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.

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

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

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