

# COMP 6231 Distributed Systems Design

## BIGQUERY



### CONTRIBUTORS:

- DHANANJAY NARAYAN (40164521)
- ISHIKA DHALL (40164795)
- SARVESH RAYTER (40162295)
- SHUBHAM VASHISTH (40164794)

# WE WILL BE DISCUSSING

PROBLEM  
STATEMENT

BIGQUERY

DISTRIBUTED  
DESIGN OF  
BIGQUERY

DEMONSTRATION

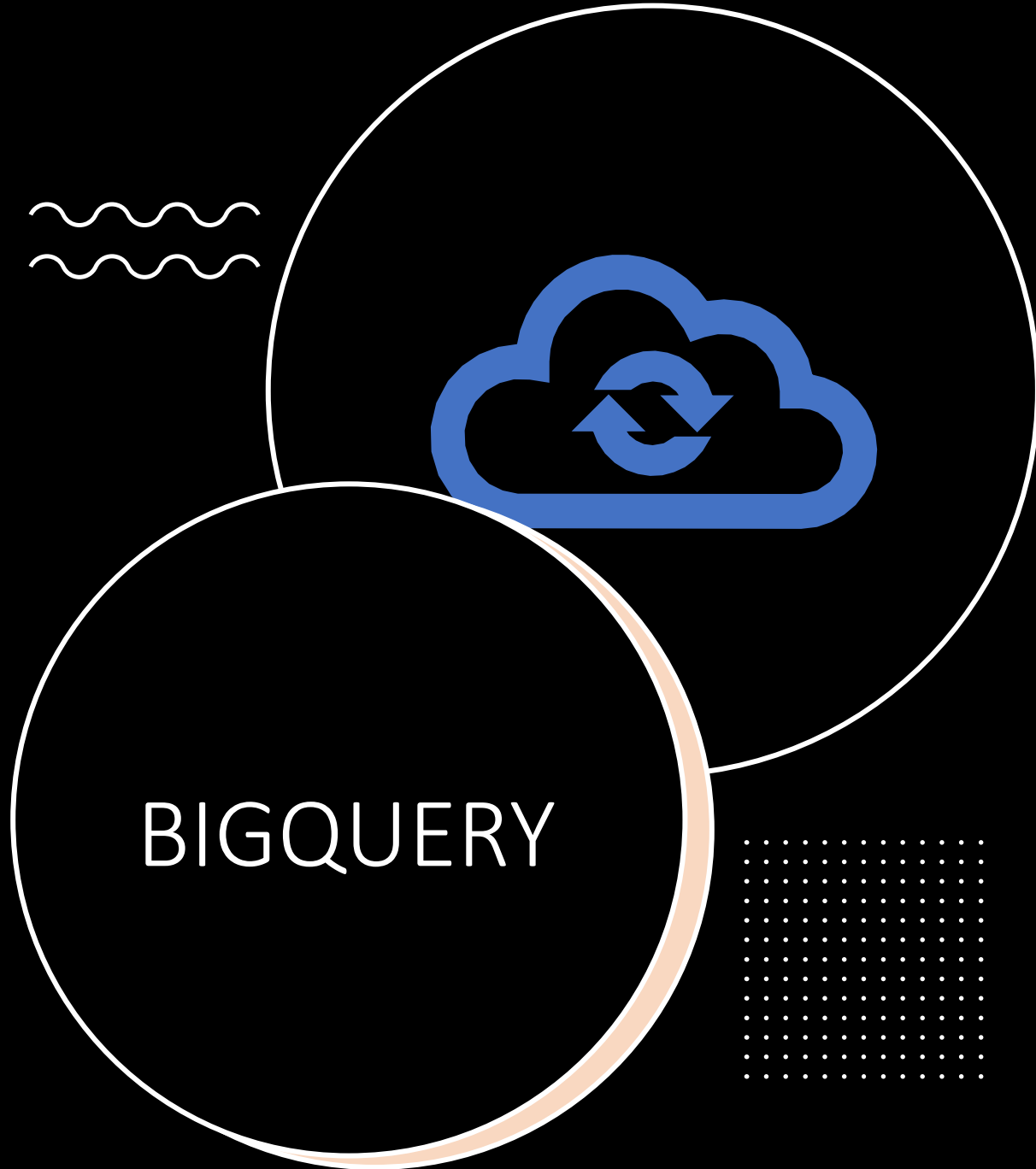
FINDINGS

Q/A

PROBLEM  
STATEMENT

**EPIDEMIOLOGY OF COVID-19 USING  
GOOGLE BIGQUERY**



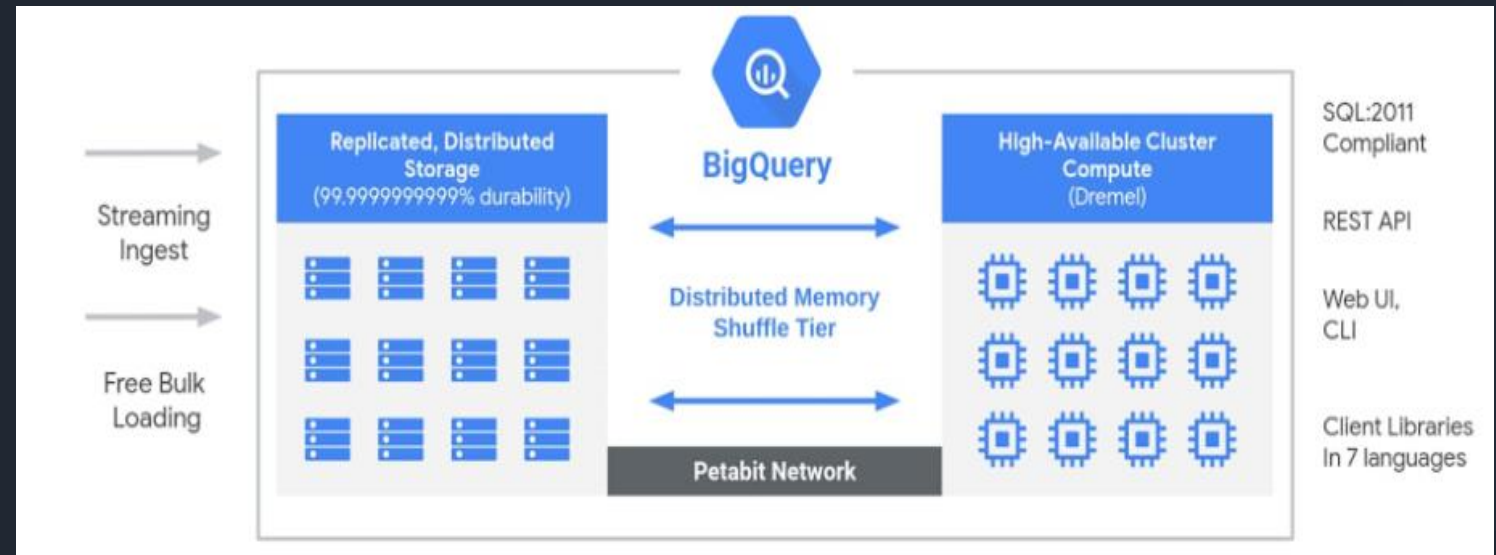


- An enterprise data warehouse
- Part of Google Cloud Platform
- REST-based web service
- Fully-managed
- Serverless
- Scalable analysis over petabytes of data
- Platform as a Service
- Supports querying using ANSI SQL



# DISTRIBUTED ASPECTS

- Decoupling serverless architecture
- Compute: Borg
- Query execution: Dremel
- Distributed File system: Colossus
- Naming: Key-value pairs
- Dremel to third-party buyers : REST API
- Fault Tolerance
- Security



A hand with a blue sleeve points at a line graph on a screen. The graph has a blue background with a red line showing an upward trend. The word "DEMONSTRATION" is written in white capital letters across the center of the image.

# DEMONSTRATION

---

The background of the slide features a magnifying glass held over a bar chart. The chart has blue and green bars grouped by quarter (Q1, Q2, Q3, Q4). A vertical line is drawn through the chart, passing behind the text. The word 'FINDINGS' is written in large white letters across the middle of the slide.

# FINDINGS

## PROS

- Allows you to run complex analytical SQL-based queries under large sets of data
- Faster Query execution

## CONS

- Pay-per-use but still expensive premium Google Cloud Storage

ANY QUESTIONS?

---