



Big Data Research & Analytics Lab

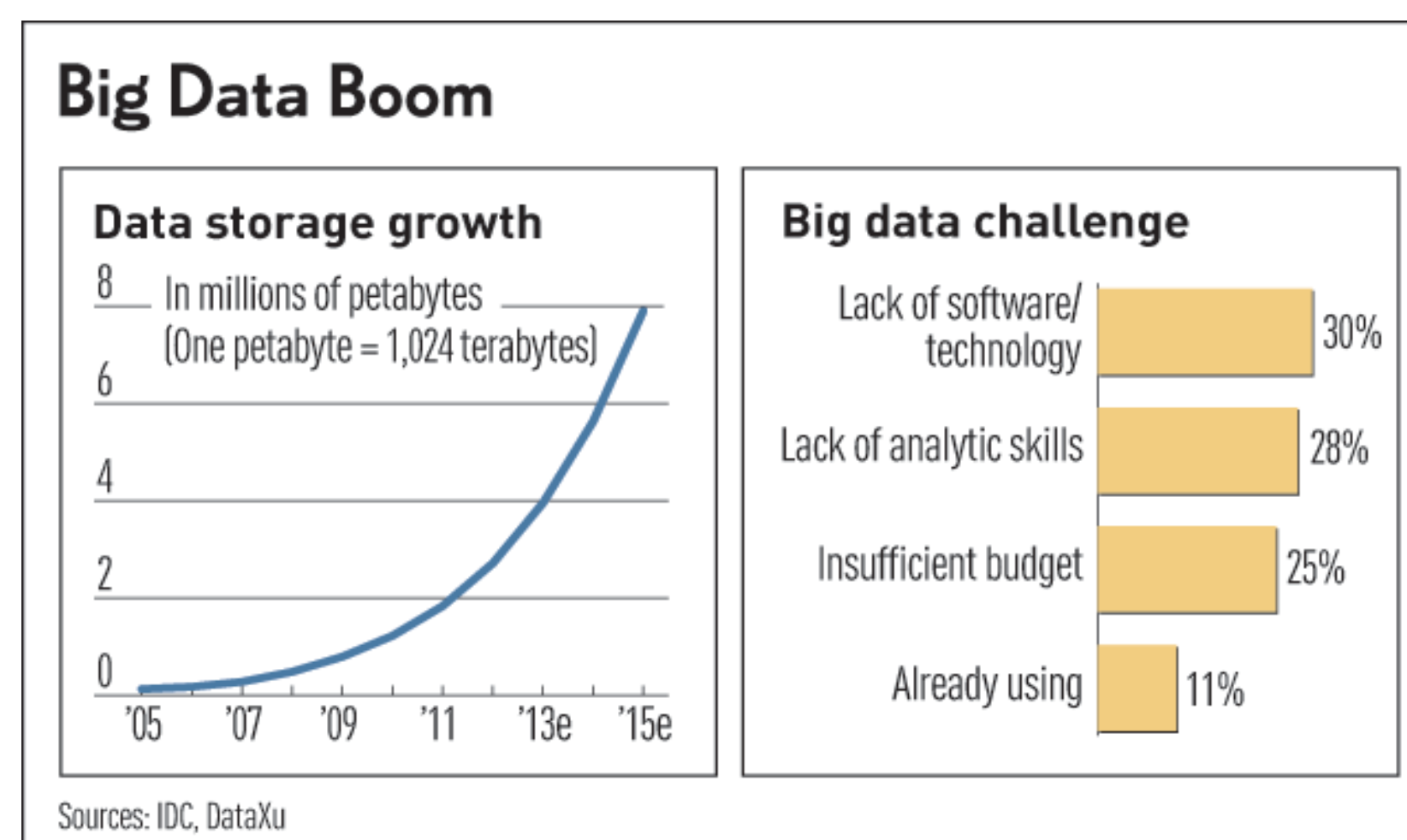
Dept. of Electrical and Computer Engineering



What is Big Data?

No single standard definition...

“**Big Data**” is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it...



The Bottleneck is in technology

New architecture, algorithms, techniques are needed

Also in technical skills

Experts in using the new technology and dealing with big data are needed

Why COTS components?

- **COTS – Commercial Off-The-Shelf**
- Cheapest Commercial Big Data system will incur an initial cost in excess of \$500K
- UAB has a number of computer systems which are refreshed every couple of years and can be obtained at little cost
- DIY test-bed can be setup with a small investment (<\$2K) in hardware and by using open-source software
- Test-bed can be scaled up quickly as more systems comes online

Infrastructure setup

Hardware

- **Head node (2):** 16 core, 72 GB RAM, 6 TB HDD, Gigabit Ethernet, Nvidia Quadro GPU (4 screen visualization station)
- **OpenMP nodes (5):** 16 core, 64 GB RAM, 2 TB HDD, Gigabit Ethernet
- **Processing nodes (8):** 2 -16 core, 8 - 32 GB RAM, 2TB HDD, Gigabit Ethernet
- **Storage Node (1):** 4 core, 4 GB RAM, 3TBx4 = 12TB, Dual Gigabit Ethernet
- **Networking (2):** 24 port Gigabit Ethernet switch

Software

- **Platform:** Unix
- **Big Data Software:** Hadoop, Pig, Spark
- **Other Analytics Software:** MATLAB, R, AFNI
- **Schedulers:** SGE, Condor, Yarn
- **Monitoring:** Nagios

System Status/Results

- Theoretical throughput ~ 2 Tflops
- 128 cores, 300 GB RAM
- 50TB processing scratch space (online)
- 50TB storage space (online)
- 48 Gbps network throughput (online – tested with iperf ~960 Mbps, node to node)
- Software stack – Testing with AFNI, Hadoop

Current Projects

Brain Mapping for the prediction of Parkinson’s Disease (Funded)

Partners – Dept. of Neurology (SOM), UAB IT Research Computing

Grants Funded – 4

MJFF - \$152K + 195K, 15 months

NIH - \$929K, 5 years

New Collaborations – ORNL, NICS

South Big Data Hubs - (Funded)

Partners – Georgia Tech, RENCI –UNC Chapel Hill

Grants Funded - 1

Contact Us

Thomas Anthony
Director
Big Data Research & Analytics Lab
tanthony@uab.edu

Murat Tanik, PhD
Professor & Chair
Dept. of Electrical Engineering
mtanik@uab.edu

“We swim in a sea of data
... and the sea level is rising
rapidly.”

Pew Research Center’s Internet & American Life
Project - July 2012