CS 327E Class 12

December 2, 2019

Announcements

- CIS Survey: **Your voice matters**.
- Milestone 12: [Presentation Schedule](#)
- GCP credits: check your balance and request second coupon if needed.

1) What infrastructure components does BigQuery depend on?

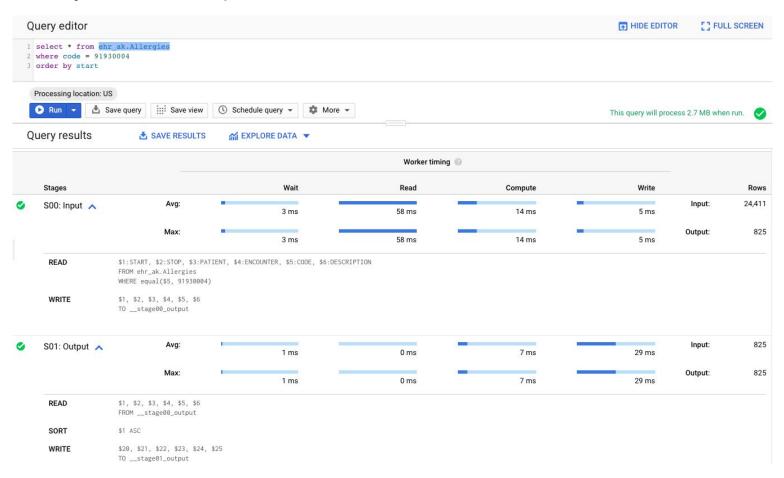   A. Colossus
   B. Borg
   C. Jupiter
   D. All of the above

2) What do the authors mean by "BigQuery is truly a serverless database"?

A. BigQuery is simple to use
B. BigQuery is a front-end app
C. BigQuery only needs a small amount of compute

3) What component orchestrates a BigQuery query?

A. Airflow
B. Composer
C. Dremel
D. Colossus

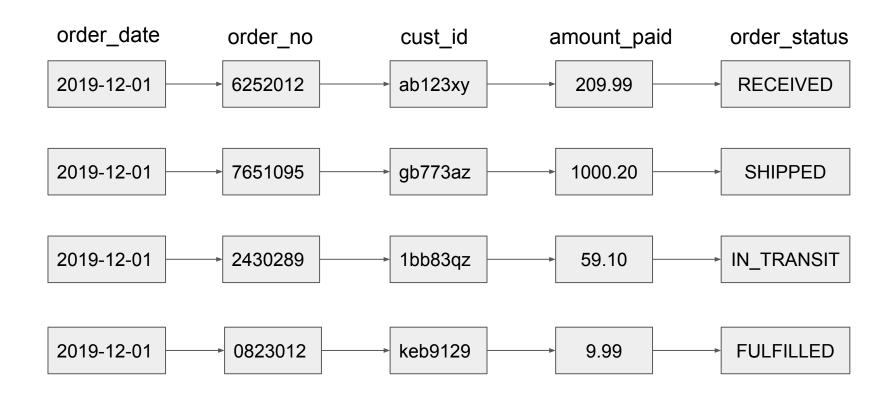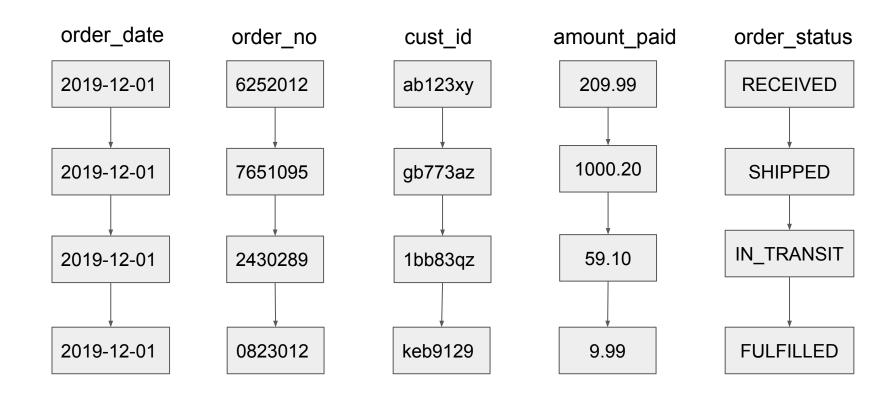# Query Plan Example

4) What type of file format does BigQuery use to store its records?

A. Row-oriented file format
B. Column-oriented file format
C. Parquet

Row-oriented file format

order_date	order_no	cust_id	amount_paid	order_status
2019-12-01	6252012	ab123xy	209.99	RECEIVED
2019-12-01	7651095	gb773az	1000.20	SHIPPED
2019-12-01	2430289	1bb83qz	59.10	IN_TRANSIT
2019-12-01	0823012	keb9129	9.99	FULFILLED

Column-oriented file format

order_date	order_no	cust_id	amount_paid	order_status
2019-12-01	6252012	ab123xy	209.99	RECEIVED
2019-12-01	7651095	gb773az	1000.20	SHIPPED
2019-12-01	2430289	1bb83qz	59.10	IN_TRANSIT
2019-12-01	0823012	keb9129	9.99	FULFILLED

5) What type(s) of failures does Borg protect BigQuery from?

A. Machine failures
B. Disk failures
C. Network failures
D. All of the above

Hands-on Exercise

git clone https://github.com/cs327e-fall2019/snippets.git

or

git pull origin master to pull down the latest code

5 Relevant Files:

01_college_workflow.py + 02_college_workflow.py
Student_single.py + Student_cluster.py + Takes_single.py

Practice Problem

Run and fix `02_college_workflow.py`

Practice Problem

Run and fix `02_college_workflow.py`

What was the cause of the error?

A. Syntax error
B. Logic error
C. All of the above