**Cognizant POV on KP A20 Analytical Digital Platform-**

1.       Use Cases - Data Integration, Data Quality, Data Consumption, Data Governance, Platform

2.       Data Management –

a. Ingestion and Storage - Data Zones (raw/cleansed/curated), Analytical DB’s – Snowflake, Synapse,

b.       Compute – Databricks, HDInsights, Hive

c.       Access and discovery - Semantic Layer, APIs enablement, External Tables

3.       Shared Services

a.       Governance – Metadata - Data catalog, Security

b.       Platform – Roadmap Alignment, Interoperability, Maintainability

1. Inferenced from the Evaluation between Snowflake and Synapse (High Level Industry Knowledge, project implementation etc.)
2. Plug in additional technology platforms to meet user needs
   1. Layer 0 / Layer 1/ -- Azure Corporate guidance
   2. Layer 2 – Tenant (Plug and Play) API , Data as a Service . Semantic Layer,

**Evaluation Criteria Outline –**

|  |  |
| --- | --- |
| **Primary Criteria** | **Secondary Criteria** |
| **1. Cloud** |  |
|  | **1.1 Cloud Integration** |
|  | **1.2  Elasticity** |
| **2. Performance** |  |
|  | **2.1 Response time & Latency** |
| **3. Storage & Data Distribution** |  |
|  | **3.1 Storage type/formats** |
|  | **3.2 Data Storage** |
| **4. Security, Maintenance & High Availability** |  |
|  | **4.1 Security** |
|  | **4.2 DR / HA** |
|  | **4.3 Maintenance & Work load management** |
| **5. Interoperability/Compatibility & Integration** |  |
|  | **5.1 Native compatibility** |
|  | **5.2 Query capability** |
|  | **5.3 Supported Programing Languages** |
| **6. Cost & Resource availability** |  |
|  | **6.1 Cost** |
|  | **6.2 Resource availability** |
| **7. Vendor** |  |
|  | **7.1 Vendor presence, credibility, training & support** |

A20 Platform Data Management Use Cases

|  |  |
| --- | --- |
| Dataset | Structured and Semi Structured Dataset |
|  | Real Time, Near Time and Batch Ingestion |
| Data Zones | Multi-Tier Data Storage to support Multiple business needs (Raw/Refined/Enriched) |
|  | Data Quality and Enterprise level data conformation |
|  | User Driven Business Data need (Tenant Zone) |
| Data Platform | Scalable and Available Architecture |
|  | Efficient and Performance Oriented Process |

Data Zone Storage (raw, refined, enriched) -

Options:

|  |  |  |  |
| --- | --- | --- | --- |
| Parameters / | Object Store (ADLS) | DB Store (Snowflake) | File Store with Structure (ADLS + Delta Lake) |
| Data Formats Supported | All data formats | JSON, CSV, tables, Parquet, ORC … | All formats, Parquet |
| Structured | Yes | Yes | Yes |
| Semi Structured | Yes | Yes (Store data as is) and parse using SQL  Supported format - **JSON, Avro, ORC, Parquet, or XML** | Yes - Schema Enforced – requires parsing and load as Structured |
| Unstructured | Yes | No | No |
| Access Methods |  |  |  |
| Near Real time (>100ms) |  |  | Yes |
| Support for Real Time < 100 MS |  |  | Yes |
| Connectors |  |  |  |
| Security |  |  |  |
| Authentication |  |  |  |
| Data Management |  |  |  |
| Data Access |  |  |  |
| Cost |  |  |  |
| Change Data Capture |  | Logic | Out of Box |
| Metadata |  |  |  |

Critical Recommendation:

Data Processing (Compute) -

Options:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Data Bricks  ADF + Spark | DB Store (Snowflake) | Synapse Studio (Future) | Informatica PaaS |
| Separation of Storage and Compute | Yes | Yes | TSQL + Spark + Data Management |  |
| Elastic Clustering | Yes | Yes |  |  |
| Workload Isolation | Yes | Yes | Yes |  |
| Multi Cluster Support | Yes |  |  |  |
| Authentication |  |  |  |  |
| Online Scaling |  |  |  |  |
| Complex ETL logic support | Yes |  |  |  |
| GUI/Programming |  |  |  |  |
| Performance |  |  |  |  |
| Cost |  |  |  |  |
| Priority Workload Schedule |  |  |  |  |
| Cashing/In-Memory Processing Support |  |  |  |  |
| CI/CD |  |  |  |  |
| Metadata |  |  |  |  |
| Data Management | External Integration |  | Seamless Integration |  |
| Data Quality |  |  |  |  |
| Data Lineage |  |  |  |  |
| Metadata Management |  |  |  |  |

Critical Recommendations:

Business Data – Tenant Zone Storage

Options:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Snowflake | Synapse Analytics (SQL DWH New version) | Hive (Context of Current KP Architecture) |
| Relational SQL Support |  |  |  |
| DML Support |  |  |  |
| Compression |  |  |  |
| Concurrency |  |  |  |
| Query Flashback |  |  |  |
| External Tables |  |  |  |
| Materialized Views |  |  |  |
| Performance Management and Monitoring tools |  |  |  |
| Data Migration Complexity |  |  |  |
| Reporting Tool Support |  |  |  |
| Statistics collection and Partitioning |  |  |  |
| Data Encryption |  |  |  |
| Row/Column Level Security |  |  |  |
| Automatic Statistics Update |  |  |  |
| Metadata |  |  |  |

Critical Recommendations