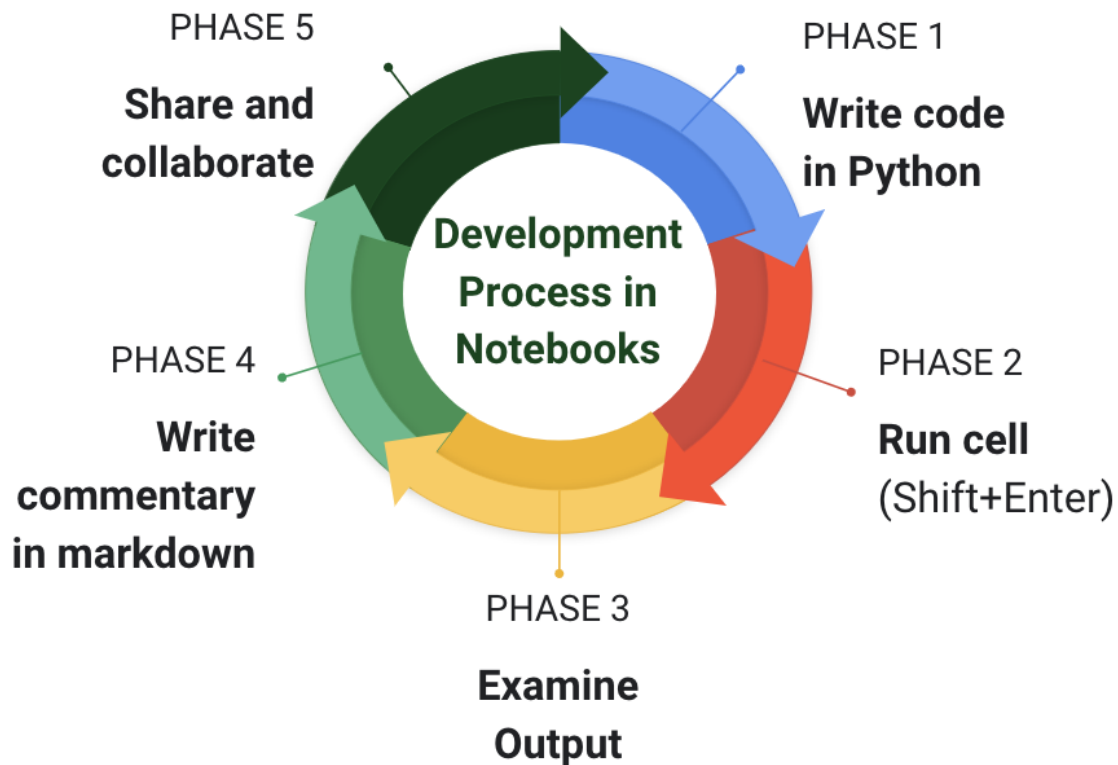


AI Platform Notebooks

Besides BigQuery, the other piece of technology you will use in the first lab is AI Platform Notebooks. AI Platform Notebooks is the next generation of hosted notebook on GCP, and has replaced Cloud Datalab. This is a managed Jupyter notebook environment that you can use to run Python code. It handles authentication to GCP, so that you can easily access BigQuery.



AI Platform Notebooks are developed in an iterative collaborative process. You can write code in Python, hit the run button, and the output shows up right on the page itself. Along with the code, you can write commentary in Markdown format and share the notebook with your colleagues.

The screenshot shows the AI Platform Notebooks interface. On the left, a sidebar contains a file explorer and a list of notebooks. The main area displays a Jupyter notebook with three sections: Markup (yellow), Code (blue), and Output (red). The Markup section contains text explaining how to use BigQuery. The Code section contains a SQL query. The Output section displays the results of the query as a table.

Markup

Let's see what a few rows of our data looks like. Any cell that starts with `%bigquery` will be interpreted as a SQL query that is executed on BigQuery, and the result is printed to our notebook.

BigQuery supports two flavors of SQL syntax: legacy SQL and standard SQL. The preferred is standard SQL because it complies with the official SQL:2011 standard. To instruct BigQuery to interpret our syntax as such we start the query with `#standardSQL`.

There are over 1 Billion rows in this dataset and it's 130GB large, so let's retrieve a small sample

Code

```
[7]: %bigquery --project $PROJECT
#standardSQL
SELECT
  *
FROM
  `nyc-tlc.yellow.trips`
WHERE RAND() < .000001 -- sample a small fraction of the data
```

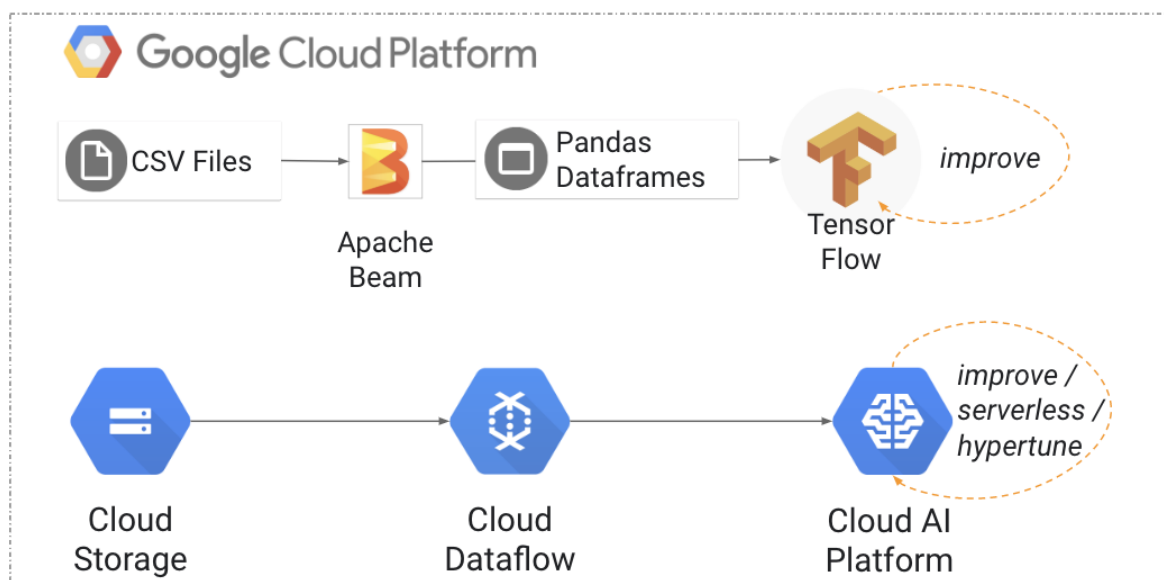
Output

	vendor_id	pickup_datetime	dropoff_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	rate_code	passenger_count
0	VTS	2010-04-26 21:40:00+00:00	2010-04-26 21:46:00+00:00	-73.972370	40.765112	-73.962437	40.755690	1	
1	CMT	2012-11-21 11:45:08+00:00	2012-11-21 11:56:27+00:00	-73.982350	40.746167	-74.003612	40.747571	1	
2	VTS	2012-10-13 22:14:00+00:00	2012-10-13 22:25:00+00:00	-73.986398	40.722425	-73.971102	40.753387	1	
3	CMT	2013-03-07 10:46:00+00:00	2013-03-07 10:54:25+00:00	-73.982214	40.778827	-73.990947	40.760982	1	

This is what the interface looks like. Notice how there are code sections interleaved with markup and output. This interleaving is what makes this style of computing so useful. Data analysis and machine learning are commonly carried out in notebooks like this. The code is in the blue section. You can execute the code by either clicking the Run button or by pressing Shift + Enter. The red section displays output from the command. Notice that the output here isn't just command line output; it's charts and tables as well. The yellow section contains markup, so you can explain why you're doing what you're doing in plain-text.



AI Platform
Notebooks



AI Platform Notebooks work with the same technologies that you're comfortable with, so you can start developing now, and then work on scale later. For example, we'll be doing an exercise where we read from a .csv file. You could then process in Pandas and Apache Beam before training a model in TensorFlow, and then improve the model through training.

Eventually though, when you are ready to scale, you can use Google Cloud Storage to hold your data initially, process it in Cloud Dataflow on an ephemeral cluster, and then run distributed training and hyper-parameter optimization in Cloud AI Platform.