1

Agenda

- Analyst Perspectives
- Basic Functionality
- Performance Features (Hands-On)
- Usability Features (Demo)
- Discussion
  - Pricing
  - Takeaways

MCKNIGHT CONSULTING GROUP

2

Basic Functionality

MCKNIGHT CONSULTING GROUP

3

Databases and DDL

- Project ID
  - Dataset ID
    - Table
- Supports CREATE OR REPLACE
- Supports multi-line strings*

Data Type	BigQuery
INT, INTEGER	INT64
BIGINT	INT64
DECIMAL(38,9) NUMBER(76,38)	NUMERIC BIGNUMERIC
FLOAT, DOUBLE, REAL	FLOAT64
CHAR(x) VARCHAR(x)	STRING*
DATE	DATE
DATETIME	DATETIME
TIMESTAMP	TIMESTAMP
VARIANT	STRUCT or JSON only

MCKNIGHT CONSULTING GROUP

4

Other DDL Features

- `` `project_id.dataset_id.table_name` ``
  - Dataset ID is **required** with GUI interactive queries (No USE equivalent)
  - Dataset ID not required with CLI and JDBC (TdBench) as long as Default Dataset is specified in the config.
- Case **sensitive** identifiers
- Constraints are **not** supported, except for NOT NULL, which is enforced but can be overridden when loading with an ALLOW_FIELD_RELAXATION schema modification.
- Rounding mode for NUMERIC and BIGNUMERIC
- Supports the following CTAS variants:
  - CREATE TABLE AS SELECT
  - CREATE TABLE LIKE (new table, same schema, no data)
  - CREATE TABLE COPY (same schema, same data)
  - CREATE TABLE CLONE (new clone table, same schema, lightweight, diff-only table)

MCKNIGHT CONSULTING GROUP

5

Loading Data

1. **Batch load** from GCS (Console & CLI only*)

2. **Scheduled** via BigQuery Data Transfer Service

3. **Stream** into MVs using Datastream & Dataflow

4. Via third-party ETL

*no SQL support (e.g., COPY)

MCKNIGHT CONSULTING GROUP

6

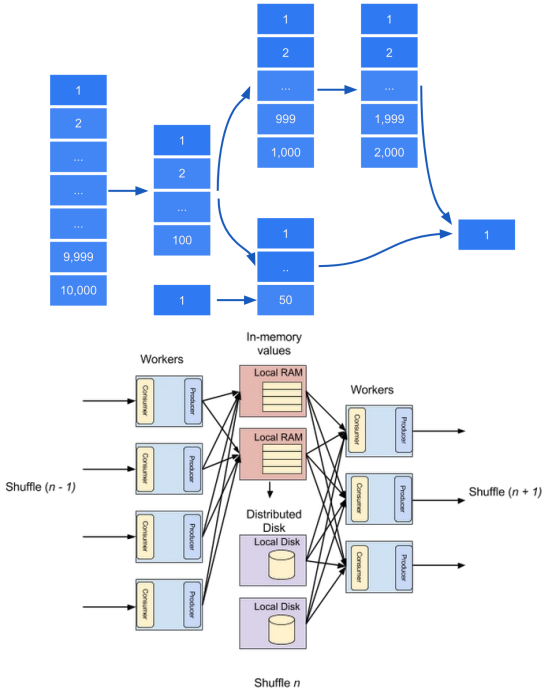
Performance Features

MCKNIGHT CONSULTING GROUP

7

BQ Architecture & Slots

- Compute = Dremel
- Storage = Colossus
- Network = Jupiter
- 1 slot = 1 virtual CPU
- Distributed shuffle architecture
  - Similar to Map-Reduce shuffle, except data is repartitioned in-memory and shuffled rows are available upon creation



8

Slots versus vCPU

BigQuery		Snowflake		Redshift	
Size	Per Hour	Size (vCPU)	Per Hour	Size (vCPU)	Per Hour*
100 slots	$4	X-Small (8)	$3	1 ra3.4xlarge (12)	$3
300 slots	$12	Medium (32)	$12	4 ra3.4xlarge (48)	$13
600 slots	$24	Large (64)	$24	8 ra3.4xlarge (96)	$26
1,200 slots	$48	X-Large (128)	$48	14 ra3.4xlarge (168)	$46
2,400 slots	$96	2X-Large (256)	$96	29 ra3.16xlarge (348)	$95

9 3/8th slots : 1 vCPU ratio **~7 slots : 1 vCPU ratio**

MCKNIGHT CONSULTING GROUP

*rounded off

9

Clustering and Partitioning

Clustering

Uses:

- Commonly filter on particular columns
- Commonly filter on columns that have many distinct values
- Tables larger than 1GB

Limits:

- Up to 4 columns
- Uses only the first 1,024 characters of a STRING to cluster
- If you alter an existing non-clustered table to be clustered, the existing data is not clustered

Partitioning

Uses:

- Can partition by:
  - Time-unit columns
  - Ingestion time
  - Integer range
- Need partition-level management (e.g., partition expiration, partition deletion, etc.)

Limits:

- Single column
- Cluster instead if:
  - Partitions are < 1GB
  - >4,000 partitions per table

NOTE: Can be combined. Data is partitioned first, then the data is each partition is clustered.

MCKNIGHT CONSULTING GROUP

10

Transparent Materialized Views

- Can be partitioned and clustered
- Preview feature: max_staleness
- Compute and storage costs to maintain
  - Uses the slots you've paid for
  - Pricing gotcha: If no slots are reserved, refreshes are billed by bytes processed

Not supported:

- Computing or filtering based on an aggregated value
- Left/right/full outer joins
- Only 20 MVs per table reference
- Cannot be nested
- Query rewrite for max staleness

MCKNIGHT CONSULTING GROUP

11

BI Engine

- In-memory analysis service
- Caches frequently queried data
- Separate paid hourly SKU ($0.0416 per GB per hour)
- Currently a max of 250GB per project
  - Was 100GB until 2022
  - Max 100GB when used with Looker Studio
- Can now specify preferred tables
  - Also new in 2022
- Improved support for different query types (e.g., JOINs)

MCKNIGHT CONSULTING GROUP

12

When to Use BI Engine

When to use

- Use BI tools
- Have certain tables that are queried more frequently
- Have pre-joined or pre-aggregated data

Limitations

- BI Engine modes:
  - Disabled: Not able to use BIE, reason given
  - Partial: Only part of the query could use BIE
  - Full: All stages used BIE
- UDFs, external tables, and STRUCT/ARRAYs not supported
- Other undocumented limitations
  - e.g., can only JOIN 5 tables

MCKNIGHT CONSULTING GROUP

13

Usability Features

MCKNIGHT CONSULTING GROUP

14

BigQuery Omni – External Tables

- Run BQ analytics on data stored in Amazon S3 or Azure Blob Storage
- Uses BigLake tables
- Extends architecture by running BigQuery compute in the other cloud
  - No data egress charges (data doesn't move)
- Limitations:
  - 20GB per query result
  - 1TB per day
  - Only in AWS US East 1 and Az East US 2
- Workflow:
  - Create an AWS IAM policy
  - Create an AWS IAM role for BQ
  - Add a trust policy
  - Create schema with location specified as other cloud region
  - Create BigLake table
  - Query as normal
- Pricing:
  - FREE until March 31 (trial period)
  - Pay on-demand rate after trial ($5 per TB)

MCKNIGHT CONSULTING GROUP

15

Time Travel

- Window set by dataset
- SQL Query via SYSTEM_TIME

  ```
  SELECT * FROM `mydataset.table1` FOR SYSTEM_TIME AS OF
  TIMESTAMP_SUB(CURRENT_TIMESTAMP(), INTERVAL 1 HOUR);
  ```

- OR restore via CLI using table decorator

  ```
  bq cp mydataset.table1@-3600000 mydataset.table1_restored
  ```

- Free if you use Logical Storage pricing
- Pay for it if you use Physical Storage pricing

MCKNIGHT CONSULTING GROUP

16

Migration Service – SQL Translation

- Free-to-use
- Dialects:
  - Azure Synapse*
  - HiveQL*
  - MySQL*
  - Netezza*
  - Oracle*
  - PostgreSQL*
  - Presto*
  - Redshift
  - Snowflake*
  - SparkSQL*
  - SQL Server*
  - Teradata
  - Vertica*

*Still in Preview

MCKNIGHT CONSULTING GROUP

17

Looker Studio

- FREE to use (only pay usual BQ and BIE charges)
- Self-service business intelligence platform
- Separate from Looker
- “Google Docs” look-and-feel

MCKNIGHT CONSULTING GROUP

18

Colab Notebooks

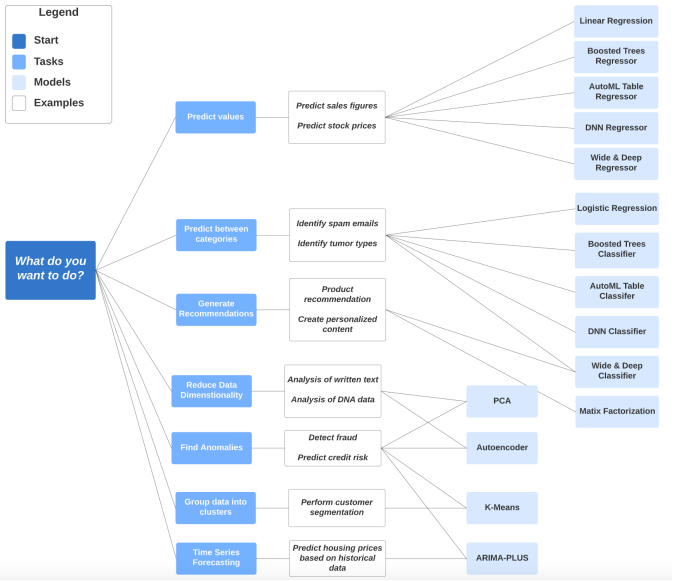
- Preview feature
- Passes BigQuery results to a notebook
- GPU-driven
- Python libraries

MCKNIGHT CONSULTING GROUP

19

BigQuery ML

- Build and evaluate models via SQL
- No Python or Java knowledge required
- Can be run from console, SQL tool, or Jupyter Notebook



20

BigQuery and Vertex AI Integrations

Import into Vertex

- Import BigQuery data into Vertex AI
- Accessing BigQuery data from Vertex AI Workbench notebooks
- Access BigQuery public datasets

Export to BQ

- Analyze test prediction data in BigQuery
- Export Vertex AI batch prediction results

MCKNIGHT CONSULTING GROUP

21

Discussion

MCKNIGHT CONSULTING GROUP

22

Pricing

Compute

	BigQuery	Omni
On-demand	$5 per TB	$5 per TB
Flex	$4.00/hr per 100 slots	$5.00/hr per 100 slots
Monthly Commit*	$2.74/hr per 100 slots	$3.42/hr per 100 slots
Annual Commit*	$2.33/hr per 100 slots	$2.91/hr per 100 slots
BI Engine	$0.0416/hr per GB	N/A

*comes with some free BI Engine

Storage[1]

	Logical[2]	Physical[3]
Active	$0.02/GB-month	$0.04/GB-month
Long-term[4]	$0.01/GB-month	$0.02/GB-month
Batch loading	FREE	
Streaming inserts	$0.01 per 200MB	
Storage API	$0.025 per 1GB	

[1] You get to choose logical or physical billing
[2] Logical = Uncompressed size (Time travel free)
[3] Physical = Compressed size + Time travel
[4] Table not modified in 90 days

MCKNIGHT CONSULTING GROUP

23