The difference between structured, unstructured and semi-structured data.
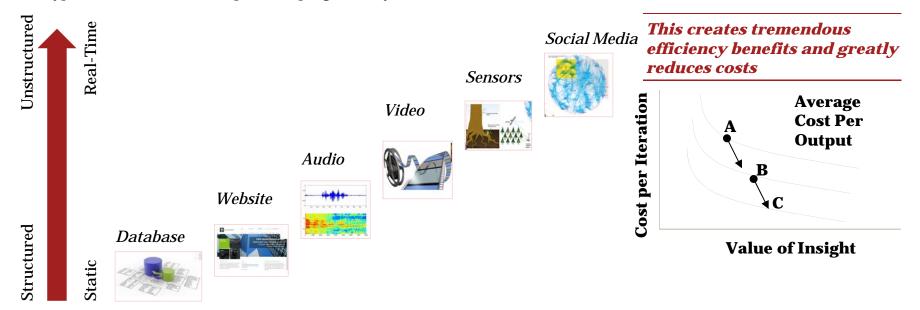
The type of data that we can exploit has progressively moved from structured to semi-structured to unstructured data

Unstructured / **Structured**

Real-Time / **Static**

Database

Website

Audio

Video

Sensors

Social Media

This creates tremendous efficiency benefits and greatly reduces costs

Average Cost Per Output

Cost per Iteration

A

B

C

Value of Insight

Structured data reside in a fixed format within a record or a file. This includes data in relational databases or spreadsheets.

Most popular platforms for structured data include:

- Oracle
- MS SQL Server
- MS Access, etc.

Example application: Traditional financial data investigation.

Big Data can be associated with structured data sources (but not exclusively).

Structured Data

- Data that reside in fixed fields.
- Relational databases spreadsheets.

Semi-structured Data

- Data that contain tags and other markers to separate data elements.
- XML or HTML tagged text.

Unstructured Data

- Data that does not reside in fixed fields.
- Books, e-mails, untagged audio, image and video data.

Business Perspective

- How do we understand the semantic meaning of the data in context of a problem?
- How do we develop an indexing scheme so that information can easily be found and utilized?

Technology Perspective

- Bandwidth and computing power required to move/analyze large sets of unstructured data around
- Shortfall of analytical talent
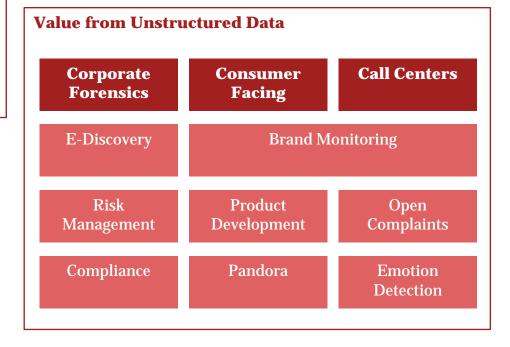- Insufficiency of current IT systems to support volume/structure

Semi-Structured

- Does not conform to a structural format like relational or other standard formats.
- E.g. XML, Email etc.

Unstructured Data

- Contains files of various formats, sizes, structure, etc.
- E.g. document collections (text), social interactions, images, video, audio, etc.

Company data from various sources:

- Files
- Audio
- Videos
- Social Media
- Emails
- Pictures

What are some business application of unstructured data?

Value from Unstructured Data

Corporate Forensics	Consumer Facing	Call Centers
E-Discovery	Brand Monitoring	
Risk Management	Product Development	Open Complaints
Compliance	Pandora	Emotion Detection