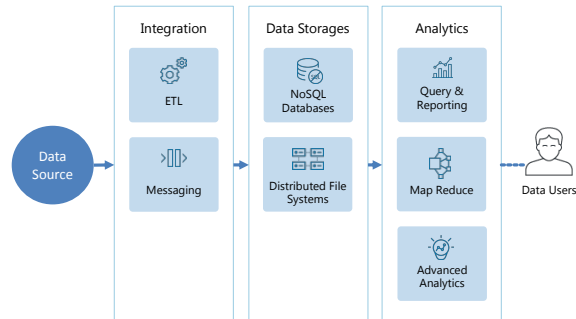## Pure Non-relational

Reference Architecture for Data Analytics

Description: This reference architecture does not rely on relational model principles. Often it is built on NoSQL storage, Hadoop, Search Engines and is highly effective for processing semi and unstructured data.
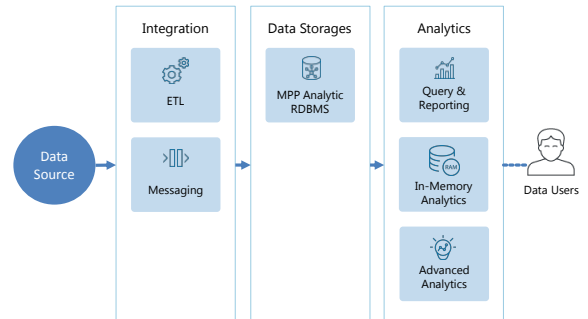

Consequences:

★★ **Ad-hoc analysis** - ad-hoc real-time query support is more difficult than in relational architecture

★★½ **Real-time analysis** – real-time with one-at-a-time processing

★★★ **Unstructured data processing** – supports easy storing and processing of semi and unstructured data

★★★ **Scalability** – can scale keeping petabytes

★★★ **Cost economy** – cost minimized due to open-source technologies

Sample implementations: Data Discovery, Data Lake, Operational Intelligence, Business Reporting

Extended Relational

Reference Architecture for Data Analytics

Description: Although this reference architecture is completely based on relational model principles and SQL-based DBMS, it intensively uses MPP and In-Memory techniques to improve scalability and extensibility.
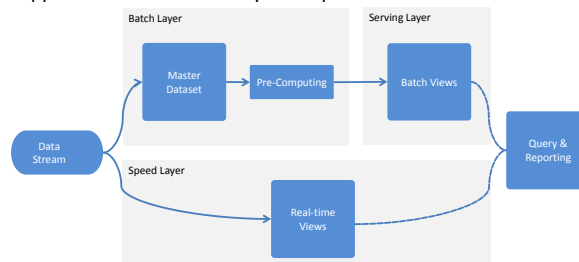

Consequences:

★★★ **Ad-hoc analysis** – supports complex ad-hoc real-time read queries

★★ **Real-time analysis** – near-real time with micro-batching technique

★★ **Unstructured data processing** – supports ingesting and querying semi-structured data such as JSON/XML

★★ **Scalability** – can run terabytes with MPP and clustering capabilities

★ **Cost economy** – MPP RDBMS license cost is quite expensive

Sample implementations: Business Reporting, Enterprise Data Warehousing, Data Discovery

Lambda Architecture (Hybrid)

Reference Architecture for Data Analytics

Description: This reference architecture enables real-time operational and historical analytics in the same solution. While the batch layer is based on non-relational techniques (usually Hadoop), the speed layer is based on streaming techniques to support strict real-time analytics requirements.
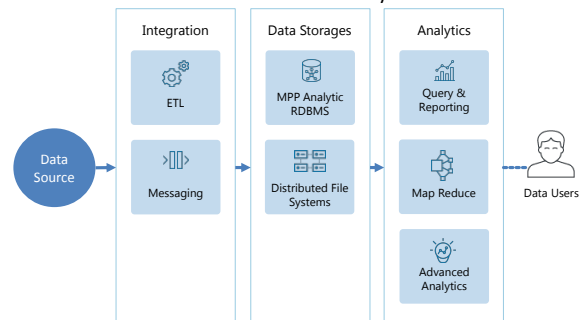

Consequences:

★★ ½ **Ad-hoc analysis** - ad-hoc real-time query support is more difficult than in relational architecture

★★★ **Real-time analysis** – streaming approach with low data latency

★★★ **Unstructured data processing** – supports processing of semi and unstructured data

★★★ **Scalability** – can scale keeping petabytes

★★★ **Cost economy** – cost minimized due to open-source technologies

Sample implementations: Real-time Intelligence, Data Discovery, Business Reporting

Data Refinery (Hybrid)

Reference Architecture for Data Analytics

Description: This reference architecture is a mix of relational and non-relational techniques. Non-relational part acts as an ETL to refine semi and unstructured data and load it cleansed into relational data warehouse for further analysis.

Consequences:

★★★ **Ad-hoc analysis** – supports complex ad-hoc real-time read queries

★ **Real-time analysis** – data latency is high due to batch processing

★★★ **Unstructured data processing** – supports processing of semi and unstructured data

★★ **Scalability** – can run terabytes with MPP and clustering capabilities

★ **Cost economy** – MPP RDBMS license cost is quite expensive

Sample implementations: Data Discovery, Business Reporting, Enterprise Data Warehousing