Concept

- store massive amounts of (un/semi)structured data
- no need for architected schema
    - no upfront costs
- typically stored in raw formats
- distributed storage and analytics for scalability and agility

Critisism Data Swamp

- low data quality due to lack of
    - schema
    - integrity constraints
    - validation
- missing meta data and data catalog for search

[[Distributed Data Storage]]