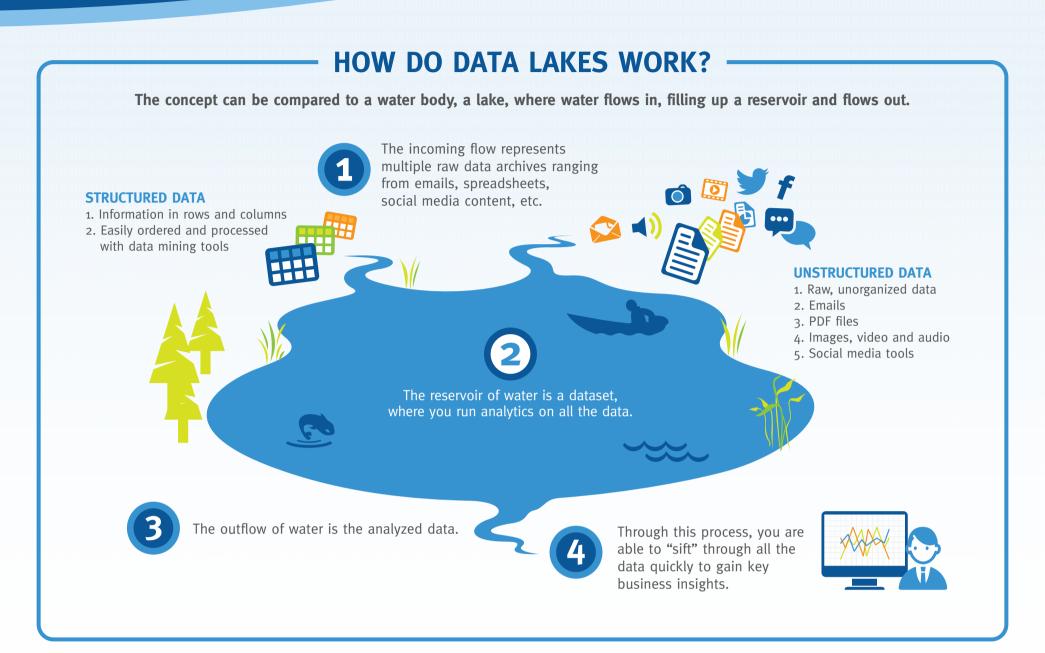# UNDERSTANDING DATA LAKES

WHAT IS A DATA LAKE?

Data lake is one place to put all the data enterprises may want to use, including structured and unstructured data

HOW DO DATA LAKES WORK?

The concept can be compared to a water body, a lake, where water flows in, filling up a reservoir and flows out.

1 The incoming flow represents multiple raw data archives ranging from emails, spreadsheets, social media content, etc.

STRUCTURED DATA
1. Information in rows and columns
2. Easily ordered and processed with data mining tools

UNSTRUCTURED DATA
1. Raw, unorganized data
2. Emails
3. PDF files
4. Images, video and audio
5. Social media tools

2 The reservoir of water is a dataset, where you run analytics on all the data.

3 The outflow of water is the analyzed data.

4 Through this process, you are able to "sift" through all the data quickly to gain key business insights.

DID YOU KNOW?

The information in the Digital Universe will grow 10 times by 2020.

In the last 10 years, companies have started using data lakes to deal with the enormous amounts of data.

Data lakes help reveal complex business issues and build predictive models to address these. Companies ranging from restaurants to mining corporations use data lake solutions in their everyday analytics.

WHO IS USING DATA LAKES?

BUSINESS & DATA ANALYSTS

Analyze reports on specific data in the organization to provide business insight

DATA ARCHITECTS

Responsible for designing, creating, deploying and managing an organization's data architecture

DATA SCIENTISTS & APP DEVELOPERS

Perform statistical analysis on big data to identify trends, solve business problems and optimize performance

---

WHY ARE DATA LAKES IMPORTANT?

BUILD APPLICATIONS

Platform for businesses to get at the data and quickly build the views, and data-driven applications they really need

FLEXIBILITY & ACCESSIBILITY

Provide flexibility and accessibility in moving large amounts of data from data warehouse to perform analytics

RETAIN DATA AUTHENTICITY

Data Lakes allow you to store and analyze the information in different formats, retaining data authenticity

SPEED

Ability to sift through immense quantities of data quickly

EXPLORE & ANALYZE

Ability to explore and analyze data to derive business value and benefit