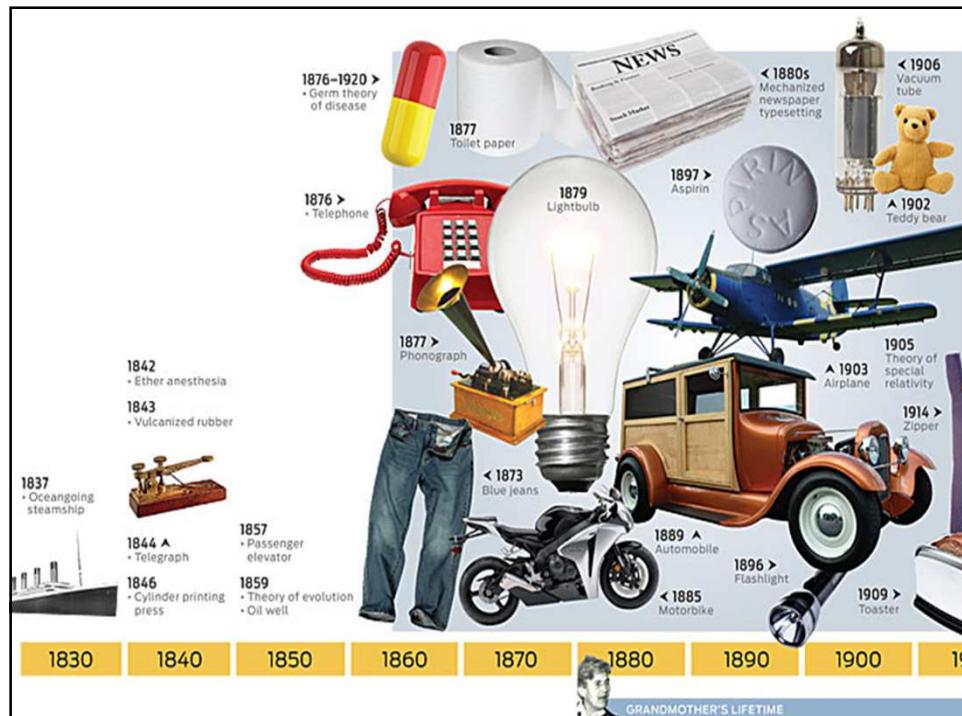


## CSC554 CL#9 Big Data Overview

- ◆ The world is changing...
- ◆ From data growth to data explosion
- ◆ What is big data?
- ◆ Challenges
- ◆ Big Data is Scaling BI and Analytic
- ◆ Big Data Ecosystem
- ◆ Solutions
  - Products
  - Vendors
- ◆ Examples of big data solutions
- ◆ Big data and cloud



## 19<sup>th</sup> Century: Industrial Revolution

**19th Century Industrial Revolution**

A word cloud centered around the Industrial Revolution, with terms like 'century', 'Industrial Revolution', 'manufacturing', 'power', 'capitalist development', and 'historians' prominently displayed.

## 20<sup>th</sup> Century: Technology Revolution

**Technology**

A word cloud centered around technology, with terms like 'technology', 'open source', 'tinkering', 'engaged', 'sense of community', and 'tension' prominently displayed.

## 21<sup>th</sup> Century: Digital Revolution (*Information Explosion*)



## Total Data Crossing the Internet

150,000,000,000,000,000 Bytes\*  
Through 2009

175,000,000,000,000,000 Bytes\*  
In 2010 Alone

\*Exabytes

Source: [http://www.idc.com/assets/doc\\_content\\_elements/networking\\_solutions/service\\_provider/visual\\_networking\\_ip\\_traffic\\_chart.html](http://www.idc.com/assets/doc_content_elements/networking_solutions/service_provider/visual_networking_ip_traffic_chart.html)



Page 10



<http://www.smartplanet.com/business/blog/business-brains/data-explosion-enough-to-fill-dvds-stretching-to-the-moon-and-back/7010/>

## Data explosion: enough to fill DVDs stretching to the moon and back

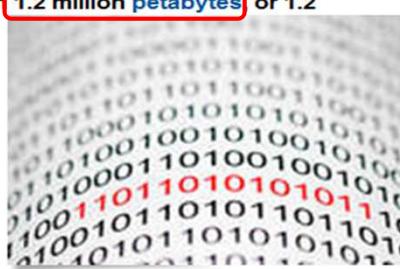
By Joe McKendrick | May 14, 2010 | [5 Comments](#)



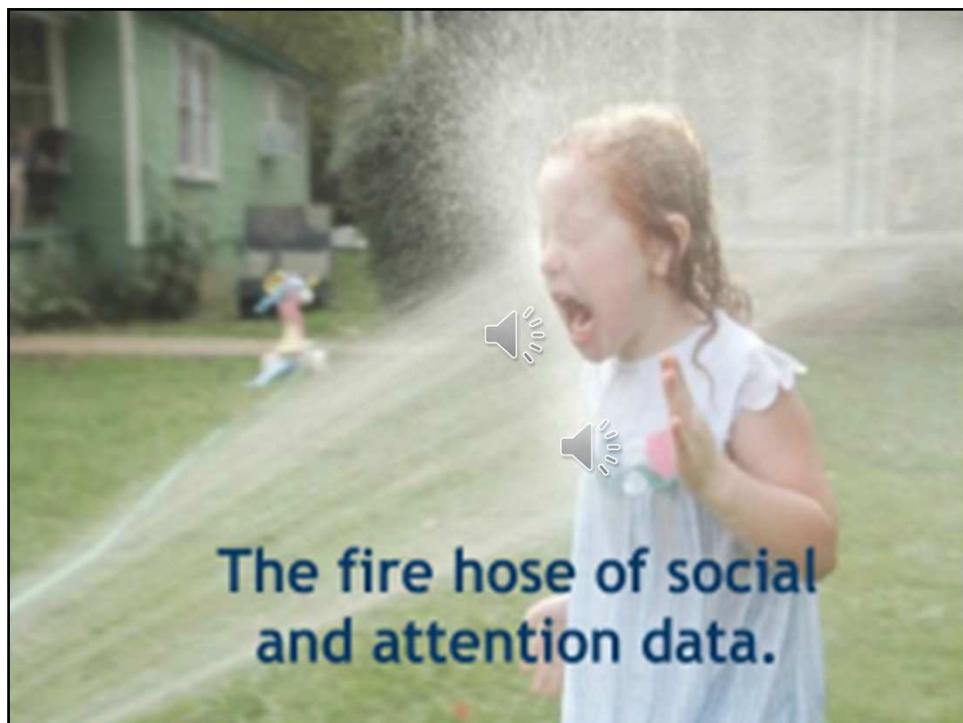
IDC has just released estimates that this year, the Digital Universe — meaning every electronically stored piece of data or file out there — will reach **1.2 million petabytes** or **1.2 zettabytes**, this year.

That's up from a measly 800,000 petabytes in 2009.

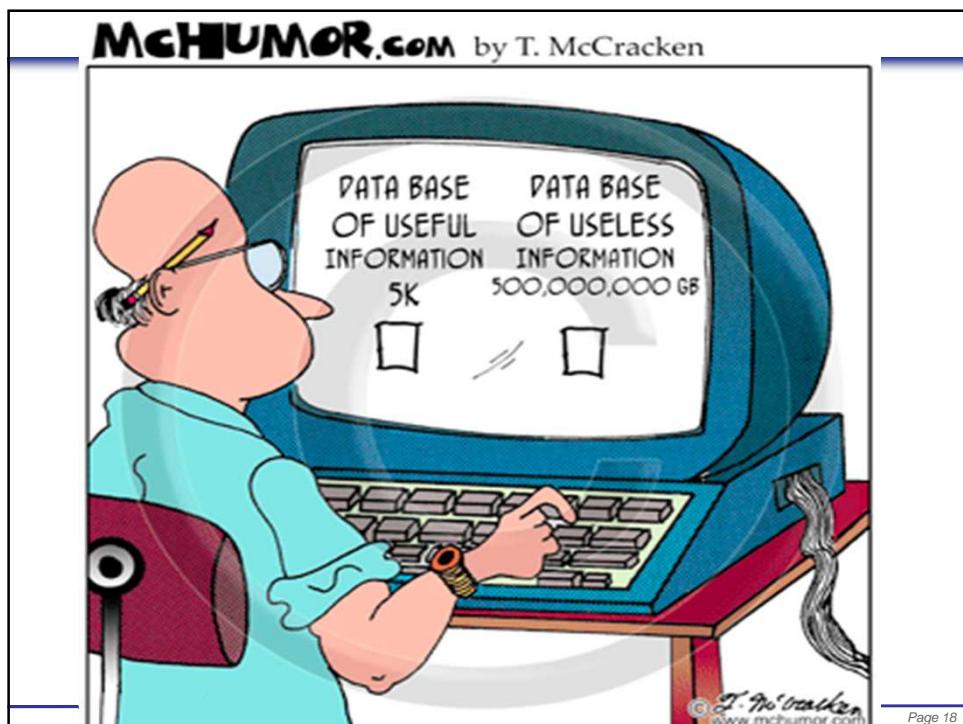
Despite an economic slowdown or meltdown in some quarters, the total amount of data still grew by 62%, IDC reports. To illustrate how much data this is, write John Gantz and David Reinsel, authors of the latest IDC report on the size of the Digital Universe, "picture a stack of DVDs reaching from the earth to the moon and back. (About 240,000 miles each way.)"



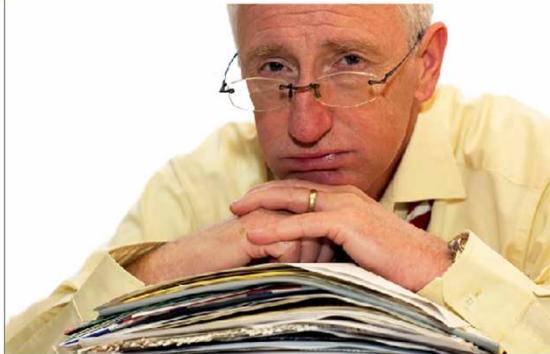
Page 12



The fire hose of social  
and attention data.



## Data, Data Everywhere!



**Are you data rich & information poor?**

In the past 50+ years, we learned **how to “store” information**  
***... and we keep everything FOREVER!***

Today and in the future, we need to figure out **how to “find”**  
**information.**

Also, we have to think about **how to “delete” unwanted information**  
***... and ONLY keep what we really NEED!***

Page 19

## It's not just data, it's your business

The Center for Information  
Technology Policy Presents

# BIG DATA

Public Policy and the  
Exploding Digital Corpus

#### A One Day Conference

November 30, 2010, 8am–5pm  
Friend Center Convocation Room

The body of digital information held by various entities is both staggering and constantly expanding. Each day we hear new reports of newly digitized “dark” archives, enhanced digital tracing techniques, data privacy breaches, and aggregated data sets. At the same time, much historically important information goes unrecorded – at least in any usable or enduring digital form. How do we reconcile the many different constituencies, technologies, uses, and norms into sensible policy? This conference will gather leading experts from a variety of domains to discuss the challenges of “big data” and the attendant policy considerations.

For more information, visit:  
<http://http://princeton.edu/events/big-data/>



Page 20

## The world is changing and becoming more...



## What will 2013 bring?

The last few years have seen a sea of changes in business intelligence (BI). The proliferation of data and advances in technologies are pushing the pace of innovation. Here are 10 trends to watch for in the months ahead.

- ◆ The **speed**, **scope**, and **scale** of change and the challenges themselves are immense.
- ◆ Changes come together – *not always be sequential*.
- ◆ Right now is a moment of **big decisions** for key decision makers.
- ◆ The challenge for leaders is trying to get the right answers.  
In order to get there ...

Leaders must first ask the right questions.

## Big thoughts on Big Data

### Put data to work

- ◆ 2011 was the beginning of the big data ...
- ◆ It moved rapidly in 2012
- ◆ Big data will only get **bigger** in 2013.
- ◆ Based on conversation and interview with SMEs, venture capitalists, and vendor execs. Here are some of their thoughts on how the world of big data and cloud computing infrastructure will shake out next year.

Information Discovery

# 10

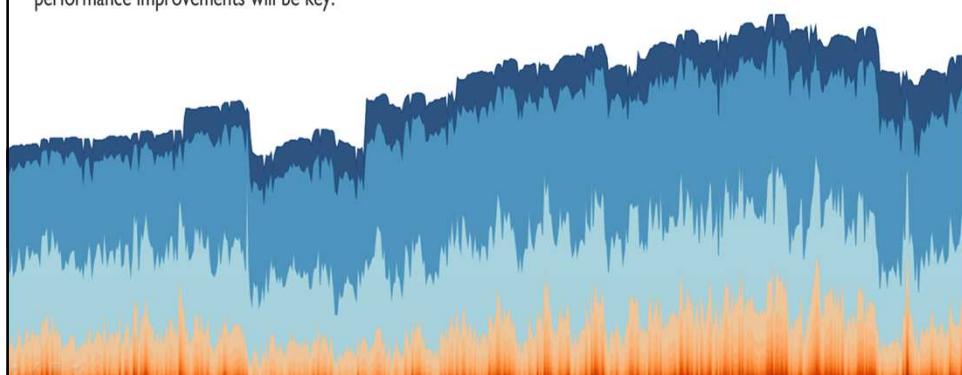
Business Intelligence Trends

Page 46

1

### Big data gets even bigger

The overarching trend is that data is growing, growing, growing. In 2012, more organizations will actually use that data, rather than just collecting and storing it. This will put pressure on vendors to deliver solutions that can really work with big data. Ad-hoc solutions, including in-memory analytics, direct access to column store databases, and performance improvements will be key.



Page 47

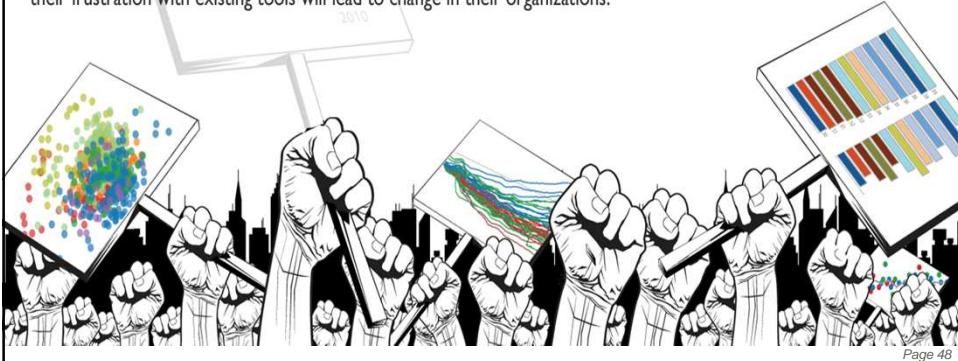
# 2

## BYOD Occupy IT

### Consumerization of IT

#### Self-reliance is the new self-service

The idea of self-service BI where IT opens up a small menu of capabilities for employees is over. Giving employees an environment where they can get the data to answer questions on their schedules will become the norm. The consumerization of enterprise software is part of the story here, but the real driver is the increasing pace of business. Business users are coming to expect that they can modify and create reports as needed. When they can't, their frustration with existing tools will lead to change in their organizations.



# 3

#### The “Consumerization of Enterprise Software” accelerates

You've already heard it: consumer software is faster, easier, and often more sophisticated than enterprise software. Why? Consumer software typically puts more thought into design and experience. People want their business analytics software to work as easily and as smoothly as their software at home. This trend is going to speed up and IT needs to be ready. Traditional enterprise software deployments beware.



Page 49

# 4

## Mobile BI goes mainstream

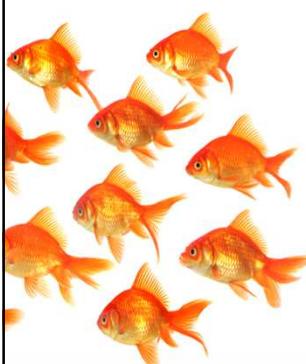
Apple claims that 92% of the Fortune 500 will be testing or deploying iPads in the 2012 timeframe. Companies are moving from the experimentation stage with mobile into production, IT-supported deployments. And the tablet finally offers a form factor that makes sense for BI. Expect to be a laggard if you're not using mobile for BI by the end of 2012.



Page 50

# 5

## Some companies start to get comfortable with social BI



Social platforms like Salesforce's Chatter already disseminate enterprise information in a more consumable, social way. Things like rapid sharing, fewer emails, and feedback in real time are starting to proliferate. Although it may take several years before most organizations are willing to support more experiential, interconnected communication with their data assets, we see social BI as an important nascent trend in 2012.



Page 51

# 6

## Companies explore the BI cloud

Lower TCO, easier setup—these factors will drive some companies to the cloud for business intelligence. In 2012 primarily small- and medium-sized businesses that don't have a lot of IT resources will adopt cloud based BI technology. Throughout the next couple years, BI offerings will also get more diverse and more mature.

Page 52

# 7

## Most jobs will require analytical skills ...leading to talent shortages

The McKinsey Global Institute released a study in 2011 predicting that by 2018, the US would face a shortage of up to "1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions." Despite the soft economy, organizations will see a tight market for analytical skills. And an increased need for smarter tools to make up the difference.



Page 53

# 8

## BI projects flourish under aligned IT & business

Companies with enlightened IT staffs and business leaders who recognize the impact of IT are achieving spectacular results from BI projects. In these organizations, IT no longer tries to squash business-driven BI projects. Instead they're asking how they can support and improve them. These organizations will continue to see gains from their alignment efforts. But in less enlightened environments, BI still represents a battleground for control.



# 9

## Interactive data visualization becomes a requirement

Graphically interesting data visualized on websites and in blogs will make business users expect that capability inside their organizations. And with the trend of bigger data, interactive data visualization will become a critical tool. Reports filled with endless rows of numbers or static, boring charts that take hours to understand and weeks to change won't cut it anymore.

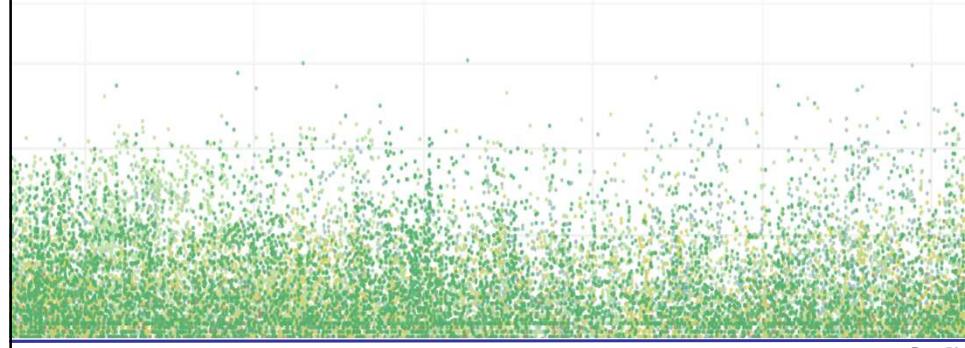


Page 55

# 10

## Hadoop gathers momentum – unstructured data isn't going anywhere

Hadoop and similar technologies are becoming one of the best ways to deal with massive amounts of data especially if it's unstructured. While still a relatively young technology, vendors like Cloudera are pushing Hadoop forward. It may not be mainstream in 2012, but we do see it getting much closer.



Page 56

### Big Data Buzz

**"The promise of big data"**

*Intelligent Utility - 8/28/11*

**"Big data, analytics get even bigger, hotter in 2012"**

*InfoWorld - 12/30/11*

**"Are you ready for the era of big data?"**

*McKinsey Quarterly - 11/11*

**"Health care is next frontier for big data"**

*Wall Street Journal - 1/19/12*

**"Big data: science's microscope of the 21st century"**

*Business Week - 11/8/11*

**"Decisions, decisions...will big data have big impact?"**

*Financial Times - 1/24/12*



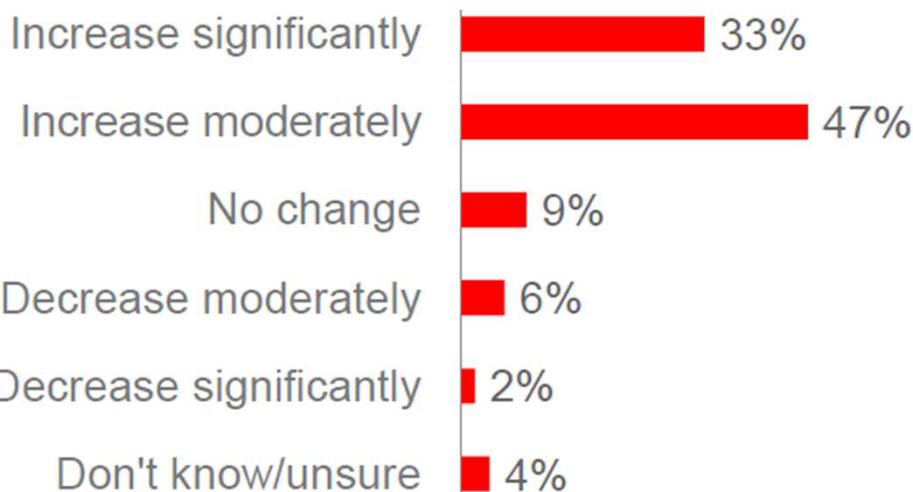
<http://www.youtube.com/watch?v=eEpxNohIRKI>

EMC Big Data



## Growth in Enterprise Unstructured Data

How will the amount of unstructured data change over next 3 years?

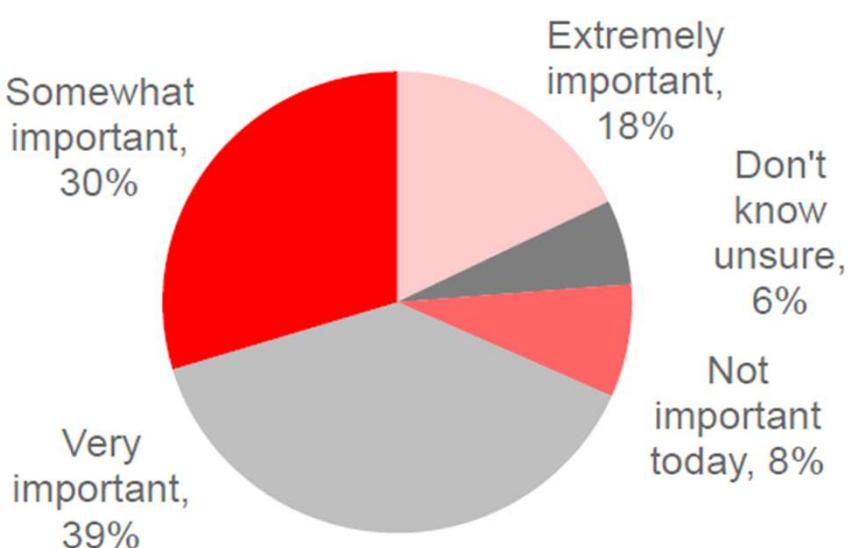


Source: Unisphere Research, 2011

Page 59

## Value of Unstructured Data

How important is unstructured data to your business?



Page 60

## Why Is Big Data Important?

US HEALTH CARE	MANUFACTURING	GLOBAL PERSONAL LOCATION DATA	EUROPE PUBLIC SECTOR ADMIN	US RETAIL
Increase industry value per year by <b>\$300 B</b>	Decrease dev., assembly costs by <b>-50%</b>	Increase service provider revenue by <b>\$100 B</b>	Increase industry value per year by <b>€250 B</b>	Increase net margin by <b>60+%</b>

“In a big data world, a competitor that fails to sufficiently develop its capabilities will be left behind.”

*McKinsey Global Institute*

Page 61

## □ What is big data?

**Big data** are datasets that grow so large that they become awkward to work with using on-hand database management tools.

Difficulties include capture, storage, search, sharing, analytics, and visualizing.

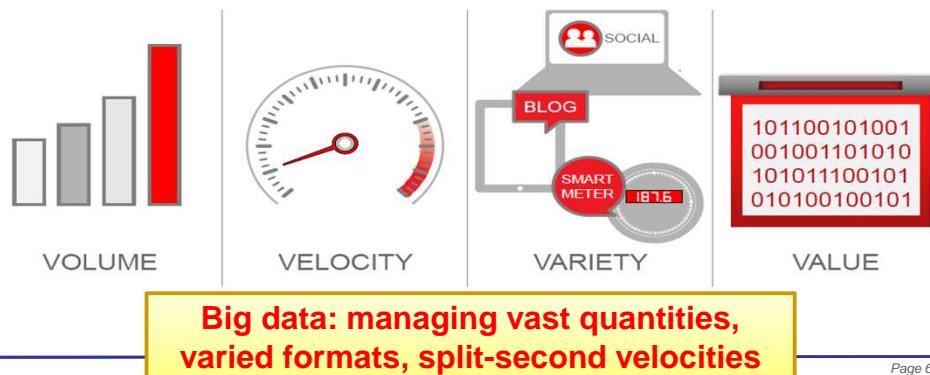
Source: Wikipedia

Page 62

## What Makes it Big Data?

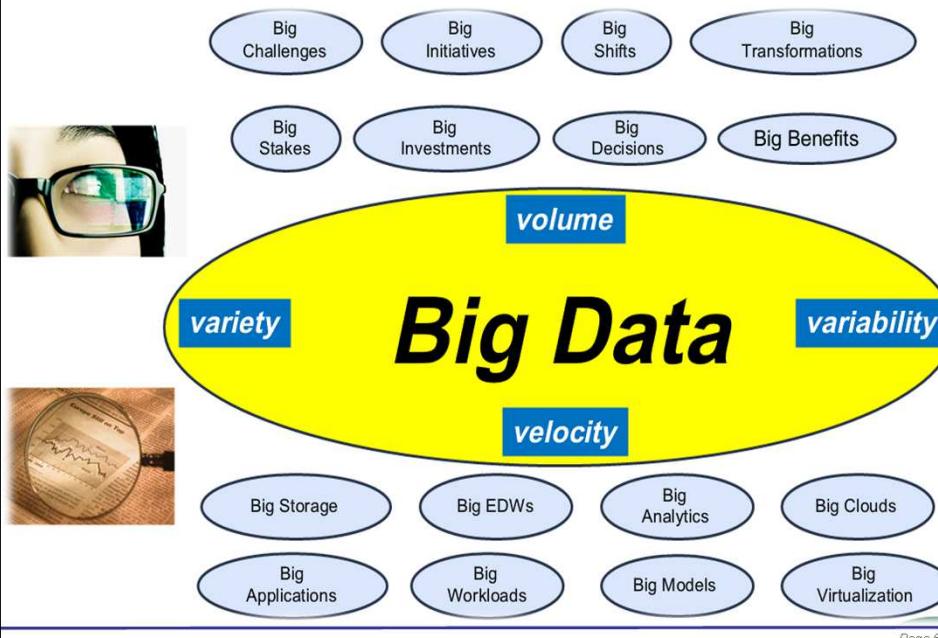
Data growth challenges as being defined three-dimensional:

- ◆ **Volume** – Big data comes in one size: **Large**. Enterprises are flooded with data, easily amassing terabytes and even petabytes of information.
- ◆ **Velocity** – Often **time-sensitive**, big data must be used as it is streaming in to the enterprise in order to maximize its value to the business.
- ◆ **Variety** – Big data extends beyond structured data, including **unstructured** data of all varieties: text, audio, video, click streams, log files and more.



Page 65

## Big Data: A Big Bursting Paradigm!



Page 66

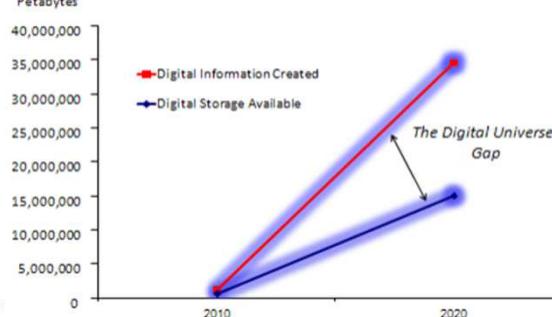
## History of Big Data

- ◆ Big data is data that is too large to process using traditional methods.
- ◆ It originated with Web search companies who had the problem of querying very large distributed aggregations of loosely-structured data.
- ◆ Google developed MapReduce to support distributed computing on large data sets on computer clusters.
- ◆ Inspired by Google's MapReduce and Google File System (GFS) papers, Doug Cutting created Hadoop while he was at Yahoo!, and named it after his son's stuffed elephant.
- ◆ Hadoop is an Apache project, written in Java and being built and used by a global community of contributors. Yahoo! has been the largest contributor to the project and uses Hadoop extensively across its businesses on 38,000 nodes.
- ◆ Doug Cutting joined Cloudera, a commercial Hadoop company that develops, packages, supports and distributes Hadoop (similar to the Red Hat model for Linux), making it accessible to Enterprise IT.

Page 67

## Big Data Characteristics

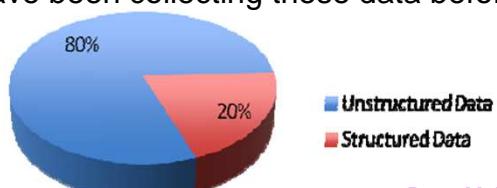
- ◆ While much of the digital content we create is simply not that important (magazines, newspapers, telephone conversations, receipts, bad pictures, etc.), the amount of data that does require permanent or longer-term preservation for a multitude of reasons is increasing exponentially.
- ◆ But as we peer into the future, we see the greatest challenges are related not to how to store the information we want to keep, but rather to:
  - Reducing the cost to store all of this content
  - Reducing the risk (*and even greater cost*) of losing all of this content
  - Extracting all of the value out of the content that we save



Page 72

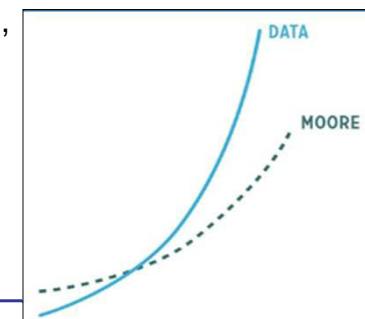
## Data Types

- ◆ Enterprise tools focus on **20%** of **structured** (*transactional*) data.
- ◆ About **80%** of the data collected by institutions is **unstructured**.
- ◆ We may have been collecting these data before, but could not process.



- ◆ Pure-play data ventures (FB, LinkedIn, Google) focus on monetizing 80% of organic, semi-structured, variable, time-sensitive, social ... data

Data Volume vs. Moore's Law



## Two types of big data



### ◆ Data in movement - streams

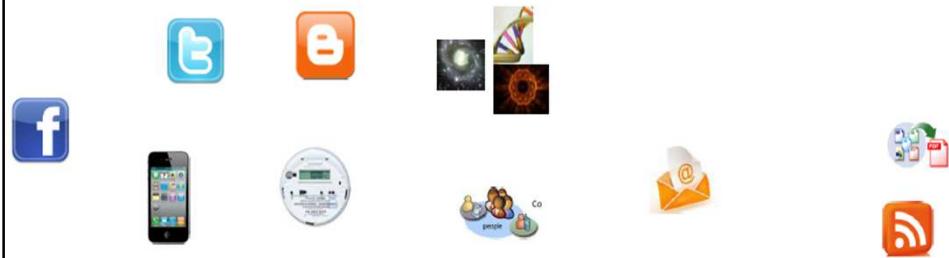
- Twitter / Facebook comments
- Stock market data
- Sensors: Vital signs of a newly-born



### ◆ Data at rest - oceans

- Collection of what has streamed
- Web logs, emails, social media
- Unstructured documents: forms, claims
- Structured data from disparate systems

## New Data Source



## Using The Data

Challenge	New Data	Transformation
Understanding online activity	Web Logs	Sessionization
Understanding failure modes	Product sensors	Identify out of range events, trends
Prioritizing responses to customers	Social Media	Sentiment analysis

## Big Data Use Cases

Today's Challenge	New Data	What's Possible
Healthcare Expensive office visits	Remote patient monitoring	Preventive care, reduced hospitalization
Manufacturing In-person support	Product sensors	Automated diagnosis, support
Location-Based Services Based on home zip code	Real time location data	Geo-advertising, traffic, local search
Utilities Complex Distribution Grid	Detailed consumption statistics	Increased availability, reduced cost, tiered metering plans
Retail One size fits all marketing	Social media	Sentiment analysis segmentation

Think about ... Data is an opportunity



Page 76

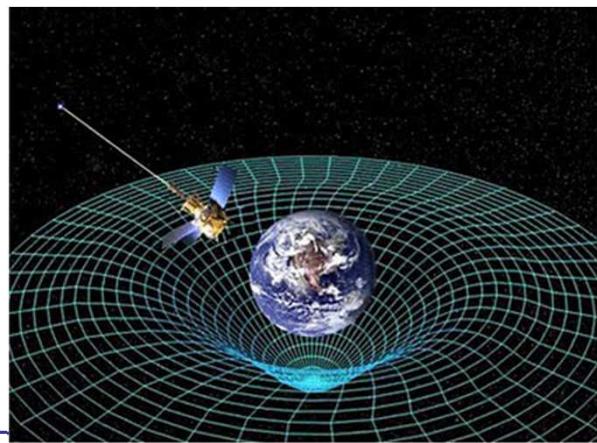
## □ Big Data Changing the Traditional Role of the Database

- ◆ Historically, the databases were used to store highly **structured business data**, which fit well into spreadsheet form and was good at answering relatively simple questions. In combination, this solved some moderately interesting use cases.
- ◆ Today, data is found in numerous forms, from **machine-generated** processes to spreadsheets, PDFs, Web logs, photos, video and so on. All that data conforms to domain-specific information models that often involve complex structures.



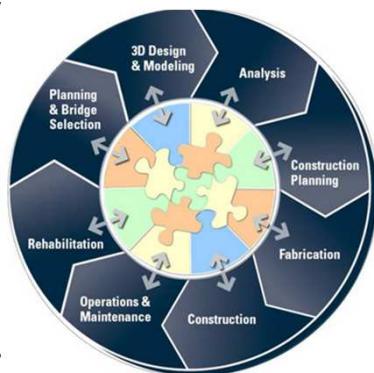
## Real-Time Paradigm vs. Predictive Modeling

- ◆ Predictive modeling, used in analytics, requires real-time processing. It is dynamic, so it requires a database that can quickly adjust to variable process states as they change under the weight of a workload. The **Big Data** is especially suited to handling the real-time requirements of predictive modeling environments.



## It's Now About the Information Model

- ◆ The old generation of business intelligence was **reactive** and centered on simple questions like, "**Who are my top 10 sales reps for the quarter?**" Answers to these were arrived at by querying the database for a specific, simple value using the standard key:value query format.
- ◆ The new generation is **proactive** and looking for answers to queries such as: "**Notice if a user has clicked on a weighted pattern of article tags over 3 disparate blogs in the last 36 hours.**" This kind of complexity is inherently handled algorithmically by rich domain modeling, which is seamlessly handled by **Big Data** technology



## Machine-generated data (MGD)

- ◆ A generic term for **information** which was automatically created from a computer process, application, or other machine without the intervention of a human.

“Data that was produced entirely by machines OR data that is more about observing humans than recording their choices.

-- Monash Research's Curt Monash

“Machine-generated data is data that is generated as a result of a decision of an independent computational agent or a measurement of an event that is not caused by a human action.”

-- Daniel Abadi, Professor at Yale

- ◆ Both exclude data manually entered by an end user. Machine-generated data crosses all industry sectors, and humans increasingly generate the data unknowingly.

### **Processing machine-generated data**

- ◆ Given the fairly static yet voluminous nature of machine-generated data, data owners rely on **highly scalable tools** to **process** and **analyze** the resulting dataset.
- ◆ Almost all machine-generated data is **unstructured** but then derived into a common structure. Typically, these derived structures contain many data points/columns.
- ◆ With these data points, the challenge lies mostly with analyzing the data. Given high performance requirements along with large data sizes, **traditional database indexing** and **partitioning limits** the size and history of the dataset for processing.
- ◆ Alternative approaches exist with **columnar databases** as only particular "columns" of the dataset would be accessed during particular analysis.

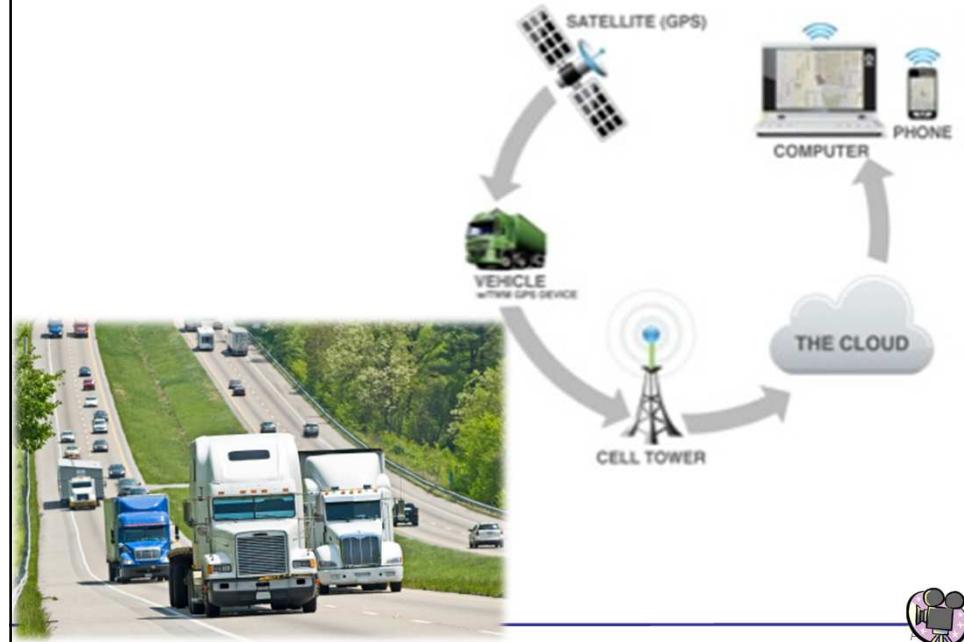
Page 86

### **❑ Usage-based Auto Insurance**

using the driving tracking device to monitor policyholders' driving patterns.



## GPS Tracking Devices



## Track kids

### Step 1

Place Amber Alert GPS on or with your child.



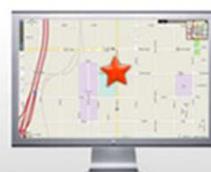
### Step 2

Call your Amber Alert GPS and receive a text message of its location.



### Step 3

View the location from your mobile phone or computer.





Page 91

The screenshot shows the NBC5.com website for WMAQ-TV Chicago. The main navigation bar includes links for Home, Search (powered by YAHOO! SEARCH), SITE, WEB, Get RSS, and E-Mail. A sidebar on the left lists Family, Coupons, Insurance, Legal Info, Kid's Health, and iVillage Parenting. A banner at the top right reads "Weekend roundup" and "Don't let another summer week waste away." Below the banner, a news article is highlighted with a red border:

**Divorce Goes In Fast Lane With E-ZPass**  
*Traffic Records Can Disprove Adulterers' Excuses*  
POSTED: 3:10 pm CDT August 10, 2007

Below the article, there is a social media sharing section with icons for Email, Print, Facebook, and Twitter. To the right, there is a "More" section featuring a graphic of two people walking on a path.

**AutoTrader.com**  
Sell Your Car  
**FAST**

**E-ZPass**

<http://www.nbc5.com/family/13867750/detail.html>

Page 92

## Case Study

E-ZPass and other electronic toll collection systems are emerging as a powerful means of proving infidelity. **That's because when your spouse doesn't know where you've been, E-ZPass does.**

**"E-ZPass is an E-ZPass to go directly to divorce court, because it's an easy way to show you took the off-ramp to adultery,"** said Jacalyn Barnett, a New York divorce lawyer who has used E-ZPass records a few times.

Lynne Gold-Bikin, a Pennsylvania divorce lawyer, said E-ZPass helped prove a client's husband was being unfaithful: **"He claimed he was in a business meeting in Pennsylvania. And I had records to show he went to New Jersey that night."**

The [Illinois Tollway](#) received more than 30 such subpoenas the first half of this year, with about half coming from civil cases, including divorces, according to Joelle McGinnis, an agency spokeswoman.

Page 93

## Google Street View Car

**March 2009 - A FURIOUS wife has called in divorce lawyers after spotting her husband's car parked outside another woman's house — on Google.**

She saw the Range Rover while using the internet giant's new Street View service to snoop on a female friend's home. The hubby had claimed he was away on business, but his missus recognized his motor immediately because of its blinged-up hubcaps.

The love cheat is not the only husband trapped by Google's new 360-degree photo search which covers 25 cities and towns throughout the country.

The lawyer said: "I was talking about the Range Rover case when another divorce lawyer came up to say his firm was dealing with the same sort of thing. People are getting caught out on Google."



<http://www.thesun.co.uk/sol/homepage/news/2350771/Cheating-husband-caught-on-Google-Street-View.htm> Page 94

### Web Server Log

- ◆ A log file (or several files) automatically created and maintained by a server of activity performed by it.
- ◆ A typical example is a **web server** log which maintains a history of page requests.
- ◆ Information about the request, including **client IP address, date/time, page requested, time spent on each page, etc.**
- ◆ These data can be combined into a single file, or separated into distinct logs, such as an access log, error log, etc.
- ◆ These files are usually not accessible to general Internet users, only to the **webmaster** or other administrative person. A statistical analysis of the server log may be used to examine **traffic patterns** by time of day, day of week, referrer, or user agent.
- ◆ Efficient web site administration, adequate hosting resources and the fine tuning of sales efforts can be aided by analysis of the web logs. **Marketing departments** of any organization that owns a website could be trained to understand these powerful tools.

Page 95

### Call Detail Record (CDR)

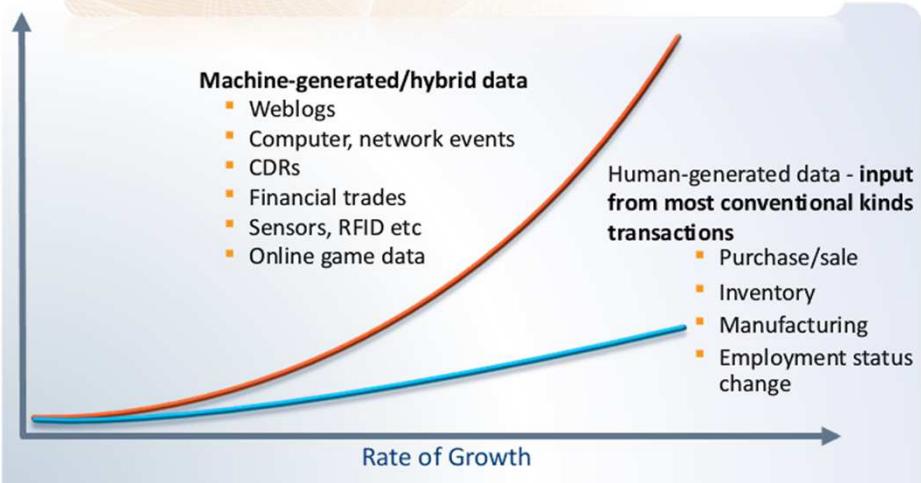
- ◆ A **data record** produced by a telephone exchange or other telecommunications equipment documenting the details of a phone call that passed through the facility or device.
- ◆ A CDR is composed of fields that describe the exchange. Examples of fields include:
  - the number making the call (calling party)
  - the number receiving the call (called party)
  - when the call started (date and time)
  - how long the call was (duration)
  - the phone number charged for the call (800 or others)
  - the identifier of the telephone exchange writing the record
  - a sequence number identifying the record
  - additional digits on the called number used to route or charge the call
  - the result of the call (whether it was answered, busy etc.)
  - call type (voice, SMS, etc.)

Page 96

## The Machine-Generated Data Problem

*"Machine-generated data is the future of data management."*

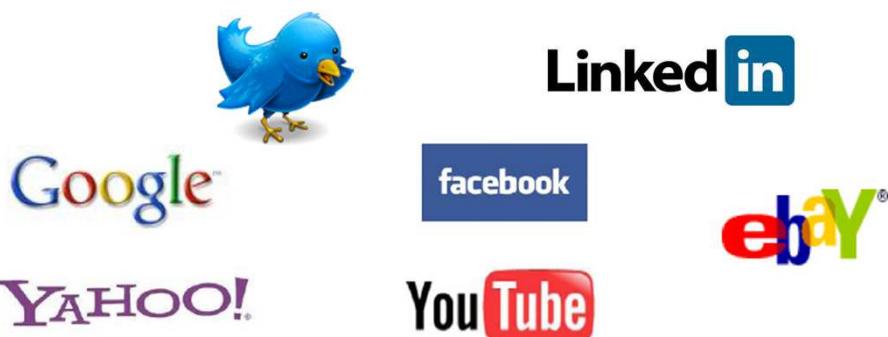
Curt Monash, DBMS2



Page 97

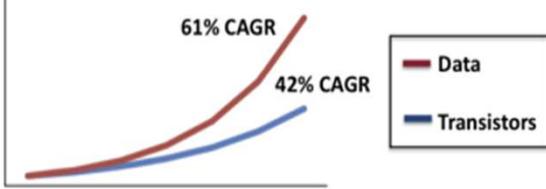
## Data-Driven on-Line Websites

- ◆ **To run the apps:** messages, posts, blog entries, video clips, maps, web graph...
- ◆ **To give the data context:** friends networks, social networks, collaborative filtering...
- ◆ **To keep the applications running:** web logs, system logs, system metrics, database query logs...



Page 99

## New Data Trends



Data is growing faster than processing power – leading to coping strategies like throwing away data or frequent archiving to tape

Circa 1975 – Transaction Data	Circa 2010 – Cloud Data
2,000 users = Huge	2,000 users = Tiny
Smaller data sets (bytes)	Extremely large data sets (petabytes)
Highly structured, relatively small data records	Unstructured, complex data blobs (images, voice, logs, video) – doesn't fit nicely into rows/columns
Absolute consistency is the primary requirement – ACID transactions	Application responsiveness/scale trumps immediate consistency

[http://news.cnet.com/8301-13846\\_3-20012164-62.html](http://news.cnet.com/8301-13846_3-20012164-62.html)

Page 100

## Challenging Times

### More data

- More online activity → more web data
- Growth of mobile → more call data, web data
- Servers/networks/sensors → lots of log/event data

### With increasing value in the details

- Target individual customers
- Identify micro-segments
- Find security threats
- Identify fraud



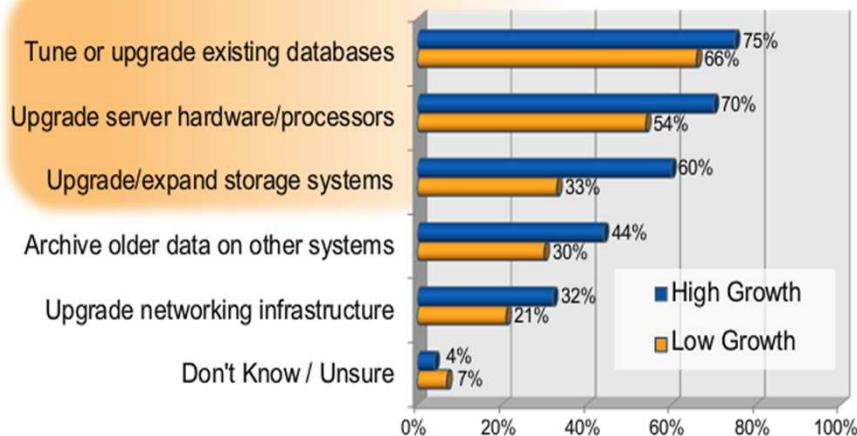
“Enterprise data growth over the next 5 years is estimated to be 650%.” Gartner

Page 101

## Current Database Technology: *Hitting the Wall*

**Today's database technology require huge effort & massive hardware.**

How Performance Issues are Typically Addressed – by Pace of Data Growth



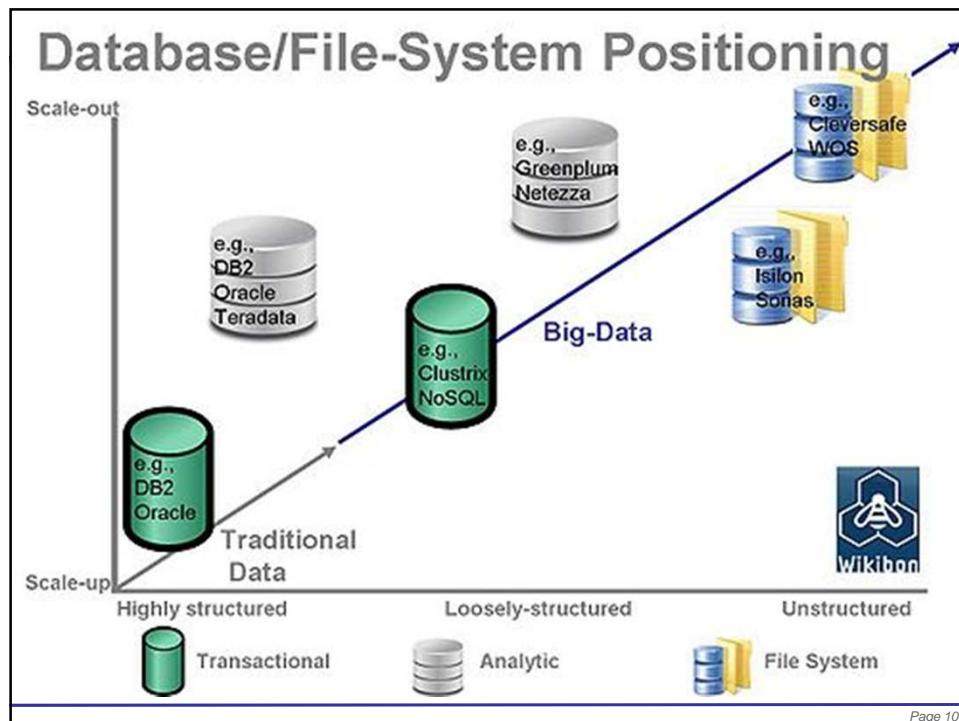
Source: KEEPING UP WITH EVER-EXPANDING ENTERPRISE DATA  
By Joseph McKendrick, Research Analyst, Unisphere Research October 2010

Page 102

## Data / Database / Information Technology of the Future

- ❖ One size **DOESN'T** fit all.
- ❖ More open source specialized database than commercial
- ❖ Many Specialized databases developed
  - Excellent at what they were designed for
- ❖ Cloud use for specialty DBMS becomes popular
  - The **cloud computing** and **Big Data** are combining with **mobility** and emerging **social** trends to re-shape the technology industry.
- ❖ Database Virtualization
- ❖ Big Data is not just about a cool technology.
  - **It's about insight**
- ❖ A different ballgame
  - **Require different skills**
  - Game-Changing Analytics
- ❖ This area is extremely fast growing, with many new entrants into the market expected over the next few years.
  - Create new job opportunities, such as, Data Scientist, etc.

Page 103



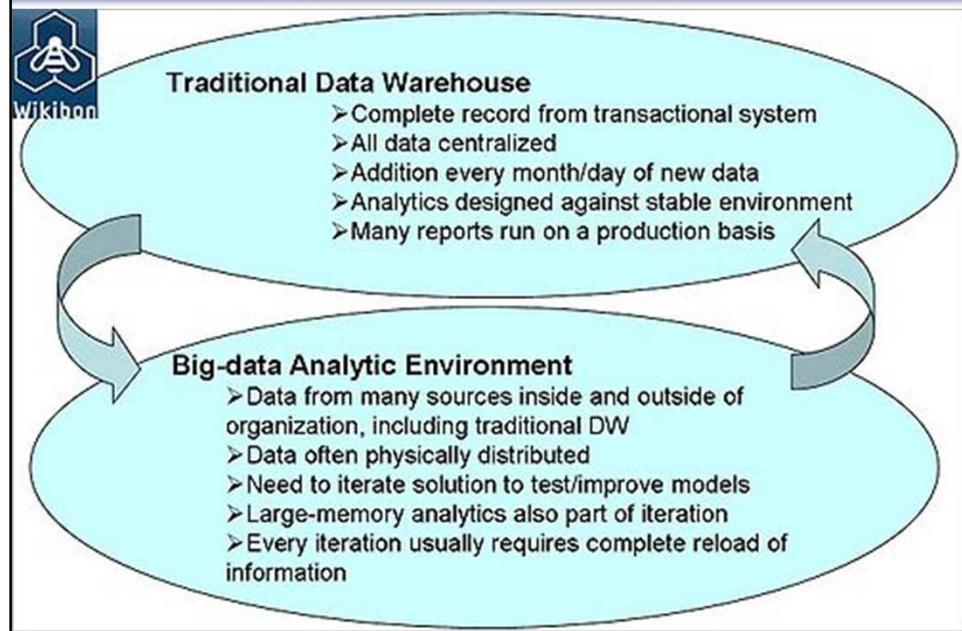
### Technical Challenges

- ◆ Moore's law?
- ◆ Yes ... but there are other factors, like ...
  - disk seek times
  - network limitations > the fallacies of distributed systems
  - ...
- ◆ Force to think out of the (*one*) box
  - automatically buys you scale and SPOF-robustness
  - ... at the cost of complexity

If all you have is a hammer, everything looks like a nail

Page 106

### Differences between traditional Data Warehouse & big data analytics.



Source: Wikibon 2011

Page 107

### New Data and Management Economics

#### Compute Trend

##### New Analytics Emerge (MapReduce, Hadoop...)

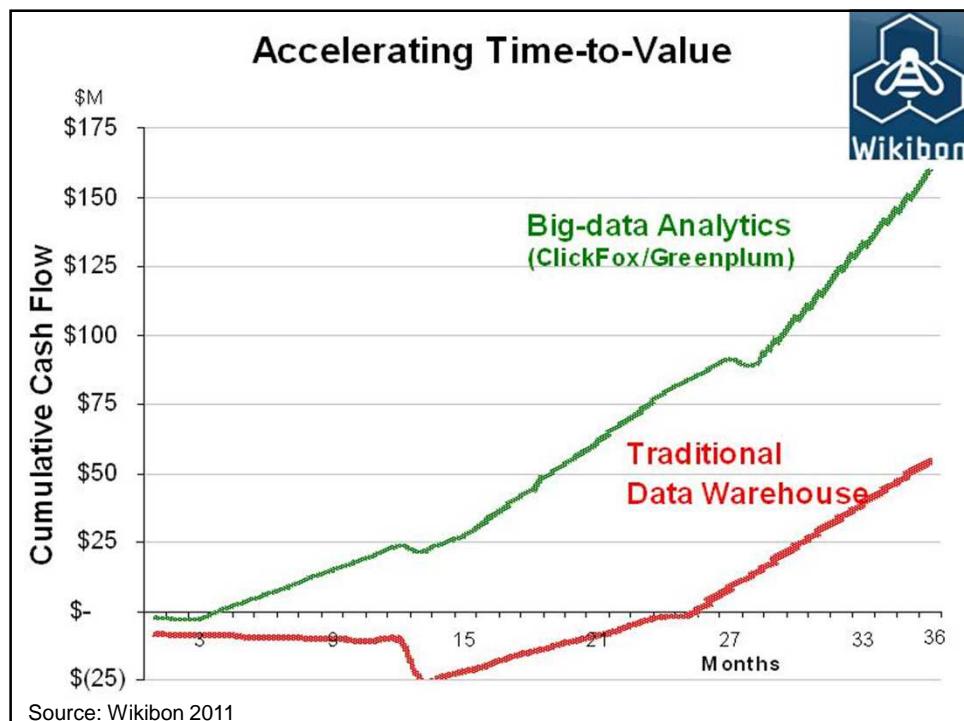
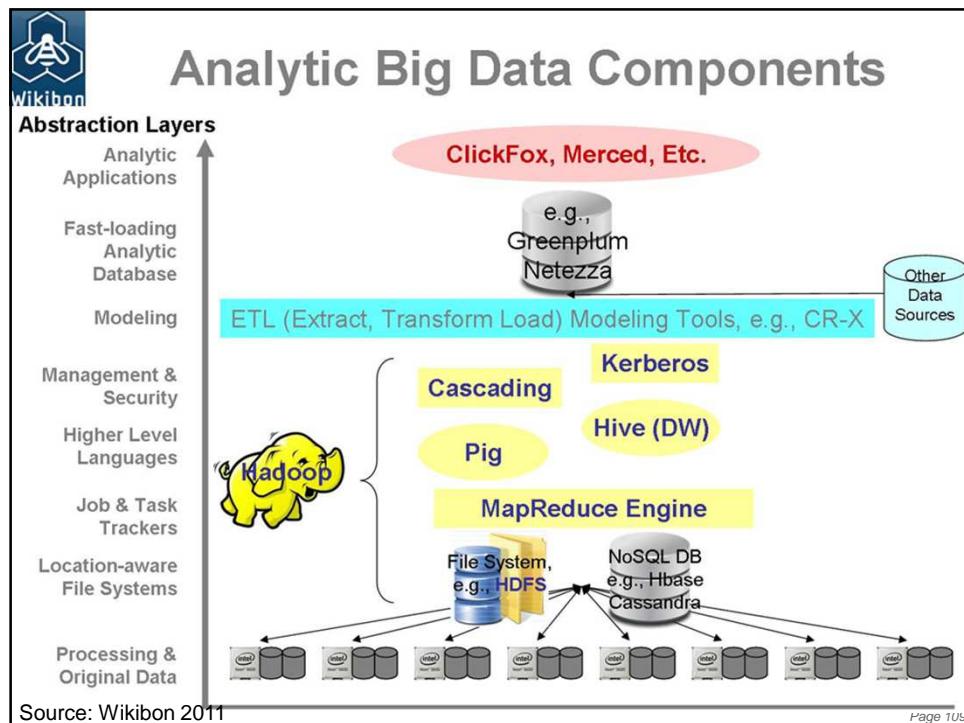


#### Storage Trend

##### Semi-structured Data (MogileFS, Bigtable, HDFS...)



Page 108



## Big Data vs. Traditional Database

Keep in Mind

### Complementary not Competitors

- ◆ A positioning of databases that support transactional and analytical workloads, and show that support of loosely structured and scale-out architectures are essential for Big-data initiatives.
- ◆ Traditional data warehouses and Big Data Analytics are complementary to each other and feed each other.
- ◆ Traditional data warehouses are a source of data for Big-data projects; if new data which is valuable on an ongoing basis during a Big-data project, it should be brought into the traditional data warehouse, cleaned up, and take advantage of the production capabilities of traditional databases.

Page 117

## Big News: Big Data – Big Deal



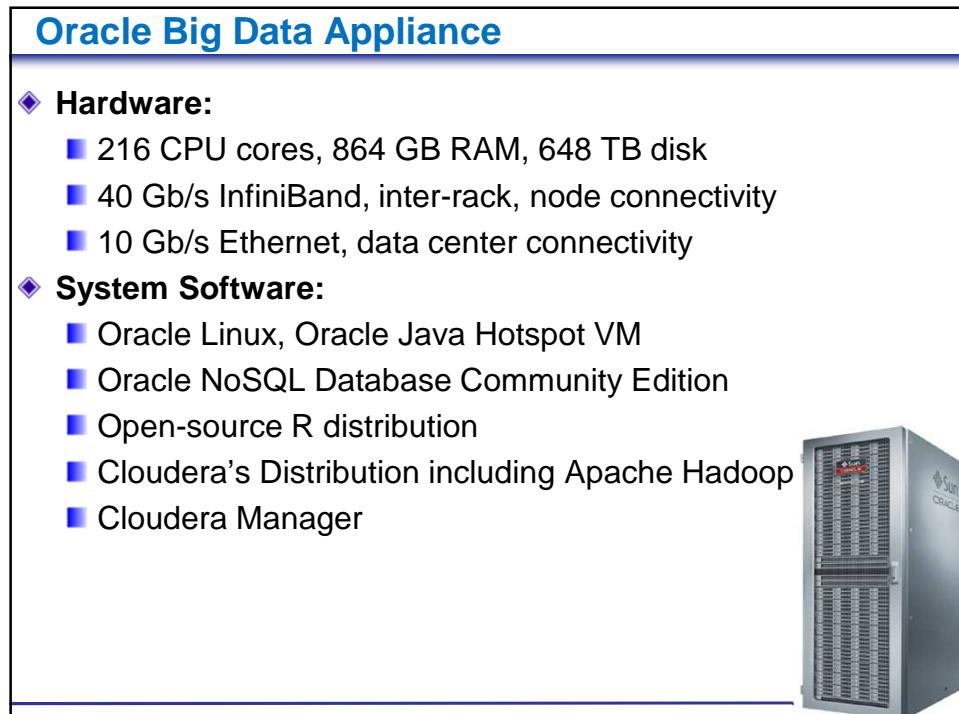
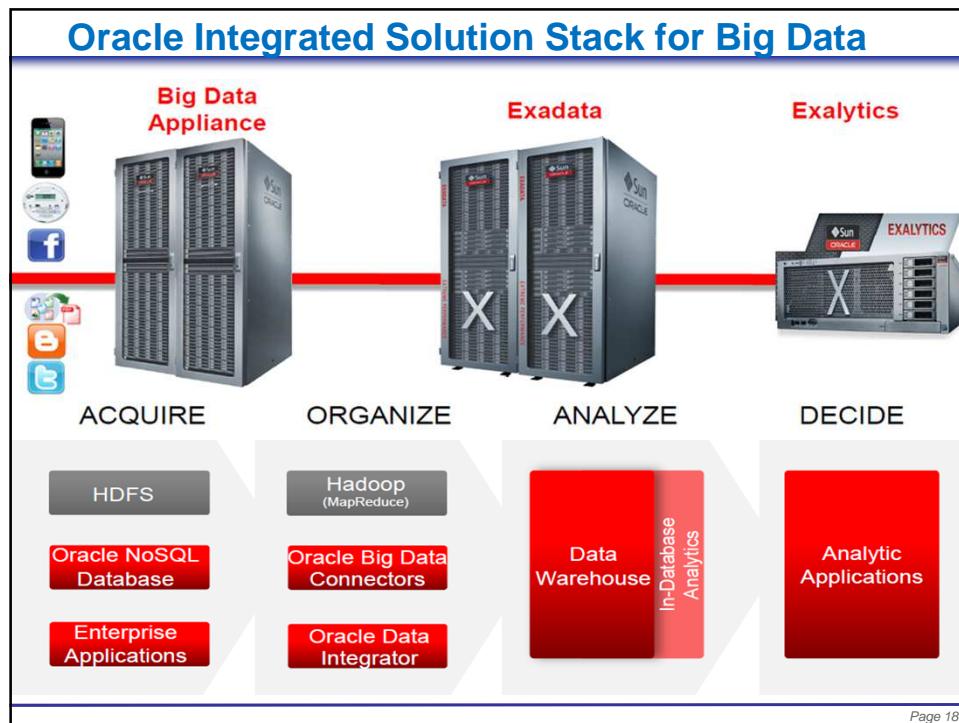
### IT 2013: It's All About Control of the Data

"This is not just about **data (technology)** but **insights**"

**Make Better Decisions Using Big Data**



Page 184



## Cloud Computing Is a Good Fit for Big Data

- ◆ The **cloud** is founded on **the utility of computing** and the "**self-healing of infrastructure.**" These tenets lend themselves well to secondary problems confronting data management on the Big Data frontier.
- ◆ **Big Data** problems are solved through **distributed software architecture** and **parallel processing**. Cloud computing inherently delivers a distributed infrastructure suitable for distributed software architectures. The ability to spread data across multiple machines and to move processes to various machines to provide efficiency is how data is best processed.



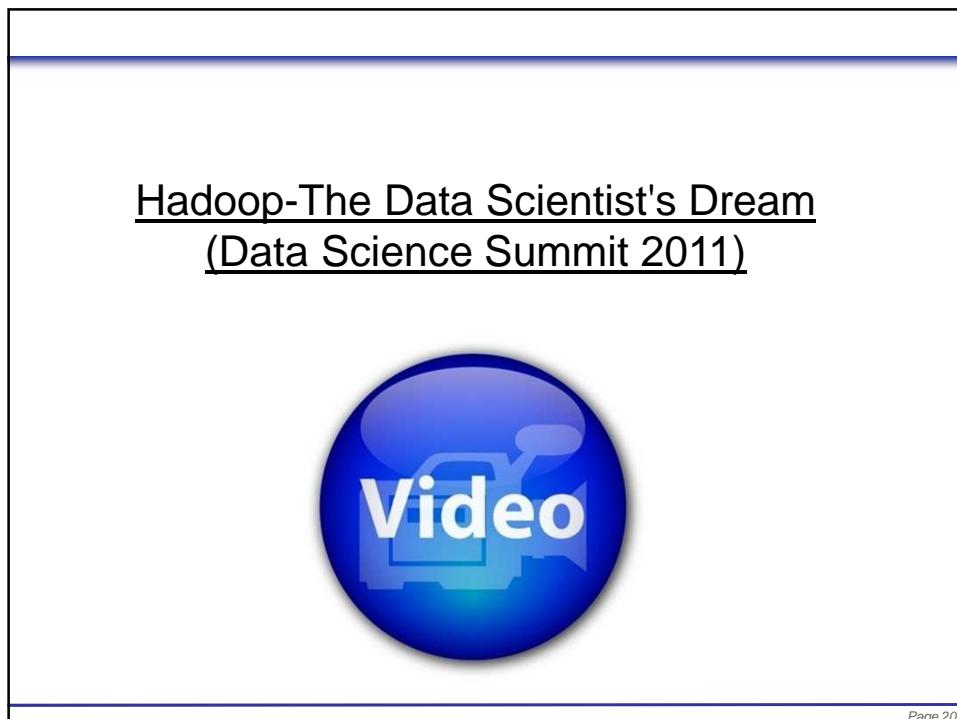
## Big Data & Cloud: Better Together



Apps can interact with the world around them to enrich the users experience by communicating with applications that run on a companies servers, applications on the cloud, or by talking directly with other apps.



Page 199



The screenshot shows the official website of the College of Computing and Digital Media at DePaul University. The header includes the DePaul logo, navigation links for CDM Home, Academics, Online Learning, Current Students, Prospective Students, About CDM, People, and Alumni. The main menu features three schools: School of Cinema and Interactive Media, School of Computing, and the Institute for Professional Development (IPD), which is highlighted with a red box and a red arrow pointing to it from the left. A yellow banner at the top displays the URL <http://www.cdm.depaul.edu>. The IPD Programs page content includes a heading "IPD Programs", a paragraph about certificate programs, sections for Big Data Technologies, Cloud Computing Technologies, and Database Technologies, each with a bulleted list of programs. The footer contains standard links for CDM Guidelines, MyCDM, Interne...