

Agenda

1. 1/31
 - a. 9-12: Setup & Snowflake
 - b. 1-4: Redshift
 - c. 4-5: Overflow/Discussion
2. 2/1
 - a. 9-12: BigQuery
 - b. 1-4: Synapse
 - c. 4-5: Overflow/Discussion
3. 2/2
 - a. 9-12: Databricks
 - b. 1-4: Overflow/Discussion

Setup

1. Consoles
 - a. Snowflake
 - b. AWS
 - c. Google Cloud
 - d. Azure
 - e. Databricks
2. CLI
 - a. Linux VM
 - b. TdBench

Snowflake

1. Analyst Perspectives
2. Basic Functionality
 - a. Snowsight vs. Classic UI
 - b. User roles
 - c. Warehouses
 - i. Create
 - ii. Auto start/stop
 - d. Databases and DDL
 - i. Cloning
 - ii. GET_DDL
 - e. Load data
 - i. COPY from S3
 - ii. Stages

- f. Running queries
 - i. Query history
 - ii. Query profiles and monitoring
- 3. Performance Features
 - a. Table design
 - i. Clustering
 - ii. Constraints
 - b. Cached results
 - c. Multi-clusters
 - d. Transparent materialized views
 - e. Search Optimization Service
 - f. Query Acceleration Service
- 4. Usability Features
 - a. External tables
 - b. Dynamic data masking
 - c. Time Travel
 - d. Semi-structured
 - e. Advanced loading data
 - i. Streaming ingest/micro-batching with Snowpipe
 - f. Snowsight Dashboards
 - g. Snowpark API
 - i. Python UDF
- 5. Pricing
- 6. Takeaways discussion

Redshift

- 1. Analyst Perspectives
- 2. Basic Functionality
 - a. Console
 - b. Provisioned Clusters vs. Serverless
 - c. Cluster Sizes
 - d. Redshift vs. PostgreSQL
 - e. Loading data
 - i. COPY command
 - ii. Spectrum/external tables
 - iii. Vacuum
 - f. Running queries
 - i. Query history
 - ii. Query profiles
 - iii. Cached results
- 3. Performance Features
 - a. Table design
 - i. Sort keys
 - ii. Distribution style
 - iii. Auto table optimizations

- iv. Redshift Advisor
 - b. Workload management and monitoring
 - c. Concurrency Scaling Clusters
 - d. Short Query Acceleration
 - e. Transparent materialized views
- 4. Usability Features
 - a. Redshift Spectrum
 - b. AutoMV
 - c. Dynamic data masking
 - d. Federated queries
 - i. External operational database
 - ii. Another Redshift database
 - e. Semi-structured and SUPER columns
 - f. Streaming ingest with Kinesis/Kafka
 - g. Python UDF
 - h. Redshift ML
- 5. Pricing
- 6. Takeaways discussion

BigQuery

- 1. Analyst Perspectives
- 2. Basic Functionality
 - a. Console
 - b. Projects
 - c. Datasets
 - d. DDL
 - e. Loading data
 - i. From Cloud Storage
 - ii. Console vs. CLI
 - iii. Streaming data
 - f. Capacity management (Slots)
 - g. Running queries
 - i. Query history
 - ii. Query profiles
 - h. Monitoring
- 3. Performance Features
 - a. Table design
 - i. Clustering
 - ii. Partitioning
 - b. Cached results
 - c. BI Engine
- 4. Usability Features
 - a. BigQuery Omni
 - b. Migration SQL translation
 - c. Time Travel

- d. Javascript UDF
- e. Looker Studio
- f. BigQueryML
- g. Colab Notebooks
- h. Vertex AI and BQ
- 5. Pricing
- 6. Takeaways discussion

Synapse

- 1. Analyst Perspectives
- 2. Basic Functionality
 - a. Console
 - b. Resource groups
 - c. Concepts
 - i. Workspaces
 - ii. Pools
 - 1. Serverless SQL
 - 2. Dedicated SQL
 - 3. Spark
 - 4. Data Explorer (Kusto)
 - d. Serverless Data Exploration
 - e. DDL and table design
 - i. Clustered columnstore index
 - ii. Segment elimination
 - f. Data loading
 - i. Blob storage via external tables
 - ii. Copy activity via Data Factory
 - g. Running queries
 - h. Query history
 - i. Query profiles
- 3. Performance Features
 - a. Cached results
 - b. Workload management groups
 - c. Monitoring
- 4. Usability Features
 - a. Dynamic data masking
 - b. External data sources
 - c. Synapse Link
 - d. SynapseML
 - e. Synapse Charts
- 5. Pricing
 - a. DWUs
- 6. Takeaways discussion

Databricks

1. Analyst Perspectives
2. Basic Functionality
 - a. Console
 - b. Ecosystem
 - c. DDL and Delta Lake basics
 - i. Unity Catalog
 - ii. Metastore-catalog-schema
 1. External tables
 2. Managed tables
 3. Views
 - d. Loading data
 - i. COPY INTO
 - ii. Conversion to Delta Lake
 - e. Databricks SQL Warehouses
 - i. Classic
 - ii. Pro
 - iii. Serverless
 - f. Running queries
 - i. Filters and Parameters
 - ii. Query history
 - iii. Query profiles
3. Performance Features
 - a. Optimizations
 - b. Z-order
 - c. Partitioning
 - d. Cached results
4. Usability Features
 - a. Merge
 - b. Selective overwrite
 - c. Visualizations
 - d. Dashboards
 - e. Alerts
 - f. History and time travel
5. Pricing
6. Takeaways discussion