



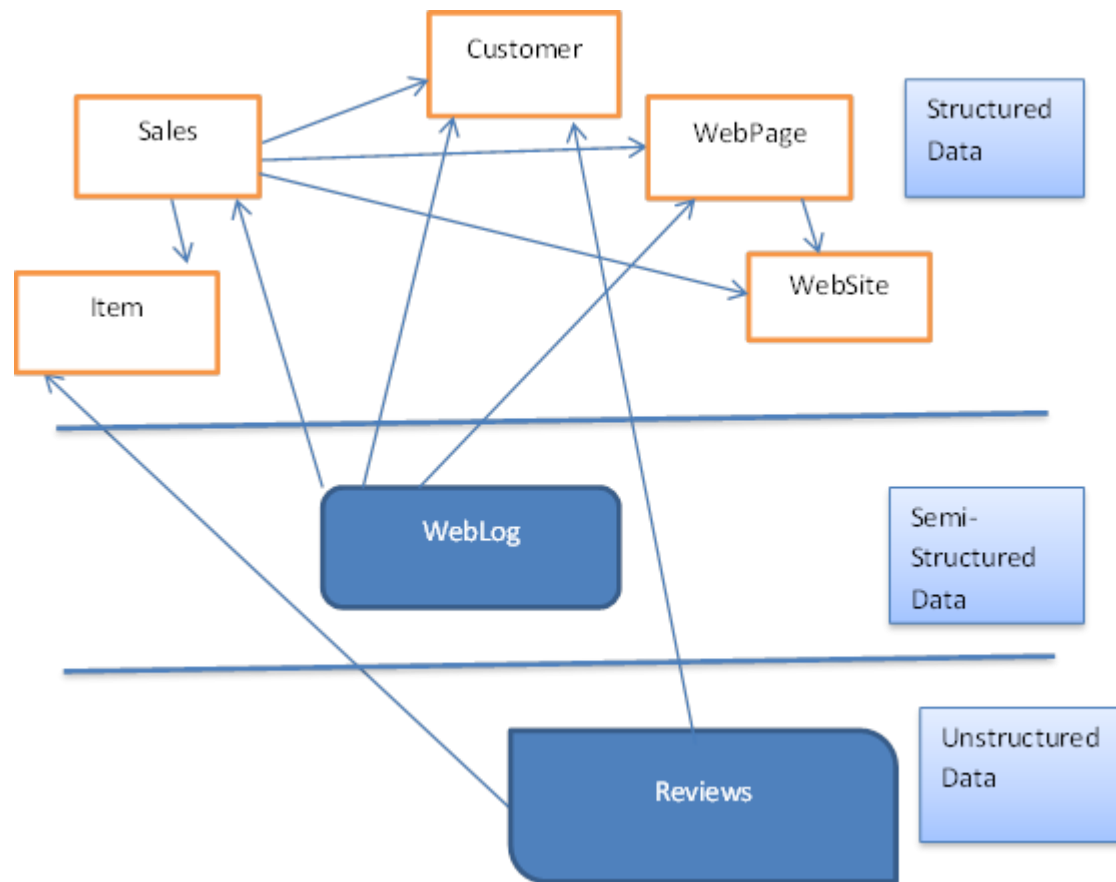
Big Data Benchmark Data Model

Ahmad Ghazal, Database Architect, Aster Data – Teradata
May 8, 2012

Big Data Benchmark Data Model

- What is Big Data?
 - Different interpretations and definitions
 - Similar view to Gartner's three V's.
 - **Volume**. Moving target. Petabytes, Exabyte's and Zettabytes
 - **Velocity**. High speed of data in/out.
 - **Variety**
 - Data sources like: transactions, internet, devices, sensor information
 - Data types: structured, un-structured and semi-structured.
- Structured data: relational data
- Unstructured data : text, video and audio
- Semi-structured: structured but not relational
 - XML, weblogs and sensor information

Big Data Benchmark Data Model



- Structured Data is snap shot of existing TPC-DS relational model
- Weblogs & Reviews are proposed additions.

Big Data Benchmark Data Model

- TPC-DS
 - Decision support benchmark from Transaction Processing Performance Council
 - http://www.tpc.org/tpcds/spec/tpcds_1.1.0.pdf
- Why build on top of TPC-DS?
- Volume :
 - No theoretical limit
 - Tested up to 100 TB
- Velocity : rolling updates
- Variety
 - Rich relational model and data, 26 tables, 7 fact tables, ...
 - Easy to add the other two sources