

1. DATA LAKE

1.1 What is data lake?

- A Data Lake is a storage repository that can store large amounts of structured, semi-structured, and unstructured data.
- It is a place to store every type of data in its native format with no fixed limits on account size or file.
- It offers high data quantity to increase analytic performance and native integration.
- The Data Lake democratizes data and is a cost-effective way to store all data of an organization for later processing.

1.2 Why Data Lake?

- With the onset of storage engines like Hadoop storing disparate information has become easy. There is no need to model data into an enterprise-wide schema with a Data Lake.
- With the increase in data volume, data quality, and metadata, the quality of analyses also increases.

1.3 References

- <https://www.guru99.com/data-lake-architecture.html>

2. DATA INTEROPERABILITY

2.1 What is data interoperability?

- Data interoperability addresses the ability of systems and services that create, exchange and consume data to have clear, shared expectations for the contents, context and meaning of that data.

3. CDISC HL7

- Clinical Data Interchange Standards Consortium - High Level 7
- **SDTM is one of the required standards for data submission to FDA (U.S.) and PMDA (Japan).**

FHIR Fast Healthcare Interoperability Resources (FHIR, pronounced "fire")