Big Data Ecosystem
Reference Architecture

Orit Levin, Microsoft
July 10th, 2013
RA Objectives

- Audience: Useful to industry, policy makers, and users (or “data owners”)
- Scope: Encompasses the whole data life cycle
- Focus:
  - Exposes projected “interoperability surfaces”
  - Assists in identifying security and privacy issues
- Addresses radically different Big Data use cases
  - An ecosystem comprised of independent stakeholders (e.g., advertising industry)
  - A stand-alone Enterprise Data Warehouse
  - Outsourcing of selected or all data transformations to SaaS providers
  - Outsourcing of data storage and/or computing to IaaS providers
  - Etc.
- Agnostic to any specific technologies
- *Does not show the mapping to NIST Cloud RA (at this time).*
Big Data Ecosystem RA

Data Sources
- Individual Data Transfer
- Government (incl. health & financial institutions)
- Industries / Businesses
- Network Operators / Telecom
- Academia

Data Objects
- Variety
- Velocity
- Volume

Data Transformation
- Data Collection
- Aggregation
- Matching
- Data Mining

Data Infrastructure
- Storage & Retrieval
- Conditioning
- Security

Data Usage
- Network Operators / Telecom
- Industries / Businesses
- Government (incl. health & financial institutions)
- Academia

Selected Data Storage and Retrieval
- Big Data Transfer
- Anonymized
- PII
- Pseudonymized

Big Data Ecosystem RA
Use Case: Enterprise Data Warehouse

Data Sources
- MS Office Documents
- Online Transaction Processing (OLTP) Systems
- Files
- Archives
- Manual

Data Transformation
- Extraction, Transformation, and Loading (ETL)
- Online Analytical Processing (OLAP)
- Managed Report Environment (MRE)
- Data Mining / Knowledge Discovery in Databases (KDD)

Data Objects
- Central Data Warehouse
- Operational Data Store
- Staging Area

Data Infrastructure
- Security Management

Data Usage
- Subject Data Mart
- Regional Data Mart
- Department Data Mart
- Application Data Mart
- Functional Data Mart

Data Objects and their relationships:
- Individual Data Transfer
- Big Data Transfer
- Selected Data Storage and Retrieval
- Big Data Storage and Retrieval

7/10/2013