

Introduction To Big Data

What is Big Data ?

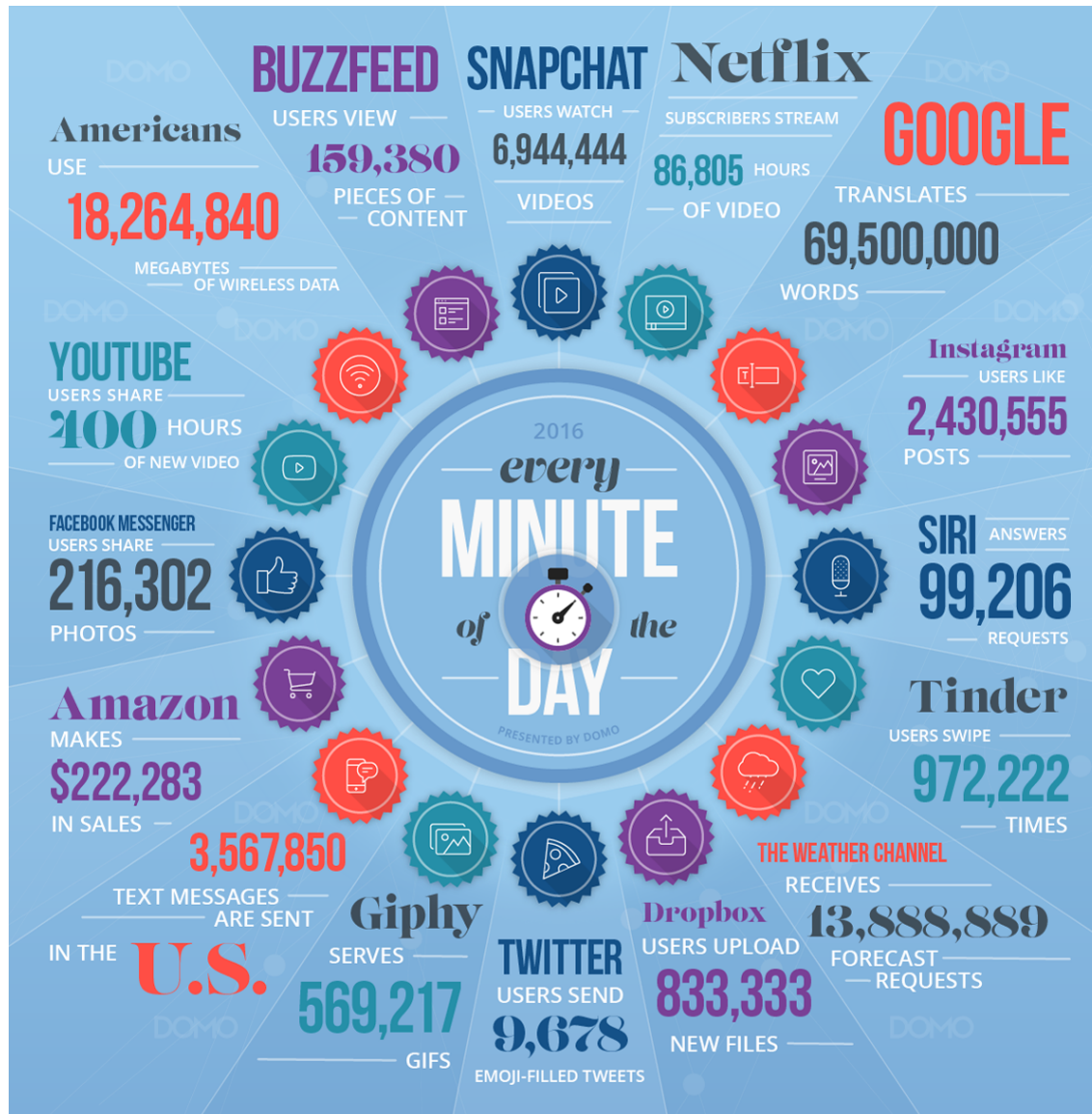
“ Big data is data that exceeds the processing capacity of conventional database systems.

**The data is too big, moves too fast,
or doesn't fit the structures of your database architectures.**

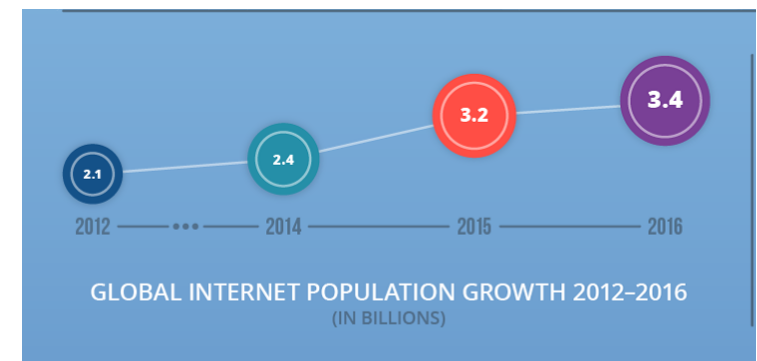
**To gain value from this data,
you must choose an alternative way to process it. ”**

Big Data Now: O'Reilly Media

Why Big Data ?



DATA NEVER SLEEP 4.0



Facebook Usage Statistics

June 2014

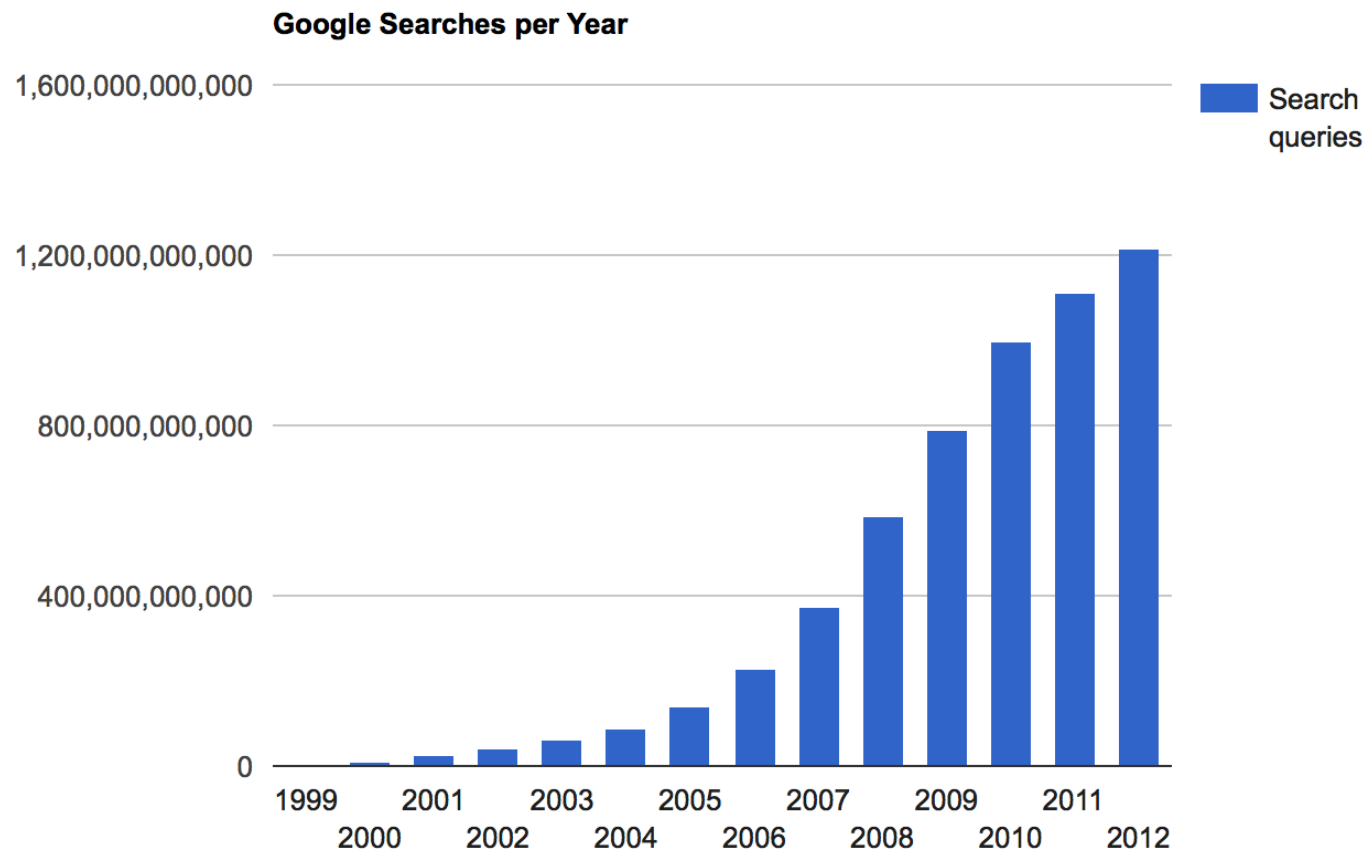
- 829 million daily active users on average
- 654 million mobile daily active users on average
- 1.32 billion monthly active users
- 1.07 billion mobile monthly active users
- Approximately 81.7% of our daily active users are outside the US and Canada

June 2016

- 1.13 billion daily active users on average for June 2016
- 1.03 billion mobile daily active users on average for June 2016
- 1.71 billion monthly active users as of June 30, 2016
- 1.57 billion mobile monthly active users as of June 30, 2016
- Approximately 84.5% of our daily active users are outside the US and Canada

Google Usage Statistics

Google now processes over **40,000 search** queries every second which translates to over **3.5 billion searches per day** and **1.2 trillion searches per year** worldwide



in 1 second, each and every second, there are...



7,370 Tweets sent in 1 second



745 Instagram photos uploaded in 1 second



56,645 Google searches in 1 second



132,881 YouTube videos viewed in 1 second



2,529,971 Emails sent in 1 second



38,625 GB of Internet traffic in 1 second

Three Characteristics of Big Data

Volume

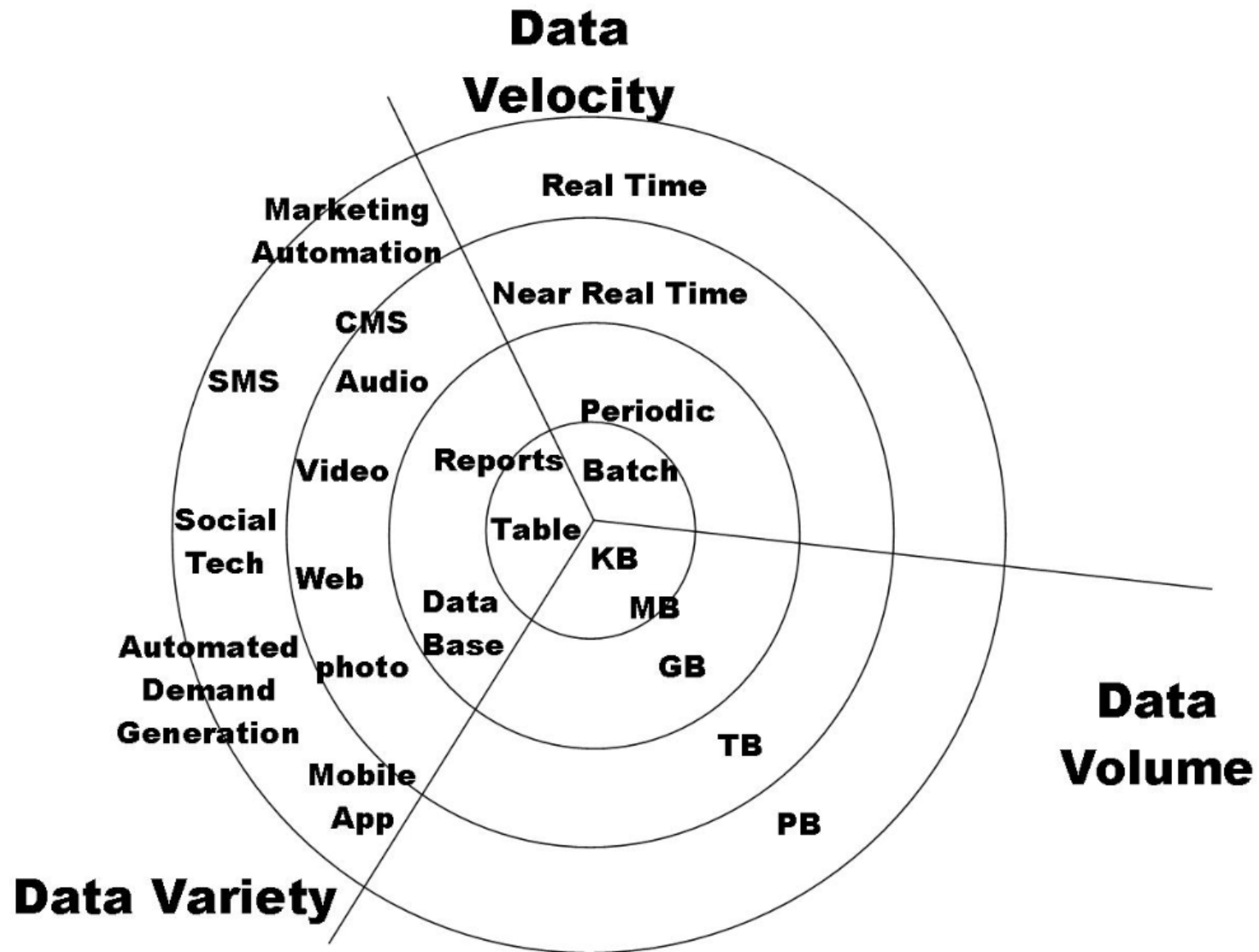
- Volumes of data are larger than those conventional relational database infrastructures can cope with

Velocity

- Rate at which data flows in is much faster.
- Mobile event and interaction by users.
- Video, image , audio from users

Variety

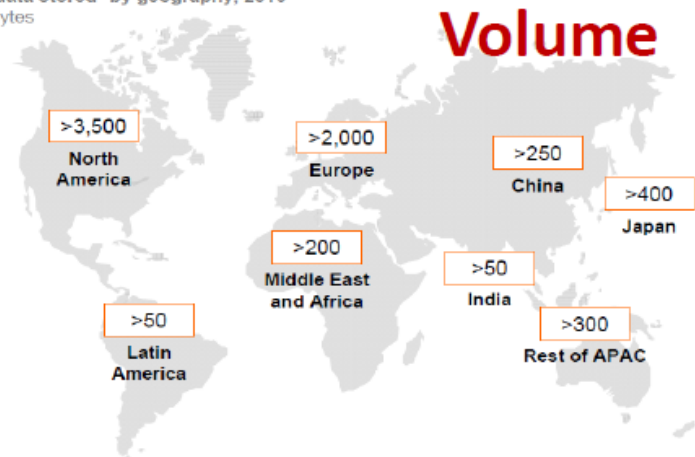
- the source data is diverse, and doesn't fall into neat relational structures eg. text from social networks, image data, a raw feed directly from a sensor source.



Big Data = Volume, Variety and Velocity (3Vs)

Amount of new data stored varies across geography

New data stored¹ by geography, 2010
Petabytes



¹ New data stored defined as the amount of available storage used in a given year; see appendix for more on the definition and assumptions.

SOURCE: IDC storage reports; McKinsey Global Institute analysis

Velocity

30 billion pieces of content are shared on Facebook every month.



4 billion hours of video are watched on YouTube each month



400 Million Tweets are sent per day
200M monthly active users



Source: IRM

Variety



People to Machine



People to People

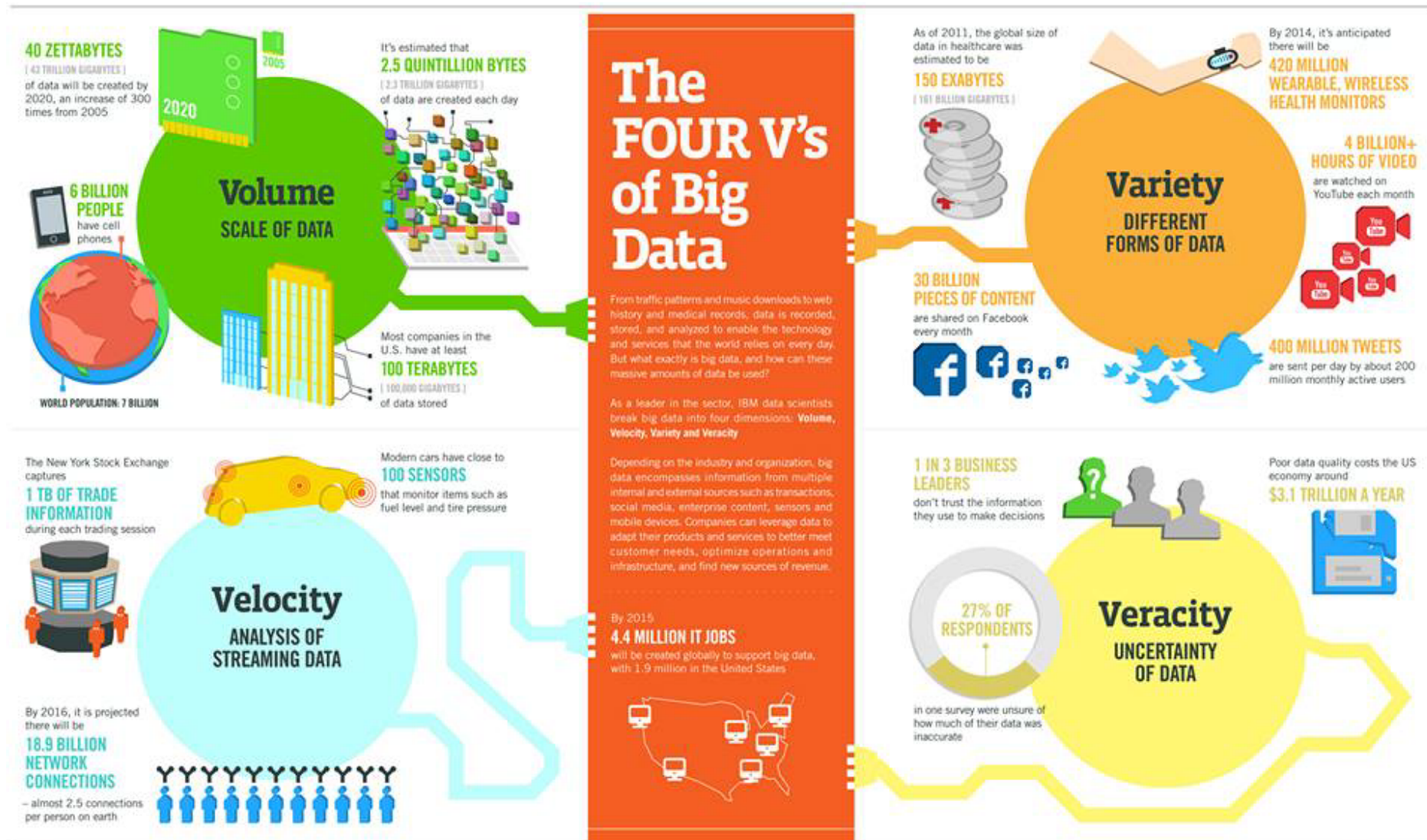
Machine to Machine



Scale

1000 kilobytes = 1 Megabyte
1000 Megabytes = 1 Gigabyte
1000 Gigabytes = 1 Terabyte
1000 Terabytes = 1 Petabyte
1000 Petabytes = 1 Exabyte
1000 Exabytes = 1 Zettabyte
1000 Zettabytes = 1 Yottabyte
1000 Yottabytes = 1 Bronobyte
1000 Bronobytes = 1 Geopbyte

4Vs of Big Data



Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, GAS

