What is a Datalake

Collection of storage instances of various data assets

Assets stored in near-exact copy of the source format

Contents stream in from multiple sources, online, to fill the lake

All data has value in a data lake, structured, unstructured, raw

A storage repository that holds a vast amount of raw data in its native format until it is neede

Inexpensively store unlimited data

Quick user access to data for power users/data scientists (allowing for faster ROI

What is a DataLake?

What is a Datalake

A storage repository that holds a vast amount of raw data in its native format until it is needed

Collection of storage instances of various data assets.

Assets stored in near-exact copy of the source format.

Large data content in a more natural state.

Contents stream in from multiple sources, online, to fill the lake.

Various users can examine, dive, or take samples from the lake.

All data has value in a data lake, structured, unstructured, raw.

Inexpensively store unlimited data

Quick user access to data for power users/data scientists (allowing for faster ROI)

true > hype < co

Raw Data Layer

Cleansed Data Layer

Application Data Layer

Sandbox Data Layer

true > hype < co