Factors for Selecting and Designing Data Stores

Considerations for designing a Data Store

Primary considerations for designing a data store:

- Type of data
- Volume of data
- Intended use of data
- Storage considerations
- Privacy, Security, and Governance needs

Type of Data

RDBMS

Best used for structured data, which has a well-defined schema and can be organized into a tabular format.

NoSQL

Best used for semi-structured and unstructured data, data that is schema-less and free-form.

Type of Data

Four types of NoSQL Databases:

Key-value

Document-based

Column-based

Graph-based

Volume of Data

Volume / Scale of data:

Data Lake

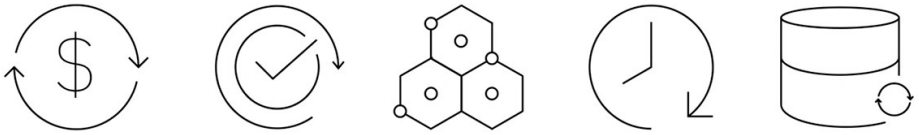
- Store large volumes of raw data in its native format, straight from its source
- Store both relational and non-relational data at scale without defining the data's structure and schema

Big Data Store

- Store data that is high-volume, high-velocity, of diverse types, needs distributed processing for fast analytics
- Big Data Stores split large files across multiple computers allowing parallel access to them

Intended use of Data

How you intend to use the data you are collecting:



Number of Transactions | **Frequency of Updates** | **Type of Operations** | **Response Time** | **Backup and Recovery**

Intended use of Data

Transactional Systems used for capturing high-volume transactions, need to be designed for high-speed read, write, and update operations.

Analytical Systems need complex queries to be applied to large amounts of historical data aggregated from transactional systems. They need faster response times to complex queries.

Schema design, indexing, and **partitioning** strategies have a big role to play in performance of systems based on how data is getting used.

Intended use of Data

The intended use of data also drives **scalability** as a design consideration.

Normalization is another important consideration at the design stage.

Optimal use of storage space

Makes database maintenance easier

Provides faster access to data