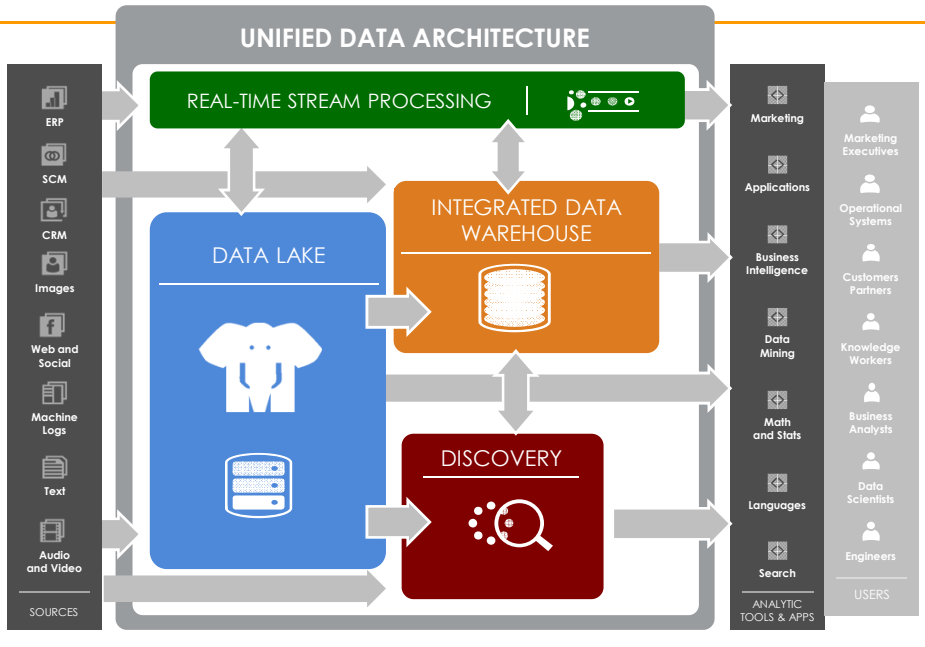
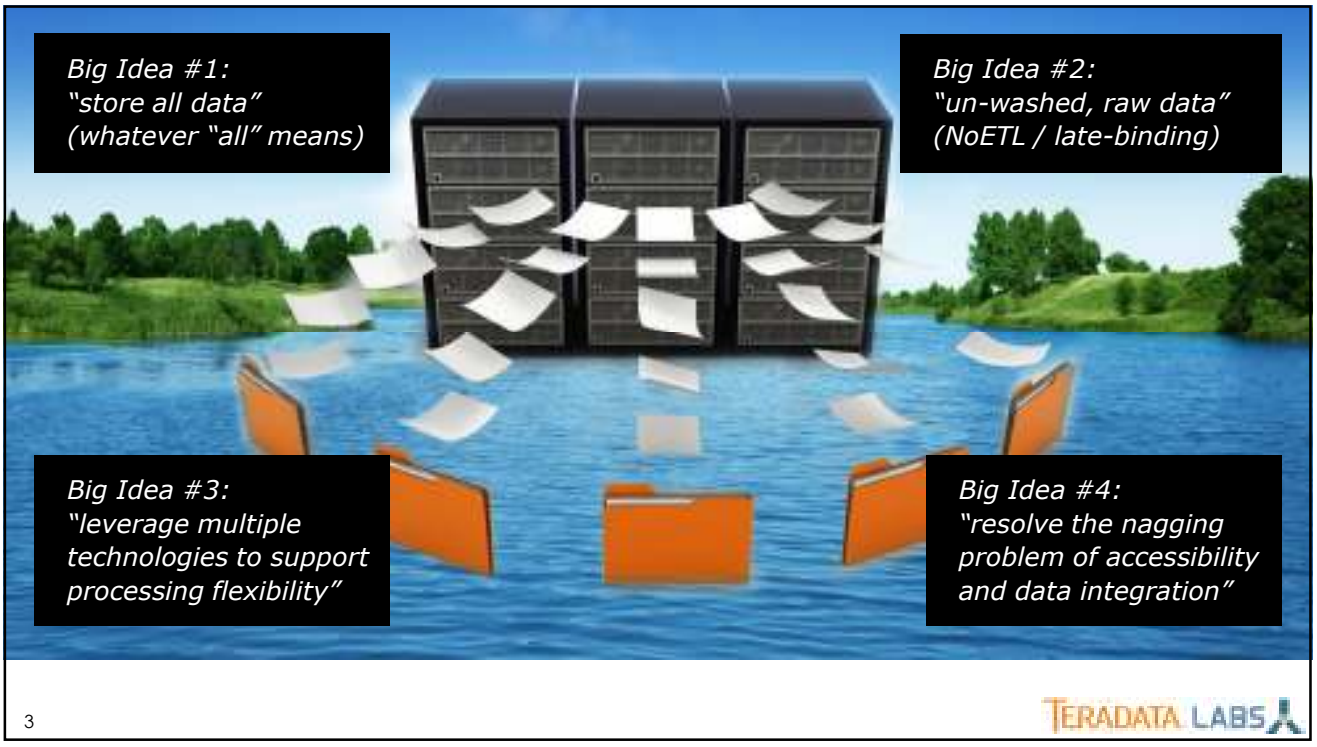
Analytic Ecosystem



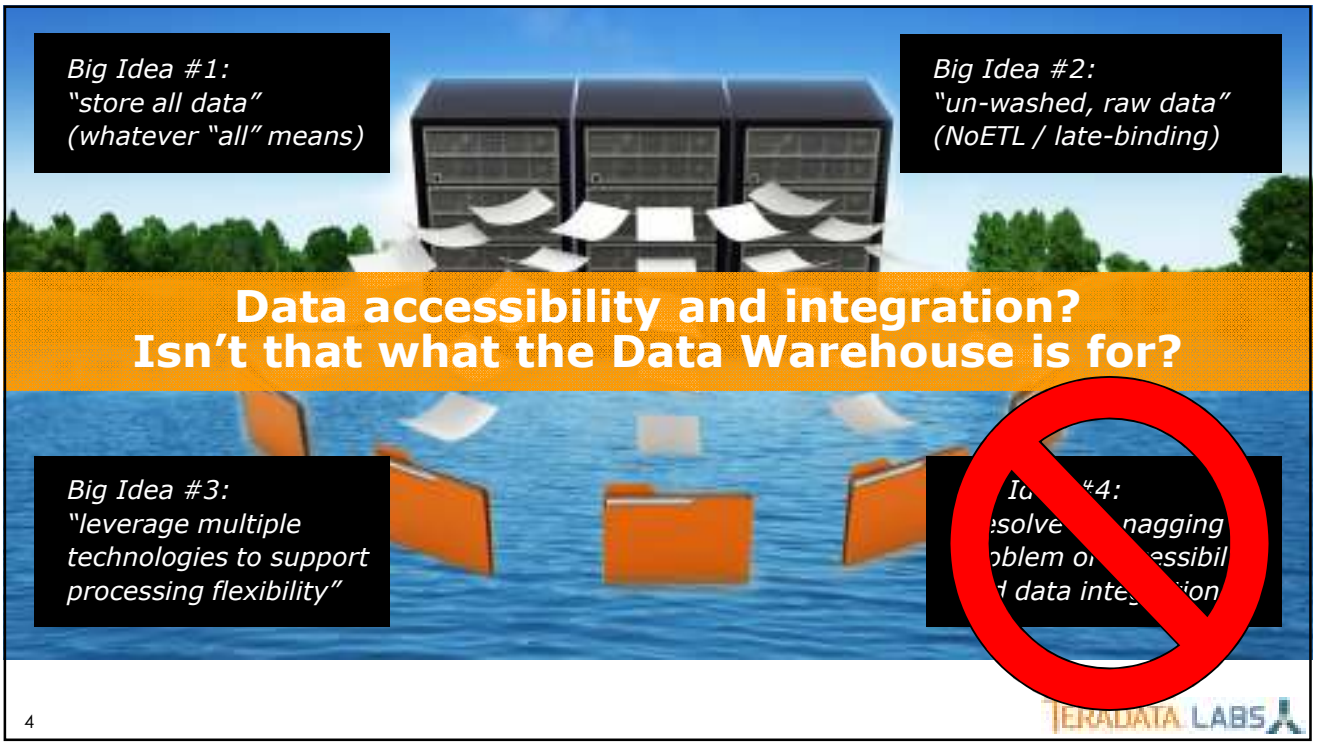
TERADATA LABS

2



3

TERADATA LABS



4

TERADATA LABS

The Data Lake as it should be: a centralized, consolidated store of raw data from multiple sources

5

TERADATA LABS

Data Lakes versus Data Swamps

Through 2018, **90%** of deployed data lakes will be **USELESS** as they are overwhelmed with information assets captured for uncertain use cases.

Gartner, Strategic Planning Assumption, Gartner BI Summit, 2015.

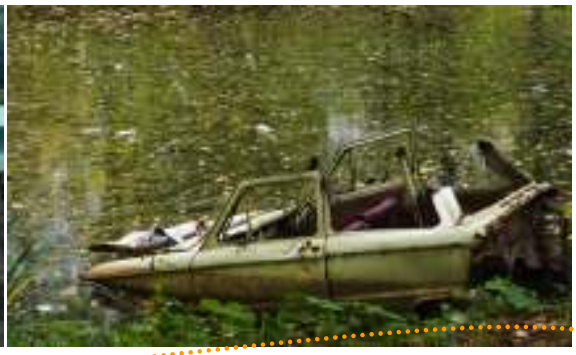
6

TERADATA LABS

Data Reservoirs versus Data Swamps

On which data asset would you rather bet your career?

Data Reservoir **…or…** **Data Swamp**



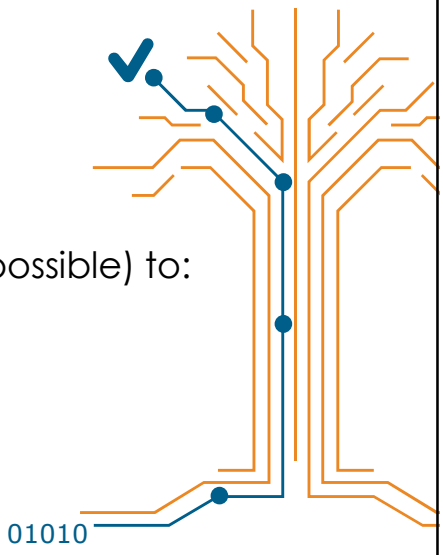
TERADATA LABS

Basic Questions of Provenance

- Who created the data asset and when?
- What is the source of the raw data used to create the asset?
- What processes were used to create the data asset?
- What are the known defects associated with the data asset?
- What algorithms were used to manipulate data?

Without provenance it is hard (sometimes impossible) to:

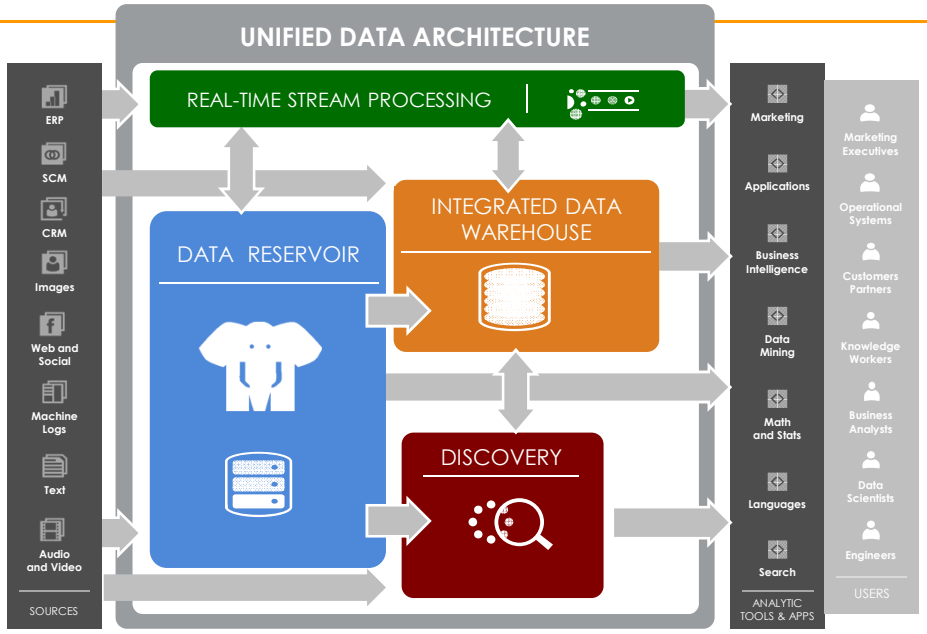
- Reproduce results,
- Solve problems collaboratively,
- Validate results with different input data,
- Understand the process used to solve a particular problem,
- Re-use the knowledge involved in the data analysis process.



Source: Hansen, Johnson, Pascucci, and Silva. Visualization for Data Intensive Science. *The Fourth Paradigm*. 2009. pp. 154-163.

TERADATA LABS

Analytic Ecosystem



9