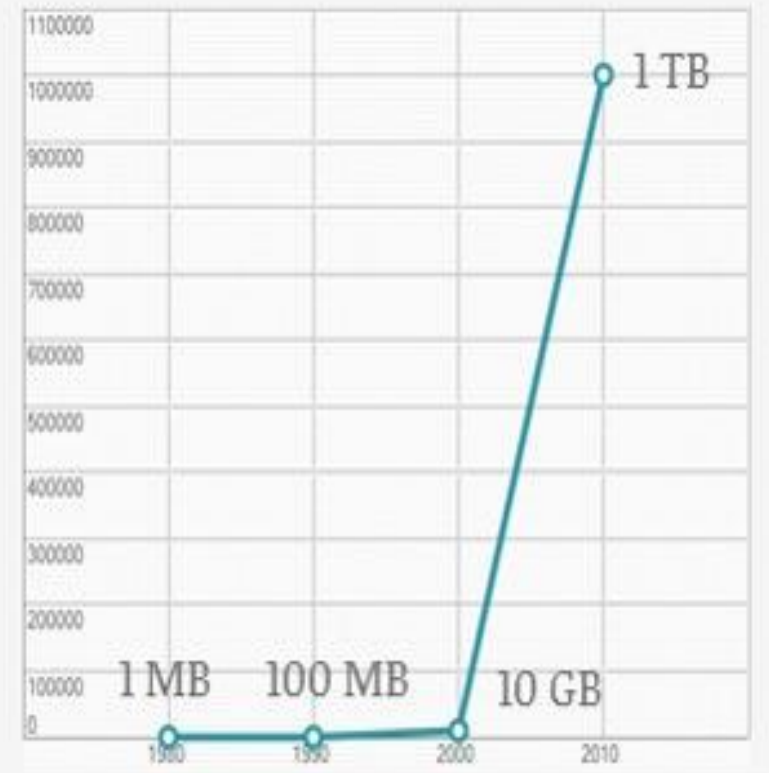


Big Data

What's the problem?

- Datasets collected or needed to be collected and analyze is so large and complex that it becomes difficult to process using traditional database management tools. Challenges include capture, curation, storage, search, sharing, analysis, and visualization
- Additional information is derivable from analysis of a single large set of related data, as compared to separate smaller sets

From 1 MB in 1980 to 1 TB in 2010

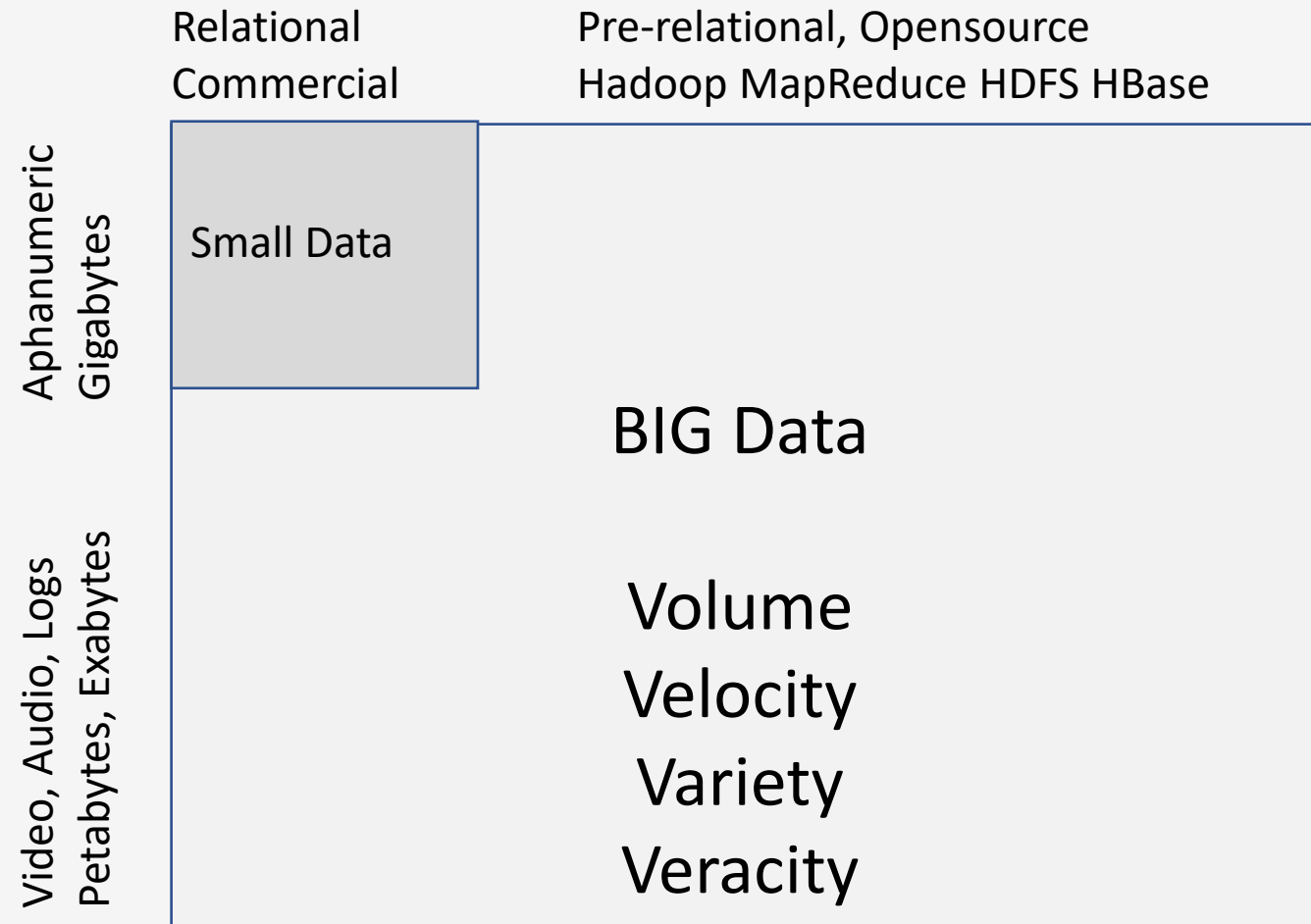


Big Data

Questions every business needs to answer in this age of Big Data

- How to use generated data as a strategic asset in real-time, to identify opportunities, thwart threats and achieve operational efficiencies
- How to organize the business to not get buried in high volume, velocity and variety of data
- How to design a 'Digital Business Strategy' around digital assets and capabilities

What is Big Data?



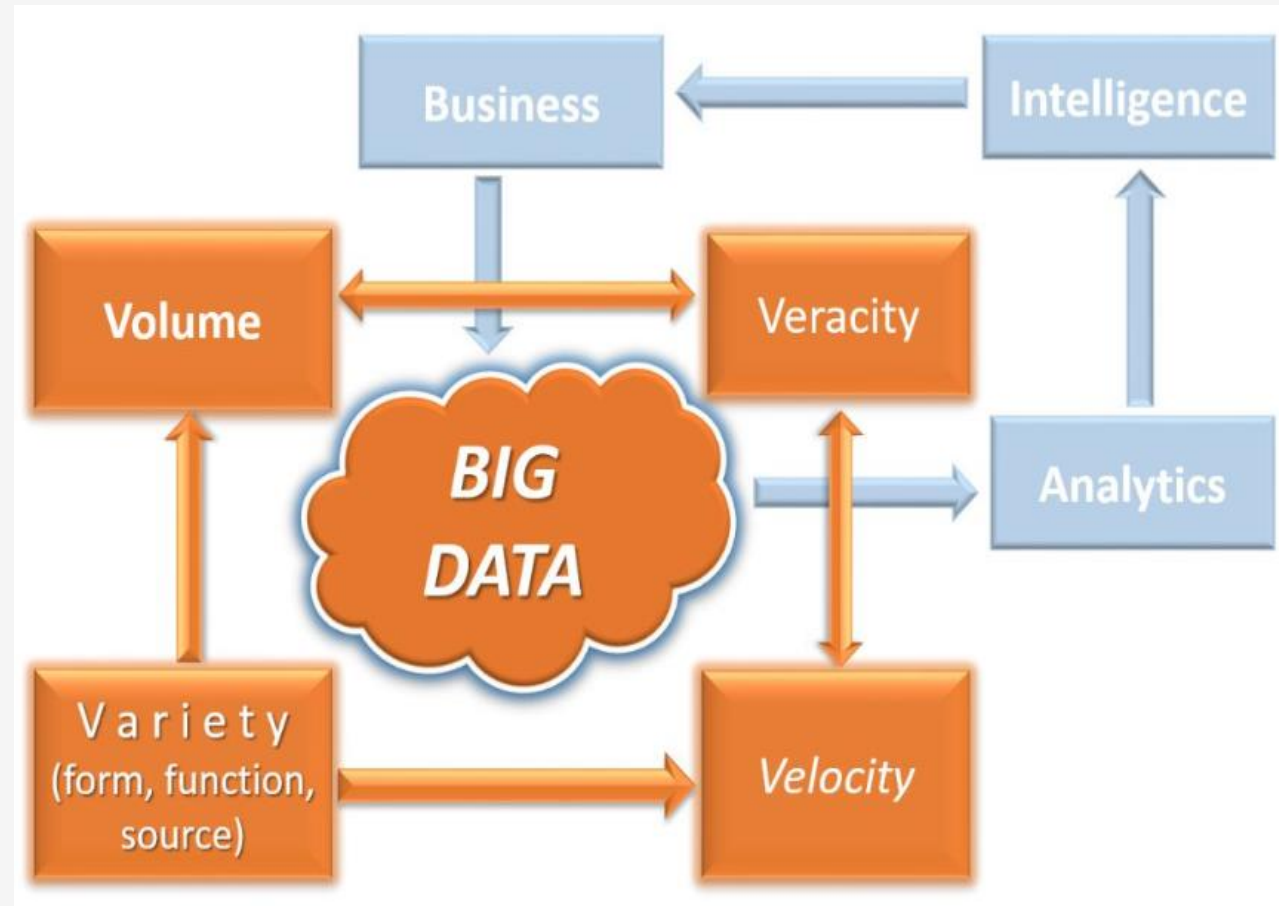
What is Big Data?

Volume

Veracity

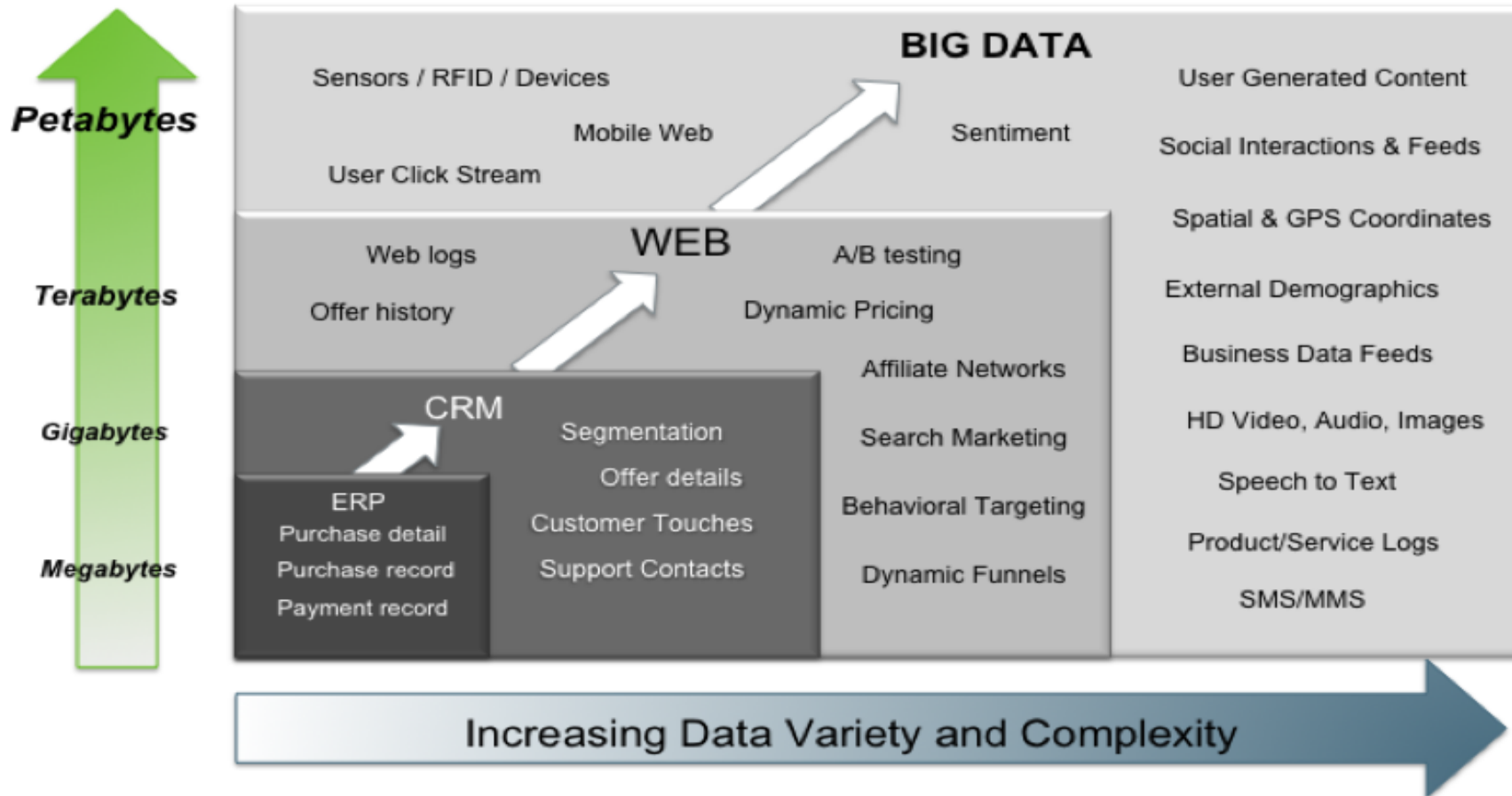
Variety

Velocity



Big Data

Big Data = Transactions + Interactions + Observations



Source: Contents of above graphic created in partnership with Teradata, Inc.

Films on Big Data

Digital nation – PBS (90 min)

<http://video.pbs.org/video/1402987791/ig>

BBC Documentary on Big Data (60 min)

<https://www.youtube.com/watch?v=tl-xxEur07Q>

Big Data Solution Technology

- Non-relational data structures
 - Hadoop and open source stack
 - Google BigFile, Semantic web, etc.
- Massively parallel computing
 - Map-Reduce algorithms
 - Spark
- Unstructured Information Management Architecture (UIMA)
 - The 'sauce' behind IBM' Watson system
 - Natural language processing

Big Data perspectives

- Economist magazine
<http://www.economist.com/blogs/dailychart/2011/11/big-data-0>
- Churchill club seminar
http://www.youtube.com/watch?v=KD_g6byn83s
- BBC Horizon on Big Data
- <https://www.youtube.com/watch?v=tl-xxEur07Q>

The What, Why and How

- Pay as you go
- IaaS (infrastructure as a service)
- SaaS (software as a service)
- PaaS (platform as a service)

Three Major Players

- Amazon AWS
- Microsoft Azure
- Google

Lower Cost of Ownership

Hardware scaled up and down dynamically

Outsource the whole IT department

Disaster recovery, Cybersecurity and Audits

Amazon AWS vs Microsoft Azure

- S3 - for storage
- EC2 - virtual computers to run computer applications
- EMR
- Amazon Glue
- RDS - for database
- Notebook
- Azure ML Studio