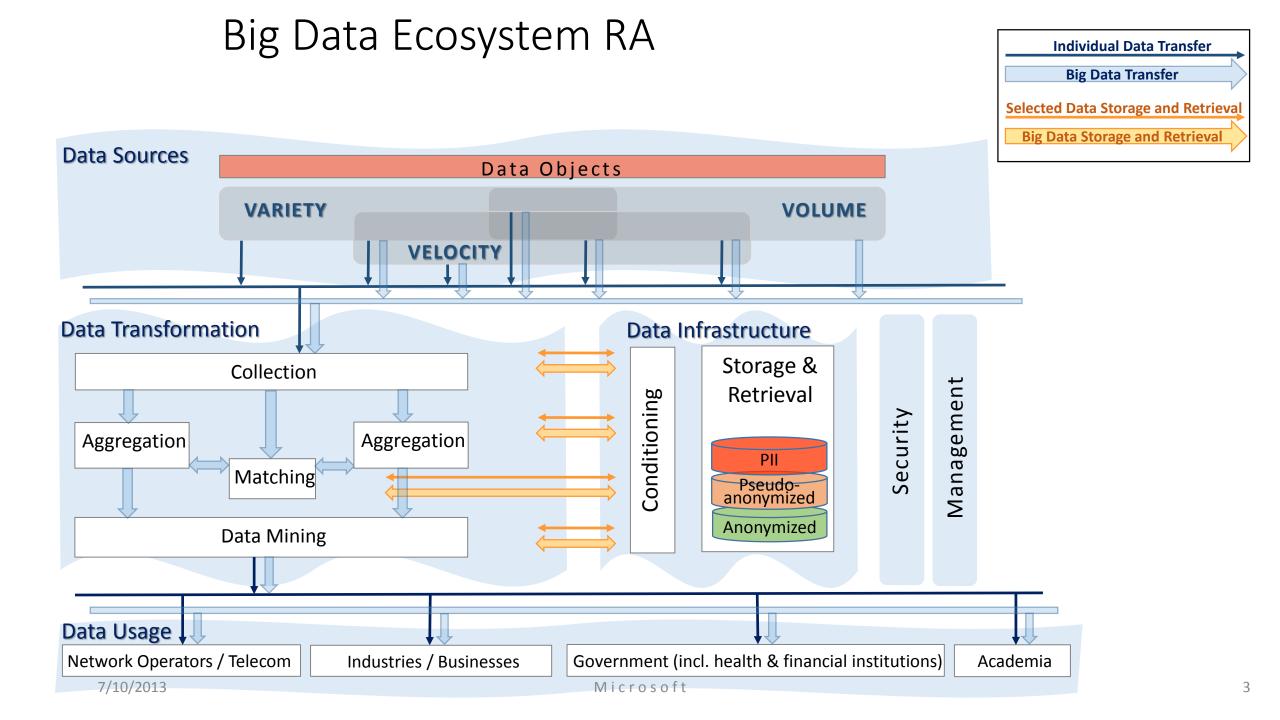
Big Data Ecosystem Reference Architecture

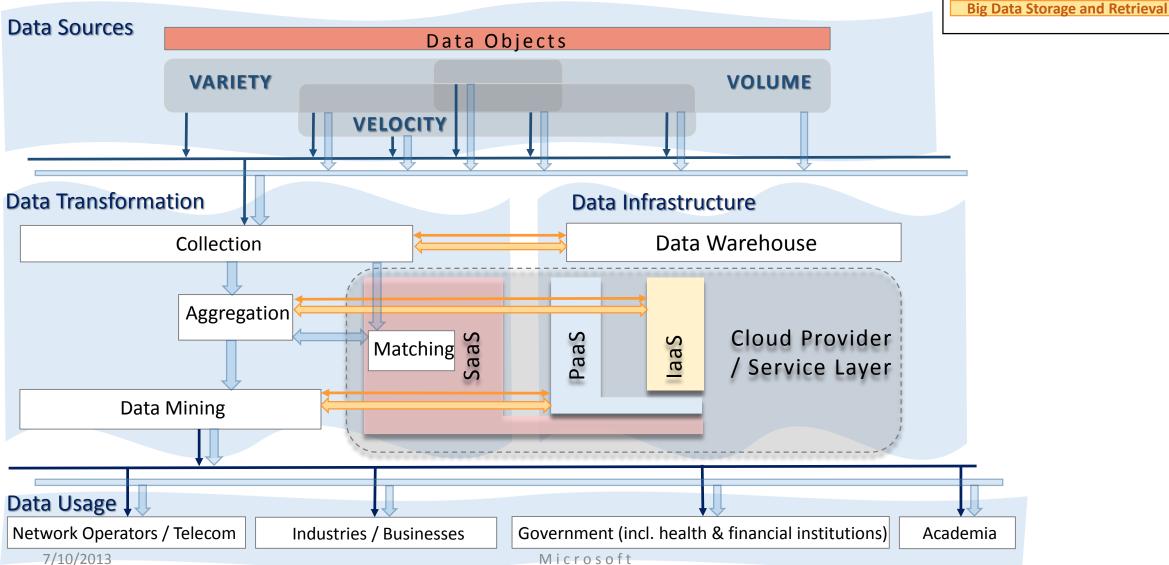
Orit Levin, Microsoft July 18th, 2013

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 laaS providers
 - Etc.
- Agnostic to any specific technologies
- Shows the mapping to NIST CC RA (slide # 4)



An Example of Cloud Computing Usage in Big Data Ecosystem



Individual Data Transfer

Big Data Transfer

Selected Data Storage and Retrieval

Use Case: Advertising **Individual Data Transfer Online Sources** Offline Sources **Big Data Transfer** Data Subject / Person 1st Party 2nd Party UI: Do Not Track (DNT) 3rd Party Other devices (Smart Grid, **Internal Records** Public Records (commons, End User devices incl. OS **Networks** surveillance, scientific, etc.) (mobile phones, etc.) government, etc.) PII De-identified **DPI** Web Browsers Aggregated **DMP Container Tag** HTTP: DNT **Match Container Tag** or Pixel request or Pixel request Collection Analytic Cookie Industries / Government, health, Network Applications (search, Appl. with customers Contextual financial institutions. Operators publishers, etc.) (communications, social **Businesses Data Collection** academia network, etc. Match Cookie Online Data Aggregator Offline Data Aggregator Match/Bridge Service **Data** DMP Cookie Management **Behavioral Data Creation Platforms Data Mining** (DMPs) Person Attribution **Users** Publisher AdNet SSP AdX **DSP** Advertising Industry Ecosystem Agency Advertiser 7/10/2013

Control

Use Case: Enterprise Data Warehouse

