Introduction to BigQuery

INTRODUCTION TO BIGQUERY

Matt Forrest
Field CTO

datacamp

What is BigQuery?

- Uses SQL
- Scalable to analyze massive datasets
- Enterprise data warehouse
- Launched in 2012 using the same tools as Google uses

BigQuery

Google Cloud Official Blog

Built in the cloud. Engineered for your enterprise.

Google BigQuery brings Big Data analytics to all businesses

Tuesday, May 1, 2012

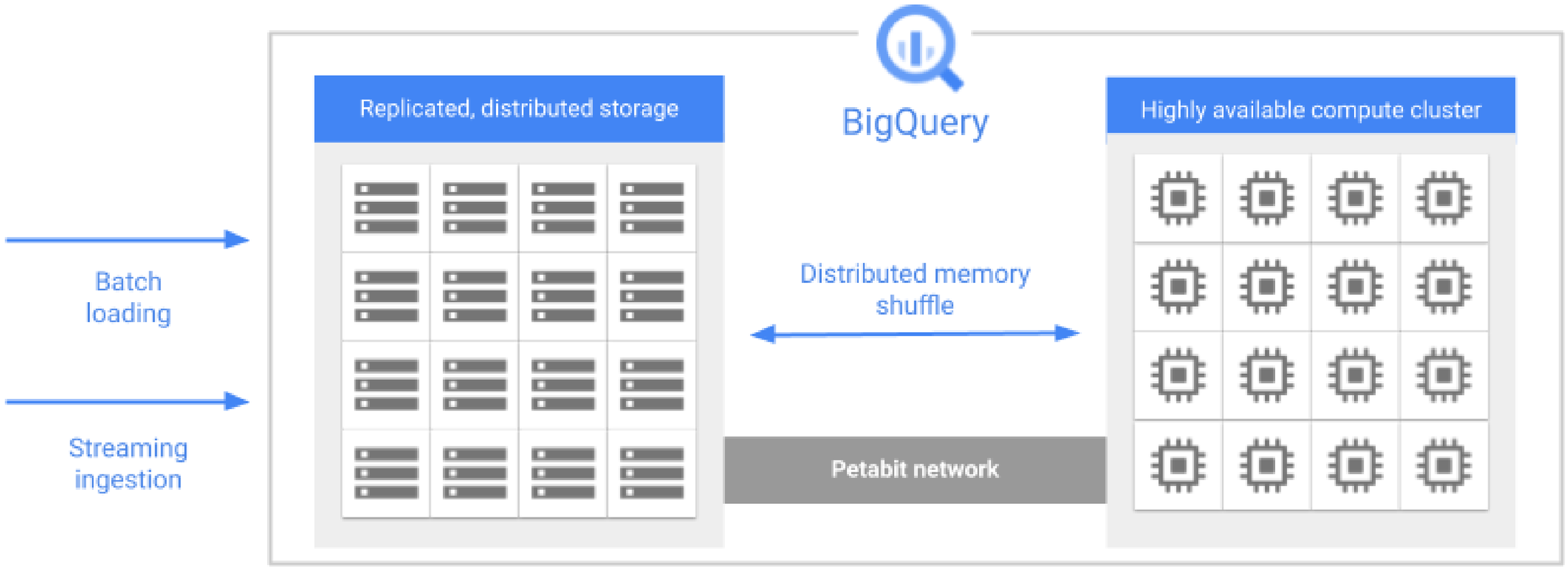
Posted by Ju-Kay Kwek, Product Manager, BigQuery

Cross-posted on the Google Developers Blog.

What makes BigQuery unique?

- Online analytical processing (OLAP)
- Separate compute and storage
- Serverless

Compute and storage



[1] https://cloud.google.com/bigquery/docs/storage_overview

Snowflake and BigQuery

Snowflake

- Popular amongst developers
- Runs on any cloud
- Specific tiers of compute resources (small, medium, etc.)

BigQuery

- Popular for analytical queries (reports)
- Only run on Google Cloud
- Completely serverless, no tiers

Redshift and BigQuery

Redshift

- Constant computation or serverless
- Suited for live dashboarding

BigQuery

- Only serverless
- Point in time analysis (once a day or hour)

Traditional SQL databases and BigQuery

SQL databases

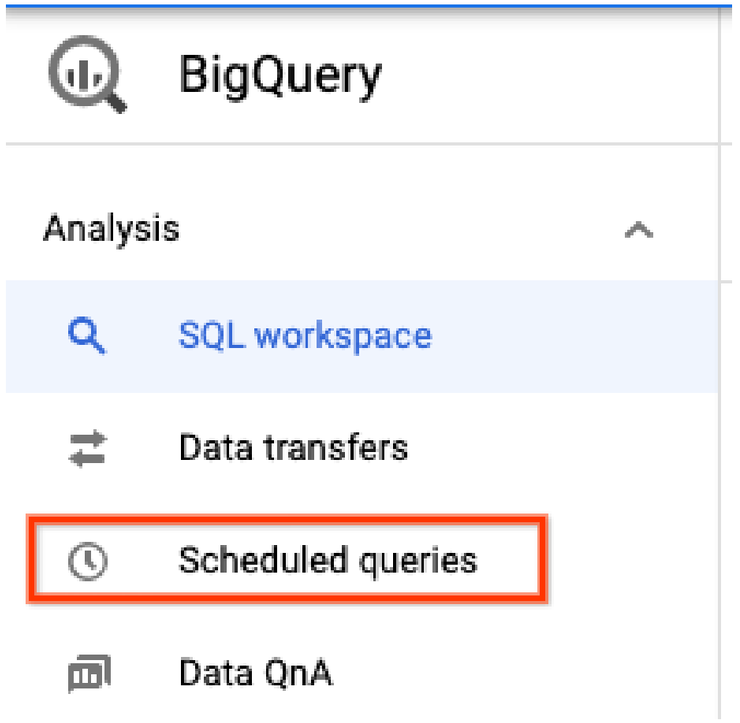
- Online transactional processing (OTAP)
- Compute and storage are linked
- Non-distributed

BigQuery

- Online analytical processing (OLAP)
- Compute and storage are separate
- Distributed query engine

How is BigQuery used?

1. Reports that run at a specific time period or day (daily ecommerce reports, quarterly reports)
2. Ad-hoc discovery (marketing campaign analysis)



Let's practice!

INTRODUCTION TO BIGQUERY

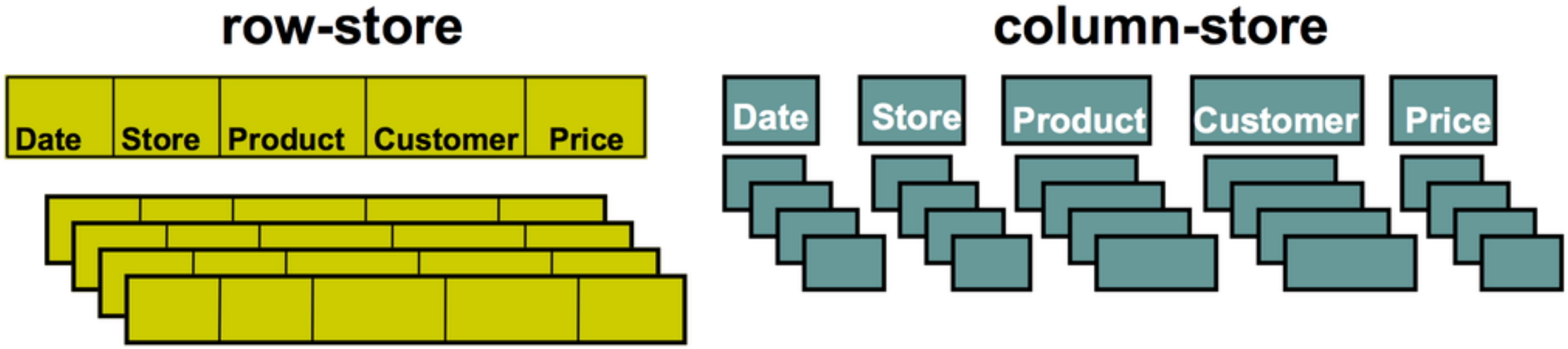
BigQuery Architecture

INTRODUCTION TO BIGQUERY

Matt Forrest
CTO and author

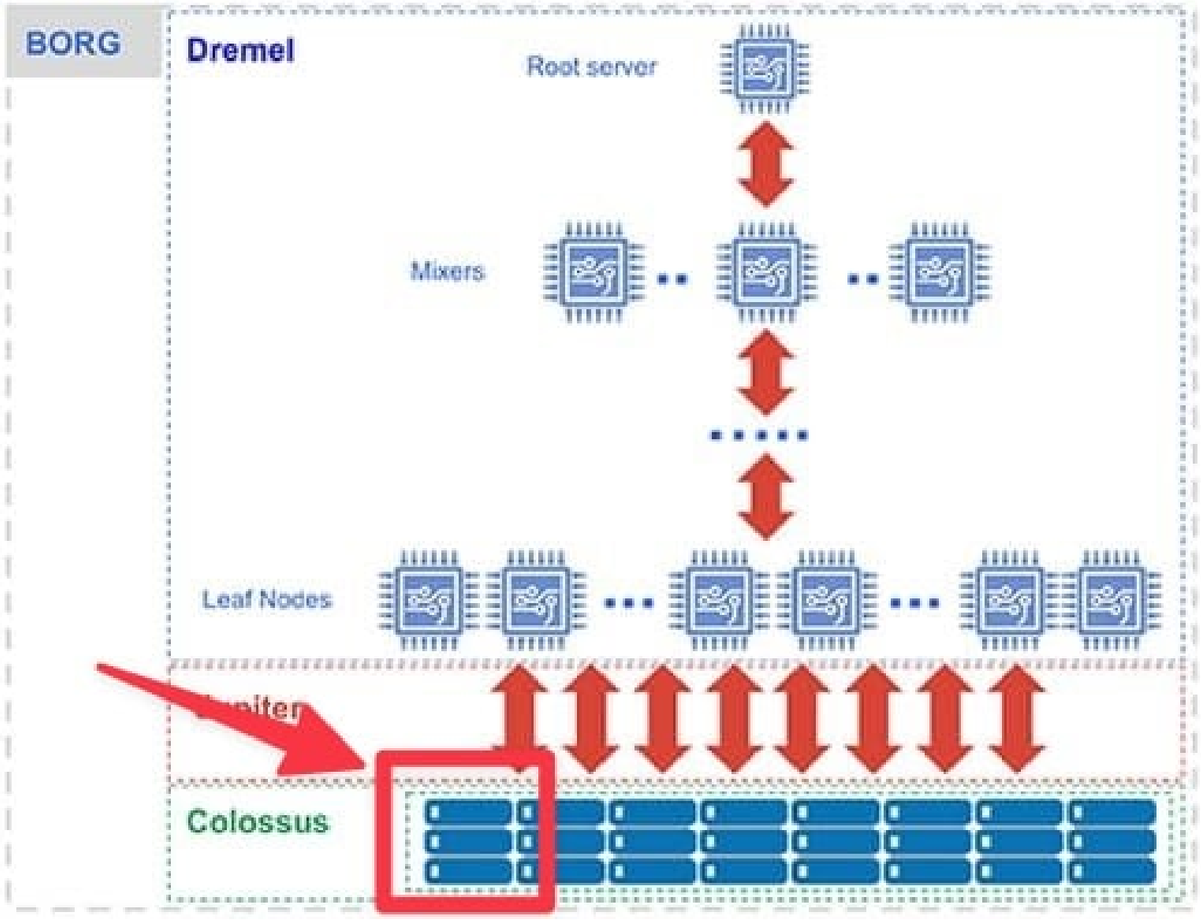
datacamp

Columnar data

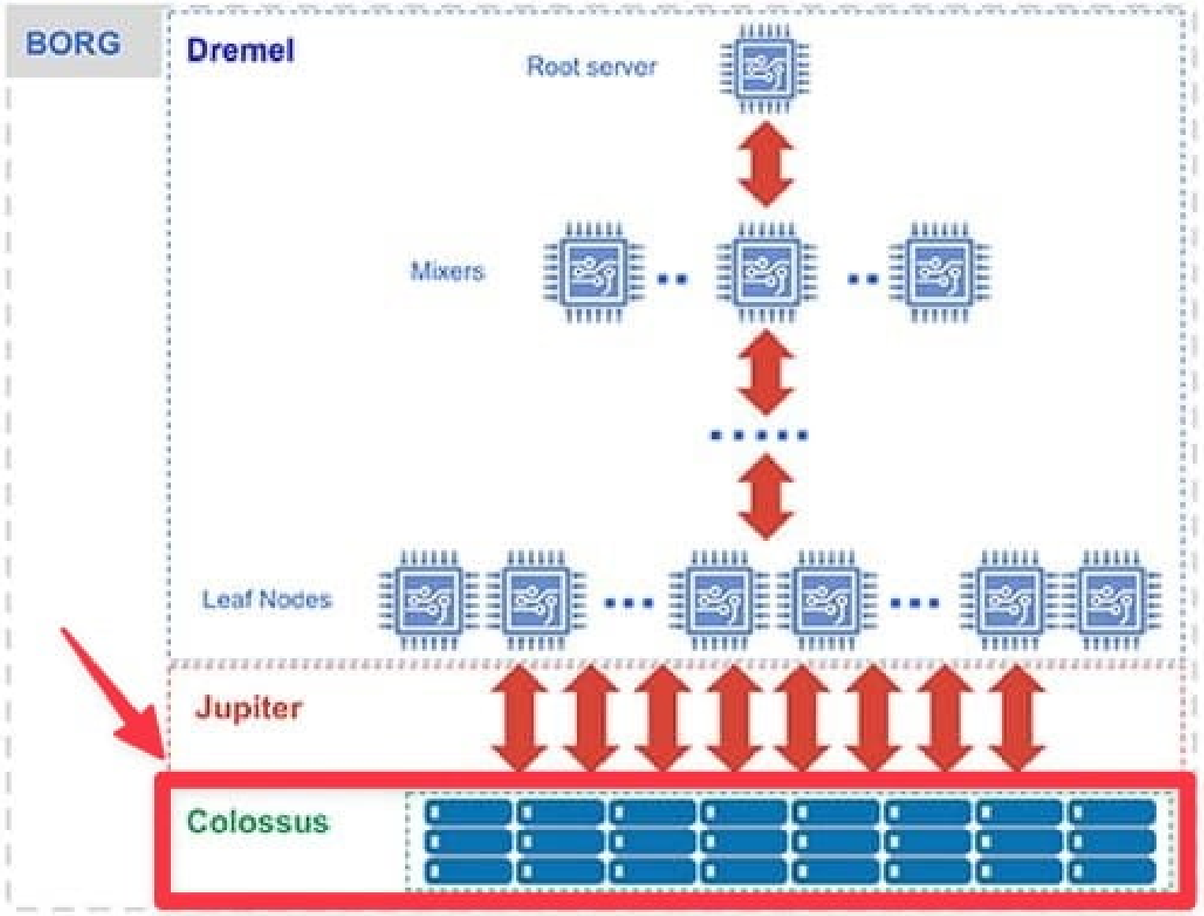


[1] http://www.primarydigit.com/blog/-a-brief-introduction-to-column-oriented-databases

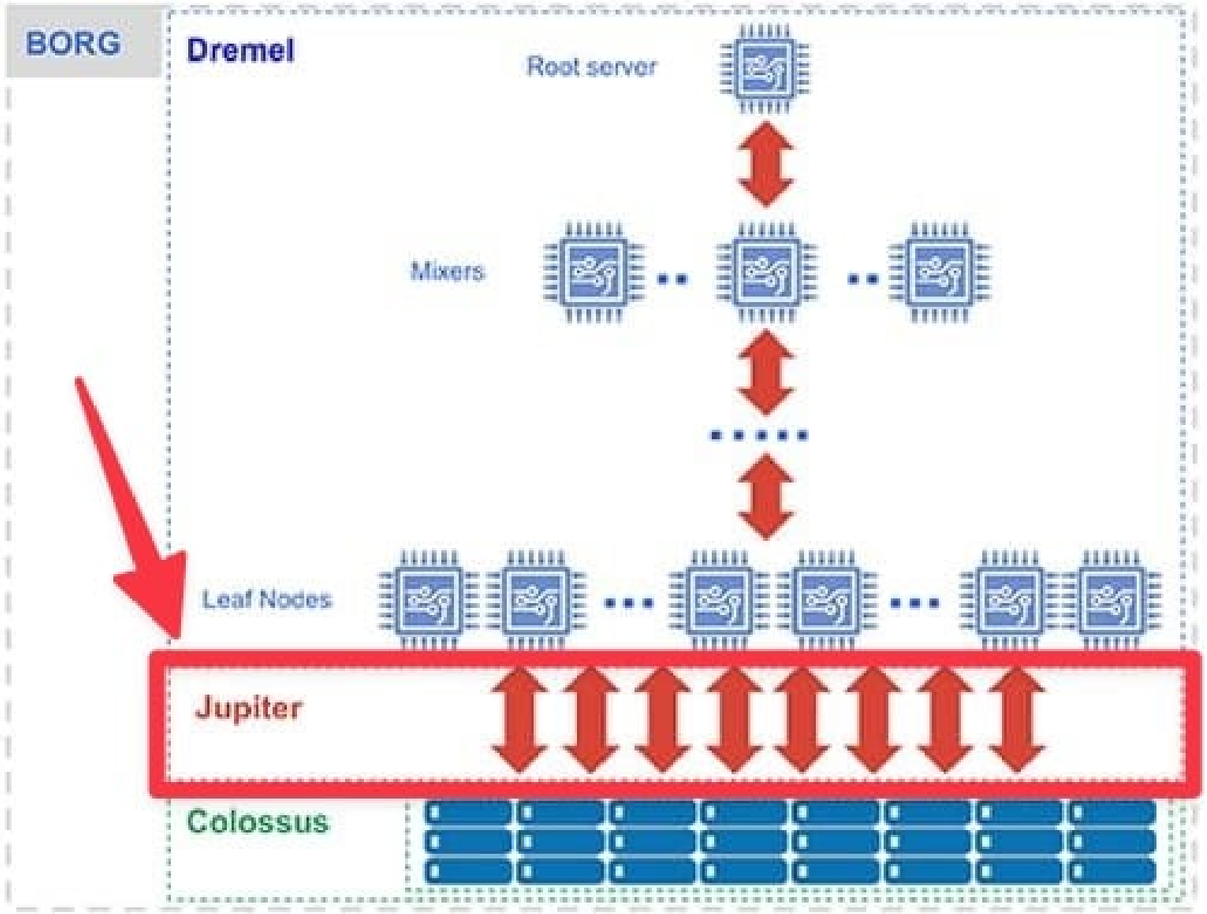
Capacitor



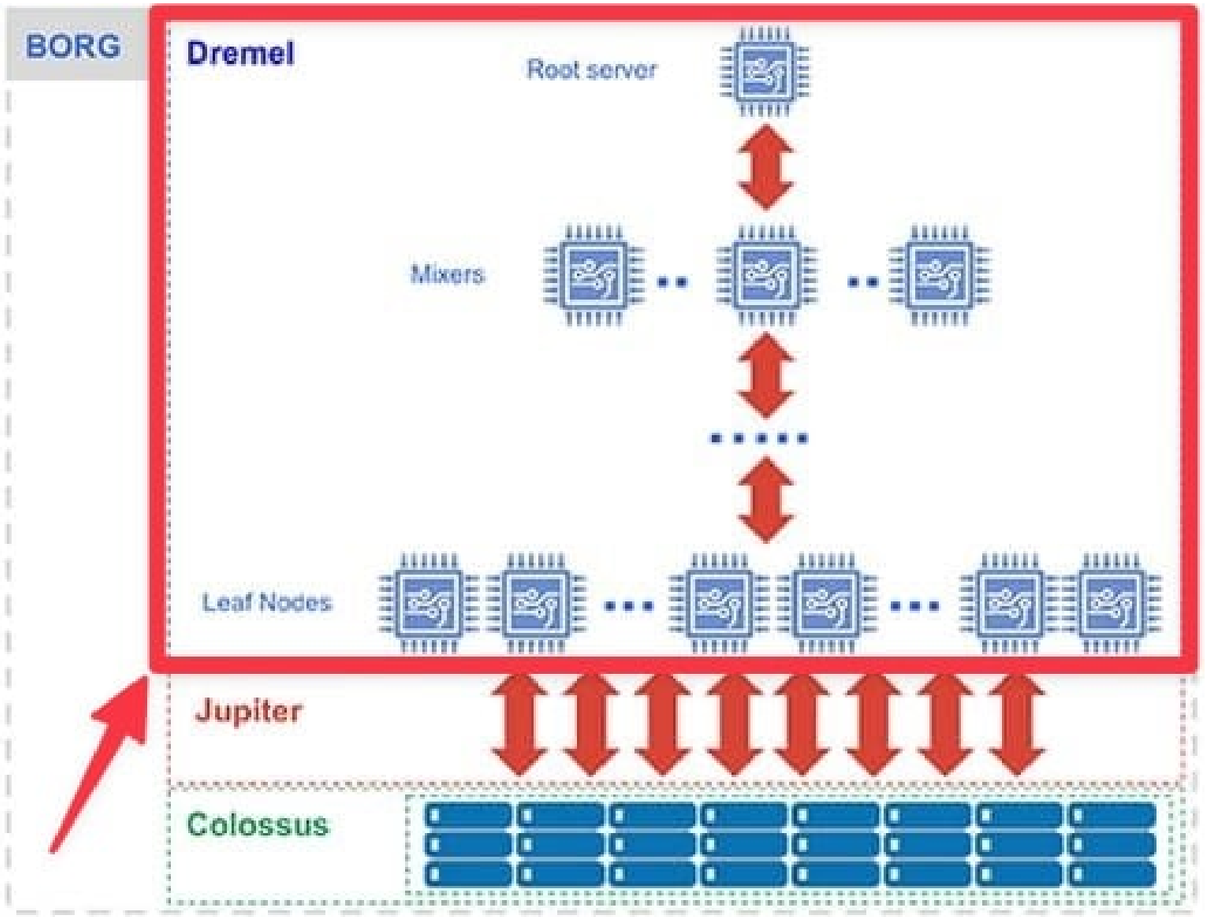
Colossus


BORG
Dremel
Root server
Mixers
Leaf Nodes
Jupiter
Colossus

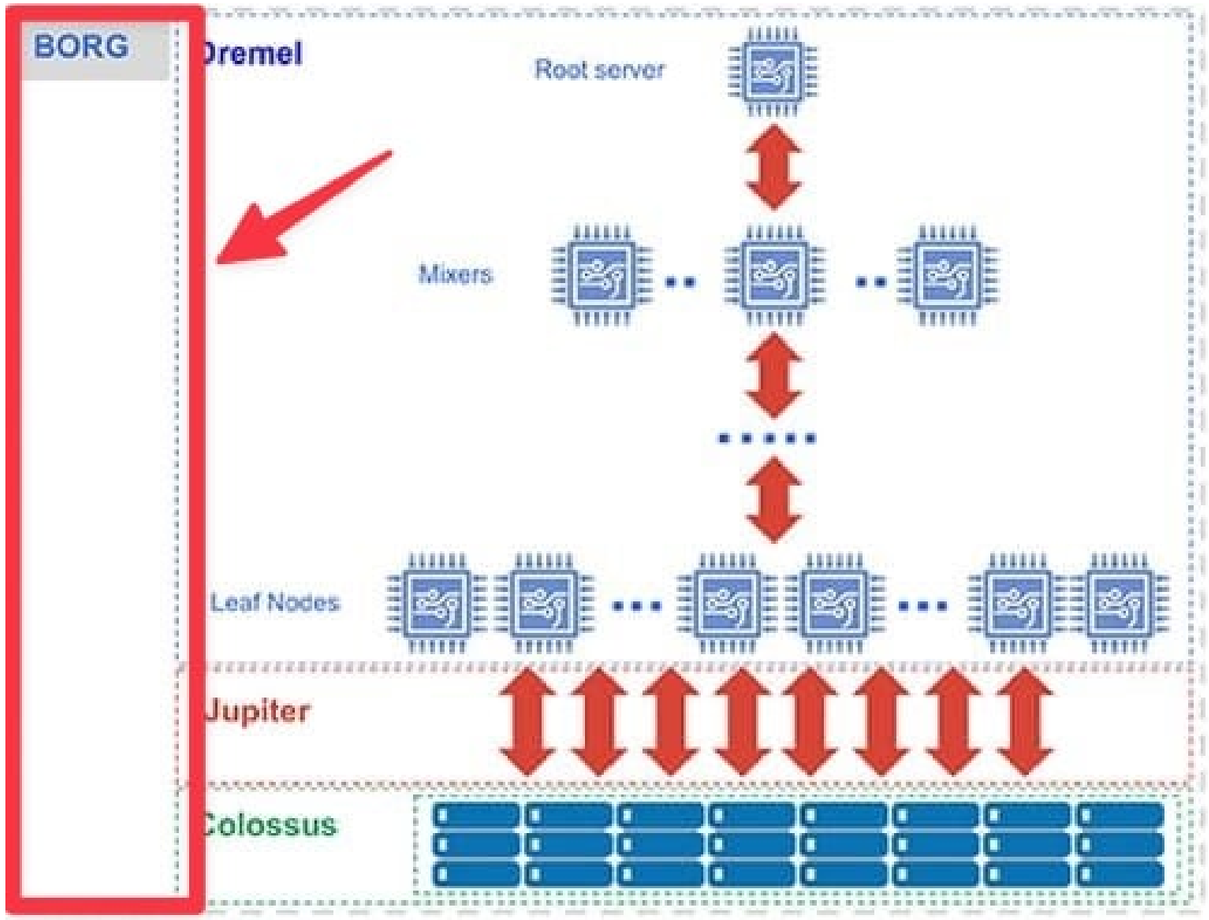
Jupiter


BORG
Dremel
Root server
Mixers
Leaf Nodes
Jupiter
Colossus

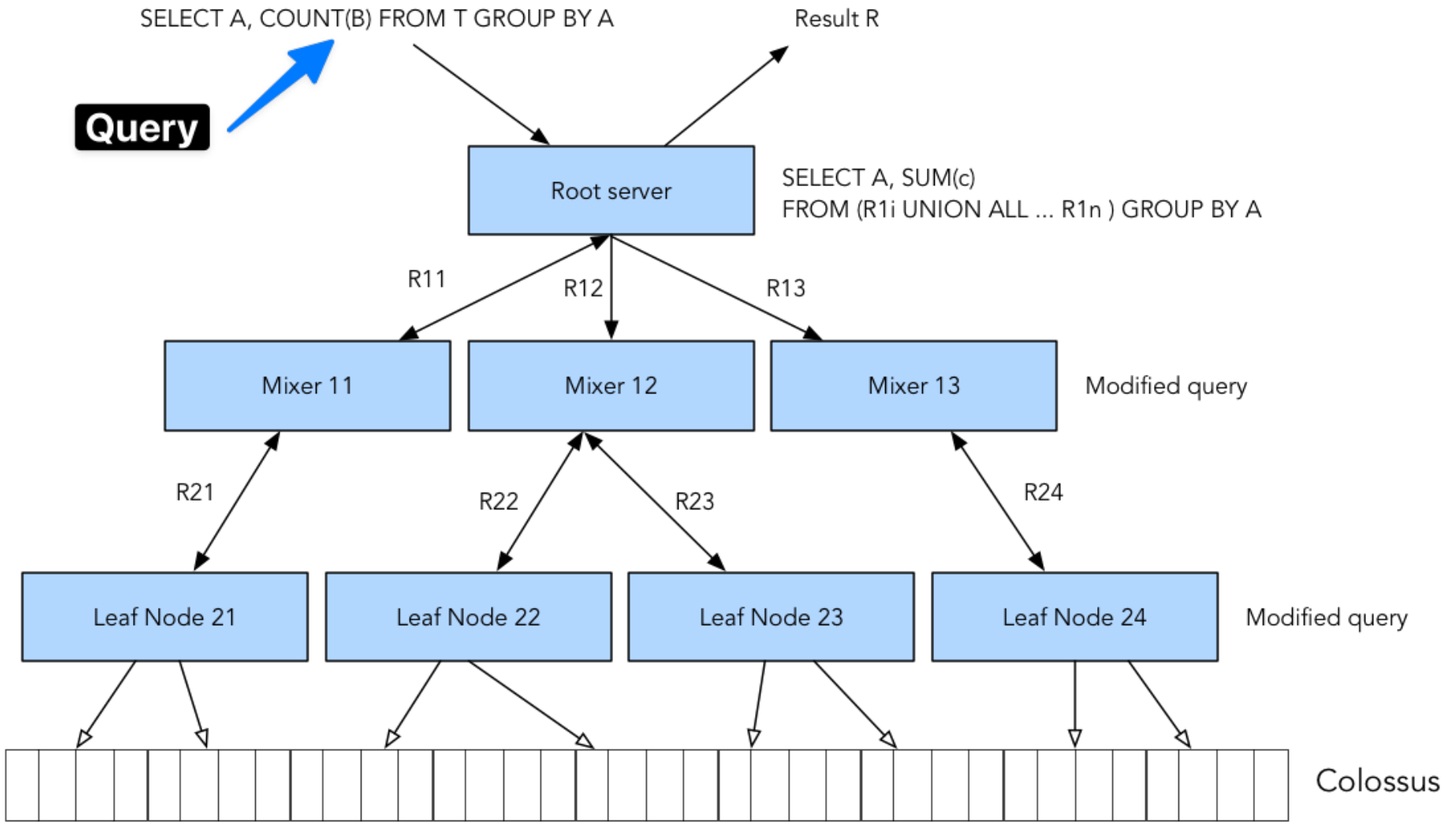
Dremel


BORG
Dremel
Root server
Mixers
Leaf Nodes
Jupiter
Colossus

Borg


BORG
Dremel
Root server
Mixers
Leaf Nodes
Jupiter
Colossus

Mixers, leaves, execution trees, and slots



[1] https://panoply.io/data-warehouse-guide/bigquery-architecture/

Categorized architecture

Storage

- Capacitor
- Colossus

Compute

- Jupiter
- Borg

Query Execution (Dremel)

- Mixers
- Leaf nodes
- Execution tree
- Slots

Let's practice!

INTRODUCTION TO BIGQUERY

BigQuery data organization

INTRODUCTION TO BIGQUERY

Matt Forrest
Field CTO

datacamp

BigQuery table names

```
SELECT
    *
FROM
    `project.dataset.table`
```

Projects

- Projects handle permissions, users, and controls
- BigQuery can have multiple projects

```
SELECT
    *
FROM
    `project.dataset.table`
```

Datasets

- Datasets are unique to a project
- Act like folders that contain individual tables
- Datasets have their own permissions

```
SELECT
    *
FROM
    `project.dataset.table`
```

Tables

- Tables are where your data resides

```
SELECT
    *
FROM
    `project.dataset.table`
```

Regions


Oregon
Salt Lake City
Las Vegas
Los Angeles
Dallas
Iowa*
Toronto
Montréal
Columbus
N Virginia
S Carolina
Querétaro
São Paulo
Santiago
Finland
Netherlands
London
Berlin
Belgium
Warsaw
Paris
Frankfurt
Turin
Zurich
Milan
Madrid
Tel Aviv
Dammam
Delhi
Doha
Mumbai
Seoul
Tokyo
Osaka
Taiwan
Hong Kong
Singapore
Jakarta
Johannesburg
Sydney
Melbourne
Current region with 3 zones
Future region with 3 zones

Working with data between regions

- Datasets cannot change regions once created
- There are methods to move/replicate data between regions
- You cannot query data in two different regions

Let's practice!

INTRODUCTION TO BIGQUERY