

Intro til Spark & DeltaLake i Synapse



Simple.
Fast.
Scalable.
Unified.

Key features



Batch/streaming data

Unify the processing of your data in batches and real-time streaming, using your preferred language: Python, SQL, Scala, Java or R.



SQL analytics

Execute fast, distributed ANSI SQL queries for dashboarding and ad-hoc reporting. Runs faster than most data warehouses.



Data science at scale

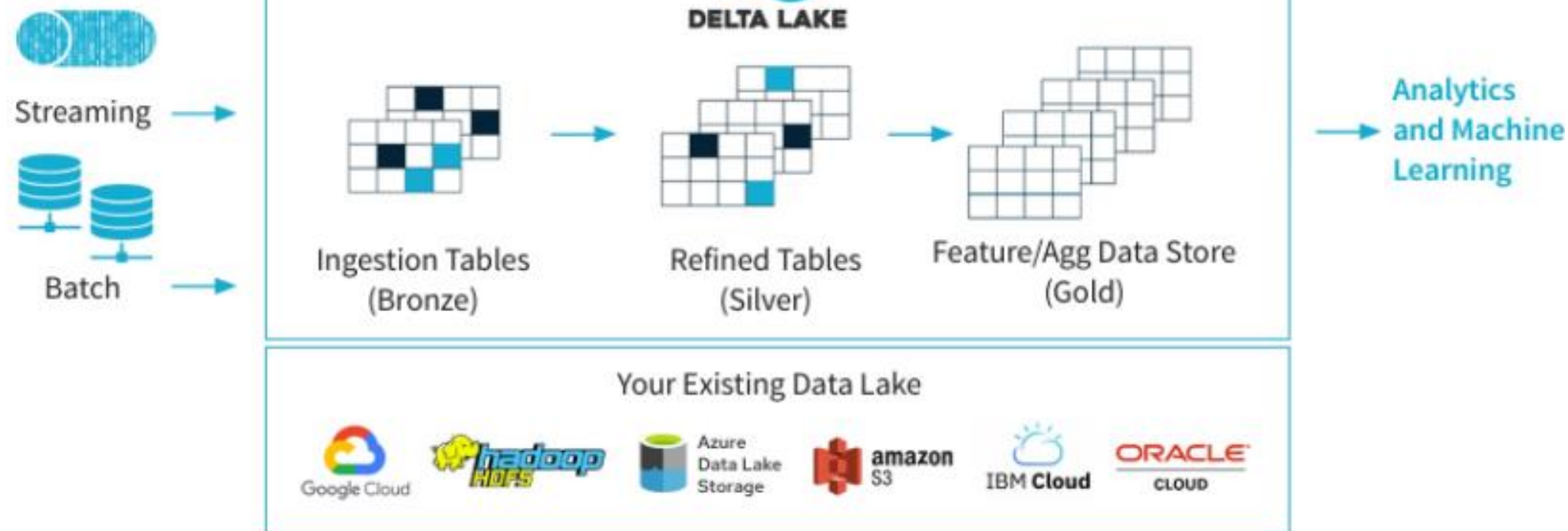
Perform Exploratory Data Analysis (EDA) on petabyte-scale data without having to resort to downsampling



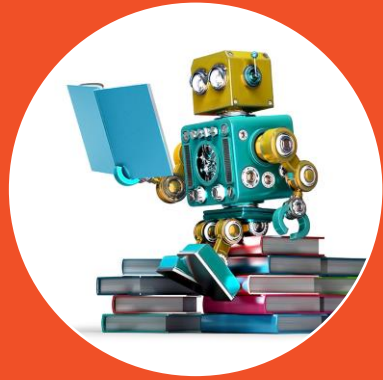
Machine learning

Train machine learning algorithms on a laptop and use the same code to scale to fault-tolerant clusters of thousands of machines.

Delta Lake



Brugsscenarier



Machine
Learning
og AI



Data Lake
House



Skalérbar
computing



Fill in the
blanks

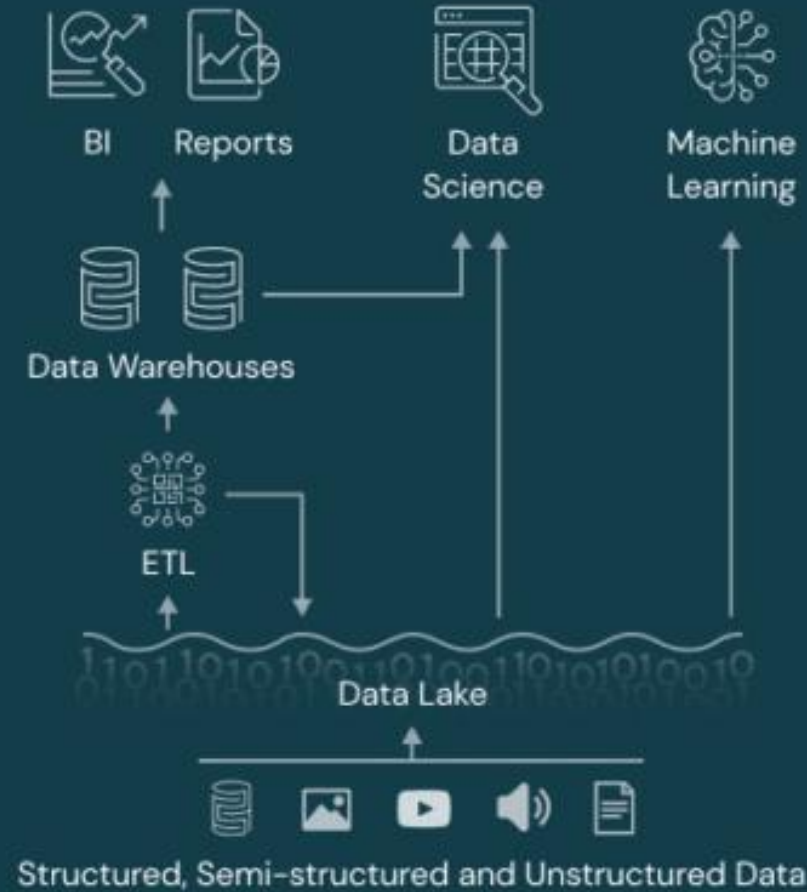
Delta Lake

Data Lakehouse

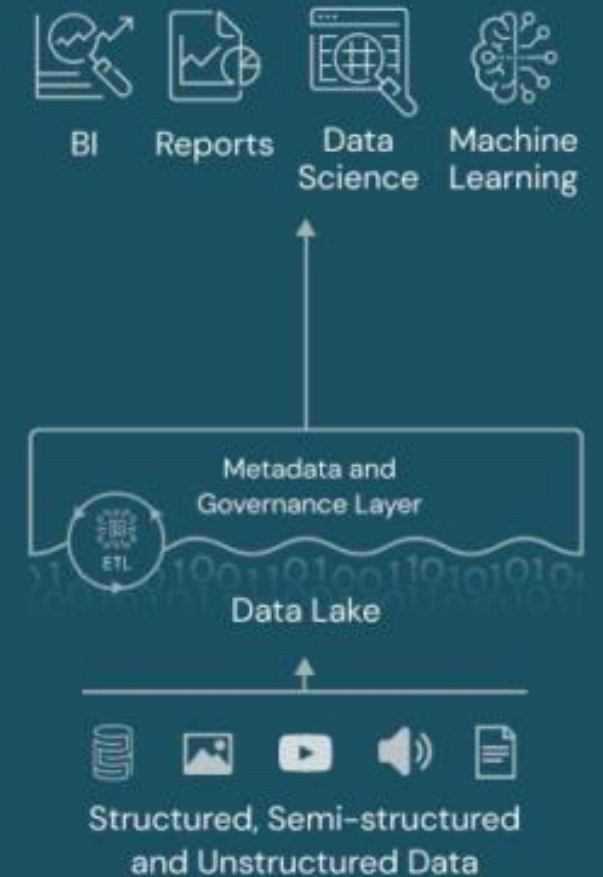
Data Warehouse



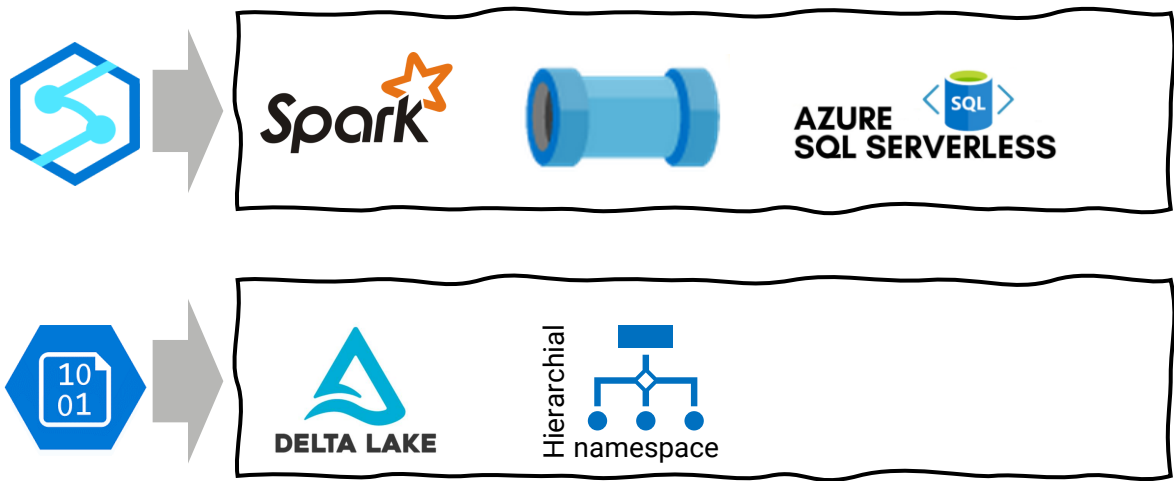
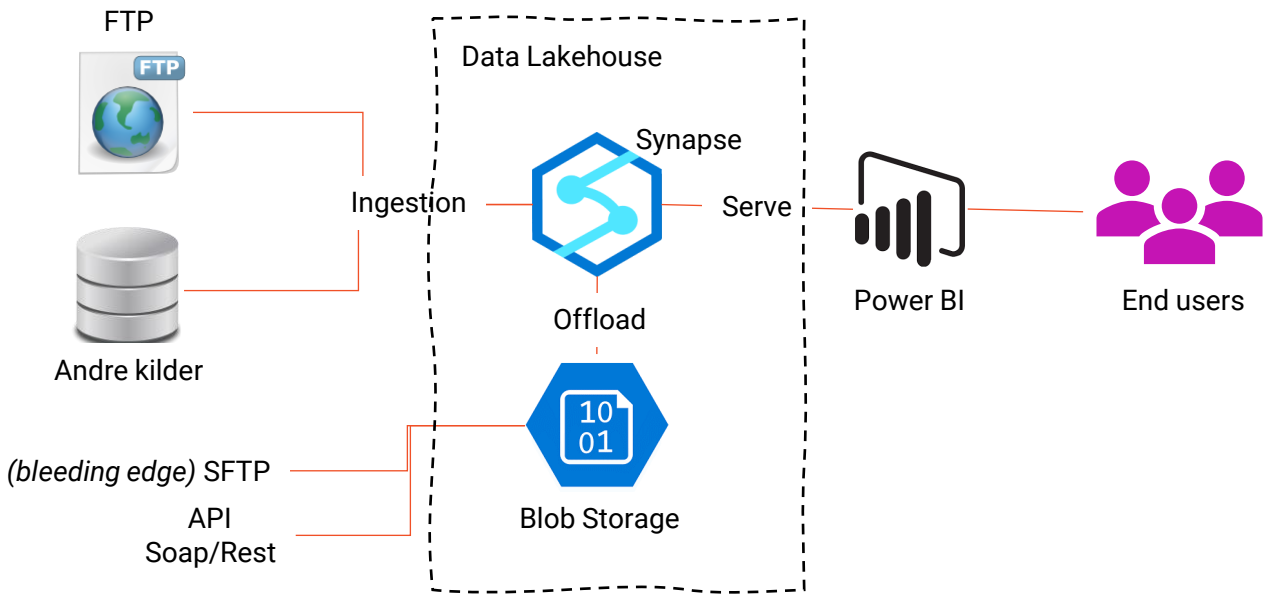
Data Lake

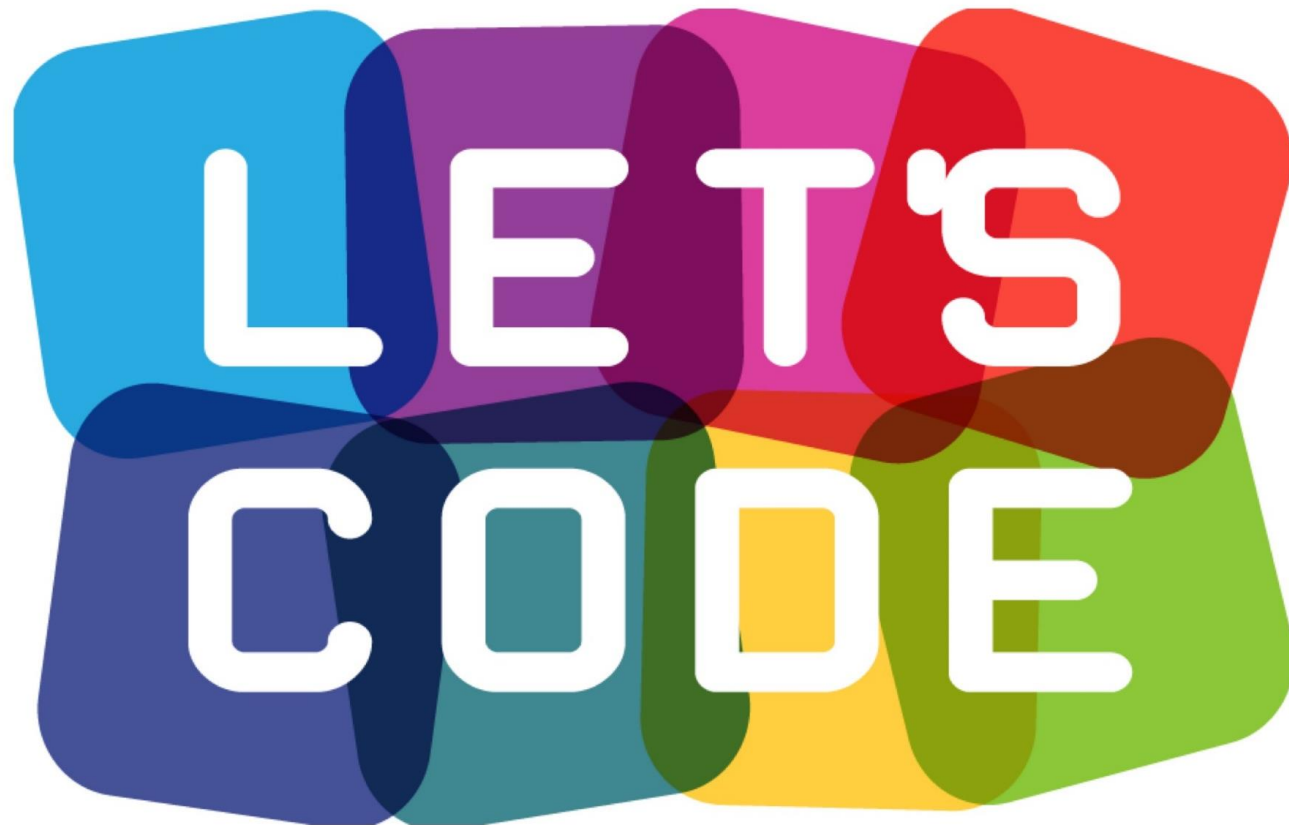


Data Lakehouse



Eks på simpel arkitektur





Alle slides og kodeeksempler ligger her: <http://fmdk.io/msbip68>

