285714292) [that occur on our SpikeySales site. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=207.4204) [going to right-click on Google Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=210.5484285714286) [Platform and open up a new tab for my web](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=211.907) [console dashboard. I'm going to use this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=214.387) [new tab to navigate over to BigQuery.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=217.03100000000006) [BigQuery, as you probably already know, is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=219.26750000000004) [the serverless big data warehouse on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=222.03400000000002) [GCP. Now data analysts want historical](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=225.2628333333334) [data and live data to be integrated and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=227.60771428571422) [present together on BigQuery for analysis.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=230.40714285714293) [BigQuery is the serving layer for my Kappa](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=233.6) [Architecture. Under the spikey-da project,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=236.44177777777767) [click on Create dataset, and create a new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=237.88633333333334) [dataset called spikey\_orders. Click on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=240.45385714285712) [Create dataset button here. And once this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=243.86857142857147)[dataset has been created, let's create a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=246.08342857142858) [new BigQuery table under this dataset. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=248.36914285714283) [going to call this table the orders table.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=250.59312500000004) [Now instead of starting off with a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=253.36628571428568) [completely empty table, I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=256.056) [populate it with a few orders. Now this is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=257.912) [how typically it is in the real world.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=260.4131818181817) [Maybe an analyst or a data scientist has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=262.347) [started using BigQuery, and that's when](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=264.555) [people realize that that's a great option](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=266.315) [for the serving layer, and they start](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=268.499) [trying to move all of their data to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=270.275) [BigQuery. I'm going to pre-populate this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=272.803888888889) [table with a few records, which I'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=274.339375) [upload from my local machine. The file](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=276.1035714285714) [format of my original record is the CSV](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=278.3957142857142) [format, so choose the CSV format, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=281.74566666666675) [let's select a file from our local](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=284.2114999999999) [machine. This file is the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=286.65766666666667) [existing\_data.csv file. Existing\_data.csv](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=289.27866666666665) [contains orders that are already present](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=291.02139999999997) [in our BigQuery table. The table name is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=294.4936666666667) [orders. The next step is to specify a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=298.2374000000001)[schema for this table. I'm going to slide](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=299.99455555555556) [this over to Edit as text, and here is my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=302.339) [schema in the text format. We have the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=305.1689999999999) [order\_id, which is of type string, and the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=307.62271428571427) [order\_size, which is of type float. Click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=309.572) [on the Create table button here, and our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=312.7325714285714) [orders table will soon be created and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=315.38075000000003) [pre-populated with a few existing records.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=317.58214285714286) [I've just gone with two pre-existing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=320.229) [records in our example here. The records](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=321.843) [that you see on BigQuery here will give](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=324.85625) [you a sense of the final format of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=326.9271111111111) [data that we want to store in BigQuery. We](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=329.47659999999996) [have the order\_id and the order\_size.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=331.7787142857143) [We'll now set up a cloud Dataflow](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=334.057) [pipeline, which does some simple](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=336.03362500000003) [processing on the historical batch data,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=337.5434285714286) [as well as the incoming streaming data, and stores it in BigQuery in this format.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=6&mode=live&start=339.5388571428572)

[Implementing Integrated Processing for Batch and Streaming](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live)

[We are now ready to get started with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=1.683) [writing our cloud Dataflow code in Python,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=3.215) [which will integrate batch as well as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=6.373) [stream processing. In order to write code,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=7.996) [I'm going to first activate Cloud Shell.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=10.618) [This is the ephemeral VM on the GCP that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=13.008) [comes preinstalled with all of GCP's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=15.088) [command line tools and utilities that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=17.443) [we'll need. Click on this pencil icon that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=19.779999999999994) [you see here at the center of your screen.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=21.975000000000005) [This will launch code editor. Code editor](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=24.028) [is a browser-based IDE offered on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=27.294) [Google Cloud Platform. This makes it very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=29.944599999999998) [easy for you to write your code and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=31.687777777777775) [configuration files right here within your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=34.16149999999999) [GCP web console. This is the home](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=36.786500000000004) [directory of my Cloud Shell virtual](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=38.95049999999999) [machine. I'm going to right-click here and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=40.87200000000001) [create a new folder within which I'm going](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=42.95425000000001) [to write my code. I'll call this folder](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=45.19150000000001) [kappa\_arch. Within the kappa\_arch folder,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=49.0922857142857) [let's create a new file that will contain](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=50.93766666666666) [my batch and stream processing code. I'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=53.46833333333332) [call this batch\_stream.py. I've expanded](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=56.559) [my code editor to be full-screen. I'm now](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=59.65175) [going to paste in the entire code and walk](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=62.07888888888889) [you through it step-by-step. This code is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=64.84866666666666) [written using Python 3. First comes all of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=67.55685714285715) [the import for the libraries that we'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=69.929) [need. Cloud Dataflow uses the apache\_beam](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=72.0895) [APIs, so there are a bunch of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=75.14466666666668) [beam-specific imports as well. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=77.08257142857143) [spikey\_schema variable here holds the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=79.75239999999998)[schema of the BigQuery table to which we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=81.42899999999999) [want to finally write out data from the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=83.60000000000001) [historical batch records, as well as from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=86.45057142857141) [the incoming stream. Since this is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=89.12574999999998) [simple demo to show you how batch and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=91.08000000000001) [streaming architectures can be integrated,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=92.94766666666666) [the actual processing that we do on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=94.53999999999999)[incoming data is going to be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=97.3205) [straightforward. The only processing we do](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=99.2315) [is to extract the order\_id and the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=100.76722222222224) [order\_size from individual records. If](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=103.571)[you're working with data in the real](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=106.16575) [world, there's, of course, much more](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=107.249) [processing and cleaning up that you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=108.781) [have to do. The class](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=110.20728571428575)[ExtractOrderIdAndOrderSize inherits from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=113.66899999999998) [the beam.DoFn, which can be implemented in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=115.26814285714286) [parallel across your dataset. The process](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=118.23600000000002) [function is what you define when you use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=120.66199999999999) [apache\_beam in order to process individual](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=122.81085714285713) [records one at a time. The element is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=124.88971428571429) [single record. We extract the order\_id and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=127.76266666666669) [the order\_size by splitting the record on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=130.0041428571429) [the comma. The input records from our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=132.72157142857142) [historical data, as well as from the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=135.58299999999997) [incoming stream, is a comma-separated](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=137.54188888888893) [value. The order\_size is available as a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=139.245) [string. We need to convert it to a float](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=141.681) [before we save it out to BigQuery. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=143.4438000000001) [next step is to return the data in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=145.726) [JSON format because this is the format](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=147.796) [that BigQuery recognizes. The keys for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=149.91437500000004) [this JSON correspond to the column or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=152.57700000000006) [field names of our BigQuery table. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=155.39585714285712) [next step is to see how the beam pipeline](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=158.05177777777777) [is set up. The beam pipeline is defined in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=159.989125) [the run function. The first thing we do is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=162.51659999999998) [to parse the command line arguments to get](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=165.43099999999998) [the GCS bucket which contains our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=168.02414285714283) [historical data and the Pub/Sub topic to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=170.47966666666667) [which live sales are published. Both of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=173.67425) [these are required parameters for our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=176.69857142857137) [Dataflow job. We'll pass in other](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=179.2328) [arguments on the command line, which are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=180.9348571428572) [required by the beam pipeline itself. Once](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=182.55842857142858) [we've extracted these arguments,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=185.62460000000002) [instantiate the PipelineOptions, and set](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=186.955) [up the pipeline. Here are two flags that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=188.855) [have to be set to true for this pipeline](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=191.86174999999997) [to work. Save\_main\_session has to be true](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=194.31139999999996) [so that a global namespace is pickled and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=196.27499999999995)[loaded onto our cloud Dataflow workers.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=199.5598571428571) [This makes what we've declared globally](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=202.358) [available to individual workers that may](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=204.471) [be running on different machines. If you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=206.22642857142853) [don't set save\_main\_session to true,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=209.19214285714284) [individual workers which are processing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=210.895) [jobs in parallel won't be able to access](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=213.075) [the libraries that we've imported in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=215.551) [global namespace. And the streaming flag](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=217.329) [has to be set to true because we're going](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=219.879) [to be processing streaming data from a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=221.511) [Pub/Sub subscription. We'll be setting up](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=223.823) [two parallel processing paths within this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=226.92842857142858) [cloud Dataflow code, one to process the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=229.1865) [historical records, the second to process](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=231.99466666666663) [streaming data. The historical records are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=234.4236666666667) [available in a GCS bucket. We use beam.io](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=236.29199999999997) [.ReadFromText and read in from Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=240.1961666666667) [Storage. This is a Dataflow connector that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=242.82649999999998) [you'll used to work with Cloud Storage.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=244.53842857142862) [Once we've read in the CSV records, we'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=247.243) [apply the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=250.5675)[ParDo(ExtractOrderIdAndOrderSize) in order](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=252.6565) [to extract granular information from each](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=254.45266666666666) [record. This is the pipeline step where we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=256.73657142857144) [format the data before we write it out to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=259.51187500000003) [BigQuery. And once we have the formatted](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=262.3258000000001) [data, we can use the Dataflow connector to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=264.3340000000001) [write out to BigQuery beam.io](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=267.0586250000001)[.WriteToBigQuery. We'll write out our data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=270.3196666666667) [to a BigQuery table. Notice how we specify](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=272.9238888888889) [the table to which we want to write. We](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=275.03499999999997) [have the project ID followed by a colon,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=277.352) [then the dataset ID, then a dot, then the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=280.05549999999994) [table ID. Here are some additional flags](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=283.3605714285714) [you can pass into the WriteToBigQuery](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=285.23)[method. We specify the schema for the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=287.64680000000004) [BigQuery table to which we are writing.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=289.78271428571435) [The create\_disposition says we don't want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=292.137) [to create this table. It already exists.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=293.997) [And we want to append data to it. We don't](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=296.009625) [want to overwrite data. This completes the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=299.13177777777764) [processing of our historical batch data.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=301.0759999999999)[Let's now move on to the live data that we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=303.516) [receive from our Pub/Sub topic. In order](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=305.462) [to read from a Pub/Sub topic, we use the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=308.267) [Dataflow connector for Pub/Sub. Beam.io](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=310.91900000000004) [.ReadFromPubSub simply takes in the topic](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=312.79049999999995) [that you want to read from, creates a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=315.68944444444446) [subscription, and reads messages from that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=317.781) [topic. The message that you received from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=320.93442857142855) [the Pub/Sub topic is in the UTF-8 format,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=322.426) [which will need to be decoded. We then](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=325.439) [apply the same processing to our streaming](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=327.99875) [data that we apply to our batch data. We](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=329.8538888888889) [extract the order\_id and the order\_size.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=333.30600000000004) [Before this stage comes all of the cleanup](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=335.928) [for batch and streaming data. This is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=338.016) [where the pipeline gets integrated. And](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=340.2017142857143) [once data is in the right format, we use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=342.56112499999995) [the same beam.io .WriteToBigQuery function](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=345.45266666666663) [to write live data out to BigQuery as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=348.15599999999995) [well. Once both batch as well as streaming](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=350.981125) [data are moved to the serving layer, they](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=353.7735714285714) [can be treated in exactly the same way.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=355.86699999999996) [And this completes the code for our simple](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=358.11349999999993) [processing as a part of our Kappa pipeline](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=360.821) [Architecture. We simply run the pipeline and wait until finish.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=7&mode=live&start=364.059)

[Executing the Pipeline for Batch and Stream Processing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live)

[We are now going to run this Dataflow job](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=1.756) [on our local Cloud Shell virtual machine.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=3.616) [But for that, we need service account](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=6.058) [credentials. Use the three-dot menu on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=8.708374999999997) [top right to upload the service account](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=10.550333333333338) [credentials file onto our Cloud Shell VM.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=13.370000000000001) [This file is now present in our home](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=16.822)[directory. Let's cd into the kappa\_arch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=18.967777777777766) [directory where we have our Python code.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=21.428500000000007) [I'm now going to use the mv command to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=24.597) [move this credentials file into the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=27.003599999999988) [kappa\_arch directory so that I have all](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=29.184333333333328) [files present under my current working](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=31.427749999999996) [directory. Here's the batch\_stream.py and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=33.57599999999999) [our service account credentials file. In](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=36.03333333333333) [order for the code that you run on the GCP](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=39.1016) [to pick up the right service account](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=40.899625) [credentials, you need to set up this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=42.675)[environment variable called](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=44.506) [GOOGLE\_APPLICATION\_CREDENTIALS and point](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=46.227) [it to your service account JSON file. Code](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=48.18375) [running on the GCP will look for this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=51.55777777777778) [environment variable in order to get the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=53.044) [credentials with which it needs to run. We](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=55.507) [are now going to run our Dataflow program](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=58.259625) [on the local Cloud Shell VM. For that we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=60.04857142857143) [need to install the beam libraries first,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=62.97775) [which you can do with a simple pip](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=64.39) [install. Once beam has been installed,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=66.6362857142857)[let's go ahead and run this Dataflow code](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=69.24057142857143) [on our local Cloud Shell VM. I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=71.92800000000001) [run this code locally so that you can see](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=74.47966666666666) [the various print statements I've pasted](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=76.334) [in when I process the batch data. When you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=78.2562222222222) [run locally, apache\_beam uses the direct](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=81.66914285714286) [runner to run your Dataflow code. So we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=83.89514285714286) [call python batch\_stream.py, and here are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=87.33914285714287) [the various input arguments that you need](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=90.24285714285715) [to specify to your Dataflow code. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=92.26828571428571) [input\_file argument contains the GCS](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=95.24316666666667) [bucket where we've uploaded our historical](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=97.209) [batch data. The input\_topic argument](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=99.308) [points to the Pub/Sub topic on which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=102.872) [messages will be published for our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=104.685) [live\_sales. The project that we are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=107.03442857142855) [working in is spikey-da. And the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=108.83100000000003) [temp\_location that Dataflow can use for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=110.47614285714285) [staging is our spikey-historical bucket.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=112.33449999999999) [Wait for a little bit, and you'll find](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=115.112) [that the batch data that is present in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=116.86311111111107) [your Cloud Storage bucket is immediately](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=118.7095) [available and will be processed by our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=120.902) [program. It's now waiting on stream data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=123.99600000000004) [from Pub/Sub. If our batch data has been](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=126.94675000000002) [processed, that means it should be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=129.39357142857136) [available in BigQuery. Let's switch over](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=130.78324999999995) [to the BigQuery web UI, and go to the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=132.33237499999998) [orders table. I'm going to run a simple](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=135.0236666666667) [query here which simply selects stars from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=136.9736666666667) [the orders table. Click on Run to execute](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=139.6981428571428) [this query, and you'll find we have 32](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=142.79885714285714) [records in this table. The 30 records from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=145.67380000000006) [our batch data have been processed and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=148.32) [added to BigQuery. Two of these were](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=150.7055) [pre-existing records, and 30 are from our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=153.29866666666666) [Cloud Storage bucket. Now let's switch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=155.91937500000006) [over to the Browser tab where we are on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=157.83450000000005) [our Pub/Sub Topics page. Here is the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=160.46550000000002) [live\_sales topic. This is what our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=163.00459999999998)[Dataflow program is listening to. Click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=165.02414285714283) [through and let's now publish a message](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=166.99385714285717) [that will represent a live sale. As people](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=169.44671428571428) [buy products on SpikeySales.com, their](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=172.71757142857146) [order ID and the price that they paid are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=174.92966666666666) [published to this topic. That's what we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=178.3012222222222) [are assuming here. This is the message.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=180.35450000000003) [The order ID is OOO so that it's easy for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=182.36644444444443) [us to identify it. Publish this message.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=185.22811111111105) [Our Dataflow program is running. It'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=187.245) [process it. Let's head over to BigQuery](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=189.10842857142856) [and query our table once again. Notice](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=190.96214285714285) [that we now have 33 records. The new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=193.71085714285715) [record from Pub/Sub has been processed by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=196.95028571428574) [the same Dataflow code and added to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=198.87985714285716) [BigQuery. Let's go back to Pub/Sub and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=201.48699999999997) [publish a second message here. This](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=204.14914285714286) [contains the order AAA. Hit Publish, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=206.8262) [let's go execute our query on BigQuery](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=209.55924999999996) [once again. And you'll find that this new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=211.8195) [record has been added as well. I was](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=214.01325000000003) [having a lot of fun here, so I tried this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=215.97580000000008) [with one more message. I published ZZZ,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=218.50420000000005) [and once this was published, I switched](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=220.72814285714284) [over to BigQuery, executed the query to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=223.31199999999995) [SELECT \*, and ZZZ is now part of my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=225.85344444444442) [serving layer. With this we've](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=229.32566666666665) [successfully written Dataflow code that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=231.20559999999998) [integrated the processing of batch as well](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=233.3582) [as streaming data. And with this demo, we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=236.22199999999995) [come to the very end of this module where](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=238.46100000000004) [we discussed how it's important for us to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=240.669) [integrate batch and stream processing. We](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=242.94042857142858) [saw that data architectures are now being](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=244.829) [designed around machine learning rather](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=247.19633333333334) [than how data is stored. This means that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=249.15862499999997) [stream processing has become integral to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=252.138) [any design. Processing live streams from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=254.6946) [the real world is no longer an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=257.14271428571425) [afterthought or an add-on. We saw that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=259.5785) [there are a number of aspects of designing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=261.7416666666667) [data architectures that are now](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=263.78766666666667) [standardized. However, we have to plan for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=265.52) [a future in which real-time prediction for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=268.5501428571429) [machine learning models is critically](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=271.664) [important. And as things stand today, we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=273.90900000000005) [haven't really built our data processing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=276.23514285714276) [pipelines around the best practices](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=278.5755000000001) [involved for stream processing. An](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=280.501) [important step in this direction would be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=282.477) [to move towards an integrated pipeline for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=284.8844999999999) [batch as well as stream processing. In the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=287.20271428571436) [next module, we'll continue to talk about](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=290.3796666666667) [data, but we'll focus on designing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=293.2445000000001)[migration and disaster recovery strategies on the GCP.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=cf7c0c25-dc98-4af7-8191-bb8b126d2603&clip=8&mode=live&start=295.6785714285714)

[Designing Migration and Disaster Recovery Strategies on the GCP](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live)

[Module Overview](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live)

[Hi, and welcome to this module on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=0) [Designing Migration and Disaster Recovery](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=2.86125) [Strategies on the GCP. You'll find that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=5.3492) [when you're migrating to the cloud from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=8.358) [your on-premises data center or moving to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=10.351) [a new cloud provider, your migration has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=12.398500000000004) [become less data-centric because data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=14.462) [migration is easy. It is more](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=16.902) [code-centric. You have to think more](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=20.411142857142856) [deeply about how you migrate your code.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=22.19711111111111) [And code migration often tends to be very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=24.6175)[specific to your use case. In the case of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=28.1125) [migrating data, make sure that when you're](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=31.385499999999997) [moving to the GCP, avoid prematurely](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=33.50714285714285) [committing to a storage or compute choice.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=35.49833333333333) [To make sure that you're making the right](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=38.10014285714285) [technology choice once you have your data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=40.999) [on the cloud, keep things simple. Get your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=42.99825000000001) [data into Cloud Storage buckets or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=45.31757142857142) [BigQuery first, and then worry about what](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=47.70366666666667) [you need to do later. As far as disaster](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=50.1075) [recovery goes, you'll see that you'll only](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=53.17133333333333) [need to focus on disaster recovery plans](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=55.049250000000015) [for your on-premises installation. You](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=57.138) [don't really have to worry about disasters](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=60.27571428571429)[affecting the GCP because Google has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=61.94916666666666) [automatically taken care of that for you.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=64.44562499999999) [Almost every service on the GCP will](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=66.97) [provide you with options that you can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=68.864) [configure in order to keep your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=70.666) [applications running even in the face of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=72.726) [disaster. But what's really critical for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=75.63449999999996) [you to protect when you're working on an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=77.42571428571426) [on-premises installation is your data.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=79.763) [Leverage the Transfer Service to make sure](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=82.178) [that your data is backed up to the Google Cloud.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=0&mode=live&start=84.25)

[Migrating Structured Data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live)

[Let's say that your organization has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=1.304) [decided to move from an on-premise data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=2.861) [center to the cloud, specifically the GCP.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=5.285) [If you have a fully functioning technology](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=8.965) [set up on-premises, here are some of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=10.925) [things that you'll have to consider](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=13.269) [migrating--structured data, unstructured](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=14.625) [data, your big data processing, client](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=16.5684) [applications, ML models, and the scripts](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=19.337200000000003) [and glue that you have holding everything](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=22.334875000000004) [together. The migration strategies that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=25.226333333333336) [you choose for each of these might be very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=27.031) [different and specific to your org. Here](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=29.565999999999992) [I'm going to talk about how you might want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=32.037333333333336) [to migrate structured data and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=33.77733333333333) [unstructured data. You'll find that cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=35.504) [providers are making it increasingly](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=38.472) [simple for you to migrate your data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=40.53866666666667)[because what they've realized is that once](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=43.16000000000001) [you have your data in the cloud, moving](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=45.045375) [compute over can be done piecemeal](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=47.480625) [application by application. Let's start](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=50.086999999999996) [off with seeing how you can migrate your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=51.9735) [structured data to the GCP. Now the GCP](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=53.876000000000005) [does not have an equivalent of the AWS](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=56.81380000000001) [Database Migration Service, so the easiest](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=59.88671428571427) [way for you to get your data to the Google](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=61.89942857142859) [Cloud is to migrate to Cloud Storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=64.59360000000001) [buckets. On the GCP just like with any](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=67.66833333333335) [other cloud provider, there are a number](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=69.90044444444443) [of services that you can use to store your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=71.73485714285717) [data. Now before you commit to any](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=74.48400000000002)[technology, make sure that you understand](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=76.52542857142855) [the entire taxonomy of GCP storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=77.913) [services. Don't jump into committing to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=80.62499999999999) [any of them during the process of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=82.63737500000002) [migration. The easiest way is to migrate](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=85.32719999999999) [your data into Cloud Storage and BigQuery.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=87.76374999999997) [BigQuery is Google's data warehouse where](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=91.154) [you can store structured data. Once you've](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=93.448) [got your data on the GCP, it's best to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=96.369) [adopt a wait-and-watch approach to see](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=99.43299999999999) [what service makes sense for your use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=101.61671428571427) [case. A good question to ask right now](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=103.78733333333335) [would be, Why Cloud Storage and BigQuery](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=106.15025) [and not any of the other services? This is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=107.999875)[because these are stable, proven, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=111.06) [widely used, and they make it very easy](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=113.16844444444445) [for you to get data into them. These](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=115.156) [services have been built with the express](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=118.07428571428572) [purpose that you'll need to import huge](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=119.77728571428572) [amounts of data into them. In fact, you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=121.90766666666667) [can simply store your data in Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=124.23166666666668) [Storage and query it using BigQuery. Let's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=125.94314285714286) [say on your on-premises data center you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=129.40585714285714) [have data stored in different services,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=131.43300000000002) [maybe a relational database, HBase.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=133.449)[Individual services differ from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=136.345) [open-source or competing cloud versions.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=138.369) [And this means that if you try to directly](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=141.42) [migrate data to the managed versions of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=143.763) [these services, this might prove](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=146.907) [problematic. Let's consider one such](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=149.71866666666665) [example for structured data. Let's say](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=151.38100000000003) [you're using Redis as a cache for the data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=153.14966666666666) [that you served from your on-premises data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=156.02777777777774) [center, and you want to use Google Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=159.00266666666658) [Memorystore for Redis. Memorystore is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=160.988)[managed Redis on the GCP, and it's a newly](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=163.89042857142857) [launched product, and it runs Redis](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=167.06799999999996) [version 3.2. But if you look at AWS](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=169.35499999999996) [ElasticCache, that runs version 5. Now you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=172.5755714285715) [might get lucky and this mismatch in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=174.931) [versions may not cause any issues for you,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=176.96600000000007) [but who knows? Why take the chance? Move](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=179.954) [to Cloud Storage, and then we'll see.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=182.644375) [Another example here that might be more](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=185.146) [relevant is that of Google Cloud SQL for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=186.847) [PostgreSQL. Let's say you're trying to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=191.86122222222232) [migrate from AWS. AWS runs version 10.6 of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=194.16174999999993) [PostgreSQL. GCP runs 9.6, and the current](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=198.40183333333331) [latest version out there in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=201.68042857142862) [open-source world is version 11. All of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=203.62828571428565) [these version mismatches can cause major](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=206.21057142857146) [headaches. It's better for you to bite the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=208.9028) [bullet, move to Cloud Storage buckets or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=211.00025000000002)[BigQuery, and then slowly move to the service of your choice.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=1&mode=live&start=213.57199999999992)

[Storage Options on the GCP](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live)

[Let's take a quick look at the services](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=0.866) [for storage that GCP has to offer. The two](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=2.9708888888888882) [kinds of data that you might want to store](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=6.789666666666666) [are unstructured data and structured data.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=8.870666666666667) [The best choice for unstructured data in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=11.466333333333333) [my opinion is Cloud Storage buckets. This](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=14.387857142857143) [offers blob or object storage, and it has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=17.02)[pay-as-you-go functionality. It's elastic](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=21.325714285714287) [storage that grows to meet your needs, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=23.122) [you only pay for what you store. If you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=25.546) [want block storage for your unstructured](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=28.305999999999997) [data, you'll go with persistent disks.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=29.8302) [Persistent disks need to be provisioned,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=32.069) [and you pay for what you provision, not](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=34.1015) [for what you use. And persistent disks on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=36.452666666666666) [the GCP happen to be zonal resources,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=39.38450000000001) [which means you can only have instances in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=41.36528571428572) [the same zone use data stored in your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=44.218999999999994) [persistent disks, whereas data stored in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=47.146333333333345) [Cloud Storage buckets can be accessed from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=48.94800000000001) [anywhere. If you want to store structured](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=52.00783333333332) [data on the GCP, you have more choices](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=53.80200000000001) [available. You first have to decide](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=56.6225) [whether you need to run SQL queries on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=58.232624999999985) [this data or its NoSQL. Let's say your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=61.34200000000001) [data is relational in nature, and you know](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=64.36314285714286) [that you need to run SQL queries. You'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=66.0491111111111) [then need to consider whether you want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=68.1882857142857) [use this data for transaction processing,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=70.5635) [or OLTP, or for analytical processing, or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=72.38733333333334) [OLAP. If you want your data to be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=76.55925000000002) [automatically updated, you want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=78.99549999999998) [transaction processing and ACID](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=80.54100000000001) [properties, you'll choose to go with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=82.17642857142857) [either Cloud SQL or Cloud Spanner. Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=84.59071428571428) [SQL is the familiar SQL RDBMS, which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=88.058)[supports both MySQL as well as PostgreSQL.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=90.75825) [When you use Cloud SQL on the GCP, though,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=93.994) [you can configure it in high availability](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=95.866) [mode. Google will automatically provision](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=98.43762499999997) [a failover replica to make sure that your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=100.2362) [data is always available. Cloud Spanner,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=102.584) [on the other hand, is an interesting](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=105.08324999999999)[proprietary offering from Google. There is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=106.8142) [no equivalent on other cloud providers.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=109.77325) [This gives you transactional properties](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=111.856) [and ACID support at scale and in a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=113.73183333333336) [globally distributed way. Cloud Spanner is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=116.58325000000004) [expensive, but it can horizontally scale](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=118.928) [to support huge amounts of data. If you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=121.29371428571429) [want SQL support for analytical](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=124.97657142857143) [processing, Google has one choice for you,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=126.338) [and that's a great one. You can go with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=128.428) [the serverless BigQuery warehouse. If you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=131.65849999999998) [have NoSQL data to store, you have two](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=133.99622222222223) [options. You can go with BigTable, which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=136.39414285714284) [offers you a columnar store, kind of like](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=138.8355) [HBase but far more reliable and scalable,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=141.09728571428573) [or you can choose Cloud Datastore where](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=143.4492857142857) [data is stored in the form of documents which have attributes or properties.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=2&mode=live&start=147.026375)

[Migrating Unstructured Data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live)

[We've discussed moving structured data to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=1.15) [the GCP. Let's talk about migrating](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=3.14275) [unstructured data. Unstructured data, it](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=5.993666666666667) [turns out, is very straightforward to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=7.885) [migrate. A very useful bit of software](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=10.421666666666665) [that GCP makes available is the Transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=12.699999999999998) [Service. This is what you'd to use to get](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=15.240666666666666)[unstructured data from another cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=16.998) [provider such as AWS or that you have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=18.978) [exposed behind an externally accessible](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=21.506) [URL. The Transfer Service is a great](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=24.99933333333334) [option for you to move unstructured data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=27.227999999999994) [into your Cloud Storage buckets. Another](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=29.452285714285708) [good option to get data into your GCS](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=32.27175) [buckets is to use gsutil. This is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=34.70883333333333) [command line utility that you can use to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=37.517999999999994) [work with GCS. The Transfer Service you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=39.840799999999994) [find is the migration tool of choice to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=42.65185714285714) [get unstructured data into your Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=45.23033333333333) [Storage buckets. You can migrate or back](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=47.823714285714274) [up data from your on-premises data center](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=49.74244444444445) [or from another cloud provider such as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=51.830000000000005) [AWS. The Transfer Service doesn't just](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=55.13239999999999) [work with other cloud providers. You can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=56.905857142857144) [also use this to move your data from one](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=59.21077777777777) [Cloud Storage bucket to another. For](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=62.217) [example, for data archival purposes, you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=64.95383333333334) [might want to move your data from a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=67.22588888888889) [multi-regional bucket to a coldline or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=68.729) [nearline bucket where storage costs are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=71.241) [lower. In just a little bit, we'll see a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=74.76300000000002) [number of hands-on demos for how we can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=76.6412) [use the Transfer Service. The Transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=78.87875) [Service works with an AWS S3 bucket as the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=81.13139999999999) [source of data, any URL location, HTTP or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=84.15722222222226) [HTTPS as long as it's externally](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=87.93288888888888) [accessible. You can also configure the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=90.88440000000003) [source to be another GCS bucket. The sink](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=92.40724999999999) [for the Transfer Service is always a Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=95.02957142857143) [Storage bucket. That's where you want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=98.55725) [get your data to. If your data is present](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=101.13377777777781) [on-premises and it's not available behind](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=104.09085714285716) [an externally accessible URL, you might](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=105.98599999999998) [prefer to use gsutil to copy from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=108.23985714285716) [on-premises to GCP. If your data is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=111.82642857142858) [available behind a publicly accessible](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=114.5845) [URL, or you're transferring from AWS,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=116.01616666666666) [you'll prefer the Transfer Service. Google](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=118.7675) [wants to make it as easy as possible for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=121.631) [you to get your data to the cloud, so it](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=123.93322222222221) [offers an off-line data migration service](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=126.30014285714286) [as well using something called a transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=128.1882857142857) [appliance. The transfer appliance is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=130.308) [high-capacity NAS or network attached](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=133.7585714285714) [storage supplied by Google on request.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=136.505) [You'll connect the transfer appliance](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=139.577) [within your data center and copy over all](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=141.219) [of the data that you want to migrate into](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=143.715) [it. You'll then ship the device to Google,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=146.2394999999999) [and Google will take care of uploading all](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=148.6) [of this data into a Cloud Storage bucket.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=150.795) [Google offers support for up to a petabyte](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=154.227) [of data migration using the transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=156.827) [appliance. This kind of off-line data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=160.42957142857148) [migration is a great choice if you're](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=162.39250000000004) [working from a location where the internet](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=164.14257142857147) [connections are poor and there is a lot of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=166.27599999999998) [data that you need to transfer. Google](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=168.4048999999999) [wants to make sure that your data is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=170.49571428571426) [secure. Data is encrypted at the moment of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=173.32999999999998) [capture. So if all other migration](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=176.27800000000008) [strategies don't work for you, the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=178.06714285714287) [transfer appliance is something that you can fall back on.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=3&mode=live&start=179.9322857142857)

[Transfer Data from a Regional to a Multi Regional Bucket](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live)

[If you want to bulk transfer data to your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=2.1) [Cloud Storage buckets on the GCP, you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=3.536) [use the Transfer Service. You can also use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=6.3952) [the Transfer Service to move data within](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=8.892000000000003) [GCP buckets, and that's what we'll see in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=10.59657142857143) [this demo. I've logged in to the GCP web](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=13.459333333333337) [console. I'm headed over to Cloud Storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=15.979125)[buckets. Here I'm going to create a new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=18.55857142857142) [bucket, which will contain the source data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=20.71624999999999) [that I want to transfer. I'm going to call](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=22.4195) [this a spikey-storage-asia bucket. This is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=24.30377777777779) [going to be a regional bucket that is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=27.28922222222222) [located in Asia. You can imagine that the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=29.149875) [customers of SpikeySales were originally](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=31.9462) [located in Asia, but they're now spread](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=33.571) [all over the US, which is why they want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=36.077888888888886) [move their data to a multi-regional](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=38.85077777777777) [bucket. Create this regional bucket that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=41.148111111111085) [will contain the source of our data. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=43.12683333333334) [going to use the Upload folder button here](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=46.0885) [to upload some source data that makes up](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=48.048875) [the static assets of my product catalog. I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=50.802749999999996) [selected the images folder, and I want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=53.98622222222222) [upload everything within that folder,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=55.724) [which contains a bunch of images of shoes.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=58.689)[Let's go into the images folder and see](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=62.065) [what's in there. Within the images folder,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=64.089) [we have a shoes folder. And within the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=65.936) [shoes folder, we have a bunch of JPEG](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=68.331) [images of shoes. These are the shoes that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=70.648) [make up our product catalog. Click on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=72.71749999999997) [Buckets link here to head back to the main](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=75.22371428571427)[page for Cloud Storage buckets. I'm now](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=77.56644444444449) [going to create a new bucket that is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=80.3318888888889) [multi-regional bucket. And I want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=83.377) [transfer data from my regional bucket to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=85.19899999999998) [this spikey-image-store multi-regional](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=87.67520000000002) [bucket. A multi-regional bucket is spread](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=89.82) [across multiple regions within a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=92.47966666666669) [continent, and it's typically used for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=94.83039999999998) [hot, frequently-accessed data. This](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=96.906) [multi-regional bucket is in the US. Click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=99.16857142857143) [on the Create button to create this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=101.46133333333333) [bucket. Once our source and destination](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=103.79466666666663) [buckets have been created, let's use the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=105.58716666666665) [transfer service, which you can do by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=107.95928571428571) [clicking on the Transfer (to your left)](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=109.77225000000003)[navigation pane, and transfer the data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=111.76516666666664) [from our spikey-storage-asia to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=114.37957142857142) [spikey-image-store. We haven't created any](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=117.38499999999999) [transfer yet. There are no jobs to show](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=119.84366666666668) [here. Click on the Create transfer button](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=121.69642857142858) [to configure your job. We start off by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=123.56700000000001) [selecting the source from where we want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=125.608) [transfer data. As you can see here](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=127.745) [onscreen, the source can be another GCS](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=129.961) [bucket, an Amazon S3 bucket, or just](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=132.456) [object URLs. We'll start off with the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=135.7771111111112) [Google Cloud Storage bucket. That's our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=137.99742857142857) [source. We want to transfer images from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=140.49640000000002) [spikey-storage-asia. That's our source GCS](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=143.5534285714286) [bucket. Hit Continue here, and you can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=146.04719999999995) [then move on to configuring the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=148.15333333333334) [destination bucket to which you want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=149.919) [transfer data. Because we have just one](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=152.5726666666667) [other bucket here, the destination bucket](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=154.843375) [has been pre-populated for you. This is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=156.48099999999997) [our multi-regional spikey-image-store. If](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=159.66799999999998) [you have other buckets available, you can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=161.99133333333333)[always use the Browse button to select one](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=163.6005) [of these buckets. Now there are a number](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=166.0864) [of transfer options that you can choose.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=168.79212500000006) [If you want to save on storage costs, you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=170.4551111111111) [might want to delete objects from the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=172.96550000000002) [source once they've been transferred, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=175.14228571428572) [that's the option I'm going to go with.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=176.89166666666665) [Click on the Continue button here. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=179.33999999999997) [next step is for you to set up the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=181.3526666666667) [configuration options for your transfer.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=184.15787500000002) [Let's say you have some kind of external](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=186.181)[program that's constantly uploading images](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=187.654) [to the spikey-storage-asia, that's your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=190.286) [source bucket, you might want to schedule](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=192.247) [this transfer on a daily basis. This way,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=194.546) [all of the data from your source bucket](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=197.91966666666664) [will always be transferred over to the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=199.65414285714286) [destination on a daily basis. For our demo](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=201.85966666666667) [here, we'll go with the Run now option.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=204.594) [Now it's time for us to give our transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=206.60755555555556) [job a meaningful name. It's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=210.15066666666667) [GCP\_regional\_to\_Multi-regional\_Once. Click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=213.36)[on the Create button, and that will set up](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=215.095) [your transfer. When I run this demo, this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=217.6155) [transfer took a couple of minutes to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=219.94699999999995) [process completely. In the meanwhile,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=221.98833333333334) [let's head over to the Configuration tab](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=223.332375) [and take a look at the current](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=224.92774999999997) [configuration of this transfer. You can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=226.9318) [see that the source bucket is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=229.08300000000003) [spikey-storage-asia. Spikey-image-store is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=230.199) [the destination bucket. When we switch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=233.13500000000002) [back to the operations page, it seems like](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=235.97450000000003) [the transfer is complete. You can select](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=238.10585714285716) [this transfer. This will bring up a panel](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=240.01659999999998) [with additional information. It still](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=242.55166666666665) [seems to be in progress. I wait a couple](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=244.0282857142857) [of minutes more, select this once again,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=246.02385714285708) [it has finally completed. Ten out of 10](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=248.49100000000004) [files have been transferred. We can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=251.792) [confirm the completion of this transfer by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=253.58014285714287) [heading over to Cloud Storage buckets.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=255.35299999999998) [Click on Browser here, and let's click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=257.648) [through to the spikey-image-store. You'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=260.113) [see that there is an images folder here](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=263.167) [that did not exist before. And within](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=265.2265) [images, there is the shoes folder. And](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=267.54525) [within shoes are all of our JPEG files.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=269.91322222222226) [Let's head back to the main Buckets page](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=273.707) [by clicking on the Buckets link here. If](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=275.8794444444444) [you remember, we had specified that we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=278.499) [want to delete objects from source after](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=280.081) [they have been transferred successfully.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=282.833) [If you click through to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=285.052) [spikey-storage-asia, you'll find that it's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=285.981) [completely empty. The transfer was](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=288.107)[successful. All the objects from the source bucket have been deleted.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=4&mode=live&start=290.6603333333334)

[Creating and Uploading Data in an S3 Bucket](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live)

[If you have a lot of your data stored in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=1.136) [Amazon S3 buckets, and you want to move](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=3.025) [these over to the GCP, that's also](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=5.029) [possible using the Transfer Service. In](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=7.772999999999996) [this demo, let's see how we can use the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=10.1943) [Transfer Service to transfer data between](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=11.781) [S3 and Google Cloud Storage. Here I am](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=13.96) [logged into my AWS account as cloud.user](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=17.84033333333333) [at loonycorn.com. I'm now going to create](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=20.829) [an S3 bucket. Click on the Services](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=24.100000000000005) [drop-down and head over to S3. I'm going](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=26.145) [to create a new bucket which will hold my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=28.498111111111104) [source data. I have no buckets configured](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=32.091571428571434) [at this point in time. I click on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=34.02112500000001) [Create bucket option here and set up a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=36.43950000000002) [bucket called spikey-images. I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=39.622666666666646) [go with the default values for all other](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=41.446000000000005) [options here. My bucket is located in US](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=43.2375) [East (Ohio). Click on the Create button,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=46.35099999999999) [and you'll be all set up with your bucket](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=48.05599999999999) [here, the spikey-images on S3. I'm going](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=50.34414285714285) [to click through to spikey-images, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=53.995999999999995) [within this, I'm going to create a new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=55.39933333333333) [folder called images. This folder is what](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=58.29342857142856) [is going to contain the static assets of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=60.38422222222222) [my product catalog. My product catalog](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=63.1175) [images are not really private. I don't](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=65.76087499999998) [need any additional encryption. Go ahead](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=67.97733333333333) [and hit Save and create this folder.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=70.22044444444445)[Within the images folder, I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=73.498) [create another subfolder called bags.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=74.865) [These are the bags that are available on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=78.066) [SpikeySales.com. Hit Save, and you have a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=81.16822222222217) [subfolder as well. Click through to the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=83.67466666666662) [subfolder, and within here, I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=85.82200000000002) [use the Upload button to add some image](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=87.78644444444447) [files to this subfolder. Click on the Add](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=90.34228571428572) [files button and choose files from your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=93.07499999999997) [local machine. Within the bags folder](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=95.22659999999998) [here, I have a number of bag images. I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=96.7)[select these images and upload them to my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=100.00177777777778) [S3 bucket in the bags subfolder. Observe](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=102.008) [that some of these images have a double](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=106.05675) [underscore prefix. Imagine that these are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=108.20133333333332) [outdated images that we don't want to copy](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=110.64228571428572) [over to our GCS bucket. Let's head one](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=113.44049999999999) [level up back to the images folder. And](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=116.31619999999998) [within here, I'm going to create another](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=118.82514285714285) [subfolder called shoes. These will contain](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=120.76875) [the shoes for my catalog. Click through to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=124.30350000000001) [the shoes subfolder, and I'm going to use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=127.49249999999998) [the Upload button once again in order to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=129.38566666666665) [upload the images of some shoes. Click on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=131.5906) [Add files here. Choose files from your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=134.45880000000002) [local machine, here are all of my shoe](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=136.583) [images, and upload them to your AWS S3](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=138.84933333333336) [bucket. In order for GCP's Transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=142.2773333333334) [Service to access data from your S3](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=144.49085714285712) [bucket, you need to set up an Identity and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=146.71762499999997) [Access Management user for programmatic](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=149.64674999999994) [access to your S3 bucket. Use the Services](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=151.75024999999997) [dropdown on AWS, and head over to IAM.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=155.01250000000002) [This is the main dashboard for IAM. Within](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=157.74099999999999) [this, you want to select the Users tab,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=160.702125) [and select the big blue button here to add](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=163.20625000000004) [a new user. I'm going to call this user](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=166.72557142857144) [spikey-migrate-data because that's what](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=169.198) [we're going to use this user for, to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=171.093)[migrate data to the GCP. And the access](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=173.32366666666667) [time is programmatic access. Click on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=176.66200000000003) [Next button here, and let's assign some](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=179.206625) [permissions for this user by attaching](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=180.927) [existing policies directly. Select S3 from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=183.67000000000002) [the search pane to filter policies that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=187.16012499999997) [directly have to do with S3 buckets. There](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=189.60699999999997) [is an existing policy that allows me to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=193.400125) [give this user full access to S3 buckets,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=195.31275) [to all resources. This is what I'm going](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=198.2755) [to go with. Select the checkbox, and click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=200.74675000000002) [on the blue button that says Next. We](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=202.92412500000006) [won't assign any tags to this user. Go](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=206.07275) [ahead and click on Review so that we can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=208.67399999999998) [review the user and the permissions that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=210.744) [we assigned to this user. This user has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=212.603375) [AmazonS3FullAccess. Click on the Create](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=217.35099999999997) [user button here, and you will have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=219.47557142857139)[successfully created an AWS user with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=221.21157142857138) [programmatic access to your S3 buckets.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=223.0492857142857) [This is the equivalent of a service](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=225.1435) [account on the GCP. In order for the GCP](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=227.2276) [to access your S3 bucket, you need to copy](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=230.38622222222227) [over the access key ID of this user and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=232.59587499999995) [the secret access key of the user as well.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=235.59900000000007) [Make sure you copy these over and save](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=237.93) [these safely somewhere. Click on the Close](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=239.697) [button here. At this point, we are done](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=242.5036666666667) [with whatever configuration we needed to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=244.45625000000004) [do on AWS and are ready to switch over to our GCP web console.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=5&mode=live&start=246.17899999999997)

[Transferring Data from S3 to GCS Buckets](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live)

[Here we are in the main Google Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=0.5) [storage buckets page on our GCP web](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=2.918) [console. We are working in the spikey-da](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=5.15625) [project. Let's click through to the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=7.9385714285714295) [spikey-image-store and take a look at what](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=9.119) [data we have stored here at this point in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=11.089) [time. Under the images folder, we have a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=13.5109) [subfolder called shoes, and within that,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=15.559) [we have a bunch of shoe images. This](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=17.638) [spikey-image-store is going to be our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=20.821857142857144) [destination bucket for the transfer that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=22.351)[we are going to configure from our Amazon](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=24.229857142857142) [S3 bucket. Click on Transfer using the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=27.162555555555556) [left navigation pane, and let's create a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=29.241714285714284) [new transfer job. This time the source of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=31.95962500000001) [our data is an Amazon S3 bucket. It's not](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=34.89099999999999) [on Cloud Storage. Select that as our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=38.47699999999999) [option, and specify the name of your S3](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=41.251000000000005) [bucket. It's spikey-images. Observe that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=43.84380000000001) [the bucket URL on AWS starts with s3. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=46.283600000000014) [next step is to configure the GCP to have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=51.17655555555555) [access to the S3 bucket, which we'll do by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=53.592125) [specifying the access key ID that we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=56.12222222222222) [copied over from AWS S3 and also the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=58.237) [secret access key. Paste both of these in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=61.6432222222222) [here. These are the keys for the user that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=64.72725000000001) [we created on AWS S3, which gives the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=66.95055555555555) [Transfer Service programmatic access to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=69.98187500000002) [the contents of our S3 bucket. This time](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=72.27533333333332) [we'll specify a few additional file](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=75.21324999999997) [filters to determine which images we want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=77.80614285714283) [to transfer over. I only want to transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=80.87128571428572) [files with some specific prefixes. Click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=83.38500000000002) [on Add item here. I only want files with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=85.43400000000001) [the prefix images. Remember, that's the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=89.30725000000001) [name of our top-level folder on S3. That's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=91.33128571428571) [the prefix to all of our image data. I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=93.51211111111112) [also want to exclude certain files from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=96.71400000000001) [the transfer. Click on Add item here. I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=99.354) [want to exclude all of the files that have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=101.88933333333334) [the double underscore prefix. Remember](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=104.448) [that folders on S3, as well as Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=107.198375) [Storage buckets, are just part prefixes.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=109.78) [So we want to exclude files which are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=112.57) [under images, bags, and have the double](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=114.45444444444439) [underscore prefix. Hit Continue here. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=116.672) [next step is to specify a destination for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=119.9325) [this transfer, which is the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=122.53649999999999) [spikey-image-store bucket. I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=124.094) [select a few other transfer options here.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=127.46775) [I want to override destination with source](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=129.17133333333334) [even when identical, which means that if I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=131.8618) [have an image or a file with same name in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=134.48700000000002) [the source, as well as the destination, I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=137.05277777777783) [finally get the file from the source. And](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=139.7025) [to reduce my storage costs, I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=142.427) [delete objects from source once they've](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=144.211) [been transferred over successfully. With](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=146.179) [that done, let's click on Continue and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=149.09187500000002) [configure our transfer. I want this job to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=151.058) [run right away. I don't want to schedule](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=153.68700000000007) [it. Click on Run now, and choose a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=155.6535) [description. This is an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=158.07499999999996) [AWS-to-GCP-image-storage-job. Click on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=161.27499999999998) [Create button here, and wait for the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=163.07849999999996) [transfer to be kick-started. You'll find](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=165.13733333333334) [that because we are transferring over from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=166.7613333333333) [S3, this transfer takes far longer. It](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=168.4475) [took about five or six minutes for this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=170.77450000000002) [transfer to be completed. And, finally, I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=173.57000000000005) [see that 22 files have been transferred](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=175.77133333333333) [over. Let's head over to the main Browser](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=177.4617142857143) [tab, which lists all of our Cloud Storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=180.202) [buckets, and take a look at the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=182.63725000000002)[spikey-image-store. We click through to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=185.6155) [the images folder, and notice we now have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=187.69514285714288) [the bags as well as the shoes subfolders](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=189.56149999999997) [in here. When I click through to bags,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=192.57900000000004) [observe that only those images that did](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=195.23675) [not have the double underscore prefix have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=197.25025000000005) [been transferred over. Those were the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=199.64200000000005)[options that we had configured. And if you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=201.7656666666667) [go back to the images folder and click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=204.094) [through to the shoes subfolder, you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=205.94966666666667) [find that all of the existing shoe images](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=208.253125) [that we had have been overwritten.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=210.48685714285716) [Remember, we chose to keep the destination](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=212.544) [object rather than the source object.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=214.624) [Let's head over to the tab where we have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=217.154) [our AWS management console open. And from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=218.735) [the Services menu, let's choose S3](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=222.289) [buckets. I want to take a look at the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=224.611) [source bucket to see where the GCP has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=226.8476) [deleted the images within it. Click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=228.86999999999995) [through to spikey-images. Click through to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=231.0005) [the images folder. And notice here that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=233.60250000000002) [the shoes subfolder has been completely](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=235.93057142857148) [deleted. Only the bags subfolder exists,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=238.44575000000003) [and if you click through to bags, you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=240.55687500000005) [find only those images that have the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=242.85342857142862) [double underscore prefix. These were the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=244.91450000000003) [images that were not transferred over.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=247.32866666666672) [They still exist in our source bucket.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=249.315)[They are not present on GCP. The GCP](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=251.483) [Transfer Service worked exactly like how we had configured it.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=6&mode=live&start=254.73233333333334)

[Transferring Data from an Externally Accessible URL](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live)

[Now it's quite possible that all of your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=1.079) [data is not available with another cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=2.334) [provider. Instead, it's on your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=5.107749999999999) [on-premises data center. You can use the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=7.249714285714285) [Transfer Service with this as well. All](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=9.434) [you need to do is make your data available](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=10.965999999999998) [at an externally accessible URL. Here we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=13.659285714285714) [are on the main page for our Google Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=16.983999999999998) [Storage buckets. We have two buckets](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=19.043666666666667) [created, the spikey-image-store and the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=20.7782) [spikey-storage-asia. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=24.105199999999996) [spikey-image-store has been the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=25.1666) [destination for most of our transfers, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=26.337) [within that, we have the images folder,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=28.874375) [the shoes folder, and a bunch of images](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=30.564) [for shoes. In order for Transfer Service](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=34.06080000000001) [to transfer data from an externally](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=36.499285714285705) [accessible URL, you need to upload a TSV](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=38.87833333333333) [file to a Cloud Storage bucket. This TSV](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=41.32714285714285) [file will contain a bunch of tab-separated](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=44.505250000000004) [URLs pointing to objects that are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=47.0315) [accessible and that will be transferred by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=49.84028571428571) [the Transfer Service. The very first line](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=52.0945) [of this file, TsvHttpData- 1.0, is the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=54.3897142857143) [format specifier. After the first line](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=58.824666666666666) [comes the URL records, which point to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=62.18049999999999) [objects that you want transferred to the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=64.55571428571429) [GCP. Here we'll transfer just the one](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=67.476) [object, a puppylinux distribution. You can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=70.03428571428573) [use the same file to specify multiple](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=73.35175) [objects to transfer. In addition to the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=75.43599999999999) [URL itself, you can specify the size of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=78.24100000000001) [the object that you want transferred. You](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=80.26100000000002) [can also optionally specify an MD5 hash.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=82.694)[Those are optional. We've left them out in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=85.914) [this simple demo. The exact format for the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=89.00500000000002) [TSV file specification is available at](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=91.38914285714284) [this link here. In order for the Transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=93.94114285714285) [Service to know what URLs to access to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=96.446) [transfer data from, this TSV file should](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=98.837) [be placed in a location that is accessible](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=100.937) [to the Transfer Service. I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=103.44100000000002) [upload this into my spikey-storage-asia](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=105.50914285714285) [bucket. Here is the urls.tsv file, which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=107.666) [I'm going to open up and upload to this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=110.90774999999998)[bucket here. I'm going to use the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=114.2396666666667) [three-dot menu over on the right to edit](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=116.15411111111108) [the permissions for this particular object](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=118.182) [so that it's universally accessible by all](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=120.06733333333332) [users. I'm going to set User to allUsers,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=123.4005) [and allUsers will have read access. Click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=126.08) [on the Save button here, and you'll find](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=129.36837500000001) [that this TSV file is now publicly](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=130.91811111111113) [accessible. It has a public link.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=133.66288888888897) [Right-click on this, and copy the link](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=136.328) [address over. We're going to need it when](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=138.068) [we set up our transfer job. Click on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=139.856) [Transfer using the left nav pane, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=142.5095) [let's create a new transfer job. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=144.69557142857144) [source for this particular transfer is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=147.7412857142857) [going to be a list of object URLs. And we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=149.147) [need to specify the URL of the TSV file,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=152.546) [which we uploaded to our Cloud Storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=154.63171428571428) [bucket. This is the public link to the TSV](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=156.921) [file. The destination for this particular](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=159.6982857142857) [transfer is once again going to be the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=161.9317142857143) [spikey-image-store Cloud Storage bucket.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=165.10199999999998) [And this time, I'll choose the last](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=167.69) [configuration option that we haven't tried](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=169.186) [out yet. I want to delete the object from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=171.178) [the destination if there is no](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=173.698) [corresponding version in the source. Now](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=175.245) [this can be a pretty scary option because](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=178.183) [it means that all of the contents that are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=179.905) [currently in my bucket will be deleted](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=182.09724999999997) [unless there is a matching version in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=184.1697142857143) [source. Click on the Continue button, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=186.88314285714284) [let's go ahead and schedule this job to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=189.0877142857143) [run right away. I'm going to call this the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=191.83799999999997) [URL-to-gcp-bucket-job. Click on the Create](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=196.73942857142862) [button here, and that will kick-start your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=198.08849999999998) [transfer. You'll need to wait patiently](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=200.93800000000005)[for a couple of minutes. Your transfer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=202.61899999999997) [will soon be complete, and exactly one](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=204.23999999999995) [file has been transferred. Let's now head](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=206.84250000000003) [over to the main buckets page and take a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=209.31300000000002) [look at our spikey-image-store. And if you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=212.023) [click through, you'll find that all of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=214.27300000000002) [existing files have been deleted, but you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=216.0702857142857) [have the puppylinux distribution that has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=218.56799999999998) [been copied over. This is the object that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=221.38250000000002) [we had specified in the TSV file, and we now have this in our Cloud Storage bucket.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=7&mode=live&start=223.60622222222227)

[Object Lifecycle Management](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live)

[Now that we know how to get data into](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=0.81) [Cloud Storage buckets, let's now study how](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=2.5002000000000004) [object lifecycle management works with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=5.026285714285715) [GCS. Any organization will have a whole](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=8.296) [lot of data. Not all of this data is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=10.80333333333333) [equally important. You don't want all of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=13.573499999999996) [this data to be stored in multi-regional](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=15.231777777777776) [or regional buckets. For example, when you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=17.548500000000004) [work with data such as logs, for older log](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=19.903499999999994) [files, you still want them to be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=22.562375000000007) [accessible, but maybe you don't want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=24.675) [pay so much for storage. That's when you'd](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=26.497) [use object lifecycle management. Here we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=29.205000000000002) [are in the spikey-da project on the Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=31.506777777777774) [Storage buckets page. I'm going to click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=33.4615) [through to the spikey-image-store. This is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=36.0055) [the bucket that we've been playing with.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=37.731111111111105) [Click on the Upload button here in order](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=39.303) [to upload some files that we'll consider](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=41.15188888888889) [to be log data. These are logs across our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=43.430500000000016) [GCP services. Now I don't want these files](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=47.94822222222224) [to be sitting around in my multi-regional](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=50.27733333333333) [bucket for its entire lifetime. Let's head](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=52.282000000000004) [over to the main buckets page, and let's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=54.75842857142856) [take a look at the lifecycle column. In](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=56.918333333333344) [order to manage data that is stored in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=59.635555555555555) [your Cloud Storage buckets, you don't have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=61.096) [to manually delete and move these files](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=63.13) [around. You can just use lifecycle](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=65.65612499999997) [management. You can see from this column](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=67.68866666666666) [here that currently we have no lifecycle](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=69.08499999999995) [configured for this particular](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=71.46566666666665) [spikey-image-store multi-regional bucket.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=73.641) [Let's click through and configure a few](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=75.623) [lifecycle rules. We have no rules](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=77.431) [configured correctly. Click on the Add](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=79.841)[rule button, and you'll be presented with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=81.78244444444447) [a number of options. Based on certain](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=83.824) [conditions, you can choose to change how](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=86.60999999999999) [the objects within this bucket are stored.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=88.48542857142856) [I'm first going to select the age option](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=90.56485714285712) [here. When my object is greater than 1 day](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=93.98679999999999) [old, I want to change the storage class of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=96.96844444444446) [that object. You can make this condition](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=100.16600000000005) [more complicated. You only want this rule](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=102.47257142857141) [to apply if the object belongs to a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=104.48211111111112) [certain storage class. Let's apply this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=107.41785714285712) [lifecycle management rule only when the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=109.92399999999999) [object belongs to the multi-regional or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=111.96942857142857) [standard storage class. The standard](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=114.458) [storage class basically means that the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=117.05399999999999) [object takes on the storage class of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=119.02) [bucket to which it belongs. So for objects](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=121.40457142857143) [that are older than 1 day, and they belong](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=123.81066666666669) [to the multi-regional or standard storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=126.084) [class, we want to do something, so what's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=127.70657142857144) [that? Let's continue and select an option.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=130.63924999999998) [We want to change the storage class of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=133.034) [that object to be set to nearline.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=134.7406666666667) [Nearline storage is what you'd use when](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=137.693) [you plan to access that particular object](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=139.51737500000004) [around once a month or so. That's the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=142.064) [frequency of access for nearline. Nearline](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=144.7957142857143) [storage, of course, costs less. This is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=146.9475) [way for you to cut down on storage costs](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=149.33859999999996)[and still have your objects accessible.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=151.22175) [Click on Save here, and we've added an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=153.696) [object lifecycle rule. When you're working](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=155.872) [in the real world, the timelines that you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=159.0615714285714) [configure for lifecycle management will,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=161.00566666666668) [of course, be longer. Maybe it will be a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=162.785) [couple of months or maybe years. Now I've](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=165.31811111111114)[configured this to be 1 day so we can see](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=167.439) [its effect right away. We'll be able to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=170.48627272727282) [see within the span of a couple of days](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=172.48299999999998) [how this storage class for our objects](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=174.77912499999996) [will change. You can configure more than](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=177.43) [one lifecycle rule on a Cloud Storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=179.73900000000006) [bucket. I'm going to add another rule here](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=181.8195) [that says all objects in this bucket that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=183.74199999999993) [are 1 day or older and currently have the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=186.28812499999995) [nearline storage class should be changed](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=189.758) [so that they now belong to the coldline](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=192.40314285714285) [storage class. Coldline data is data that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=194.457) [you'd access, say, once a year. Coldline](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=197.627) [storage is what you'd use typically for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=200.78216666666665) [disaster recovery. Go ahead and save this.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=203.1135) [You now have two lifecycle rules](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=205.613) [configured as a part of this bucket. You](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=207.5055714285715) [can configure as many rules as you want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=210.05033333333333) [on this page. Lifecycle management is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=212.255) [way for you to automatically manage the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=214.901) [data stored in your Cloud Storage buckets](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=216.759) [without manual intervention. When I now](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=219.279) [use the Back button and head back to my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=222.2493333333333) [main Cloud Storage buckets page, you can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=224.171125) [see that for spikey-image-store, we have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=226.22566666666668)[lifecycle management enabled. It's now](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=228.971) [been about 30 hours since I configured the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=231.631) [lifecycle management rules for the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=234.22220000000002) [spikey-image-store. I'm going to click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=237.01433333333333) [through and see what its effect has been](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=238.27199999999996) [on all of the files that I've uploaded](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=239.81566666666666) [here. Here are the log files that I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=241.858) [uploaded roughly 30 hours ago, and you can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=244.53585714285717) [see that the storage class for these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=247.01666666666665) [individual objects has been updated to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=248.86166666666665) [nearline. The storage class for the bucket](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=251.42857142857142) [itself continues to be multi-regional, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=253.7732) [any new objects that you upload to the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=256.03833333333336) [bucket will have the storage class](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=258.6145000000001) [multi-regional, but older data, thanks to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=260.5196) [our lifecycle management rule, has been](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=262.85400000000004) [converted to the nearline storage class.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=264.96828571428574) [Here I am on my Cloud Storage buckets page](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=267.1395) [once again roughly 55 hours after I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=270.5237142857142) [originally configured the lifecycle](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=273.52) [management rules. I'm going to click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=275.47) [through to spikey-image-store, and I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=277.4306666666666) [going to take a look at the storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=278.6592857142858) [classes of the data that I had uploaded](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=280.93700000000007) [here. You can see that all of these log](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=283.0865999999999) [files now belong to the coldline storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=285.7839999999999)[class. Thanks to the lifecycle management](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=288.5884999999999) [rules that I configured, the storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=290.4378333333334) [classes of my files have been updated without me having to intervene.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=8&mode=live&start=292.50285714285724)

[Planning for Disaster](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live)

[You'll find that when we talk about](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=0.875) [disaster recovery on the GCP, there isn't](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=2.075) [really that much to say. In a GCP-only](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=4.883) [world, disaster recovery is largely](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=7.965857142857142) [provided. Google takes care of everything](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=10.011) [for you. You really only need to worry](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=12.27442857142857) [about disaster recovery for your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=15.51357142857143) [on-premises data center if you continue to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=16.96) [use that. One thing that you'd want to do](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=19.591) [for sure is to back up all of your data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=21.903) [using the Transfer Service to Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=24.638) [Storage buckets. You can also schedule a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=26.831000000000007) [job on a regular basis, maybe every day,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=28.92477777777778) [to make sure that your data is backed up.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=31.144999999999996) [As far as your applications are concerned,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=34.352) [it's best to use Docker containers that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=36.144) [are registered with the Google Container](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=37.95524999999999) [Registry. In this way, even if you're apps](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=40.59983333333334) [and machines go down, containers will help](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=42.84712500000002) [you get up and running quickly. If you're](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=44.691714285714305) [planning for disaster on the GCP, the one](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=48.34074999999999) [suggestion I would make is to use the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=50.60111111111111) [Deployment Manager to provision your VMs,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=52.68466666666667) [applications, and other resources. This is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=55.7264) [the Infra as Code option on the GCP. So if](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=58.7376) [you find that your data center is down,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=61.821333333333335) [and you quickly need to provision VMs, if](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=63.674285714285716) [you have the Deployment Manager and all of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=65.872) [your deployments configured using YAML](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=68.50399999999999) [files, you'll be up and running in no](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=70.55090000000001) [time. If you're planning for disaster](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=73.61410000000006) [recovery, you need to do so for your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=75.56616666666667) [on-premises DC. Google takes care of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=78.50033333333337) [things on the GCP for you. Make sure you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=80.37300000000002) [back up any data that you have on-premises](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=82.78399999999998) [using the Transfer Service. You'll simply](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=85.176) [back up all of your data to Cloud Storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=87.51422222222223) [buckets. Hopefully all of your code is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=90.754) [running on GitHub or cloud source](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=92.91699999999999) [repositories so you don't have to worry](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=95.30585714285712) [about backing up code. However, getting](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=98.12449999999998) [from code to your final application is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=100.08057142857142) [long process. Make sure that your code is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=103.14499999999998) [packaged in containers, and containers are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=105.6175) [what you're running to host your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=107.66833333333334) [applications. Register your container](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=111.2586) [images with the Google Container Registry,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=112.78049999999999)[and this will ensure that you don't lose](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=114.754) [your packaged applications in the case of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=117.23814285714289) [disaster. And, finally, if you want to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=120.57849999999998) [keep your data secure so that it's not](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=122.40799999999997) [intercepted and viewed by third parties,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=124.5905714285714) [you might want to encrypt your data using](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=126.41050000000001) [keys. And your keys can be managed by the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=130.171) [GCP using the Key Management Service. Make](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=132.84388888888893) [sure that you don't store your customer](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=135.70928571428573) [managed encryption keys on your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=137.412) [on-premises data center because in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=139.377) [case of disaster, you'll lose the keys](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=141.33985714285717) [that you need to decrypt your data. Store them on the Google Cloud.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=9&mode=live&start=143.31700000000004)

[Module Summary](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live)

[And on this note, we come to the very end](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=1.086) [of this module where we discussed data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=2.815) [migration and disaster recovery. We saw](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=4.791) [that migration has become less](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=8.391857142857143) [data-centric and more code-centric. In](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=10.014) [fact it's very easy for you to get your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=12.4456) [data to the Google Cloud or to any other](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=14.449) [cloud provider for that matter. Code, on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=16.441) [the other hand, tends to be harder to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=19.092999999999996) [migrate and specific to your use case.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=20.894625) [When you're moving your data to the GCP,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=24)[it's better to avoid prematurely](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=26.001) [committing to storage or compute choice](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=27.914333333333328) [because you're not really sure what](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=30.22842857142858) [technology will suit your needs. It's best](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=32.534142857142854) [if you keep things simple. Simply migrate](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=35.40755555555555) [your data to Cloud Storage buckets or to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=38.062) [BigQuery, and then adopt a wait-and-watch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=40.43737499999999) [approach. As you get more familiar with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=43.549) [the storage services that the GCP has to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=46.03299999999999) [offer, you'll soon figure out which one](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=48.492428571428576) [suits your use case best. As far as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=50.46724999999998)[disaster recovery is concerned, you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=53.15774999999999) [find that every product or service on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=54.921) [GCP has an option that you can configure](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=57.472333333333324) [for failover in the case of disaster.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=60.20299999999999) [Focus on disaster recovery plans for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=62.6755) [protecting your on-premises installation](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=64.99879999999999) [rather than worrying about disaster](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=66.797) [affecting the GCP. And, finally, in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=69.16799999999999) [case of disaster, what is really critical](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=71.78457142857141) [for you to back up is your data. If your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=73.63688888888886) [data is located in your on-premises](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=76.57688888888889) [installation, make sure that you leverage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=78.7084) [the transfer service to back it up to the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=80.75800000000001) [Google Cloud. In the next module, we'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=83.82614285714288) [talk about machine learning. We'll see the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=85.85028571428573)[suite of products available on the GCP for us to build powerful ML models.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=d49dee69-ff63-458d-a3f7-f111288a88d9&clip=10&mode=live&start=87.91971428571432)

[Designing Robust ML Workflows on the GCP](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live)

[Module Overview](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live)

[Hi, and welcome to this module on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=0) [Designing Robust Machine Learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=2.9872500000000004) [Workflows on the GCP. As things stand](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=4.8227142857142855) [today, the use of machine learning models](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=7.56275) [for a variety of use cases not limited to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=9.65275) [just technology is becoming increasingly](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=12.840142857142858) [common, which means when you design your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=14.9465) [data architecture, you have to think about](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=17.904857142857143) [machine learning first. Data architectures](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=20.319) [are being designed around machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=22.230999999999998) [rather than storage. Machine learning is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=24.7165) [becoming increasingly democratized, which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=27.596200000000003) [means everyone uses models, and almost](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=29.452199999999998) [everyone builds models. Before you jump](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=32.8315) [right in, carefully consider the entire](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=35.62100000000001) [machine learning suite that the GCP has to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=37.767399999999995) [offer. There are many options for you to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=40.81662499999997) [choose from to harness the power of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=42.492250000000006) [machine learning. You don't always have to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=45.023750000000014) [write your own code and build your own](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=46.695999999999984) [custom model. Even just a few years ago,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=49.14500000000001) [the only option that you had to train your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=51.15099999999998) [own custom model on your dataset was to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=53.405000000000015) [write code. That's no longer true. So](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=55.99914285714284) [don't jump directly into TensorFlow and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=58.06357142857143) [Cloud ML Engine for distributed training](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=60.215) [and prediction. Consider the other options](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=61.944) [first. You might find that other services](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=64.88828571428571)[like ML APIs, AutoML, or BigQuery ML give](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=67.3515) [you what you need without having to do too](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=70.32275000000003) [much work. Choose to build your own custom](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=73.57822222222225) [model using TensorFlow or any other framework once you're sure of your ground.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=0&mode=live&start=75.70999999999997)

[Google Cloud AI Products: Making ML Accessible to All](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live)

[In this clip, we'll introduce the suite of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=2.199) [artificial intelligence products available](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=4.109) [on the GCP, and you'll see how many of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=6.447) [them have been geared for a world where](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=8.820599999999999) [almost everyone will use models and almost](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=11.124142857142859) [everyone builds models. You'll see from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=14.4204) [this suite of products that Google has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=16.832625) [taken a significant step towards](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=18.55371428571429) [democratizing artificial intelligence. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=20.233) [ML products on the GCP can be divided into](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=23.554111111111112) [two broad categories. You have pre-trained](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=25.900750000000002) [models that have been trained on Google](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=28.202714285714283) [data and built by Google engineers. Or if](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=30.696000000000005) [you have your own datasets and your own](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=33.87133333333334) [models, you can use Google Cloud to build](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=35.652875) [your custom models. But the thing is for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=38.765) [many of the common use cases, you don't](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=41.683000000000014) [have to build your own models at all. Use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=43.303875000000005) [cases such as natural language processing,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=45.892) [image detection, label detection, all of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=48.56450000000001) [these are available using pre-trained ML](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=51.14785714285715) [APIs. At the other end of the spectrum](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=54.672000000000004) [under the custom models category when you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=56.797666666666686) [have your own dataset and your own very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=58.654428571428554) [specific use case, you might want to use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=61.36900000000001) [Cloud ML Engine with either scikit-learn,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=63.85166666666666) [XGBoost, or TensorFlow to build your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=66.06825) [custom model. ML Engine offers distributed](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=68.83314285714285) [training and prediction on the cloud. But](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=71.19133333333332) [let's say you're somewhere in between. You](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=74.09057142857142) [want to build your custom model, but you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=75.89957142857142) [don't have the programming or data science](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=77.92525) [background in order to use TensorFlow or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=79.57266666666666) [scikit-learn. Google still has products](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=82.71666666666665) [that you can use. BigQuery ML allows you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=84.66399999999997) [to use just SQL to build linear and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=87.20866666666669) [logistic regression models, and AutoML](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=90.03800000000001) [allows you to use a very simple-to-use UI](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=91.771) [to build models for common use cases.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=94.63985714285715) [We'll now examine each of these categories](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=97.5485) [in some detail starting with the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=100.04157142857143) [pre-trained ML APIs. This is what you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=101.833) [use if you know nothing about machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=104.774) [learning but still want to harness its](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=106.801) [power. You just have to know a little bit](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=109.75499999999998) [of programming to use pre-trained ML APIs,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=111.23033333333329) [which are available for a number of common](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=114.167) [use cases. For example, the Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=116.357) [Speech-to-Text API allows you to convert](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=118.629) [audio to text. This API's incredibly](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=121.14) [powerful. It recognizes 120 languages and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=124.61914285714282) [its variants. If you have text, and you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=127.95685714285717) [want to convert it to speech, you can use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=130.07333333333332) [the Cloud Text-to-Speech API. All of these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=132.25024999999994) [are neural networks built and trained by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=135.35699999999994) [Google engineers. There is the Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=139.14449999999994) [Vision API that allows for image detection](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=141.43300000000002) [and recognition. This allows you to label](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=144.0498) [and characterize images into thousands of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=146.28985714285716) [predefined categories. These are just a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=149.287) [few examples of pre-trained models that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=151.38757142857145) [Google makes available to us behind](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=153.12800000000001) [easy-to-use APIs. There are APIs which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=155.5453333333333) [deal with video natural language](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=157.71242857142863)[processing, language translation, and for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=159.99433333333334) [many of the most common use cases, you can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=162.12357142857138) [get by by just using these APIs. These](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=165.46300000000005) [APIs have been built with the idea that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=168.76000000000002) [everyone will need to use models. Let's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=171.03014285714286) [move on to the next category of AI](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=173.80644444444442) [products from Google, that is BigQuery ML.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=175.418) [This is another step in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=178.249) [democratization of building machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=179.9864285714286) [learning models. BigQuery is Google's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=182.21340000000004) [serverless data warehouse and is an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=184.753) [incredibly popular product. Many](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=187.3221428571429) [organizations have their data stored in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=188.92866666666666) [BigQuery. Once they have the data,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=190.86857142857144) [organizations will want to harness the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=193.398) [power of machine learning to extract](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=195.148) [insights. That's where the Google Cloud AI](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=197.90885714285722) [suite of products come in. But Google has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=200.13375000000005) [very cleverly merged both of these and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=202.69942857142857) [developed Google Cloud BigQuery ML.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=205.88242857142865) [BigQuery ML allows you to build linear and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=209.11) [logistic regression models in SQL without](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=211.816) [ever leaving BigQuery. BigQuery ML is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=214.878) [very interesting and innovative offering.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=218.31771428571432) [Google realized that BigQuery is used by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=220.1272) [both business analysts, as well as data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=223.07528571428574) [scientists. The Cloud AI suite of products](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=226.02439999999996) [that they offered were used largely by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=227.86449999999994) [data scientists but needed to become more](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=230.42033333333333)[accessible to business analysts, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=233.135) [that's where BigQuery ML comes in, which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=234.8416666666666) [brings in the best of both worlds.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=237.22375) [BigQuery ML brings machine learning models](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=240.314) [to where your data is stored rather than](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=242.879) [moving your data to where ML models can be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=244.847) [developed. Another way to build custom](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=248.36240000000006)[models that have been trained on your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=250.0037142857143) [dataset without writing any code is to use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=251.8718571428572) [AutoML. AutoML presents a nice user](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=255.50471428571424) [interface where you can upload your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=257.9332857142856) [training data, and behind the scenes it](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=259.42542857142854) [uses transfer learning plus neural](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=261.82366666666667) [architecture search to find the best](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=264.0835714285714) [machine learning model for your use case.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=266.607) [Cloud auto ML is a suite of machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=269.584) [learning products that allows even](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=271.453) [developers with limited machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=273.195) [expertise to train high-quality models.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=275.101) [Currently AutoML supports just a number of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=278.269) [curated use cases for translation, natural](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=280.77) [language processing, and vision. AutoML](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=283.7392) [has been built to reflect the same trend](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=286.232125) [of the democratization of machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=288.25683333333336) [like pre-trained ML APIs. AutoML helps](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=291.2022) [make AI and ML accessible to all.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=294.17633333333333) [Pre-trained ML APIs offer you no](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=297.219) [customization. You have to work with the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=299.6987142857144) [model that Google gives you. With AutoML,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=301.41866666666664) [you start with the pre-trained model,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=303.23400000000004) [which you can then customize on your dataset. That is transfer learning.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=1&mode=live&start=304.9873333333333)

[Cloud ML Engine for Building and Training Custom Machine Learning Models](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live)

[At one end of the spectrum, you want it to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=1.573) [be simple to use and build machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=3.353) [learning models. At the other end of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=6.051888888888889) [spectrum, you want to be able to train](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=8.124333333333333) [your custom model on your dataset with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=10.288888888888886) [distributed training. That's when you'd](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=13.318) [use Cloud ML Engine. Cloud ML Engine is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=15.627142857142859) [the managed service you'd use on the GCP](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=18.21171428571429) [to build and deploy custom models. If you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=20.684111111111104) [are a developer or a data scientist with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=24.80711111111111) [machine learning expertise, and you want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=26.7485) [to be able to build and train your own](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=28.5868) [models on the cloud, the GCP has a suite](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=30.394) [of products and services that you can use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=32.942800000000005)[for the entire production machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=35.11600000000002) [workflow. Machine learning typically](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=38.512666666666654) [starts off with the data preparation and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=39.96999999999999) [model development phases. There are a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=42.7202857142857) [number of frameworks that you can choose](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=44.723749999999995) [from to build your custom model. Currently](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=46.2665) [supported are TensorFlow, Keras, XGBoost,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=48.5944) [and scikit-learn. If you're looking to use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=51.573) [deep learning to build your machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=54.48987499999999) [learning models, and you want to construct](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=55.84185714285714) [your neural network, that's when you'd](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=57.815000000000005) [choose to use TensorFlow. TensorFlow was](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=60.079857142857136) [originally developed at Google before it](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=62.456599999999995) [was open source. It has tight coupling](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=64.653) [with ML Engine, and it's relatively](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=66.89000000000001) [complex to use. A simpler way to use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=69.04999999999998) [TensorFlow is to use the Keras library. It](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=71.37877777777777) [once again is for deep learning, but it's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=73.42899999999999) [much easier to work with than TensorFlow](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=75.87850000000002) [directly. If you're not looking to use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=79.00899999999999) [deep learning to build neural networks,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=80.893) [you can use other libraries and frameworks](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=82.624) [on the GCP such as scikit-learn and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=84.95499999999998) [XGBoost. Cloud ML Engine can help you with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=88.53066666666666) [the building and deployment of models](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=91.09874999999998) [across all of these frameworks. Here is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=93.1528333333333) [the complete ML workflow that you can work](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=96.28375) [with using Cloud ML Engine. Here](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=98.67214285714287) [critically important is the phase for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=100.90857142857143) [model training and evaluation. This is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=102.7) [where Cloud ML Engine works with the ML](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=105.30499999999999) [framework of your choice, whether it's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=107.589)[TensorFlow, scikit-learn, or XGBoost,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=109.69200000000001) [very, very closely. Cloud ML Engine as we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=111.87580000000001) [mentioned earlier is very closely](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=114.94899999999998) [integrated with TensorFlow. When you're](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=116.6968) [using TensorFlow as your framework, you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=118.64300000000001) [can configure training and hyperparameter](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=120.468) [tuning very easily. As a developer, all](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=122.703) [you need to do is to configure the right](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=125.57840000000002) [parameters for training and hyperparameter](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=128.8585) [tuning, and Cloud ML Engine takes care of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=130.82866666666666) [the rest. It picks up your config](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=134.38400000000004)[parameters and runs training and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=136.42387499999998) [hyperparameter tuning in a distributed](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=137.85733333333332) [manner. ML Engine takes care of all of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=139.8535) [these steps of deployment, prediction, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=143.21028571428573) [managing the different versions that you've deployed of your model.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=2&mode=live&start=145.35799999999998)

[Introducing Dataprep Flows and Recipes to Wrangle Data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live)

[In this demo, we'll see how we can use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=1.522) [Cloud Dataprep, a third-party tool](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=3.074) [provided by Trifacta, in order to clean,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=5.614) [transform, and prepare your data for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=8.348857142857142) [machine learning. Cloud Dataprep is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=10.748571428571429) [visual service for cleaning your data, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=12.686857142857141) [it makes many mundane tasks easy. Here we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=14.59485714285714) [are in the spikey-da project. Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=18.731857142857145) [Dataprep needs a bucket where we upload](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=20.600285714285715) [data. I've headed over to Browser here](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=22.766571428571428) [under Storage, and I'm going to create a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=25.37325) [new bucket called spikey-data-prep. This](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=27.9522) [bucket can be a multi-regional bucket](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=30.82542857142857) [located in the US. You can accept the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=32.412) [default settings. Click on Create, and we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=34.052) [have our bucket. Dataprep is a third-party](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=36.292) [tool, but it's fully integrated with the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=39.193) [GCP and, in fact, uses Dataflow in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=41.063) [back end to clean data. Use the Navigation](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=43.47455555555554) [menu here, and go to Dataprep. Dataprep is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=46.828) [subject to Trifacta's terms of service, as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=50.47571428571429) [well as GCP's terms of service. Click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=53.36075)[Agree, and let's get started. You need to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=56.40566666666666) [be an owner on your current project to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=59.428666666666665) [allow Trifacta to access project data.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=61.576) [Click on Allow, and sign in with your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=64.21) [current GCP account. Because Dataprep is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=66.071) [third-party tool, if this is the first](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=69.41) [time you're accessing Dataprep in this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=71.31400000000004) [current project, there are a number of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=73.00850000000003) [hoops that you need to jump through. Click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=74.76300000000003) [Accept, and let's move forward. Dataprep](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=77.47816666666667) [needs a Cloud Storage bucket to which it](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=79.59342857142857) [saves and uploads files. Go ahead and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=82.18466666666667) [change this to the spikey-data-prep bucket](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=84.56257142857142) [that we've just created. Click on Continue](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=86.71366666666667) [here, and we're finally ready to get](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=90.097) [started. When you work with Dataprep,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=93.05071428571428) [you'll be creating flows to wrangle your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=94.59214285714283) [data. All your flows are accessible using](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=97.48433333333335)[this icon on the left pane. This library](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=99.53585714285711) [page is what you'll use in order to input](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=102.99050000000001) [data to get started. And, finally, all of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=105.41685714285713) [your data processing and cleaning jobs](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=108.24771428571431) [that you run using Dataprep will be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=110.11542857142857) [available in the Jobs page. Let's start](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=112.70812499999998) [off by heading over to the flows page and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=115.4085) [creating a new flow. I'm going to call](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=117.70414285714286) [this the automobile\_dataset\_flow, and this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=119.65766666666667) [will contain the series of steps or](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=121.64200000000001) [recipes to clean my automobile price](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=124.37099999999997) [prediction dataset. Now if it's a real](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=127.59833333333333) [world dataset, you know that it won't be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=130.0055555555556) [in great shape. The automobile price](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=131.926) [prediction dataset, however, is very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=133.90239999999997) [clean. So what I've done is intentionally](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=135.82225) [introduce a bunch of discrepancies in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=137.939) [data. I've also split the dataset into two](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=141.22987499999996) [so that we have some experience joining](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=143.50600000000006) [data from two different datasets. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=145.74657142857143) [going to click on Choose a file here and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=148.678) [upload a file to the Cloud Storage bucket](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=150.695125) [from my local machine. This is one of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=153.07920000000001) [datasets that I'm going to be using. Once](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=155.50344444444445) [this file has been uploaded, you can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=157.303) [choose Import and Add to Flow. And this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=159.62300000000002) [dataset becomes a part of this automobile](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=162.38875) [flow. This dataset contains just half the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=164.5708) [data that I need. I'm going to go ahead](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=167.93028571428573) [and click on the Add Datasets button once](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=169.90749999999997) [again and import the second dataset. This](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=171.95714285714283) [is present in a different file. The second](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=175.26614285714285) [file that I upload from my local machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=177.9598888888889) [is called](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=180.42575) [additional\_automobile\_stats\_.csv. I import](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=182.63524999999998) [this dataset and add this to the flow as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=184.25111111111113) [well. And observe now how both datasets](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=186.63750000000002) [are visually represented as a part of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=188.7827142857143) [same flow. Both of these datasets that we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=191.67725000000002) [just imported have been uploaded to our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=194.562) [Cloud Storage bucket, and you can see from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=196.524625) [the right nav pane quickstarts about the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=199.0825714285714) [dataset itself. Let's start off by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=201.66585714285716) [selecting the basic\_automobile\_stats\_.csv](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=203.861) [dataset. When you select this, you'll be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=206.64199999999997) [given the option to add a new recipe. A](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=208.8455) [recipe is nothing but a series of data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=211.58166666666665) [cleaning and formatting steps. Once I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=213.587)[select on this additional node in this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=216.44299999999998) [Dataflow graph, I'll be given the option](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=218.61471428571429) [to edit the name of the recipe. This will](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=220.899) [allow me to know what exactly I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=224.09966666666665) [be doing in this series of steps. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=226.52944444444444) [going to join and drop columns. You can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=228.5904285714286) [now click on the Edit Recipe button that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=231.2567777777778) [shows up on the right nav pane, and this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=233.502625) [will bring up a visual UI that you can use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=236.0156) [to clean and explore your data. You can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=238.79200000000003) [use this UI to explore all of the columns](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=242.00760000000002)[that are present in this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=243.82733333333334) [basic\_automobile\_stats\_.csv file. You can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=245.374) [see that the data that it contains is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=248.99499999999998) [bunch of details about individual](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=251.2645714285714)[automobiles, and we'll use these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=252.791) [automobile features in order to predict](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=254.83850000000004) [the price of the automobile. That's the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=256.87800000000004) [objective of this data-cleaning exercise.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=259.24800000000005) [You can see at the bottom here that this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=262.032) [dataset has 16 columns of data, 244 rows,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=263.488) [and 3 datatypes--strings, floats, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=267.381) [integers. Click on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=271.7051666666665) [automobile\_dataset\_flow link at the very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=273.98659999999995) [top. This is a link that you can use to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=275.39869999999996) [get back to the main flow page, and let's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=277.29010000000005) [select the second dataset, the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=279.8673333333333) [additional\_automobile\_stats, and add a new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=281.77) [recipe. I'm just going to call this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=284.06) [Dataset II and go ahead and edit this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=286.594) [recipe. Clicking on Edit Recipe will take](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=290.3815555555556) [us to the page where we can explore our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=292.2389999999999) [dataset. You can see that this dataset](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=295.4364285714285) [also contains a bunch of information about](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=297.6247499999999)[the same automobiles. There are 15 columns](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=300.2300000000001) [of data here, 245 rows, and 3 datatypes.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=302.62757142857134) [This is the original automobile price](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=305.973) [prediction dataset that I've split into](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=307.701) [two. Use the link at the top left to head](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=310.36157142857144) [back to the main automobile\_dataset\_flow](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=312.6842) [page. I'm now ready to edit the join and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=315.3146000000001) [drop columns, a recipe to add the series](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=318.01333333333326) [of steps to clean my data. The first thing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=320.4817499999999) [I want to do is join both of my datasets](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=323.8877) [together so that I have a single dataset](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=326.11166666666674) [to work with. In this entire idea behind](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=328.2738) [democratization of the UI, see how easy it](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=331.59366666666676) [is to join columns. Select the Join icon](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=334.13672727272746) [here, and this will bring up a bunch of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=336.2405) [options. Dataprep will quickly scan the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=339.02016666666657) [other datasets that you've uploaded and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=340.8131666666667) [suggest viable join options. I have just](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=342.81916666666666) [the one additional dataset, and this is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=345.4275) [what I choose to join in. And notice how](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=347.8962857142858) [Dataprep has figured out what is the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=351.6821250000001) [column that you should be using for this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=353.83033333333344) [join. It has picked up the index column as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=355.99633333333344) [having a 100% match. By default Dataprep](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=358.13722222222214) [will try to perform an inner join, but](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=360.82599999999996) [other join types are possible to specify](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=363.1545000000001) [as well. Yes, this is indeed the right](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=365.7565714285715) [dataset that we want to join in. Click on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=368.18449999999996) [Next, and you'll now be shown all of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=370.44737499999985) [columns that will be present in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=372.5936666666665) [resulting dataset. Because of the join,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=375.0697500000001) [observe that we now have two index columns](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=376.6794999999999) [in the result. Over to your right, you can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=379.5492857142857) [select the current tab to see what columns](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=382.2757499999999) [are present in the current dataset. This](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=384.36524999999995) [is the dataset on the left side of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=387.44599999999997) [join. If you observe the data carefully,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=389.8939999999999) [you'll find that there are three columns](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=391.354) [in common across the two datasets, the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=393.1827500000001) [index column on which we perform the join,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=395.9640000000001) [the make of the car, and the average miles](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=398.287) [per gallon for each car. You can confirm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=400.351) [this by selecting on the Joined-In tab,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=403.00250000000005) [which will show you the columns that come](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=404.4818571428571) [in from the joined-in dataset. The same](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=407.268) [three columns are present. You can now use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=409.8476666666666) [this checkbox here to choose what columns](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=412.0969999999999) [you want to keep and what columns you want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=413.81399999999996) [to drop from the joined-in or the current](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=416.1705) [dataset. All three columns that are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=419.10639999999995) [duplicates here I'm just going to uncheck.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=421.02499999999986) [I don't want them duplicated in my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=422.813) [resultant dataset. That's the only change](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=424.347) [that I'll make here. I'm now ready to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=427.073)[review and accept this resulting dataset.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=428.987) [I'm going to add this joined-in to my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=431.783) [recipe. And notice how the right](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=434.7643333333332) [navigation pane is updated with the inner](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=436.5413749999999) [join. That is now part of your recipe. We](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=438.93280000000004) [now have a joined dataset with 28 columns,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=442.18475) [360 rows, and 3 datatypes. Now it's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=444.99733333333336) [possible that your dataset contains a lot](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=448.7303333333333) [of additional columns whose data you don't](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=450.6301428571429) [want to use for your ML prediction. You](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=452.5322) [can select each of these columns, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=455.0541428571429)[you'll find that Dataprep suggests that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=457.63471428571427) [you drop these columns or delete these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=459.541) [columns. You can see that this suggestion](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=462.45824999999985) [is in the right nav pane. A bunch of other](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=463.8950000000001) [suggestions are available as](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=466.198) [well--GroupBy, Move columns, Merge](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=468.0484) [columns, and so on. I'm going to delete](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=469.73500000000007) [these 16 columns to make my dataset more](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=472.3430000000001) [manageable. Don't worry about what exactly](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=475.3286000000001) [these columns are. Use this dataset to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=477.1918333333332) [explore. The resources included with this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=480.0345714285715) [course will include a final copy of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=482.30050000000006) [cleaned dataset. That's the dataset we'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=485.8102222222224) [use in the simple ML model that we'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=487.5795) [build. We started off with roughly 200](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=490.4269999999999) [rows in our dataset, but we seem to have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=492.26100000000014) [360 rows. It's possible that some of these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=494.8015000000001) [rows are duplicates. You can click on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=498.0516666666668) [drop-down for the main index column or for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=500.2877777777778) [any column, and that will bring up a menu,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=502.16450000000003) [and you can filter and remove duplicate](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=504.5202222222222) [rows. There are a number of rows](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=507.62399999999997) [highlighted here, all of which contain](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=509.5031428571428) [duplicate data. We don't want those. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=511.38371428571423) [going to use the right navigation pane](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=512.9502222222225) [that has given me the suggestion to remove](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=514.3896250000001) [duplicate rows. Click on Add, and this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=517.045) [step will be added to my recipe. My](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=519.1034444444442) [dataset has been reduced to 223 and 12](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=521.3815000000002) [columns. Click on the main flow link at](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=525.6384000000002) [the top left of your page, and this will](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=527.8240999999997) [take you back to the flows page. You can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=529.3803333333332) [see that you've successfully completed one recipe in your data preparation.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=3&mode=live&start=530.7383333333331)

[Cleaning and Formatting Data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live)

[Now that we've joined both of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=1.5) [datasets, we've dropped duplicate rows and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=3.5004999999999997) [deleted some of the columns that we don't](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=5.248285714285713) [need, we are now ready to clean our data.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=7.2535) [That is a new recipe. Click on the Edit](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=9.615) [Recipe button here, and that will take you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=12.69657142857143) [to the main page where you can play around](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=14.231000000000002) [with your data. You can clean data using](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=16.867333333333335) [icons. But what's actually easier is to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=19.562000000000005) [use this Search icon here at the top](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=21.291999999999994) [right. This will bring up a search box](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=23.849571428571423) [where you can search for the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=25.324222222222215) [transformations you want to apply. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=26.615142857142857) [going to apply the NUMFORMAT](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=28.409625000000002) [transformation to the index column to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=30.189) [convert it to an integer. It's currently a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=32.26) [floating point. NUMFORMAT is what you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=35.364399999999996) [use for any kind of numeric transformation](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=37.232) [on your data. Use the auto-suggest here. I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=39.75649999999999) [want to apply the NUMFORMAT to the index](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=42.83544444444444) [column, and I want to format it to be of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=44.983) [type integer, which I specify using the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=47.598) [hash. You can see from all of the helpful](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=50.36300000000002) [suggestions that Dataprep provides that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=52.156) [you don't really need to know these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=53.88266666666668) [formatting options upfront. It would be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=55.69542857142856) [confusing to have two columns with the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=57.962857142857146) [same name, so I'm going to call this new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=59.883666666666656) [column Indx. Click on the Add button to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=62.870142857142845) [the right, and now I have a new formatted](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=64.97050000000002) [column as a part of my dataset. I no](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=67.04439999999998) [longer need the original index column, so](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=69.57299999999996) [I'm going to select it and delete this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=71.67157142857144) [column. And I'm going to rename the Indx](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=74.74716666666669) [column to be called index. So click on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=77.44188888888887) [drop-down for the Indx column, select the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=80.42937500000002) [Rename option from the menu, and rename](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=82.83057142857139) [Indx to index. All of these steps become](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=86.02299999999998) [part of your recipe. Observe that at the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=89.16699999999997) [very top of each column, you have a little](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=91.43400000000003) [histogram and a green horizontal bar. All](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=93.56999999999998) [of this gives you information about the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=95.86671428571428) [data stored in that column. If you observe](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=98.149) [closely, you might find that there is a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=100.97649999999997) [little black or red in some of these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=102.68844444444447) [columns. That means Dataprep thinks that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=104.94800000000004) [these columns need further attention. If](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=107.10199999999998) [you select that little highlight, you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=110.49933333333334) [see that this column, the make column, has](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=112.21628571428572) [three missing values, and Dataprep makes a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=114.5584)[suggestion to delete the rows with missing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=116.75214285714287) [values. The suggestions are always on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=119.854) [right pane. I'm going to accept and add](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=121.88774999999998) [this to my recipe. As I scroll around and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=124.95839999999998) [look at other columns of data, such as the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=127.43137500000002) [fuel-type column, I find missing values](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=129.61671428571424) [there as well. I'm going to add all of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=131.63812499999995) [these, cleaning of missing values,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=134.09733333333332) [deleting rows where values are missing, to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=135.719) [my recipe. Dataprep detects not just](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=139.4438) [missing values but also values that are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=142.1643333333334)[mismatched. You can see that the price](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=144.80016666666666) [column has four mismatched values. If you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=146.64200000000002) [select that little red bar, it'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=148.78733333333338) [highlight what is mismatched. You can see](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=151.23799999999994) [that there are question marks in the price](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=153.52812500000002) [column. Well, I definitely don't want](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=155.38328571428573) [those rows. I'm going to go ahead and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=157.1445) [accept the suggestion to delete those](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=158.63049999999996) [rows. I've now gotten rid of a bunch of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=160.80599999999995) [rows where the price column is mismatched.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=162.54888888888897) [It's not an integer. You'll find that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=165.396)[there are a number of columns in this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=167.447) [dataset with missing or mismatched values.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=168.967) [You can find all of them and use Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=171.296) [Dataprep to clean your data. You can see](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=173.558) [that there were missing and mismatched](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=177.44775) [values in the price, engine-type, bore,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=178.5392) [and horsepower columns. All of these are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=181.0468) [added to be part of my recipe. I'll now](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=183.76336363636364) [use the link on the top left to take a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=186.06520000000003) [look at my flow as it currently stands.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=188.05679999999998) [We've completed a bunch of faces here. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=190.452) [going to now go ahead and add a new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=192.8897) [recipe, and this time to format the data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=194.525) [in my dataset. Click on the Edit Recipe](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=197.14900000000003) [button, and let's get started. Observe](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=200.253) [that for columns in our dataset that have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=202.99642857142857) [categorical values such as make, Dataprep](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=205.23842857142856) [shows you how many categories of data are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=207.595)[in there. There are 22 makes of cars](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=210.187) [represented in this dataset. If you hover](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=212.9262) [over that little histogram, you'll be able](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=216.03499999999997) [to see how many of each type are present.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=218.11689999999996) [You can see that there are some cars which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=220.61975) [have just a few values, but they're](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=222.84649999999996) [actually wrongly labeled. You can see that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=225.50619999999998) [we have a few values which say BMW in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=227.62019999999998) [uppercase, and we have a number of values](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=229.5703333333333) [that say bmw in lowercase. Well, that's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=232.5035714285714) [something that we'll need to fix. Use the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=235.343) [drop-down next to the make column, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=237.54825) [you'll be presented with a menu where you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=239.82850000000002) [can format the data and convert it all to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=241.75933333333333) [lowercase. The UI will show you the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=244.26274999999998) [original as well as the converted columns,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=246.368) [and if you feel that this conversion looks](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=248.993) [good, you can go ahead and add this step](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=250.856) [to your recipe. As I explored my data, I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=252.95839999999995) [found that this same problem existed in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=256.9355) [the fuel-type column as well. Gas was in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=258.976) [uppercase. I'm going to convert this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=262.07525) [column to lowercase as well. This data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=263.3697142857143) [formatting step is also part of my recipe.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=265.6468333333333) [There's one more column with uppercase](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=269.695) [data, that's body-style. Convert that to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=270.959) [lowercase as well. In each of these steps,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=273.909) [the original column is automatically](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=276.31275000000005) [dropped from your dataset. Only the new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=277.9126666666666) [column remains. Dataprep gives you a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=281.1236) [variety of ways to format your data, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=283.1219999999999) [you can search for them using the Search](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=285.0709999999999) [icon on the top right. I'm now going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=287.5533333333332) [label and code some of my data by](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=290.14039999999994) [replacing text or pattern. That's the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=292.12933333333336) [option that I choose. Go ahead and choose](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=294.65299999999996) [the column that you want replaced. You'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=296.60850000000005) [see that columns such as num-of-cylinders,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=298.58199999999994) [num-of-doors, all have values that are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=302.222) [strings. I want to label and code these. I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=304.73883333333345) [want this to have values that are not the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=307.033) [string two but the number 2. So I'm going](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=308.725) [to find or search for two and replace that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=312.513) [with the number 2. Click on the Add button](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=315.67966666666666) [here, and this will replace all of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=319.10755555555545) [instances of two, the string two, with the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=321.71314285714294) [number 2. Now because I'm going to do this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=324.86357142857133) [manually, I'm going to apply a series of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=327.10224999999986) [steps that are repetitive. I'll show you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=328.7543333333333) [what the steps are. These are steps you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=330.5992857142856) [can try out by yourself. I've already](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=332.40166666666676) [shown you one example of each. Observe](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=334.0954285714285) [that for num-of-cylinders and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=336.42566666666676) [num-of-doors, we replace matches of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=338.5033333333333) [string four with the number 4. I then saw](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=340.6695000000001) [that there were some mismatched values for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=343.2473333333334) [num-of-doors. I went ahead and deleted the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=346.2688) [rows with the mismatched values and then](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=347.9643333333334) [performed a number of other replace](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=349.92566666666676) [matches. So all instances where a number](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=352.9358000000001) [was represented in its string form, I've](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=355.015) [converted to the actual number, 3, 5, 6, 8, and 12 for num-of-cylinders.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=4&mode=live&start=357.846875)

[Preparing Data for Machine Learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live)

[Click on the flow link at the top left of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=1.581) [your screen, and let's take a look at the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=3.262) [flow as it currently stands. We've joined](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=4.753) [two datasets together, dropped the rows](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=8.5284) [and columns that we did not need. We've](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=10.161000000000001) [cleaned and formatted our data. We're now](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=11.935875) [ready to prepare this data for the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=14.1715)[purposes of machine learning. Add a new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=16.3368) [recipe called ML prep. Click on the Attic](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=18.87414285714286) [Recipe button, and let's see some of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=21.282499999999995) [features that Dataprep offers to prepare](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=23.023999999999994) [your data for machine learning. Click on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=25.016) [the Search icon on the top right to bring](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=27.754444444444445) [up the options to type in your formula.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=29.933000000000003) [I'm going to construct my own custom](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=32.649) [single-row formula to detect outliers in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=34.849) [my data. I want to see whether there are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=37.97724999999999) [any records that I have where the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=40.7108888888889) [horsepower is beyond 3 standard](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=42.71275000000001) [deviations. Dataprep has built-in formulas](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=45.12533333333334) [to calculate absolute values, average, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=47.561333333333344) [standard deviation. I've put these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=50.467999999999996) [together to set up my own custom formula](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=52.33466666666666) [to calculate the standard deviation. I've](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=55.08971428571429) [called this column the std-dev-horsepower.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=58.388000000000005) [This causes Dataprep to automatically](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=61.308) [populate the values in this column based](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=62.948) [on my custom formula. I'm going to add a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=65.15299999999999) [new step here that will conditionally test](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=68.276) [for outliers in this column. The condition](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=70.47157142857144) [that I'm about to configure is a custom](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=73.39185714285713) [formula, and it's a custom single formula,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=75.35562500000002) [which means it operates on one row at a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=77.62899999999999) [time. And the condition here is to check](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=80.90777777777775) [for outliers. I want to see whether there](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=83.65259999999999) [are any records where the standard](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=85.68766666666664) [deviation of the horsepower is greater](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=87.075) [than 3, any horsepower value greater than](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=89.20939999999999) [3 standard deviations away from the mean.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=92.083) [And, yes, indeed, there is. Here is the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=94.995) [entry for Jaguar. Once we've identified](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=98.07499999999999) [this outlier, Dataprep helpfully suggests](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=100.5738) [the action that we might want to take](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=102.353)[here. I'm going to delete rows where we](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=104.40499999999999) [have outliers. And once I've detected the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=107.56900000000005) [outlier, I'm going to go ahead and delete](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=109.82699999999998) [the new column that I just created, the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=111.80455555555557) [std-dev-horsepower. I no longer need it.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=115.1235714285714) [Another common data preparation step for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=117.14162499999998) [machine learning is the one-hot encoding](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=119.267) [of categorical values, and Dataprep makes](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=122.27900000000002) [it very easy for you to do this. In the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=124.81766666666665) [Search bar, simply search for one-hot, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=127.55199999999996) [it'll bring up the option to one-hot](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=129.81199999999995) [encode column values. So just go ahead and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=132.58339999999998) [type in the name of the column to which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=135.283) [you want to apply one-hot encoding, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=137.35049999999998) [specify the maximum number of columns to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=139.43449999999999) [create. I'll go with 50. That'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=141.58059999999998) [definitely include all of the car makes](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=143.71328571428572) [that I have. Click on the Add button here,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=145.44144444444444)[and that's all you need to do to one-hot](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=147.9798888888889) [encode the make column values. And once](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=150.56033333333335) [you have the one-hot encoded values, you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=154.05014285714287) [can go ahead and drop the original column.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=155.7355) [This dataset contains a number of columns](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=158.39733333333336) [with categorical values. I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=161.8375) [one-hot encode a bunch of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=163.80262500000003) [these--fuel-type, body-style, engine-type,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=164.942) [and drive-wheels. The steps are exactly](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=167.677) [the same as what we saw right now. This is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=169.857) [something you can try on your own. And](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=172.27622222222223) [once I have the one-hot encoded values,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=174.34) [I'm going to go ahead and delete the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=176.286) [original categorical](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=178.158) [columns--drive-wheels, engine-type,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=179.9346666666667)[body-style, and fuel-type. And with this,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=181.5085) [I'm done cleaning and preparing my data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=184.21642857142854) [for the purposes of machine learning. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=185.79128571428572) [ready to run the job to get my customized](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=187.72899999999998) [and cleaned dataset. Click on the Run Job](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=191.62042857142853) [button at the top right, and this will](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=194.09816666666666) [show you the options that you have. This](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=195.86899999999991) [is going to create a CSV file with all of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=198.4403) [the data wrangling steps applied to the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=200.866875) [original dataset. I can click on the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=202.84) [pencil icon if I want to edit settings for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=205.257) [my job. I'm going to call this file](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=207.697) [automobile\_dataset\_complete, and I'm going](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=210.073) [to accept the default values for all other](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=212.58) [options. Click on the Update button here](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=215.0946666666667) [to start your processing. This is a Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=217.01924999999994) [Dataflow job. You can see that one job is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=219.9365) [currently in progress. The jobs link on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=223.17439999999996) [the left navigation pane shows you that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=225.27850000000004) [one job is currently running. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=227.46225000000004) [three-dot menu at the extreme right of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=229.61937500000002) [your job link will give you a bunch of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=231.568)[options. Let's take a look at the actual](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=234.48669999999996) [Dataflow job that's wrangling our data. We](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=236.41049999999998) [only use the UI to set up the steps in our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=239.15688888888886) [various recipes. This was converted to a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=243.28399999999993) [Cloud Dataflow job that's now wrangling](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=245.316) [and cleaning and formatting our data to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=247.6365) [give us the final result. And this is an](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=250.40477777777778)[example of how machine learning is slowly](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=252.97150000000002) [getting democratized. All of the steps](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=255.45700000000002) [needed to clean your data can now be done](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=257.3414444444445) [visually rather than by writing code. Our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=259.9169999999999) [job took about six minutes to run. Switch](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=262.51485714285707) [over to the Cloud Dataprep tab, and you'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=265.20585714285716) [see that your job is now complete. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=268.1925555555556)[final output CSV file with the clean data](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=270.42425000000003) [should be present in your cloud storage](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=272.5248571428571) [bucket. On the GCP web console, let's head](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=274.66857142857145) [over to Cloud Storage in Browser, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=277.6446666666667) [click through to the spikey-data-prep](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=280.0129999999999) [bucket. Under jobrun, you'll find a new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=282.23) [file here, the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=284.82474999999994)[automobile\_dataset\_complete.csv. Click on](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=287.345) [this file link, and download this file to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=289.626) [your local machine. We'll now use this to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=291.58140000000003) [perform our machine learning analysis. You](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=294.1518750000001) [can open up this file and take a look at](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=296.39571428571423) [this nice clean dataset that you have to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=299.43274999999994) [work with. All you have to worry about now](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=301.32177777777775) [is to build the actual model. There is no data cleaning you need to do in code.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=5&mode=live&start=303.5985)

[Building and Serializing a Linear Regression Model Using Scikit Learn](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live)

[Now that we have our cleaned and prepared](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=1.909) [dataset, let's use this for some simple](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=3.57) [machine learning. We'll build a linear](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=5.405) [regression model to predict automobile](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=7.418) [prices based on its input features. And](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=9.455) [we'll do this in a Jupyter notebook. I've](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=12.1205) [downloaded the automobile\_dataset\_csv file](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=15.15825)[onto my local machine under the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=17.21957142857143) [automobile\_model directory. I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=19.071) [use the scikit-learn framework to build my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=22.73728571428572) [simple regression model. Here is my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=24.582749999999997) [automobile\_dataset.csv. This is the same](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=26.613) [cleaned dataset that we got from Dataprep.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=30.036499999999997) [It's just renamed so that it has a simpler](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=32.17928571428571) [name. Let's open up a new notebook in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=35.693) [Python 3 in order to write our price](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=37.92577777777775) [prediction linear regression model. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=40.276857142857125) [going to call this notebook simply](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=42.34314285714285)[automobile\_price\_prediction. Set up the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=44.102) [import statements for the libraries that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=47.14085714285715) [we need. I'm going to use pickle in order](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=48.57249999999999) [to serialize this model out to a pickle](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=51.042999999999985) [file. We'll then host this model on Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=53.733750000000015) [ML Engine. We'll build a linear regression](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=56.80011111111112) [model using scikit-learn. So import the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=59.18914285714287)[LinearRegression class from](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=61.469) [sklearn.linear\_model. If you don't have](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=64.50875) [the scikit-learn package on your local](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=66.41342857142855) [machine, you can get it with the simple](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=67.63133333333334)[pip install sklearn. Use the pandas](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=69.9968) [library to read in our CSV file,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=72.70533333333334) [automobile\_dataset.csv. You can call the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=75.99000000000001) [head function in order to get a quick](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=78.46777777777775)[glimpse of what your data looks like. Head](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=80.43100000000001) [will display the top five rows of your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=82.89666666666668) [data frame. You can use the columns](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=84.87) [variable on your data frame in order to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=87.751) [view the columns that make up your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=89.931) [dataset. One of these columns is the price](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=92.71437500000003) [column, and that is our target label of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=95.18837500000004) [what we want to predict. I'm going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=97.08177777777773) [extract the price column, convert it to a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=99.34549999999997) [numeric form, and save it in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=102.73125) [auto\_target data frame. I'll then drop the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=105.98462500000001) [price column from the original dataset and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=108.31) [create a new data frame called](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=110.41042857142858) [auto\_features. These will form the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=113.14433333333332) [features of our ML model. We'll use these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=115.0043333333333)[features to build a linear regression](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=117.24287500000001) [model that will predict the price of our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=118.651) [automobile. We don't need the index column](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=122.287) [as a part of our features. That has no](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=124.17539999999997) [predictive value. I'm going to go ahead](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=126.74014285714287) [and drop that column as well. And,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=128.56344444444446) [finally, here are the feature columns that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=130.53914285714285) [we're going to work with to build our](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=132.7848571428571) [linear regression model. The actual model](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=135.96885714285716) [building just involves using the right](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=138.52949999999998) [class from the library. I instantiate a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=140.59342857142855)[LinearRegression model and call fit with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=143.198) [my featured data. And I'll pass in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=145.621) [target labels as well. The fit function](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=148.27499999999998) [that we invoke on our sklearn linear](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=150.06825) [regression estimator will build and train](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=152.42340000000002) [the model for us. I'm now going to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=155.48271428571428) [serialize this model out to a pickle file,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=158.14399999999998) [and I'm going to take the serialized model](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=160.21100000000004) [and host it on Cloud ML Engine on the GCP.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=162.45225) [But before we move over to Cloud ML](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=165.84311111111114) [Engine, let's quickly run predict with a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=168.42057142857146) [test instance on our model. I've simply](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=170.77825) [copied over one of the rows from the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=173.09575) [original data, changed a few values, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=175.07014285714286) [you can see our model predicts the price](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=177.109) [pretty well. The original price for this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=179.82600000000002) [particular car before I changed a few](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=181.96928571428566) [feature values was around $3500. So given](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=184.01887499999995) [our small dataset of a few hundred rows,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=187.41114285714286) [we have a pretty decent model. I'll run a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=190.37844444444448) [quick ls bash command in order to check](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=193.37962500000003) [whether the serialized pickle file is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=195.92357142857148)[present in my current working directory.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=197.72471428571427) [And yes, indeed, it is. This is now what I'll deploy to Cloud ML Engine.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=6&mode=live&start=200.005)

[Deploying a Model and Using Cloud ML Engine for Prediction](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live)

[I'm going to now take this serialized](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=1.359) [linear regression model that I have saved](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=3.707) [out as a pickle file and deploy it to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=5.622) [Cloud ML Engine to use for predictions.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=7.998000000000002) [I'll first need to upload this pickle file](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=11.115) [to a Cloud Storage bucket. Head over to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=12.974) [your Cloud Storage buckets page, click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=15.429999999999998) [through to the spikey-data-prep bucket,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=17.299333333333333) [and I'm going to upload this file right](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=19.342) [here. Here is the pickle file on my local](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=21.62911111111111) [machine containing the linear regression](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=24.180999999999997)[model that I just serialized. Upload it to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=26.151333333333334) [this Cloud Storage bucket. Now use the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=29.49671428571429) [Navigation menu and go to ML Engine under](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=31.96833333333333) [Artificial Intelligence and choose Models.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=34.711000000000006) [If this is the first time you're using](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=36.5815) [Cloud ML Engine in this project, you'll be](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=38.80871428571428) [presented with an option to enable the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=41.138999999999996) [API. Go ahead and click Enable. You can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=43.410999999999994) [now create a new model. I'm going to call](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=46.04528571428571) [this offline\_model\_to\_gcp. And the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=49.30766666666668) [description is just something meaningful](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=50.739999999999995) [that makes sense to you. Click on Save](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=52.155166666666666) [here, and your model will be created. And](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=54.842) [the next step is for you to deploy a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=57.247) [version of this model. Click through to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=58.921) [the model on this page. There are](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=62.091249999999995) [currently no versions of this model](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=63.659142857142854) [deployed. Click on the Create a version](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=65.17257142857142) [link right here. This will bring up a](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=67.76474999999999) [simple form that you need to fill out. I'm](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=70.28211111111113) [going to call it offline\_model\_to\_gcp\_v1.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=71.86844444444444) [The Python version on which I've built my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=75.443) [model is 3.5. You can see from this little](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=77.225) [warning here that at the time of this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=80.597) [recording, Cloud ML Engine only supports](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=82.449) [online prediction for models built using](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=85.009) [Python version 3. The next step is to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=87.977) [choose the framework with which our model](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=90.868) [has been built. Our model has been built](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=92.619)[using scikit-learn. Select that and let's](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=94.589) [move to the next step where we select the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=97.429) [framework version. Here let's choose the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=99.509) [latest version, 0.19 .2. This is the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=102.281) [latest version of the scikit-learn](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=106.43171428571429) [framework that Cloud ML Engine supports at](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=107.61633333333333) [the time of this recording. You'll then](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=110.61914285714286) [need to choose the version of ML Engine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=113.42275) [that you want to deploy this model to.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=115.21775000000001) [I'll go with the latest once again, 1.1.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=117.2455) [Because this is just a test model, I'll](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=119.433) [have it hosted on a machine which has just](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=122.305) [a single core CPU. Now we can browse for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=125.28657142857142) [where our model file is located. This will](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=128.31588888888888) [open up a page where you can choose the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=130.4536666666667) [Cloud Storage bucket to which you uploaded](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=132.87050000000005) [your model. I'm going to select this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=135.23760000000001) [bucket here, and it'll automatically pick](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=137.1435714285714) [up the pickle file from that bucket. If](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=138.68557142857142) [you're expecting a lot of incoming](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=141.1327142857143) [traffic, you can choose to go with](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=142.349) [autoscaling your prediction deployment.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=144.101) [Click on the Save button here, and your](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=146.773) [model version will be created. Let's click](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=148.597) [through and take a look at the details.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=152.97439999999997) [Everything seems fine here. We'll now use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=154.74) [our Cloud Shell terminal window in order](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=156.9947142857143) [to use this model for prediction. On my](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=159.83066666666664) [Cloud Shell VM, I'm going to create a new](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=163.18760000000003) [directory called model\_input and cd into](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=165.05933333333334) [that directory where I'll set up a simple](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=167.21437500000002) [JSON file, which will contain the test](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=170.02157142857143) [instance that I'll pass into my model for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=172.5905)[prediction. Pass in the features for](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=175.4755714285714) [prediction as a simple list enclosed](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=177.771) [within square brackets. These are the same](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=180.05528571428565) [set of features that I passed in when I](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=182.25699999999995) [ran predict on this model on my local](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=184.31474999999992) [Jupyter notebook. Hit Ctrl+X to save this](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=187.16955555555563) [file, and we'll now set up a few](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=189.51211111111112) [environment variables. The name of the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=192.675) [model that we use for prediction is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=194.26955555555554) [offline\_model\_to\_gcp. We set that in the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=197.0523333333333) [MODEL\_NAME environment variable. The](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=199.9315) [version is offline\_model\_to\_gcp\_v1, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=201.80425) [the INPUT\_DATA\_FILE is the input.json file](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=204.42766666666665) [that we just created. We'll now use the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=206.969) [gcloud command line tool to invoke Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=210.21650000000005) [ML Engine for prediction, gcloud ml-engine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=212.65485714285717) [predict, pass in the MODEL\_NAME, the](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=214.882) [version, and the JSON instances, and there](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=217.21466666666666) [is the price as predicted by our model.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=220.01) [This is the exact same price prediction](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=223.707) [that we got for this input when we ran prediction locally.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=7&mode=live&start=225.788)

[Summary and Further Study](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live)

[And with this demo, we come to the very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=2.194) [end of this module on Designing Robust](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=3.636699999999999) [Machine Learning Workflows on the GCP. In](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=5.98957142857143) [this module, we discussed how AI is](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=9.242999999999999) [getting increasingly democratized, which](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=11.427) [means machine learning is the core](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=13.844714285714286) [component around which data architectures](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=16.123) [are designed. We saw that the GCP offers](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=18.166) [an entire suite of services for machine](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=20.575) [learning and artificial intelligence, and](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=23.243) [many of these services don't require that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=25.69) [you write any code to harness the power of](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=27.756) [machine learning. So when you need to use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=30.7465) [machine learning for your specific use](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=32.92125000000001) [case, consider the entire ML suite rather](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=34.40514285714286) [than jumping into coding up your custom](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=36.998857142857155) [model using TensorFlow or some other](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=39.754999999999995) [framework. If your use case is a specific](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=42.48959999999999) [one and requires a more complex model than](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=44.56888888888888) [BigQuery ML, AutoML, or the ML APIs can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=47.39999999999999) [offer, that's when you'll choose to](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=50.66874999999998) [embrace TensorFlow and Cloud ML Engine.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=52.649) [And on this note, we've come to the very](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=55.603) [end of this course on data architectures.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=57.514) [Now before you head out, make sure you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=60.795) [clean up after yourself so you're not](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=62.371) [charged for resources that you provision.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=64.075) [Clean up any Cloud Storage buckets that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=66.491) [you might have created, Pub/Sub topics.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=67.95662499999996) [Make sure that you delete your Cloud](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=71.141) [Dataprep workflows and any ML models that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=72.686) [you might have deployed. If you're](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=75.354) [interested in the wide variety of products](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=77.921) [and services that the GCP has to offer,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=79.557) [here are some courses on Pluralsight that](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=81.83655555555556) [you can watch, Designing and Implementing](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=83.94099999999997) [Solutions Using Google Machine Learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=86.17) [APIs, Designing and Implementing Solutions](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=88.82875) [Using Google Cloud AutoML. Both of these](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=90.87940000000002) [are courses that show you how you can](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=93.49214285714282) [harness the power of machine learning](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=95.58899999999997) [without writing the code for your own](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=97.4541666666667) [custom model. Or if you're interested in](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=100.10675000000002) [architectural patterns on the GCP,](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=102.58966666666669) [Leveraging Architectural Design Patterns](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=104.939) [on the Google Cloud is a course that you](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=106.50059999999999) [might find interesting. That's it from me](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=108.61899999999996) [here today. I hope you enjoyed this course. Thank you for listening.](https://app.pluralsight.com/player?course=google-cloud-platform-designing-scalable-data-architectures&author=janani-ravi&name=e56d6421-e8b7-413a-9829-c4f601fe627d&clip=8&mode=live&start=111.01449999999997)