

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

Query editor

HIDE EDITOR

FULL SCREEN

```
1 select * from ehr_ak.Allergies
2 where code = 91930004
3 order by start
```

Processing location: US

Run

Save query

Save view

Schedule query

More

This query will process 2.7 MB when run.



Query results

SAVE RESULTS

EXPLORE DATA

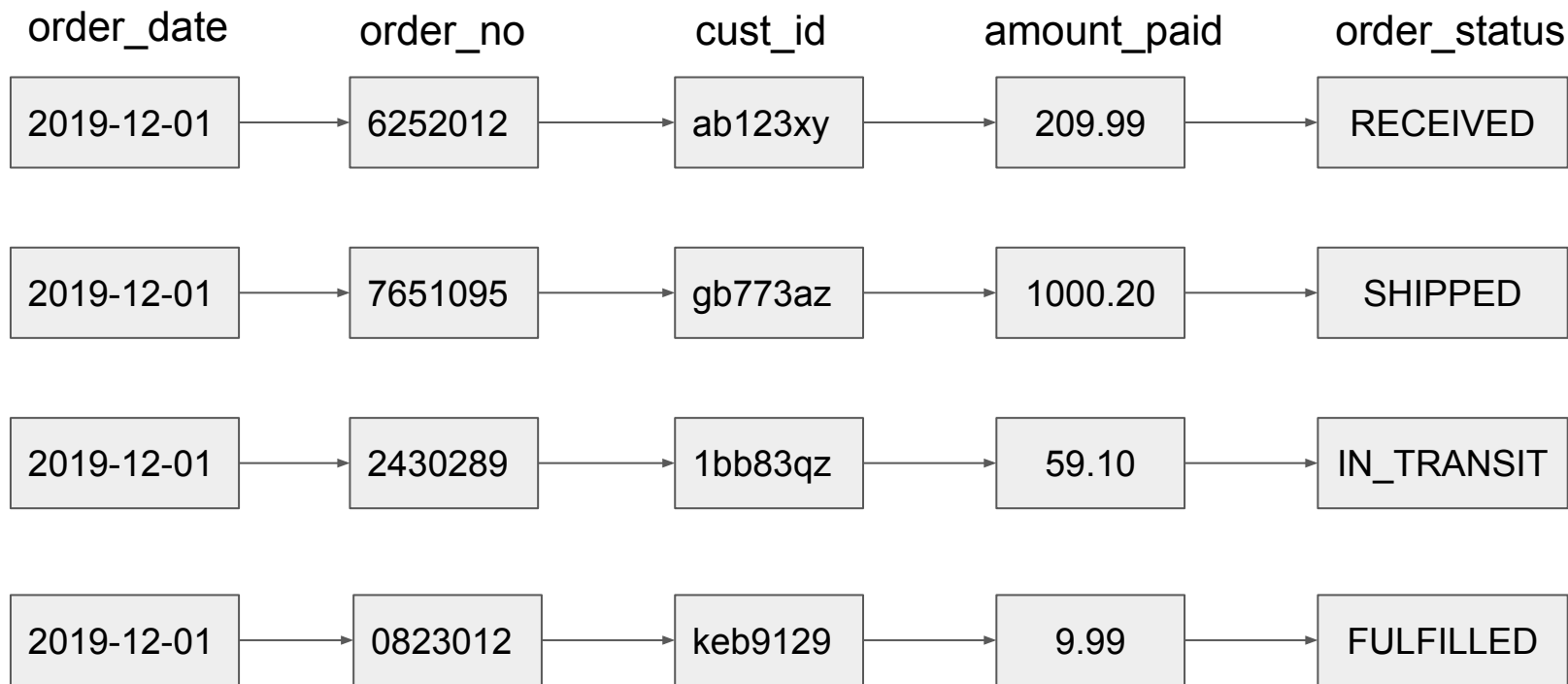
Worker timing

Stages	Wait	Read	Compute	Write	Rows
S00: Input	Avg: 3 ms	58 ms	14 ms	5 ms	Input: 24,411
	Max: 3 ms	58 ms	14 ms	5 ms	Output: 825
READ	\$1:START, \$2:STOP, \$3:PATIENT, \$4:ENCOUNTER, \$5:CODE, \$6:DESCRIPTION FROM ehr_ak.Allergies WHERE equal(\$5, 91930004)				
WRITE	\$1, \$2, \$3, \$4, \$5, \$6 TO __stage00_output				
S01: Output	Avg: 1 ms	0 ms	7 ms	29 ms	Input: 825
	Max: 1 ms	0 ms	7 ms	29 ms	Output: 825
READ	\$1, \$2, \$3, \$4, \$5, \$6 FROM __stage00_output				
SORT	\$1 ASC				
WRITE	\$20, \$21, \$22, \$23, \$24, \$25 TO __stage01_output				

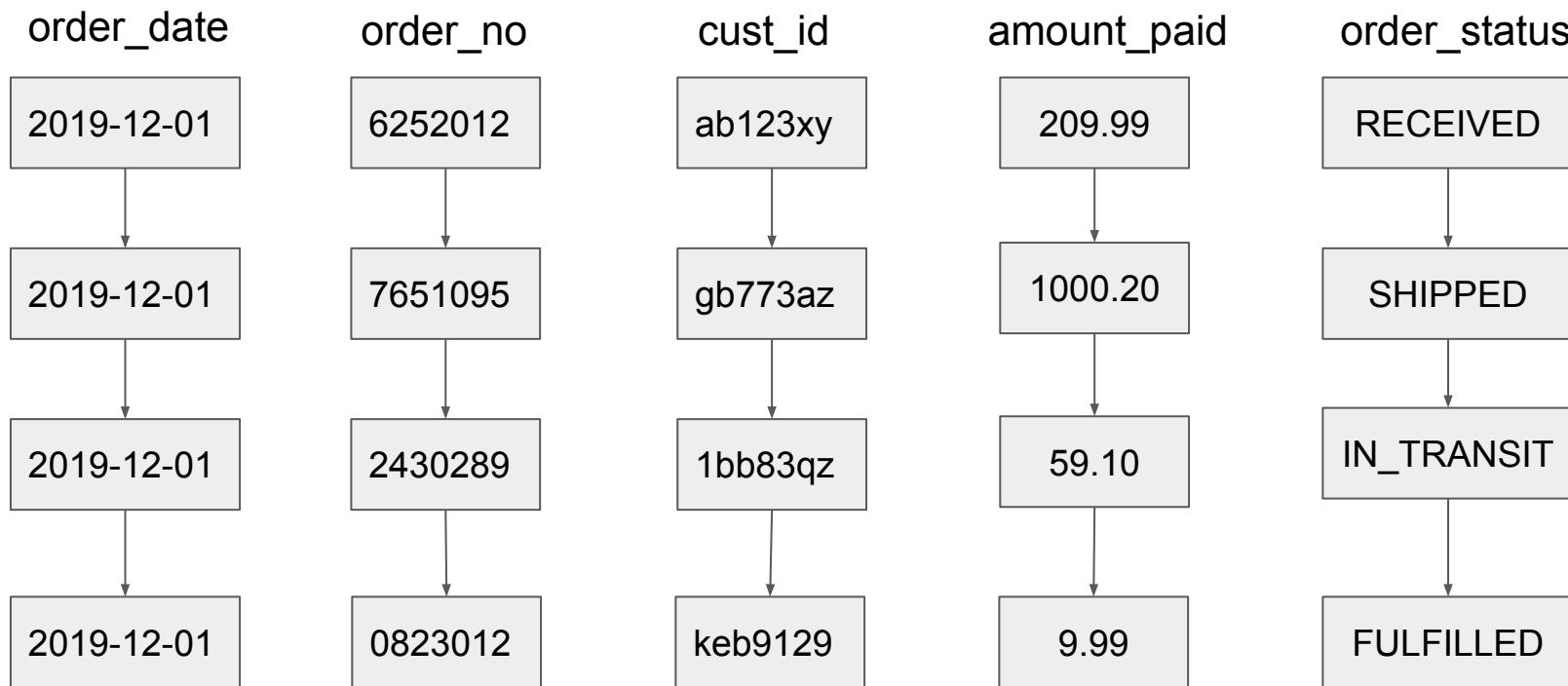
4) What type of file format does BigQuery use to store its records?

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

Row-oriented file format



Column-oriented file format



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