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

Concepts

Azure Synapse Analytics consists of:

- Resource Group
- Workspace
  - "Securable collaboration boundary"
  - Data Lake Storage Gen2 and a file system
- Linked Service
  - Blobs, ADLS Gen2, Cosmos
- Dedicated SQL Pools (DWUs)
  - Formerly SQL DW
- Serverless SQL Pools
  - Data exploration
  - On-demand, no provisioning
- Spark Pools
  - Scala, PySpark, C#, and SparkSQL notebooks
- SynapseML
  - previously known as MMLSpark
- Data Explorer
  - Also known as ADX, formerly known as Kusto
  - Uses a different query syntax KQL
- Pipelines
  - Formerly known as Azure Data Factory
  - Data integration
- Private Endpoints
  - Enterprise-grade security

MCKNIGHT CONSULTING GROUP

4

Serverless Data Exploration

OPENROWSET construct

- Source
  - BULK – files
  - DATA_SOURCE – authenticated database
- FORMAT
  - CSV
    - Parser v1 - feature rich
    - Parser v2 - faster
  - Parquet
  - Delta Lake (preview)
- WITH
  - Define schema
  - Can be omitted with CSV; returns C1, C2…

Source	Prefix	Path
Azure Blob Storage	https	<storage_account>.blob.core.windows.net/path/file
Azure Blob Storage	wasbs	<container>@<storage_account>.blob.core.windows.net/path/file
Azure Data Lake Store Gen1	https	<storage_account>.azuredatalakestore.net/webhdfs/v1
Azure Data Lake Store Gen2	https	<storage_account>.dfs.core.windows.net/path/file
Azure Data Lake Store Gen2	abfss	<file_system>@<account_name>.dfs.core.windows.net/path/file

MCKNIGHT CONSULTING GROUP

5

Databases and DDL

Dedicated SQL Pool

- Table features
  - Rowstore
    - Heap or clustered index
  - Columnstore*
    - Clustered columnstore index*
    - Ordered or unordered*
  - Distribution
    - Hash, Round Robin*, or Replicate
  - Partitions

*ASA defaults

Clustered Columnstore Indexes

- ASA CCI features now available in SQL Server 2022
- Run faster when:
  - Queries have equality, inequality, or range predicates
  - Predicate columns = the ordered CCI columns
- Slower loading
- Sometimes must be rebuilt
  - Poor segment quality
  - Use higher WLM resource class (xlargerc)

MCKNIGHT CONSULTING GROUP

6

Loading Data

1. **Batch load** from Blob/ ADLS (COPY)
2. **Scheduled /GUI** via Pipelines
3. **Stream** via Event Hubs & Stream Analytics
4. Via third-party ETL

MCKNIGHT CONSULTING GROUP

7

Performance Features

MCKNIGHT CONSULTING GROUP

8

Workload Management

- Workload Group < Workload Classifier
- Manipulates maximum memory and concurrency
- Parameters:
  - MIN_PERCENTAGE_RESOURCE – guaranteed % of memory
  - CAP_PERCENTAGE_RESOURCE – maximum % of memory allowed to give
  - REQUEST_MIN_RESOURCE_GRANT_PERCENT – Minimum allowed per request
- Max concurrency = CAP % / REQUEST_MIN %
  - For example, 100% Cap / 20% Request Min = 5 Max Concurrency

Size	Max Conc
100	4
200	8
300	12
400	16
500	20
1000, 1500	32
2000, 2500	48
3000, 5000	64
6000+	128

MCKNIGHT CONSULTING GROUP

9

Other Performance Features

- Estimated query plan (coming soon)
- Transparent materialized views
- Adaptive caching (recently use data on NVMe)
- Azure Advisor
- No auto scaling (can trigger a scaling operation with a Pipeline job)
- Result caching turned off by default
- No short query acceleration
- No separate optimization service

MCKNIGHT CONSULTING GROUP

10

Usability Features

MCKNIGHT CONSULTING GROUP

11

Dynamic Data Masking

- Configurable with Portal or T-SQL
- Do not apply to role db_owner
- Functions
  - Default() – [0, xxxx, 1901-01-01]
  - Email() – jXX@XXXX.com
  - Creditcard() – XXXX-XXXX-XXXX-1234
  - Random(min, max) – 1234
  - Partial(prefix, mask, suffix) – abcXXXyz

MCKNIGHT CONSULTING GROUP

12

External Data Sources

- Dedicated SQL Pools only
- Requires:
  - Master key
  - Database scoped credential
  - External data source
  - External file format
  - External table
- Use OPENROWSET for Serverless

MCKNIGHT CONSULTING GROUP

13

Synapse Link

- ***Near*** real time analytics over operational data in:
  - Azure SQL Database
  - SQL Server 2022
  - CosmosDb (via an isolated column store)
- Minimal operational performance impact
  - Automatically extracts incremental changes
  - Automatically replicates to Synapse dedicated SQL pool, no ETL

MCKNIGHT CONSULTING GROUP

14

SynapseML

Features

- Previously known as MMLSpark
- Installable in:
  - Synapse
  - Python/Conda
  - Scala
  - Existing Spark clusters
  - Databricks
  - Apache Livy and HDInsight
  - Docker
  - .NET

Workflow

- Data ingest – Pipelines
- Data preparation – Spark SQL Pools
- Data exploration – Serverless SQL
- Model Training –
  - Spark SQL Pools with Mlib
  - Automated ML
- Model Scoring –
  - TSQL PERDICT
  - Spark Pools

MCKNIGHT CONSULTING GROUP

15

BI Options

- Power BI integration
- Studio charts

MCKNIGHT CONSULTING GROUP

16

Discussion

MCKNIGHT CONSULTING GROUP

17

Data Warehouse Units (DWU)

- Official: "a collection of analytic resources…defined as a combination of CPU, memory, and IO…[which] represents an abstract, normalized measure of compute resources and performance."
- Increasing DWUs linearly improves performance
- "Nodes"
  - **Bold** – terms used by Microsoft engineers in 2018
  - *Italics* – MCG inferred
  - # of nodes = DWUs x 2 / 1000

DWUs	"Nodes"	Price/hr
100	*1/5th*	$1.20
200	*2/5th*	$2.40
300	*3/5th*	$3.60
400	*4/5th*	$4.80
500	**1**	$6
1000	*2*	$12
1500	*3*	$18
2000	*4*	$24
2500	*5*	$30
3000	*6*	$36
5000	**10**	$60
6000	*12*	$72
7500	*15*	$90
10000	*20*	$120
15000	**30**	$180
30000	**60**	$360

18

Pricing

Component	Price
Serverless	$5/TB processed
Dedicated	$/hour >>>
1-year Reserved	37% discount
3-year Reserved	65% discount
Storage	$23/TB-month

- Additional charges (per vCore-hour) for Synapse Link, Data Explorer, and Spark Pools
- Pipelines priced by DIU-hour, runtime-hour, and per activity run

DWUs	Price/hr
100	$1.20
200	$2.40
300	$3.60
400	$4.80
500	$6
1000	$12
1500	$18
2000	$24
2500	$30
3000	$36
5000	$60
6000	$72
7500	$90
10000	$120
15000	$180
30000	$360

MCKNIGHT CONSULTING GROUP

19