Google Cloud

Explore the Dataset

Advanced ML with TensorFlow on GCP

End-to-End Lab on Structured Data ML

Production ML Systems

Image Classification Models

Sequence Models

Recommendation Systems

Steps involved in doing ML on GCP

1. **Explore the dataset**
2. Create the dataset
3. Build the model
4. Operationalize the model

The most common ML models at Google are models that operate on structured data

Type of network	# of network layers	# of weights	% of deployed models
MLP0	5	20M	61%
MLP1	4	5M	
LSTM0	58	52M	29%
LSTM1	56	34M	
CNN0	16	8M	5%
CNN1	89	100M	

https://cloud.google.com/blog/big-data/2017/05/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu

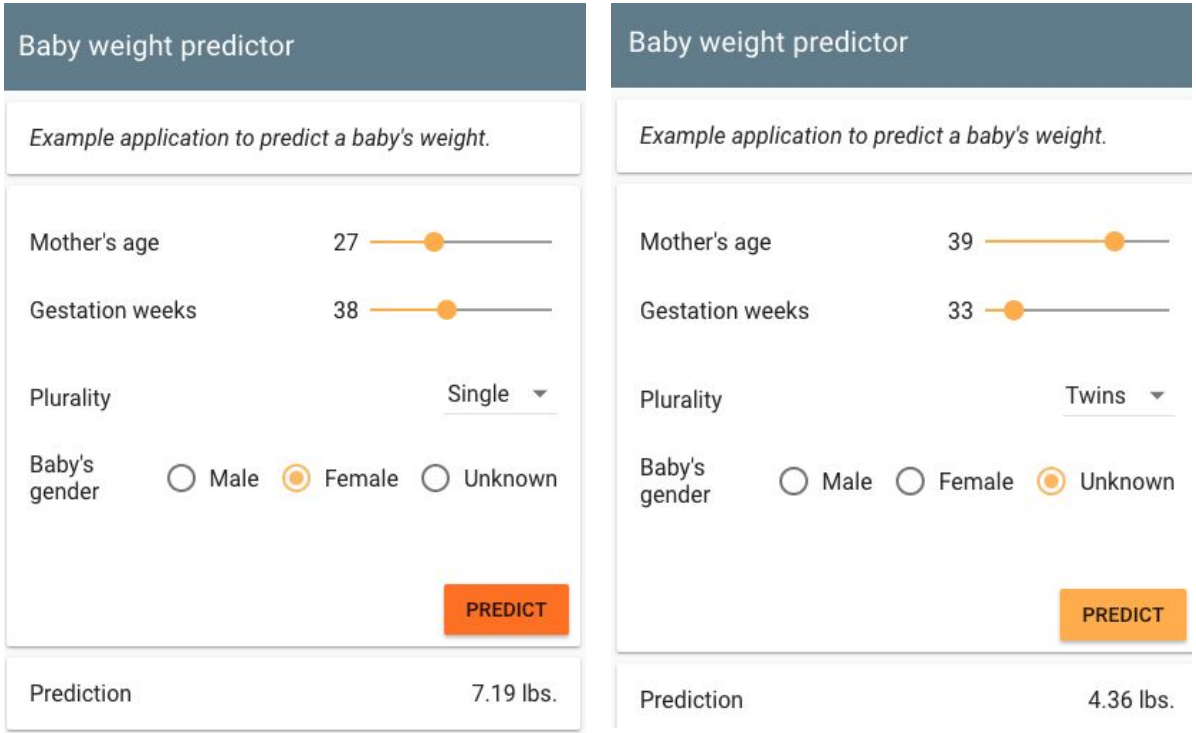
Our goal is to predict the weight of newborns so that all newborns can get the care they need

Predict the weight of newborns

Identify babies who may need special facilities

Get babies the care they need

This is what we will build



An open dataset of births is available in BigQuery

Births recorded in the 50 states of the USA from 1969 to 2008.

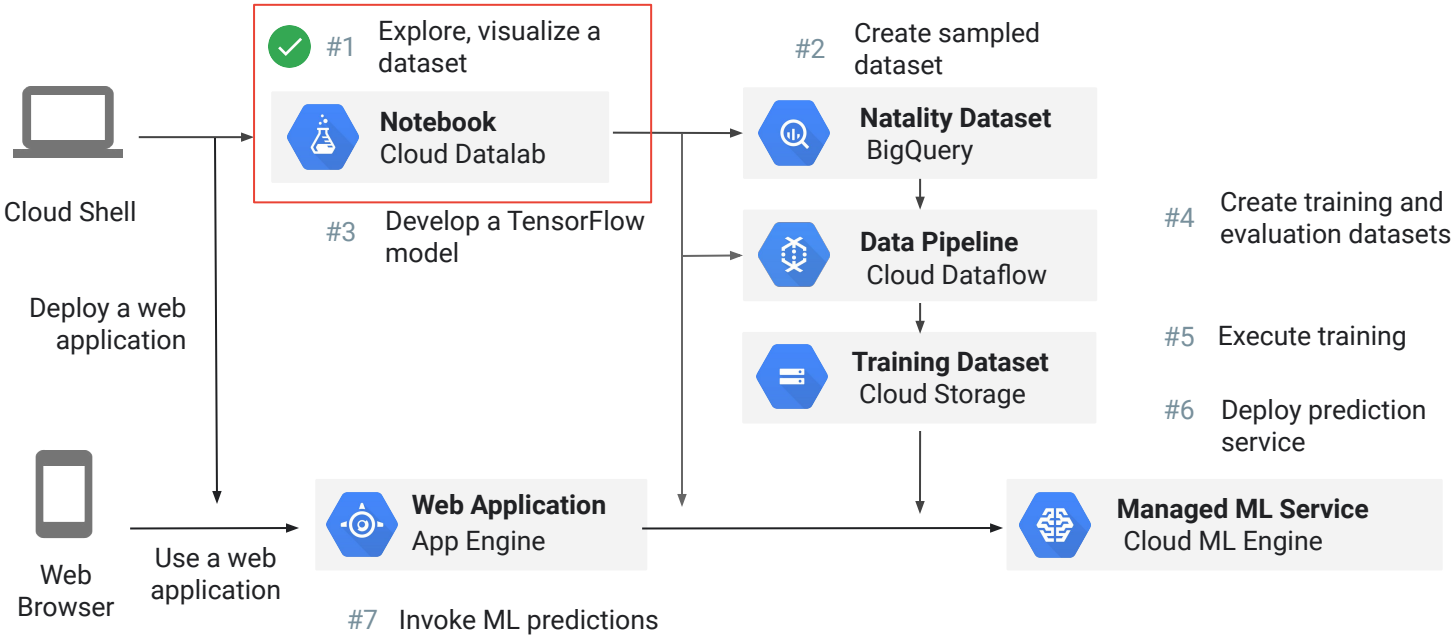
Table ID	bigquery-public-data:samples.natality
Table Size	21.9 GB
Long Term Storage Size	21.9 GB
Number of Rows	137,826,763

https://bigquery.cloud.google.com/table/bigquery-public-data:samples.natality

The data set includes details about the pregnancy

Date of birth	**year**	INTEGER	NULLABLE	Four-digit year of the birth. Example: 1975.
	month	INTEGER	NULLABLE	Month index of the date of birth, where 1=January.
	day	INTEGER	NULLABLE	Day of birth, starting from 1.
	wday	INTEGER	NULLABLE	Day of the week, where 1 is Sunday and 7 is Saturday.
Location of birth (US state)	**state**	STRING	NULLABLE	The two character postal code for the state. Entries after 2004 do not include this value.
Baby's birth weight (lbs)	**weight_pounds**	FLOAT	NULLABLE	Weight of the child, in pounds.
Mother's age at birth	**mother_age**	INTEGER	NULLABLE	Reported age of the mother when giving birth.
Duration of pregnancy	**gestation_weeks**	INTEGER	NULLABLE	The number of weeks of the pregnancy.
Mother's weight gain (lbs)	**weight_gain_pounds**	INTEGER	NULLABLE	Number of pounds gained by the mother during pregnancy.

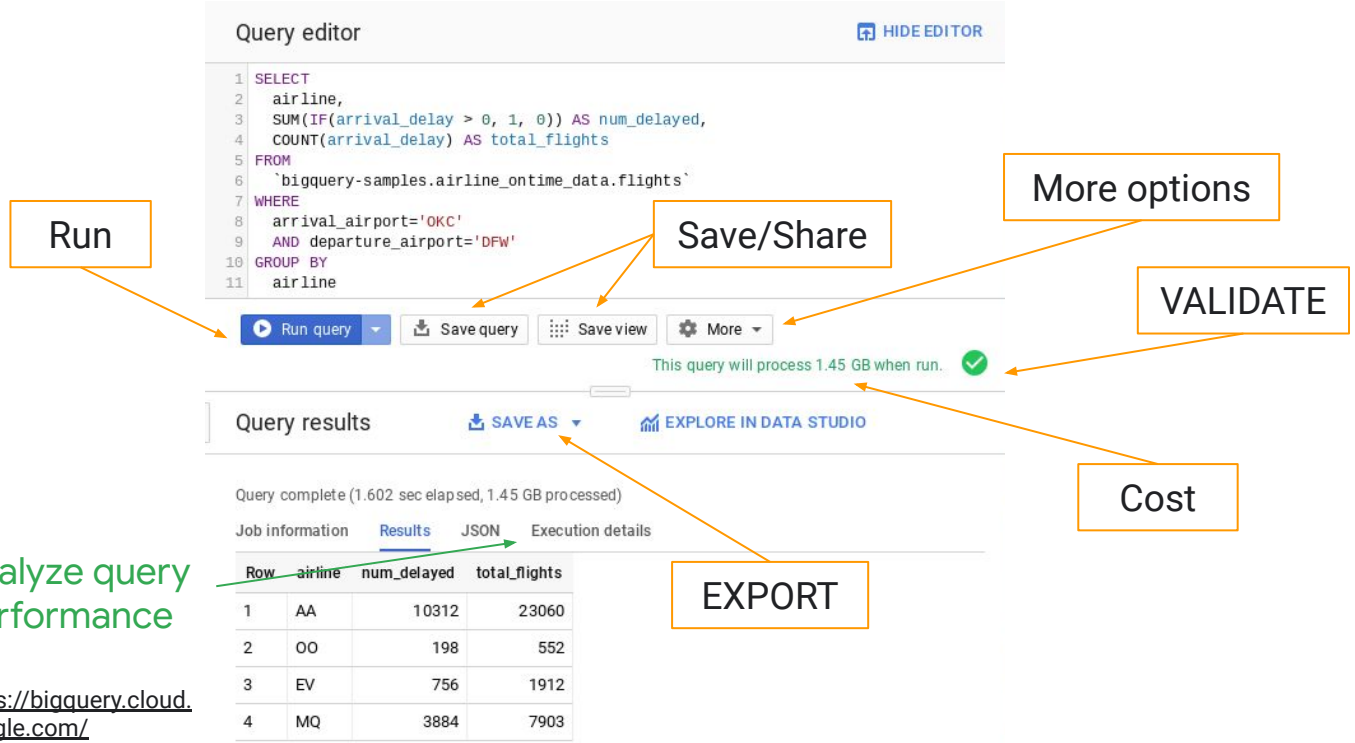
The end-to-end machine learning set of labs



BigQuery is a serverless data warehouse

1. Interactive analysis of petabyte scale databases.
2. Familiar, SQL 2011 query language and functions.
3. Many ways to ingest, transform, load, export data to/from BigQuery.
4. Nested and repeated fields, user-defined functions.
5. Data storage is inexpensive; queries charged on amount of data processed (or a monthly flat rate).

Run a query from BigQuery web UI



https://bigquery.cloud.google.com/

Demo: Query large datasets in seconds

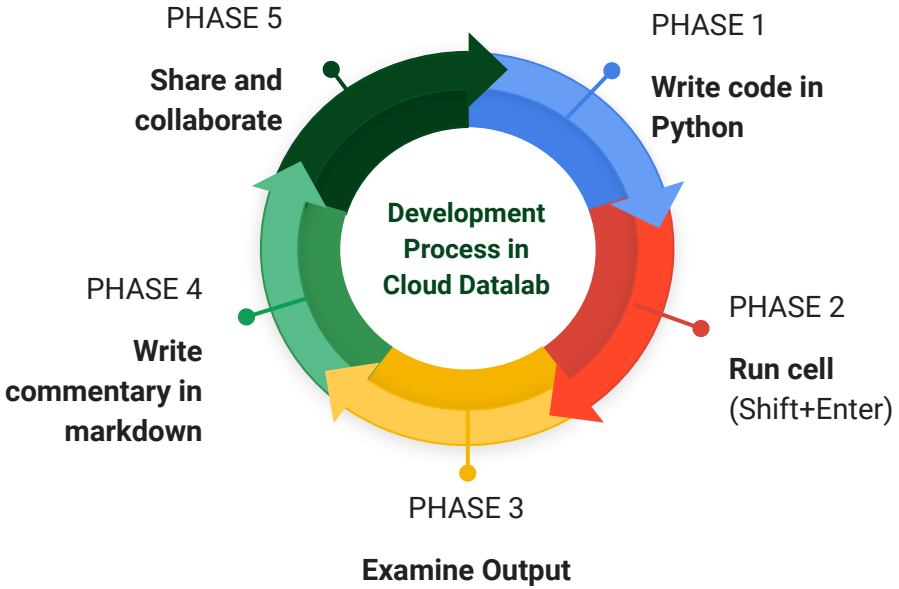
```
# standardsql

# medicare claims in 2014
SELECT
  nppes_provider_state AS state,
  ROUND(SUM(total_claim_count) / 1e6) AS total_claim_count_millions
FROM
  `bigquery-public-data.medicare.part_d_prescriber_2014`
GROUP BY
  state
ORDER BY
  total_claim_count_millions DESC
LIMIT 5;
```

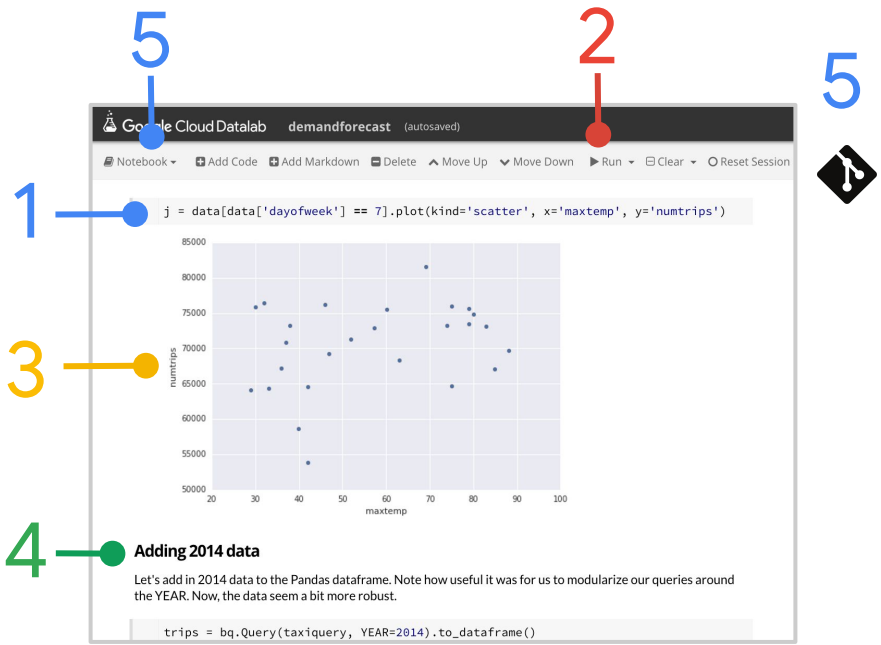
https://bigquery.cloud.google.com/savedquery/663413318684:781a98ddf2264505af2b6a8fc398a80e

Row	state	total_claim_count_millions
1	CA	116.0
2	FL	91.0
3	NY	80.0
4	TX	76.0
5	PA	63.0

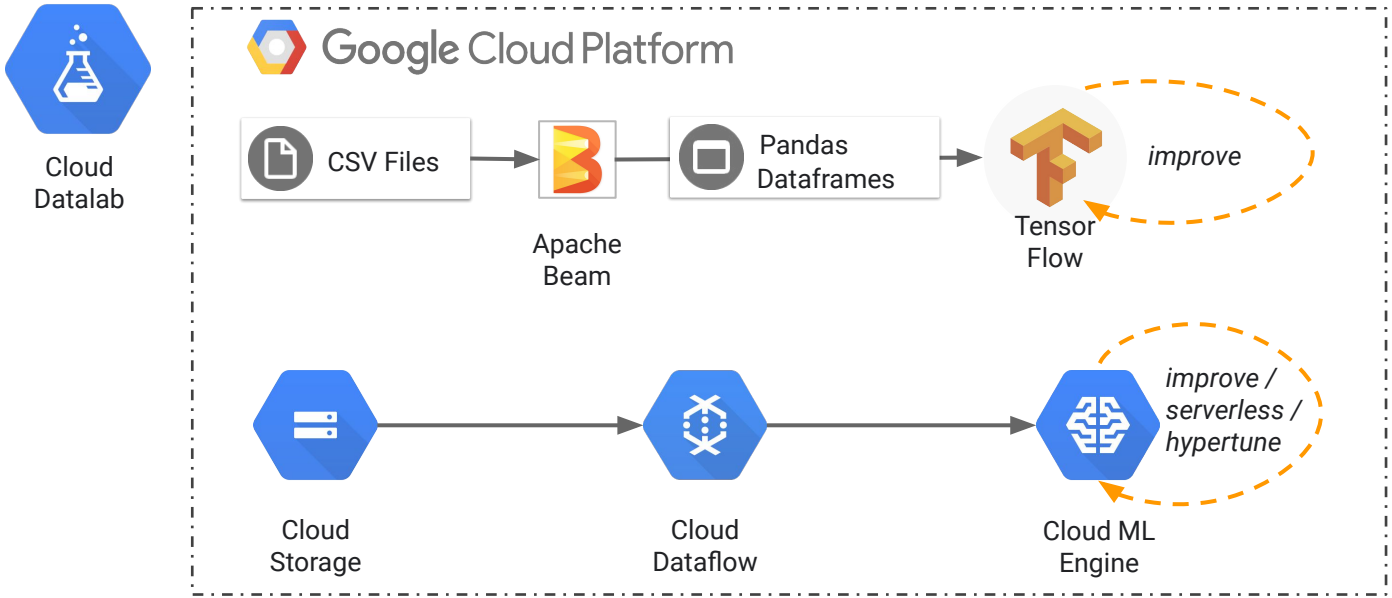
Cloud Datalab notebooks are developed in an iterative, collaborative process



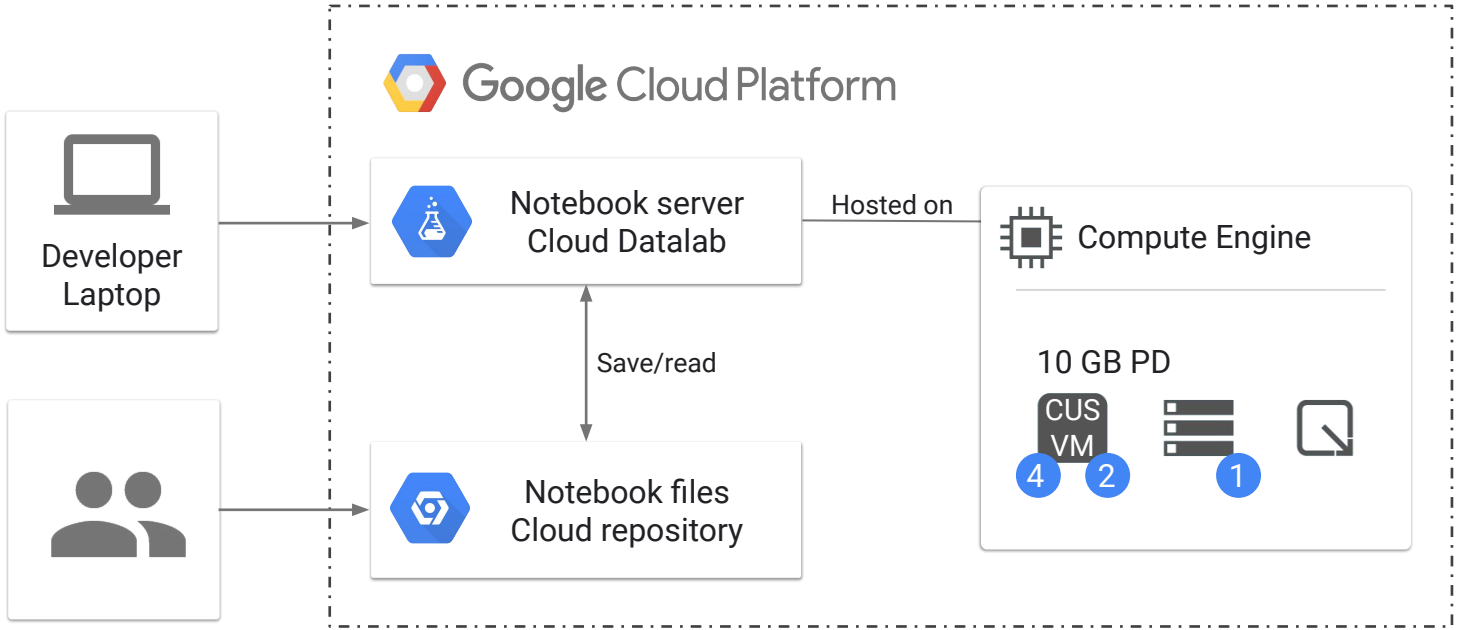
Cloud Datalab notebooks are developed in an iterative, collaborative process



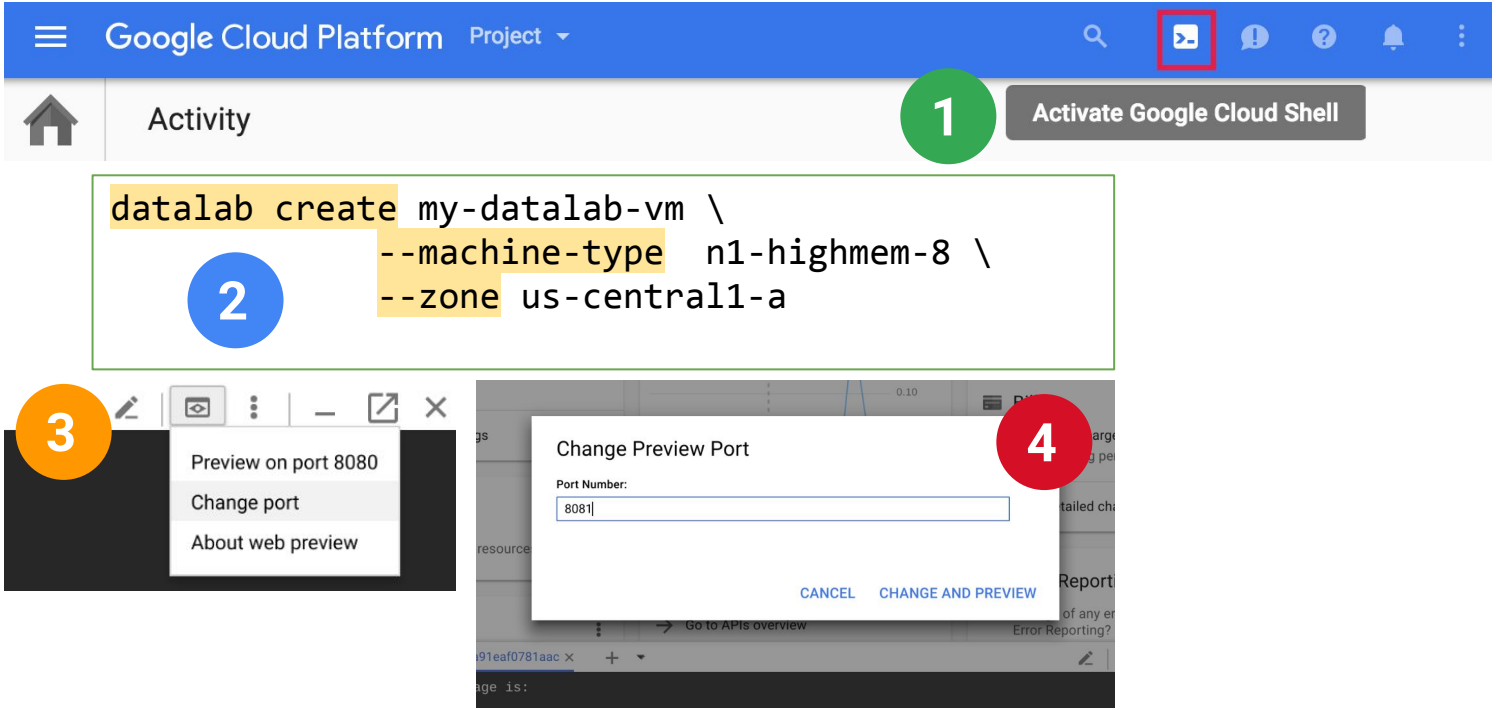
You can develop locally with Cloud Datalab and then scale out data processing to the cloud



Cloud Datalab notebooks let you change the underlying hardware



Starting Cloud Datalab in Cloud Shell is simple



Preprocessing data at scale with BigQuery + Cloud Datalab

BigQuery in Python to get a Pandas DF

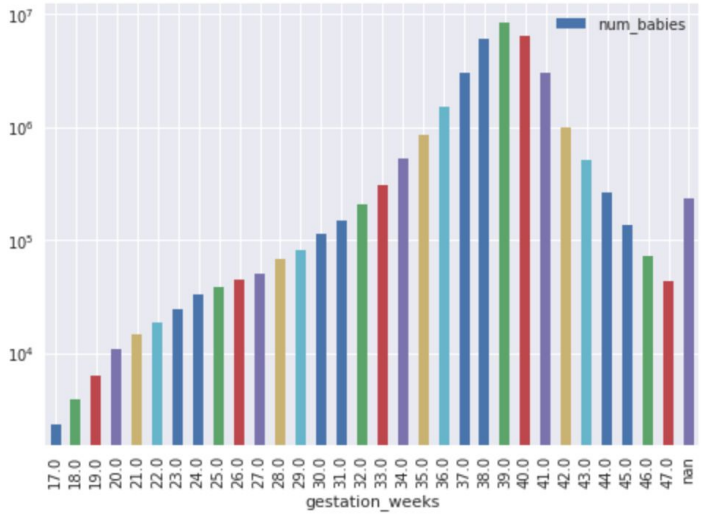
```
query = """
SELECT
  weight_pounds,
  is_male,
  mother_age,
  plurality,
  gestation_weeks,
  ABS(FARM_FINGERPRINT(CONCAT(CAST(YEAR AS STRING), CAST(month AS STRING))
FROM
  publicdata.samples.natality
WHERE year > 2000
"""
```

```
# Call BigQuery and examine in dataframe
import google.datalab.bigquery as bq
df = bq.Query(query + " LIMIT 100").execute().result().to_dataframe()
df.head()
```

	weight_pounds	is_male	mother_age	plurality	gestation_weeks	hashmonth
0	3.562670	True	25	1	30	1403073183891835564
1	3.999185	False	30	1	32	7146494315947640619

Pandas + BigQuery in notebook rocks!

```
# Bar plot to see gestation_weeks with avg_wt linear and num_babies logarithmic
df = get_distinct_values('gestation_weeks')
df = df.sort_values('gestation_weeks')
df.plot(x='gestation_weeks', y='num_babies', logy=True, kind='bar');
df.plot(x='gestation_weeks', y='avg_wt', kind='bar');
```

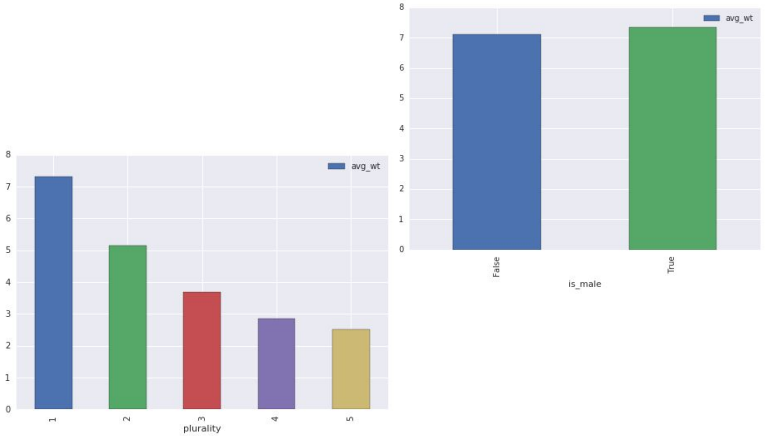


Lab

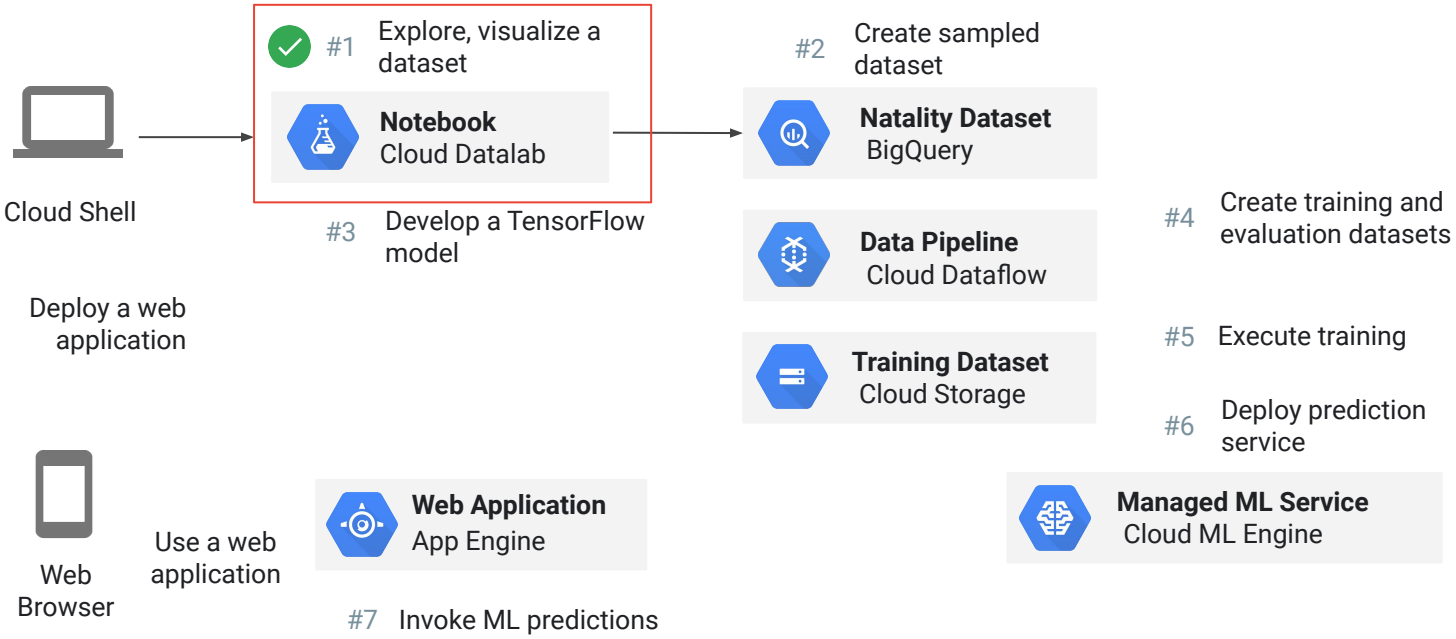
Explore a BigQuery dataset to find features to use in an ML model

In this lab, you will investigate which features have influence on what you want to predict: the baby's weight.

`publicdata.samples.natality`



The end-to-end process



cloud.google.com