

Characteristics of Big Data:

Volume

After this video you will be able to..

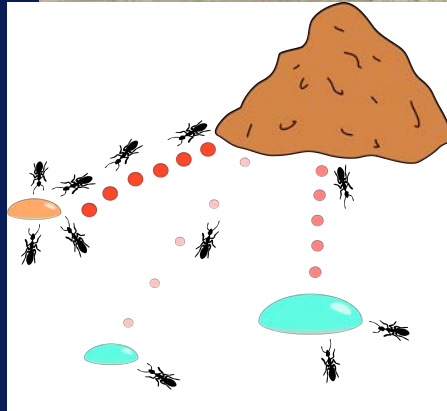
- Describe what volume of big data means and why you should care about it
- Explain why data volume is not just about storage

Volume = Size

Volume = Size



Volume = Size



Every minute...



204 Million emails

Every minute...



204 Million emails

200,000 photos

facebook

1.8 Million likes

Every minute...



204 Million emails

200,000 photos

facebook

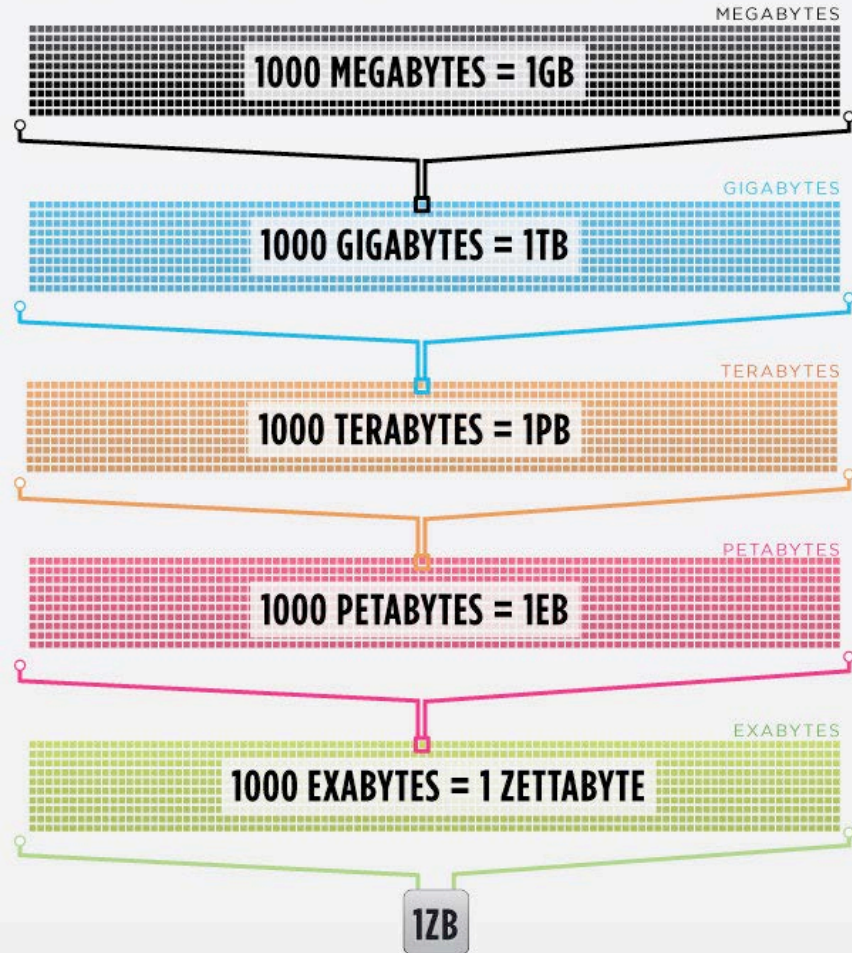
1.8 Million likes



1.3 Million video views

72 hours of video uploads

But how much data are we talking about?



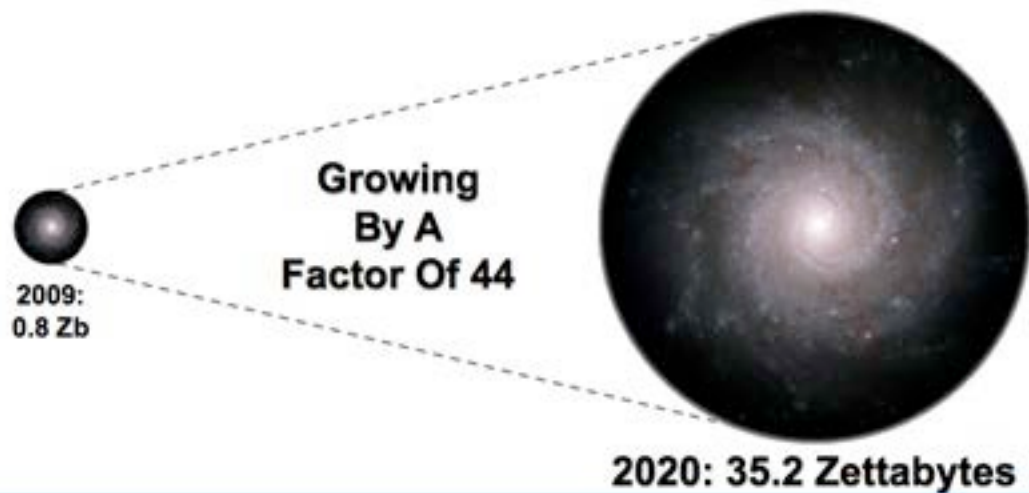
100 MBs ~= couple of volumes of Encyclopedias

A DVD ~= 5 GBs

1 TB ~= 300 hours of good quality video

LHC ~= 15 PBs a year

The Digital Universe 2009-2020



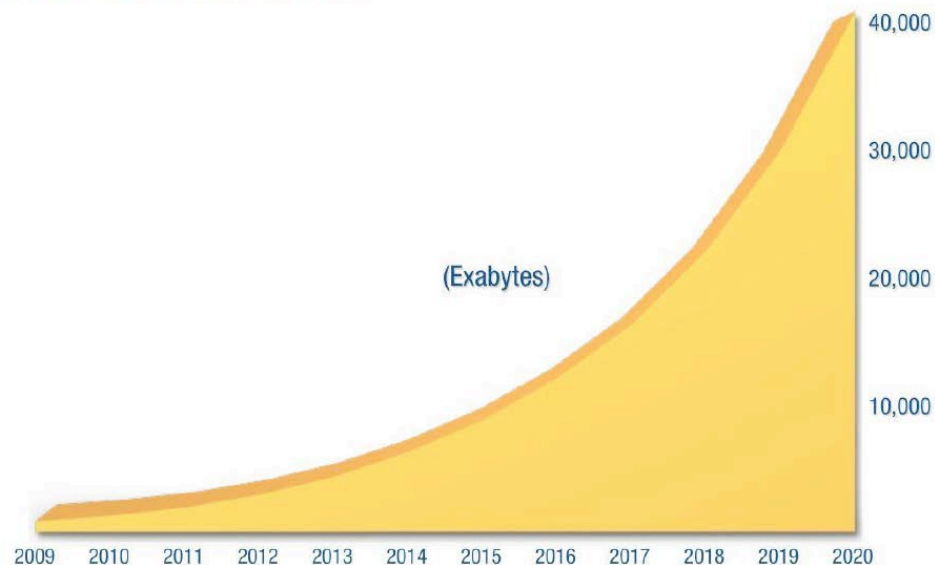
EMC²
Data Growth Study

Source: IDC Digital Universe Study, sponsored by EMC, May 2010

© 2010 EMC Corporation. All rights reserved.

Exponential data growth!

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020



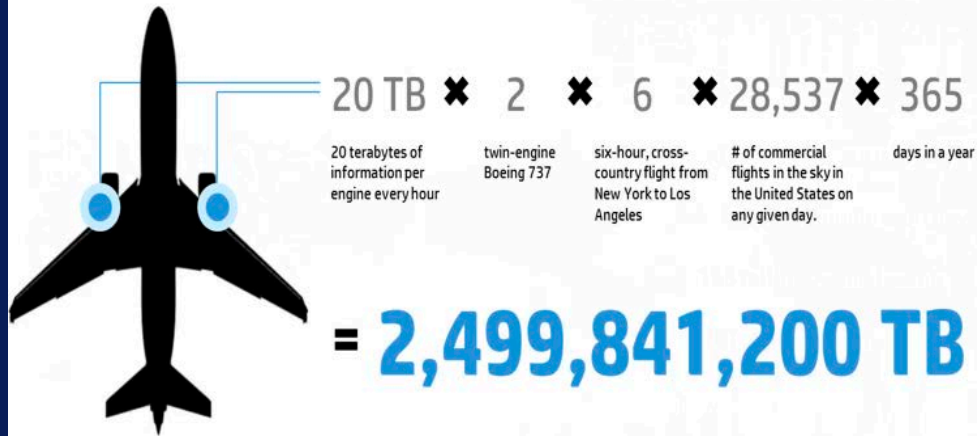
This IDC graph predicts exponential growth of data from around 3 zettabytes in 2013 to approximately 40 zettabytes by 2020. An exabyte equals 1,000,000,000,000,000 bytes and 1,000 exabytes equals one zettabyte. Source: IDC's Digital Universe Study, December 2012, <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>.

Relevance of Volume for Us?



More data = Better safety

Sensor data from a cross-country flight

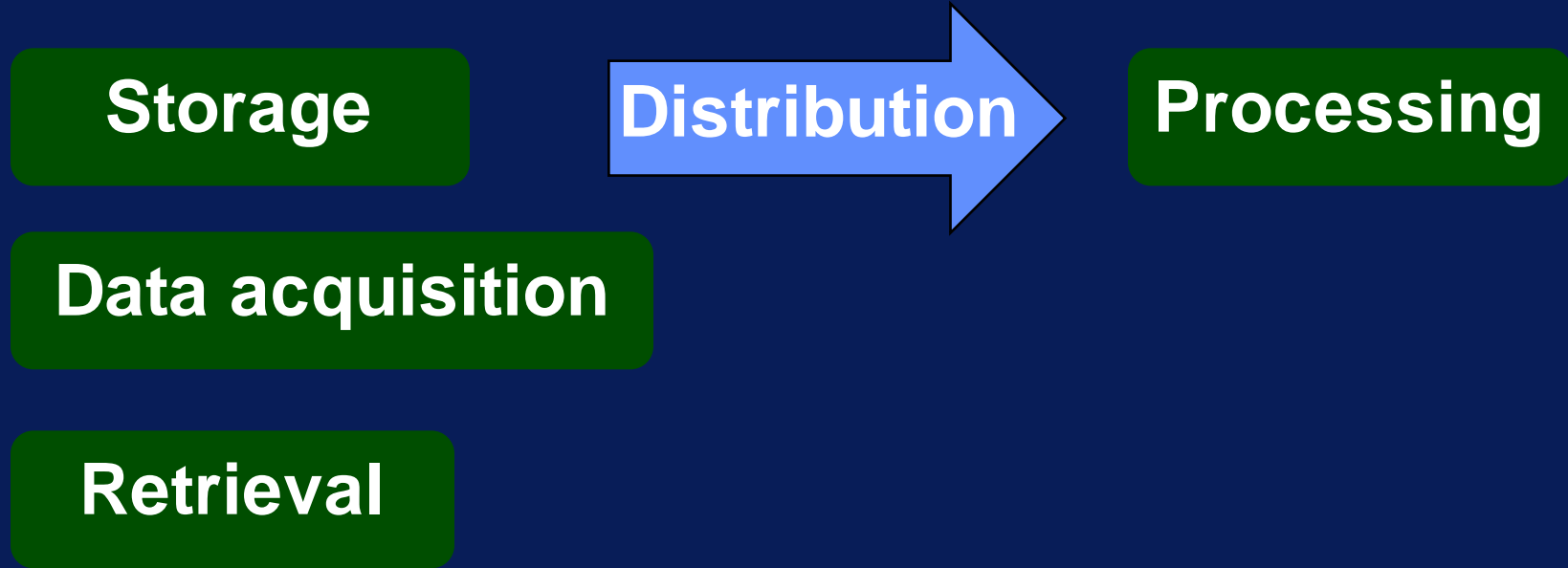


Volume



Business Insight

Challenges: Storage and more...



Processing Big Data



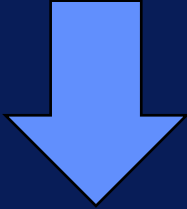
The diagram consists of three vertical green arrows on a dark blue background. The leftmost arrow points upwards and contains the word 'Volume'. The middle arrow points downwards and contains the word 'Performance'. The rightmost arrow points upwards and contains the word 'Cost'. All text is in white, bold, sans-serif font.

Volume

Performance

Cost

Volume = Size



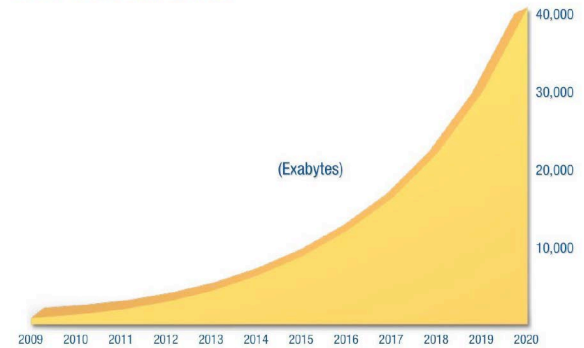
Challenges

Storage

Access

Processing

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020



This IDC graph predicts exponential growth of data from around 3 zettabytes in 2013 to approximately 40 zettabytes by 2020. An exabyte equals 1,000,000,000,000,000 bytes and 1,000 exabytes equals one zettabyte. Source: IDC's Digital Universe Study, December 2012, <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>.