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

Classic UI vs. Snowsight

	Classic	Snowsight
Databases	X	X
Warehouses	X	X
Worksheets	X	X
History/Activity	X	X
Admin	X	X
Dashboards		X

- Accounts created after Oct. 2022 can only access Snowsight
- Prior accounts can toggle back and forth

MCKNIGHT CONSULTING GROUP

4

User Roles

	Organization	Usage/Bill	Security	Users	Create	Data
ORGADMIN	X	X				
ACCOUNTADMIN		X	X	X	X	All
SECURITYADMIN			X	X		
USERADMIN				X		
SYSADMIN					X	All
PUBLIC						Public*
Custom						Limited

*Objects created in the PUBLIC role are available Account wide

MCKNIGHT CONSULTING GROUP

5

Warehouses

- 10 sizes
- New Snowpark-optimized with 16x memory than Standard (open preview)
- MCG analyst intelligence suggests 8 vCPU 16 GB RAM per Standard node
  - e.g. AWS c6g.2xlarge $0.272
- MCG *infers* 32 vCPU 256 GB RAM for Snowpark nodes
  - e.g. AWS r6g.8xlarge $1.6128

Size	Standard	Snowpark
XS	1	N/A
S	2	N/A
M	4	6
L	8	12
XL	16	24
2XL	32	48
3XL	64	96
4XL	128	192
5XL	256	384
6XL	512	768

Multiply by $ per hour for price

MCKNIGHT CONSULTING GROUP

6

Databases and DDL

- Database
  - Schema (PUBLIC default)
    - Table
- To preserve right-padded spacing, you must load it this way:
  - `"CLERK012 "`
- Supports CREATE OR REPLACE
- CLONE

Data Type	Changes To
INT, INTEGER	NUMBER(38,0)
BIGINT	NUMBER(38,0)
DECIMAL(15,2)	NUMBER(15,2)
FLOAT, DOUBLE, REAL	FLOAT
CHAR(10)	VARCHAR(10)*
CHAR	VARCHAR(1)
VARCHAR	VARCHAR(16777216)
DATE	DATE
DATETIME	TIMESTAMP_NTZ

MCKNIGHT CONSULTING GROUP

7

Other DDL Features

- <database_name>.<schema_name>.<object_name>
- Case insensitive identifiers
- Constraints are supported, but not enforced, except for NOT NULL, which is always enforced.
- SELECT GET_DDL('<object_type>' , '[<namespace>.]<object_name>');

MCKNIGHT CONSULTING GROUP

8

Loading Data

1. **COPY** from Cloud Object Storage (e.g., S3)
2. **COPY** from Snowflake staging area (file PUTs)
3. With **Snowpipe** (streaming micro-batches)
4. Via third-party ETL

MCKNIGHT CONSULTING GROUP

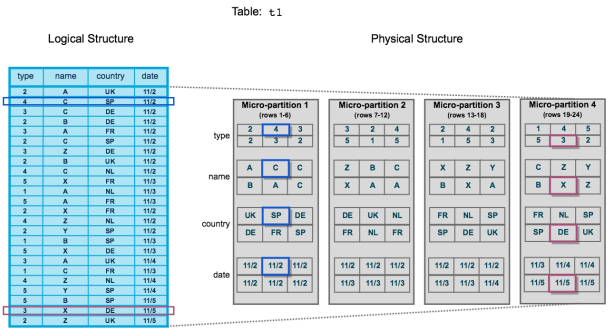
9

Performance Features

MCKNIGHT CONSULTING GROUP

10

Micro-partitions



- Created during loading
- Prunes partitions not needed for a query
- Prunes by column within remaining partition

MCKNIGHT CONSULTING GROUP

11

Clustering Keys

Clustering is best for tables that meet all of the following criteria:

- Large number of micro-partitions, i.e., > 1TB of data
- User queries are often selective
- User queries often sort the data
- A high percentage of user queries can benefit from the same clustering key(s)

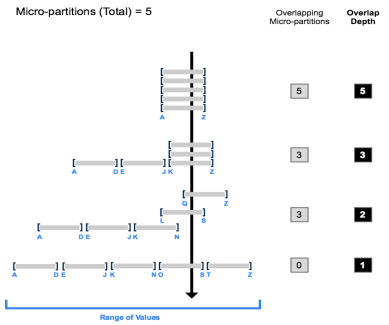
Selected cluster keys by using the following guidelines:

- Columns most actively used in selective filters
- Columns frequently used in join predicates
- Large enough cardinality to enable effective pruning on the table
- Small enough cardinality to effectively group rows in the same micro-partitions
- Clustering on more than 3-4 columns tends to increase cost over benefits

MCKNIGHT CONSULTING GROUP

12

Clustering Depth



- Before clustering, the range of values in all the micro-partitions overlap
- As the number of overlapping micro-partitions decreases, the overlap depth decreases
- The lower the better

MCKNIGHT CONSULTING GROUP

13

Multi-Clusters

- Only available on Enterprise ($3.00/hour) or higher accounts
- Up to 10 Multi-clusters supported
- Three modes (where X=maximum number of Multi-clusters allowed):
  - Multi-clusters OFF (min: 1, max: 1)
  - Auto-scale Multi-clusters (min: 1, max: X)
  - Maximized Multi-clusters (min: X, max: X)
- Two scaling policies
  - Standard (default; prevents queuing)
  - Economy (favors fully-loaded clusters over spawning new ones; allows some queueing)

MCKNIGHT CONSULTING GROUP

14

Transparent Materialized Views

- Requires Enterprise ($3.00/hour) or better account
- Can be clustered (i.e., add clustering keys)
- Do not support CLONE* or Time Travel
- Costs compute and storage $ to maintain
  - Updates can be suspended
  - However, not supported by Resource Monitors
- Runs updates on a separate cluster
  - MATERIALIZED_VIEW_MAINTENANCE
  - Theoretically will not impact workloads on other warehouses (compare to BigQuery)
- Suspended MVs or changes to base table may cause compilation errors

MCKNIGHT CONSULTING GROUP

*However, if you CLONE a schema or database that contains an MV, it will be created as well.

15

Search Optimization Service

- Still in Open Preview
- Requires Enterprise ($3.00/hour) or better account
- Purpose: To improve the performance of certain types of lookup and analytical queries that use an extensive set of predicates for filtering
- Uses:
  - Critical dashboards with highly selective filters
  - Exploring large data volumes and looking for specific subsets of data
  - Retrieving a small set of results based on an extensive set of filtering predicates

MCKNIGHT CONSULTING GROUP

16

When to use Search Optimization Service

- Tactical queries that run a few seconds or longer
- At least one of the columns accessed through the filter operation has at least 100,000 to 200,000 distinct values
- Equality or IN predicates
- Substrings and Regex expressions:
  - LIKE, LIKE ANY, LIKE ALL, ILIKE, ILIKE ANY, CONTAINS, STARTSWITH, ENDSWITH, SPLIT_PART, RLIKE, REGEXP, REGEXP_LIKE
- VARIANT, OBJECT, and ARRAY columns
- Geospatial functions

MCKNIGHT CONSULTING GROUP

17

Individual Query Performance Feature Comparison

Improves	Clustering	Materialized Views	Search Opt. Service
Equality searches	X	X	X
Range searches	X	X	X
Sort operations		X	
Substring and Regex			X
VARIANT searches			X
Geospatial			X
Extra Costs			
Compute	X	X	X
Storage		X	X

MCKNIGHT CONSULTING GROUP

18

Query Acceleration Service

- Still in Open Preview
- Requires Enterprise ($3.00/hour) or better account
- Purpose: To improve overall warehouse performance by reducing the impact of outlier queries by offloading portions of the query processing work to shared compute resources that are provided by the service
- Uses:
  - Ad hoc analytics
  - Workloads with unpredictable data volume per query
  - Queries with large scans and selective filters
- Does not accelerate queries on tables that have search optimization enabled

MCKNIGHT CONSULTING GROUP

19

Usability Features

MCKNIGHT CONSULTING GROUP

20

External Tables

- Schema on read
  - If an error occurs, it skips to the next file, but still returns rows found in the current file up until the error occurred.
- Recommended 16MB – 256MB file sizes (256-512MB for Parquet)
- Delta Lake support
- Workflow:
  - CREATE STAGE > CREATE EXTERNAL TABLE > Create cloud object storage event notification > Automatic refresh

MCKNIGHT CONSULTING GROUP

21

Dynamic Data Masking

- Dynamically rewrites the query applying the masking policy SQL expression to the column
- The column rewrite occurs at every place where the column specified in the masking policy appears in the query
- Workflow:
  - Grant masking policy privileges to custom role
  - Grant custom role to user
  - Create a masking policy (uses CASE WHEN THEN ELSE END logic)
  - Apply the policy to table or view column

MCKNIGHT CONSULTING GROUP

22

Time Travel and Fail Safe

DATA_RETENTION_TIME_IN_DAYS = 1 (default)
Can be 0 – 1 for Standard accounts
Can be 0 – 90 for Enterprise accounts and above

- Query data in the past that has since been updated or deleted.
- Create clones of entire tables, schemas, and databases at or before specific points in the past.
- Restore tables, schemas, and databases that have been dropped
- Storage costs:
  - Calculated as a percentage of the table that changed
  - Full copies of tables are only maintained when tables are dropped or truncated

23

Semi-Structured Data

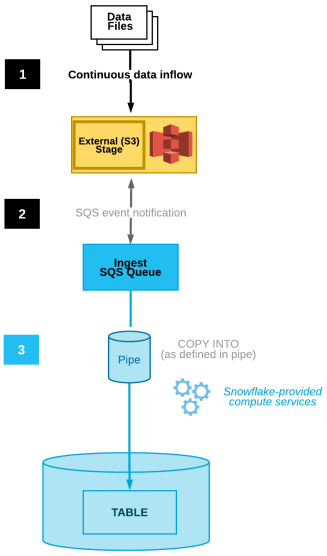
- Supports JSON, Avro, ORC, and Parquet
- Uses VARIANT data type
- Query using:
  - Dot notation: <column>:<level1_element>.<level2_element>::data_type
  - Bracket notation: <column>['<level1_element>']['<level2_element>']::data_type
  - GET_PATH

MCKNIGHT CONSULTING GROUP

24

Snowpipe

- Uses separate database objects called Pipes
- Uses separate compute
- Two methods:
  - Automatic – using cloud messaging services (SQS, Event Grid, Pub/Sub)
  - Manual – calling Snowflake REST APIs
- Supports transformations



25

Snowsight Dashboards

- Snowsight introduced a new feature:
  - BI tool inside the Snowflake GUI console
- Works like most other BI tools
  - Dashboards > Tiles > Queries
- Big limitation: Sharing
  - Can only share with fellow Snowflake users inside the same account with the same privileges as you or higher

MCKNIGHT CONSULTING GROUP

26

Snowpark API

- Supports Python, Scala, and Java
- Core abstraction is the DataFrame
- Does not require a separate compute (e.g. Spark cluster)
  - Everything is done inside Snowflake
- Can use a newly-introduced (still in Preview) cluster type built for ML called Snowpark-optimized warehouse
  - Use a different pricing structure

MCKNIGHT CONSULTING GROUP

27

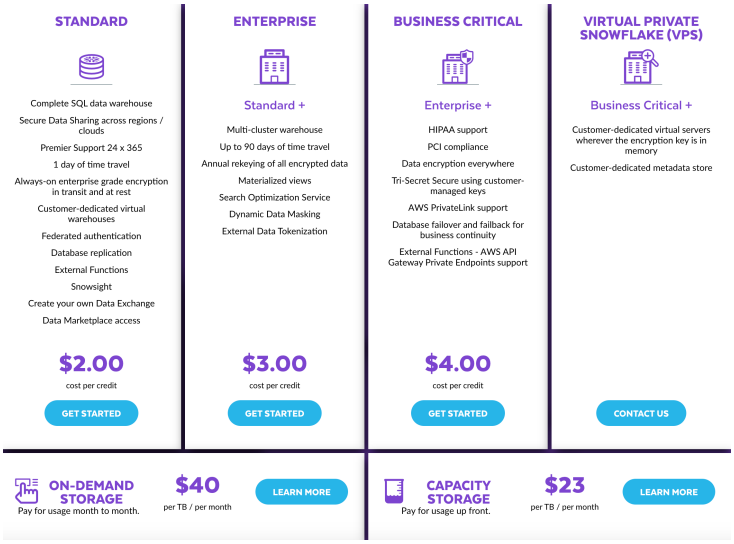
Discussion

MCKNIGHT CONSULTING GROUP

28

Pricing

- Gotchas:
  - Concurrency and price-per-performance
  - Effective Warehouses (Multi-clusters)
  - Add-on compute:
    - Automatic Clustering
    - Materialized View Refreshes
    - Search Optimization
    - Query Acceleration
  - Time travel storage
- Discounts



29

Warehouses (review)

- 10 sizes
- New Snowpark-optimized with 16x memory than Standard (open preview)
- MCG analyst intelligence suggests 8 vCPU 16 GB RAM per Standard node
  - e.g. AWS c6g.2xlarge $0.272
- MCG *infers* 32 vCPU 256 GB RAM for Snowpark nodes
  - e.g. AWS r6g.8xlarge $1.6128

Size	Standard	Snowpark
XS	1	N/A
S	2	N/A
M	4	6
L	8	12
XL	16	24
2XL	32	48
3XL	64	96
4XL	128	192
5XL	256	384
6XL	512	768

Multiply by $ per hour for price

MCKNIGHT CONSULTING GROUP

30