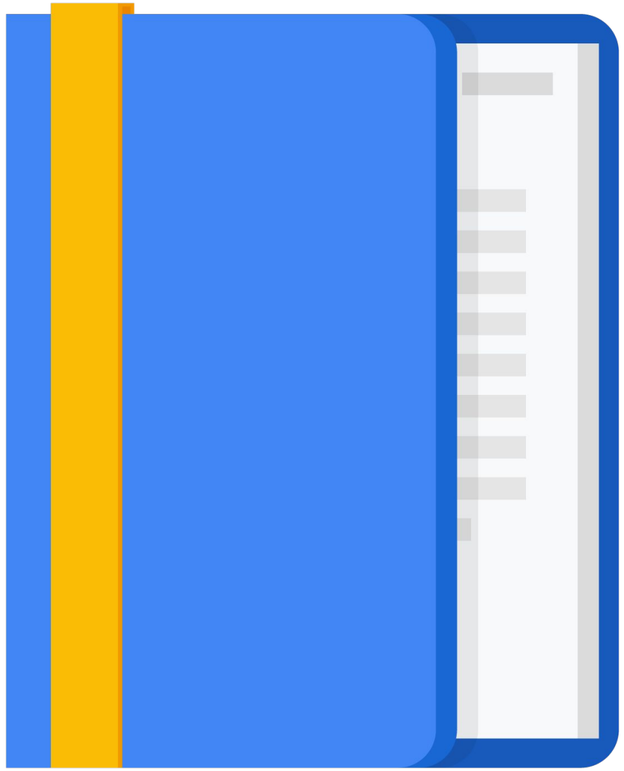
Google Cloud

Big Data Analytics with Cloud AI Platform Notebooks

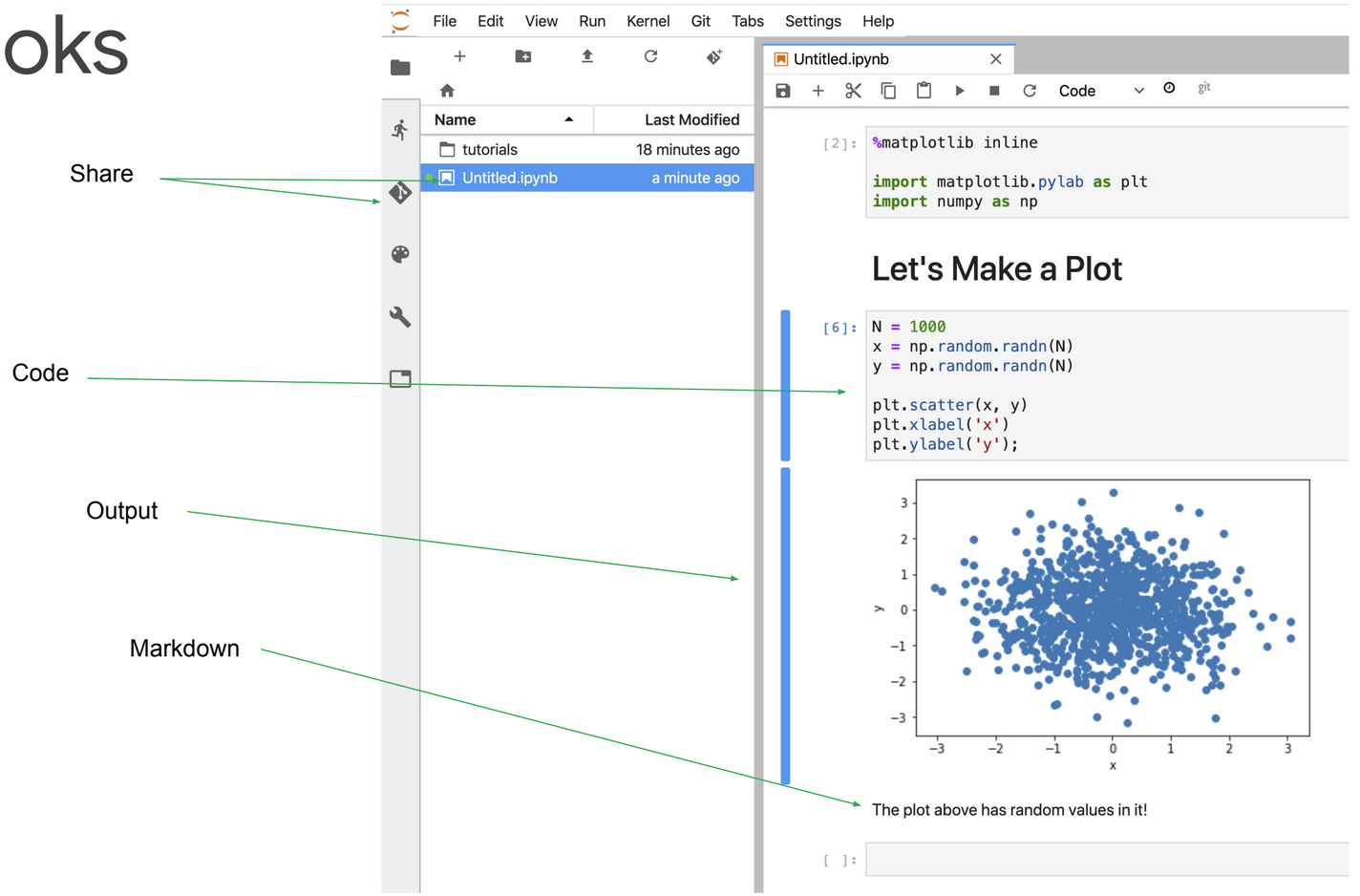
Agenda

What's a Notebook

BigQuery Magic and Ties to Pandas



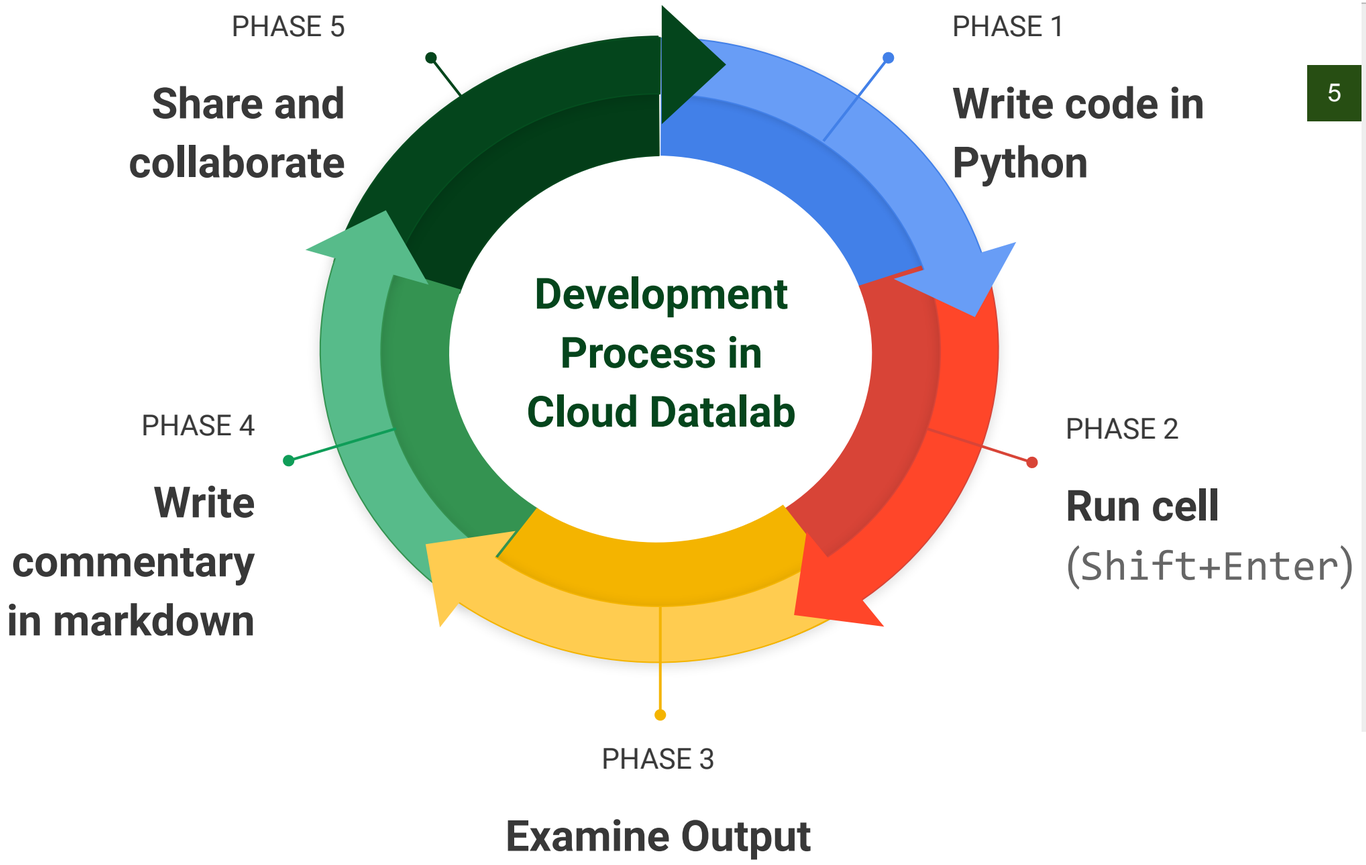
Increasingly, data analysis and machine learning are carried out in self-descriptive, shareable, executable notebooks

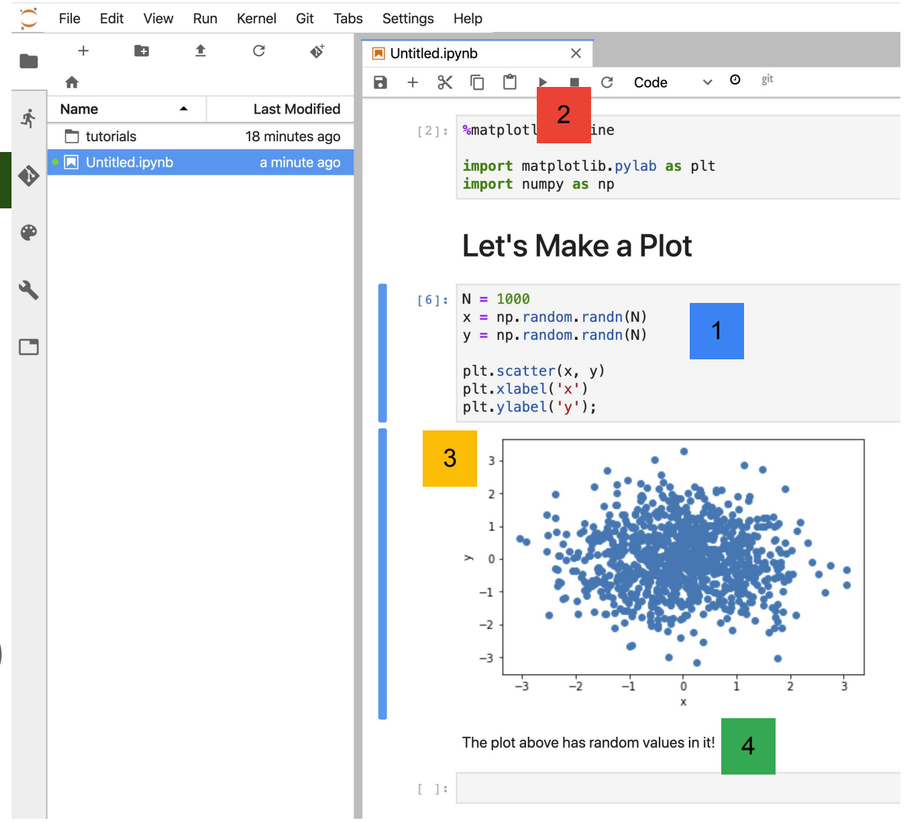


A typical notebook contains code, charts, and explanations

Google Cloud

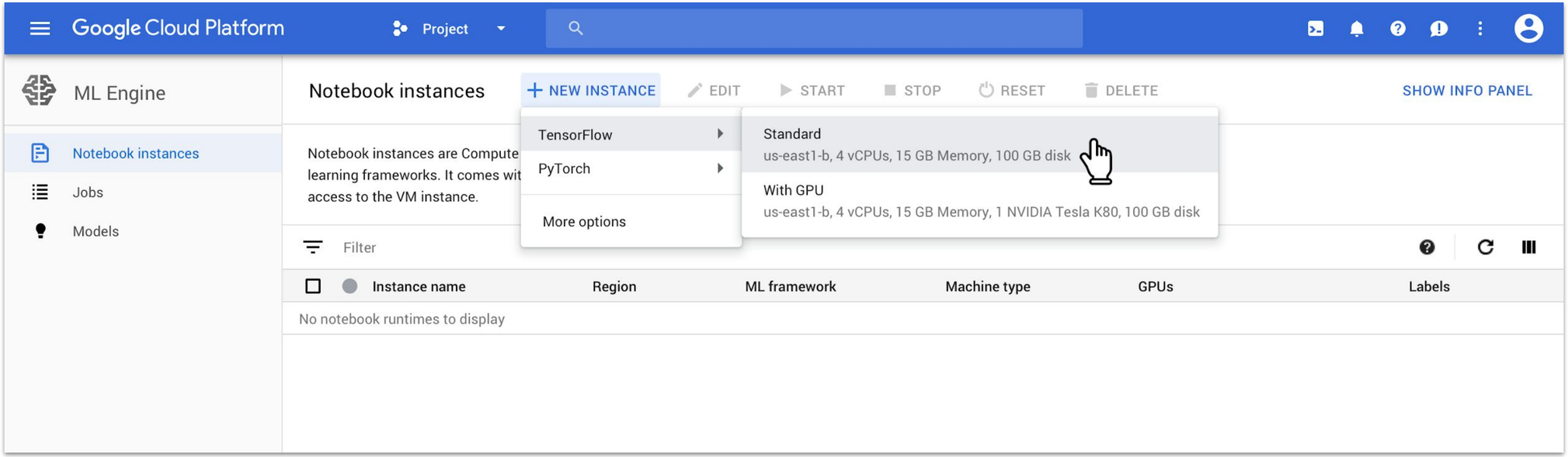
Notebooks are developed in an iterative, collaborative process





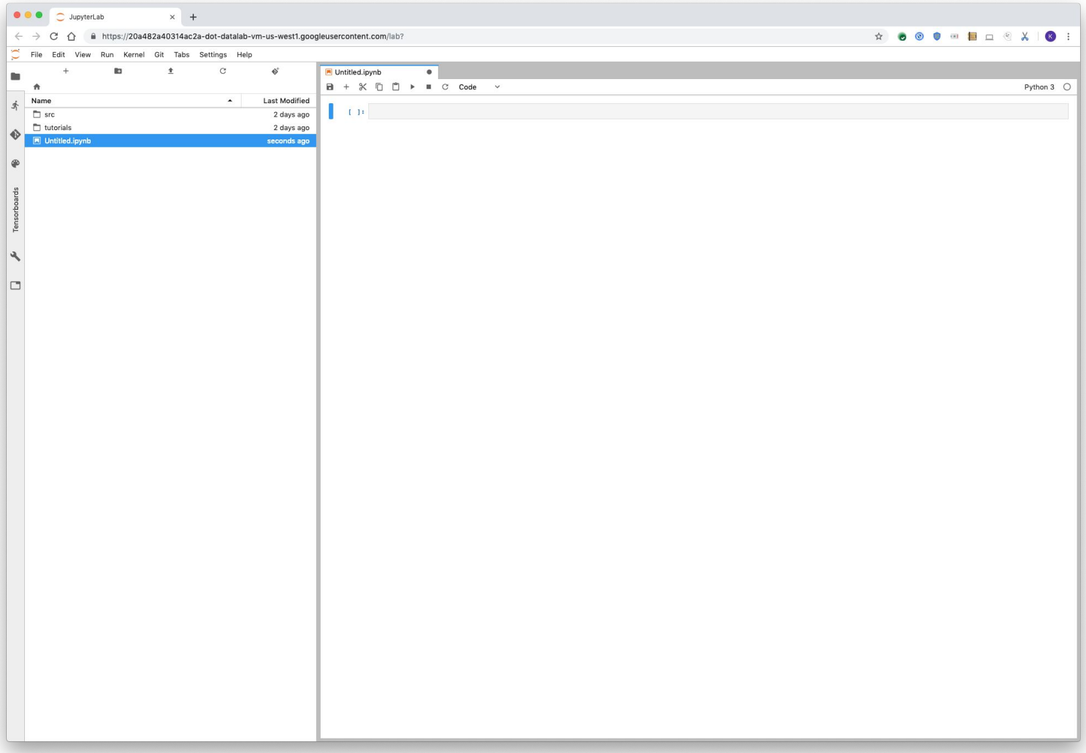
Google Cloud

Spin up a JuypterLab instance, pre-configured with the latest machine learning and data science frameworks in one click.



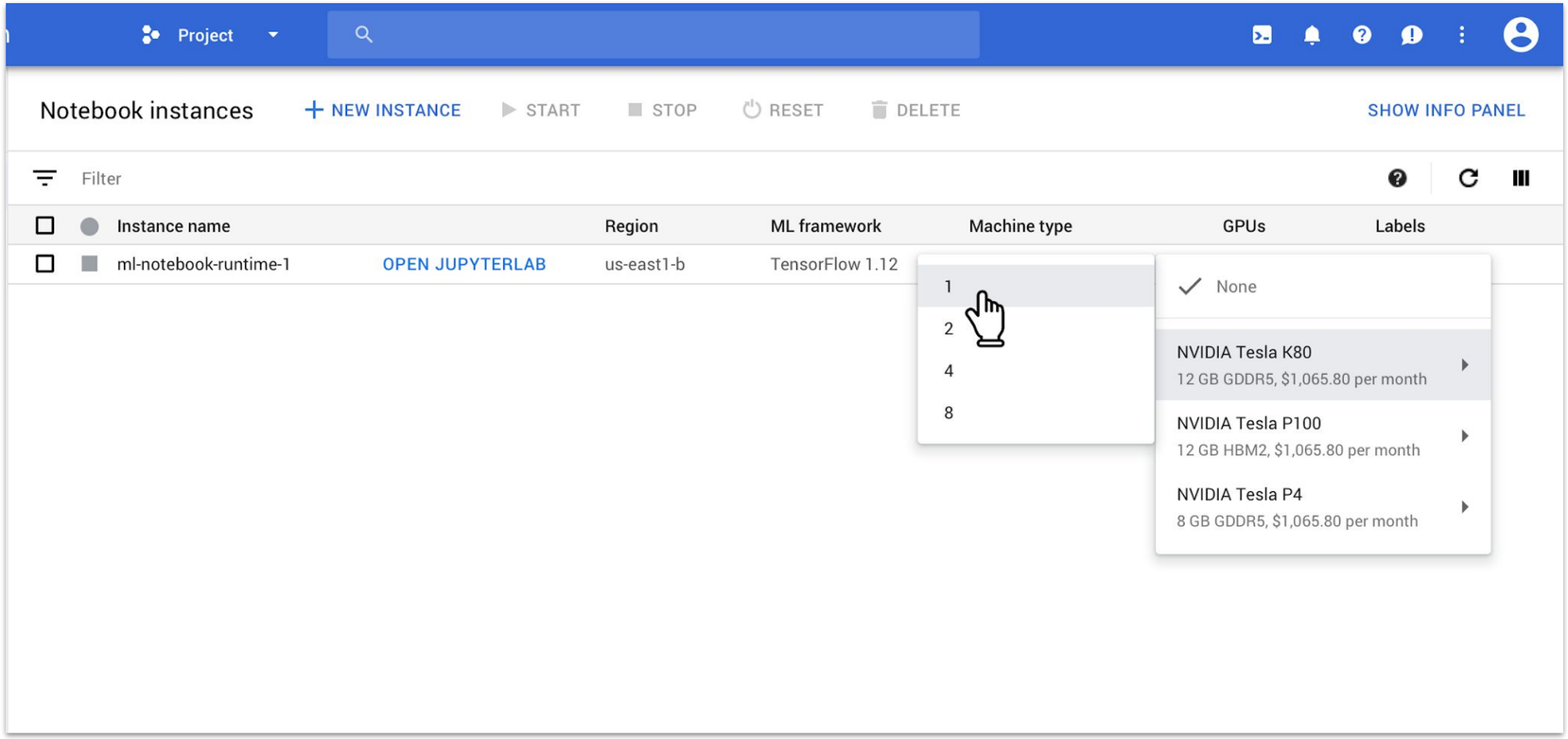
Google Cloud

AI Platform Notebooks uses the latest open-source version of the industry-standard JupyterLab

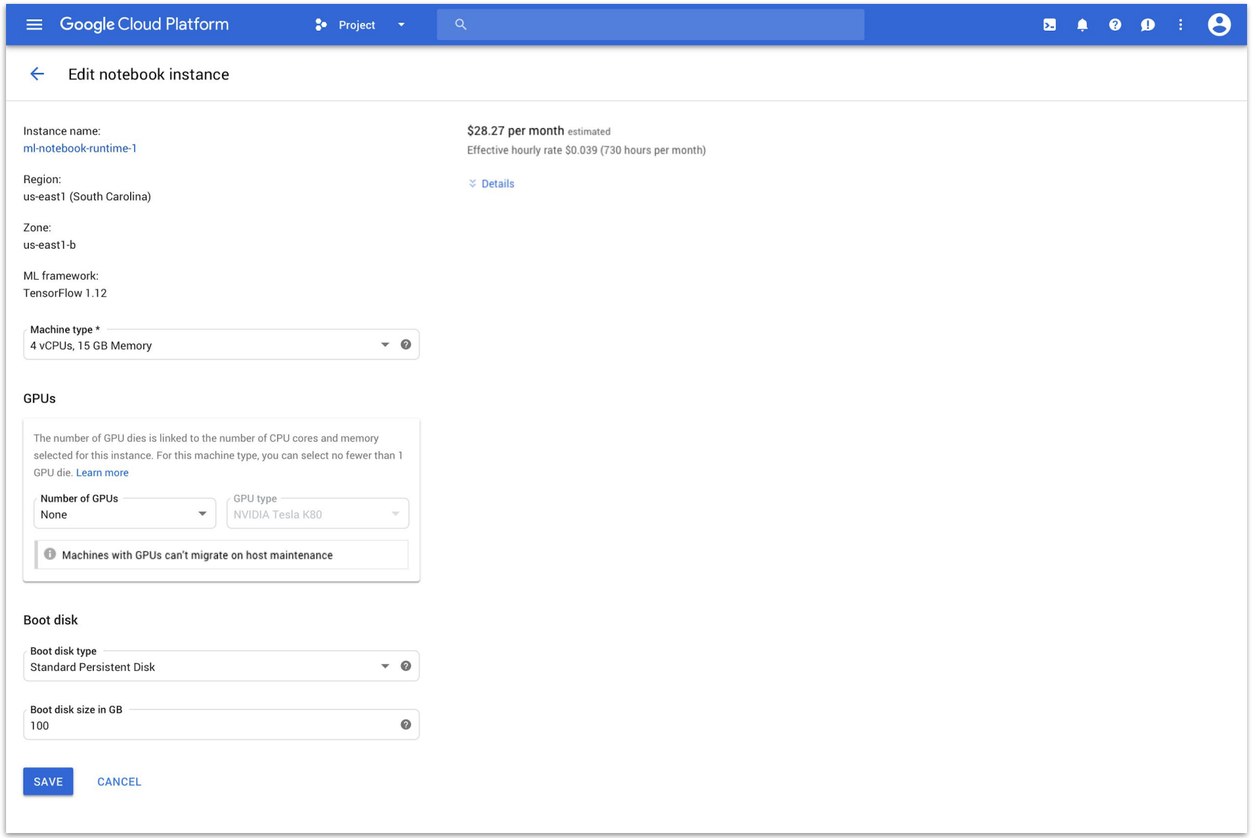


Google Cloud

You can easily change hardware including adding and removing GPUs

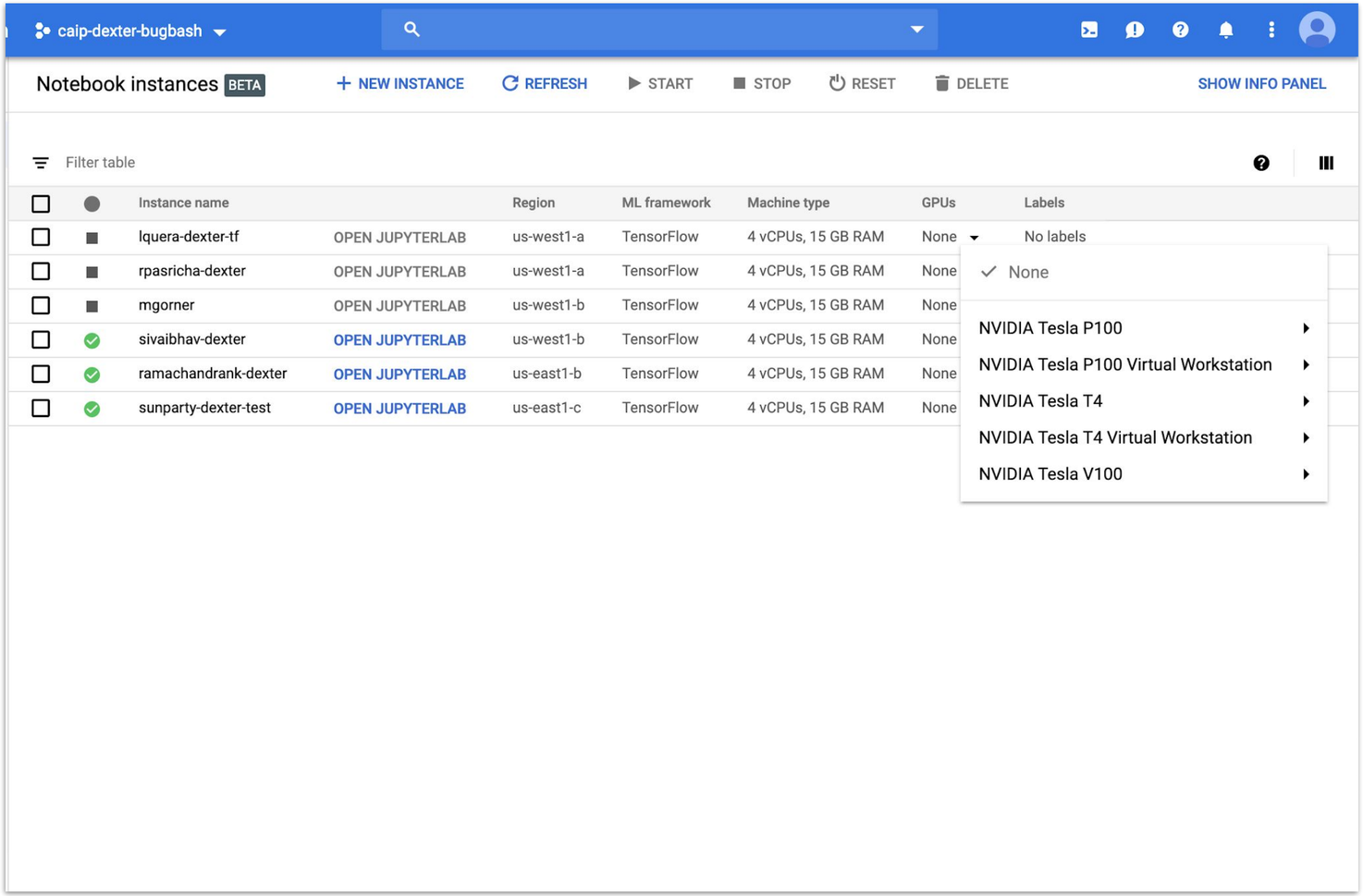


Use any GCE instance type. You can pick the hardware that makes sense, and scale up or down as needed



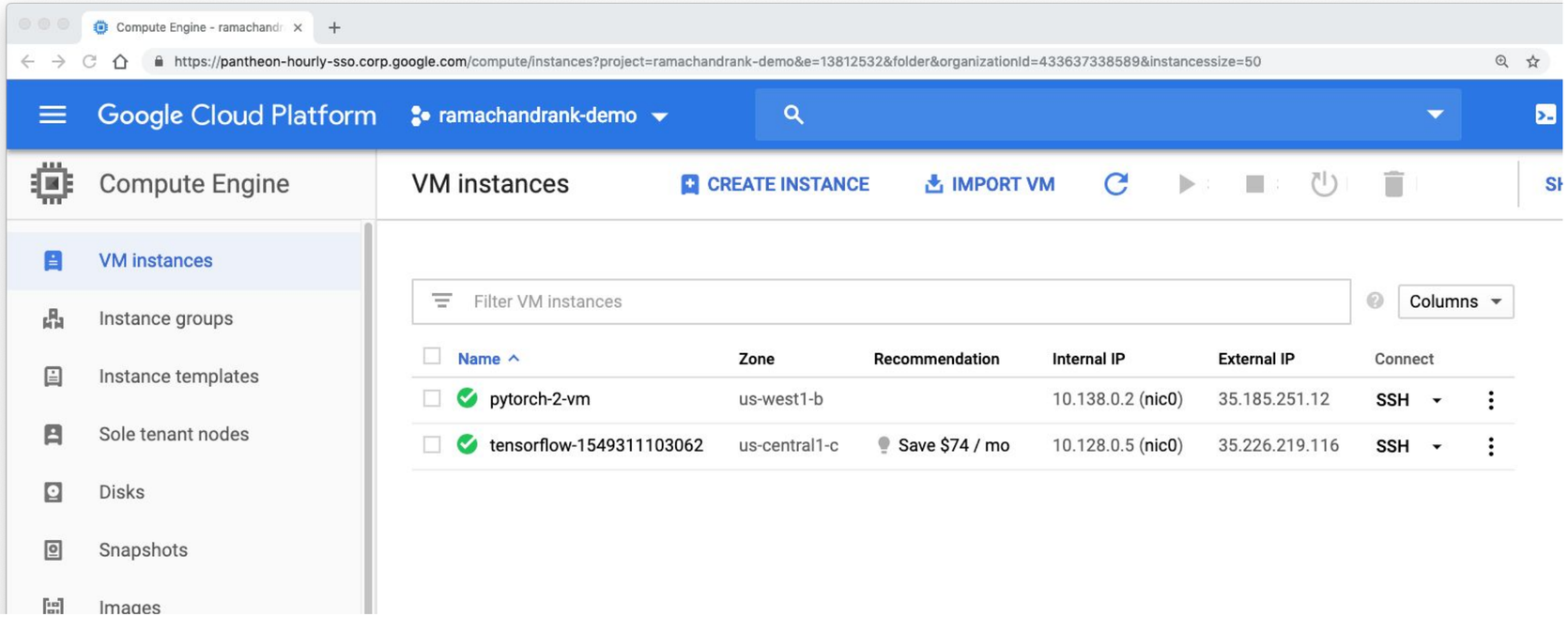
Google Cloud

You can even add and remove GPUs

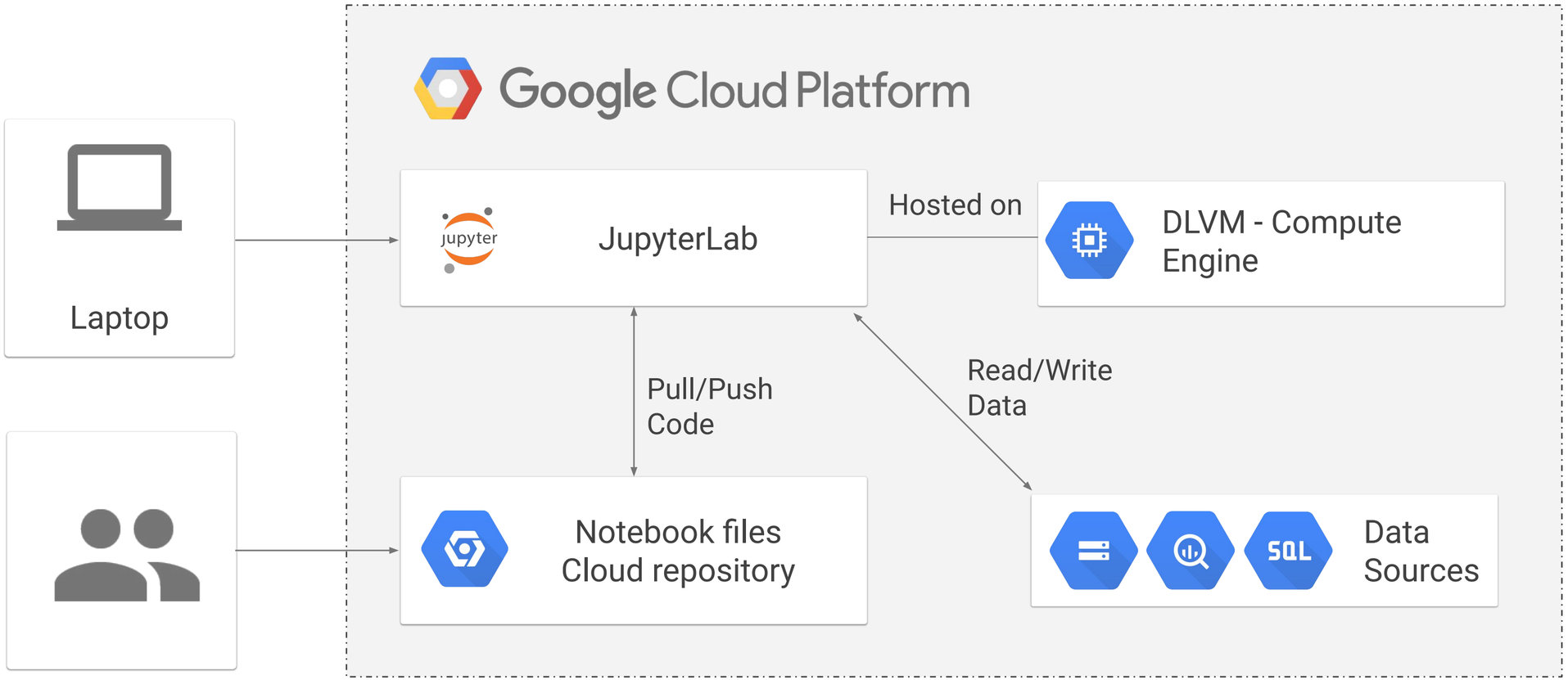


Google Cloud

Notebook instances are standard GCE instances that live in your projects



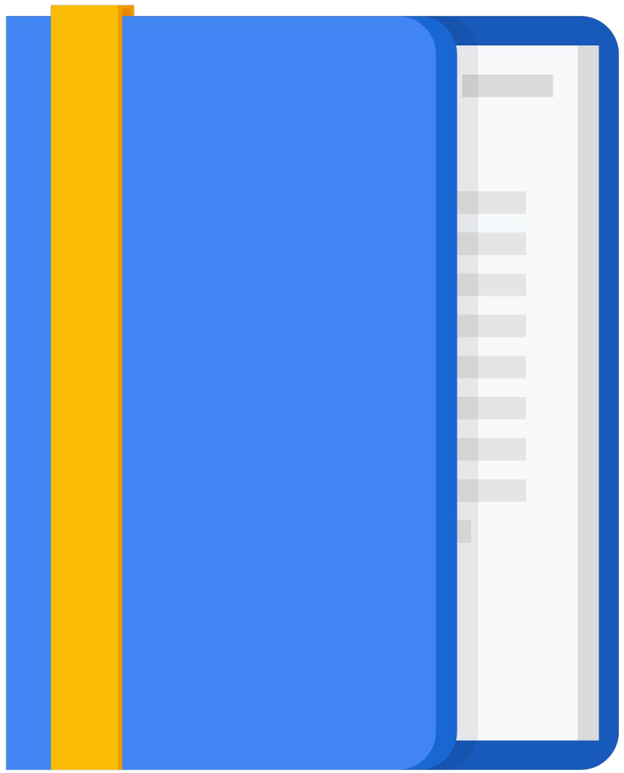
How does it work?



Agenda

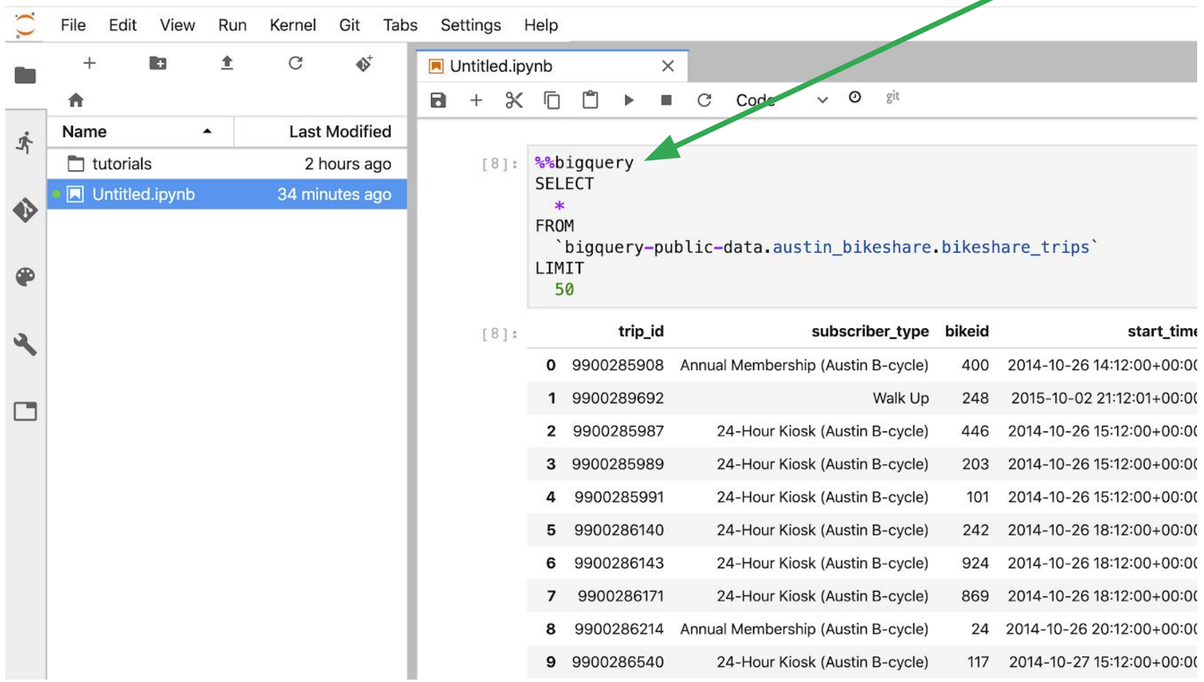
What's a Notebook

BigQuery Magic and Ties to Pandas



You can execute BigQuery commands from AI Platform Notebooks

- Useful for checking query validity
- Viewing query output
- But… can't use query output for anything



	trip_id	subscriber_type	bikeid	start_time
0	9900285908	Annual Membership (Austin B-cycle)	400	2014-10-26 14:12:00+00:00
1	9900289692	Walk Up	248	2015-10-02 21:12:01+00:00
2	9900285987	24-Hour Kiosk (Austin B-cycle)	446	2014-10-26 15:12:00+00:00
3	9900285989	24-Hour Kiosk (Austin B-cycle)	203	2014-10-26 15:12:00+00:00
4	9900285991	24-Hour Kiosk (Austin B-cycle)	101	2014-10-26 15:12:00+00:00
5	9900286140	24-Hour Kiosk (Austin B-cycle)	242	2014-10-26 18:12:00+00:00
6	9900286143	24-Hour Kiosk (Austin B-cycle)	924	2014-10-26 18:12:00+00:00
7	9900286171	24-Hour Kiosk (Austin B-cycle)	869	2014-10-26 18:12:00+00:00
8	9900286214	Annual Membership (Austin B-cycle)	24	2014-10-26 20:12:00+00:00
9	9900286540	24-Hour Kiosk (Austin B-cycle)	117	2014-10-27 15:12:00+00:00

Can use the BigQuery API in Notebooks to return query results as a Pandas DataFrame

[44]:
```
%%bigquery df
SELECT
  *
FROM
  `bigquery-public-data.austin_bikeshare.bikeshare_trips`
WHERE
  end_station_name = 'Stolen'
```

[46]:
```
print(type(df))
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
```

[46]:

	trip_id	subscriber_type	bikeid	start_time	start_station_id	start_station_name	end_station_id	end_station_name	duration_minutes
0	9900259257	Walk Up	93	2015-09-18 08:12:05+00:00	2712	Toomey Rd @ South Lamar	None	Stolen	2863
1	16898448	Walk Up	1857	2018-03-18 22:51:20+00:00	2501	5th & Bowie	None	Stolen	3806
2	9900298869	Walk Up	127	2015-10-10 19:12:38+00:00	2574	Zilker Park	None	Stolen	3632
3	9900290440	Local365	277	2015-10-02 22:12:06+00:00	2494	2nd & Congress	None	Stolen	8
4	9900322570	Walk Up	439	2015-11-01 02:12:28+00:00	2496	8th & Congress	None	Stolen	6609

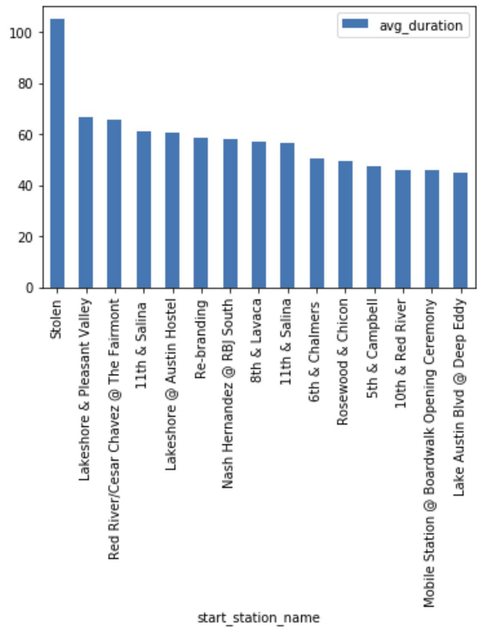
Pandas DataFrame

Google Cloud

Pandas + BigQuery in Notebook rocks!

[47]:
```
%%bigquery avg_dur_by_station
SELECT
  start_station_name,
  AVG(duration_minutes) as avg_duration
FROM
  `bigquery-public-data.austin_bikeshare.bikeshare_trips`
GROUP BY
  start_station_name
ORDER BY
  avg_duration
DESC
LIMIT 15
```

[48]:
```
avg_dur_by_station.plot(x='start_station_name', y='avg_duration', kind='bar');
```



Google Cloud

Lab

BigQuery in Jupyter Labs on AI Platform

Objectives

- Instantiate a Jupyter notebook on AI Platform
- Execute a BigQuery query from within a Jupyter notebook and process the output using Pandas

Module Summary

- AI Platform Notebooks are ideal for prototyping machine learning pipelines and models
- Notebooks integrate nicely with BigQuery and other GCP services