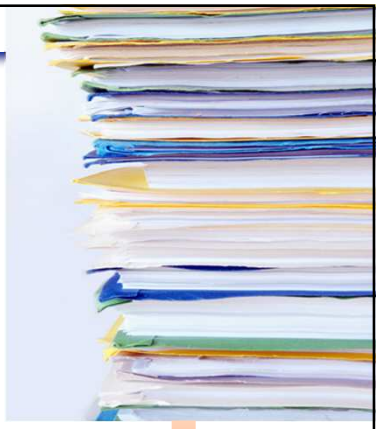
The Ancient World

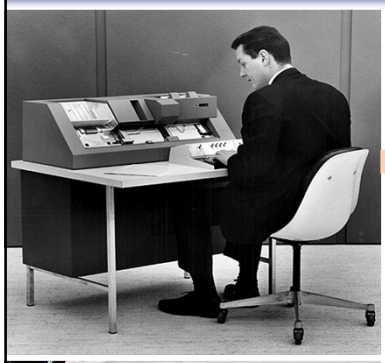
Pictograph

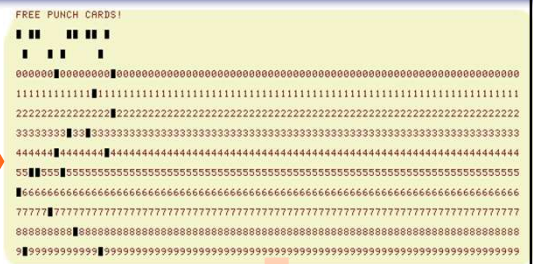
Paper was not widely used in the west until the 8th or 9th century, Ex: Hebrew Manuscript on Paper

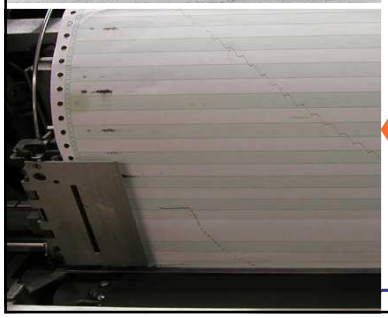


The Old World

Data in Text Format - Character Mode





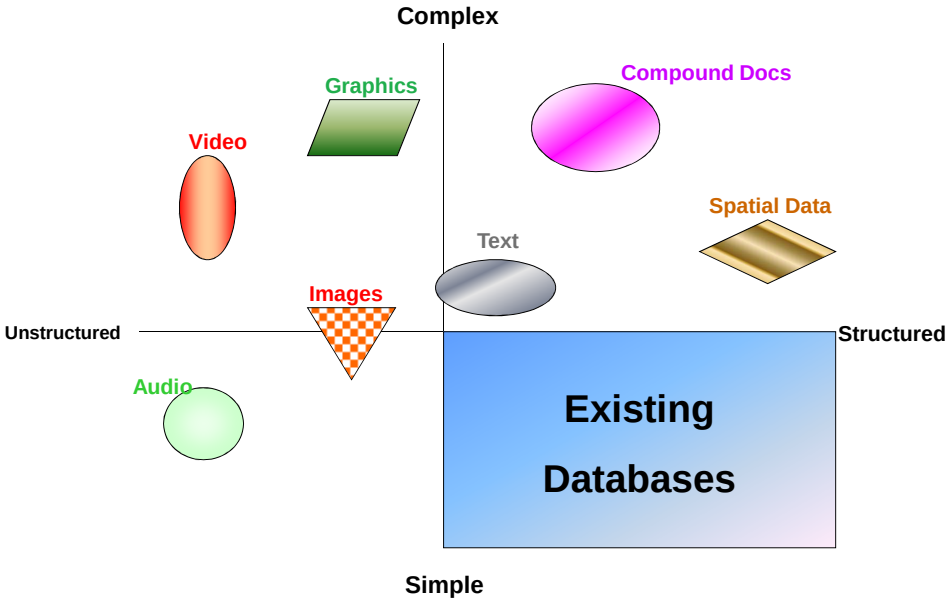


History of Information Technology

- **The "dark ages":** paper forms in file cabinets
- Computerized systems emerge
  - Initially for big projects like Social Security
  - Same functionality as old paper-based systems
- **The "golden age":** databases are everywhere
  - Most activities tracked electronically
  - Stored data provides detailed history of activity
- **The next step:** use data for decision-making
  - Made possible by different types of users
  - Identify inefficiencies in current processes
  - Quantify likely impact of decisions

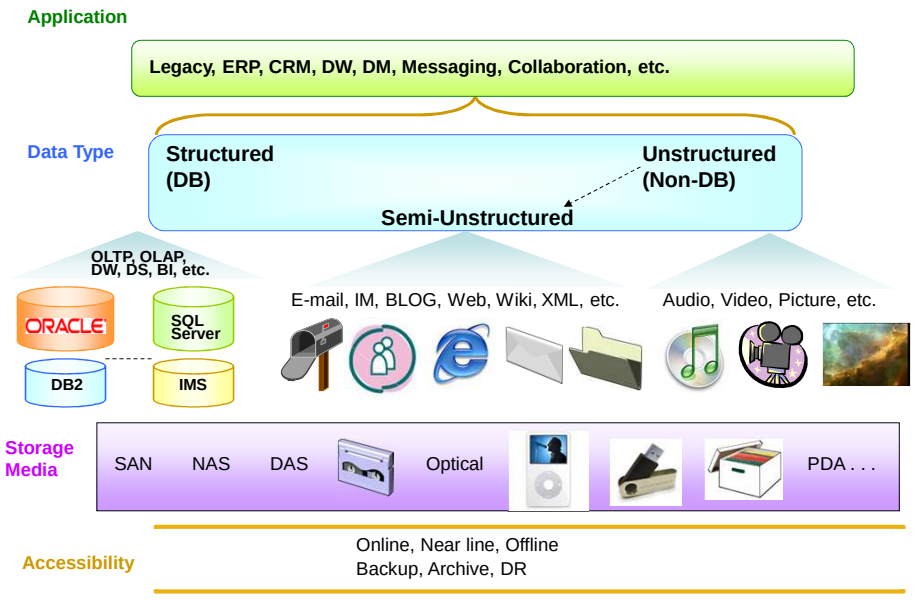
EMC: 55 sec

We Live In a World of (Data) Information in digital format



Page 6

The New World — Information in Any Format – Digital Life



Types of Data

People use them everyday …

- **Structured Data**
  - Structured data is anything that has an enforced composition to the atomic data types. Structured data is managed by technology that allows for querying and reporting against predetermined data types and understood relationships.
- **Two Categories of Unstructured Data**
  - **Bitmap Objects:** Inherently non-language based, such as image, video or audio files.
  - **Textual Objects:** Based on a written or printed language, such as Microsoft Word documents or Microsoft Excel spreadsheets.
- **Semi-Structured Data**
  - A form of structured data that does not conform with the formal structure of tables and data models associated with databases but contains nonetheless tags or other markers to separate semantic elements of records and fields within the data, such as XML, email.

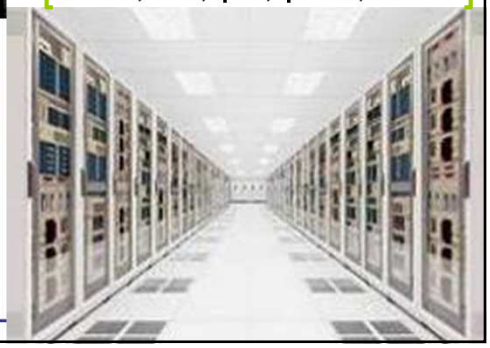
Goodbye Paper? Hello ESI (Electronic Stored Information)

- Over a thousand different formats
- Store in any device

Company
- Onsite: data center, remote office
- Offsite:
  - Physical: Iron Mountain, etc.
  - Virtual: Cloud (Amazon, etc.)

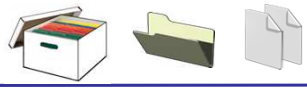
Personal
- Home PC w/burner, webcam, screen capture, etc.
- Google storage, Gmail, etc.
- DVD, USB, ipod, iphone, etc.

- 1 format
- 2 sizes (letter, legal)
- Store in box

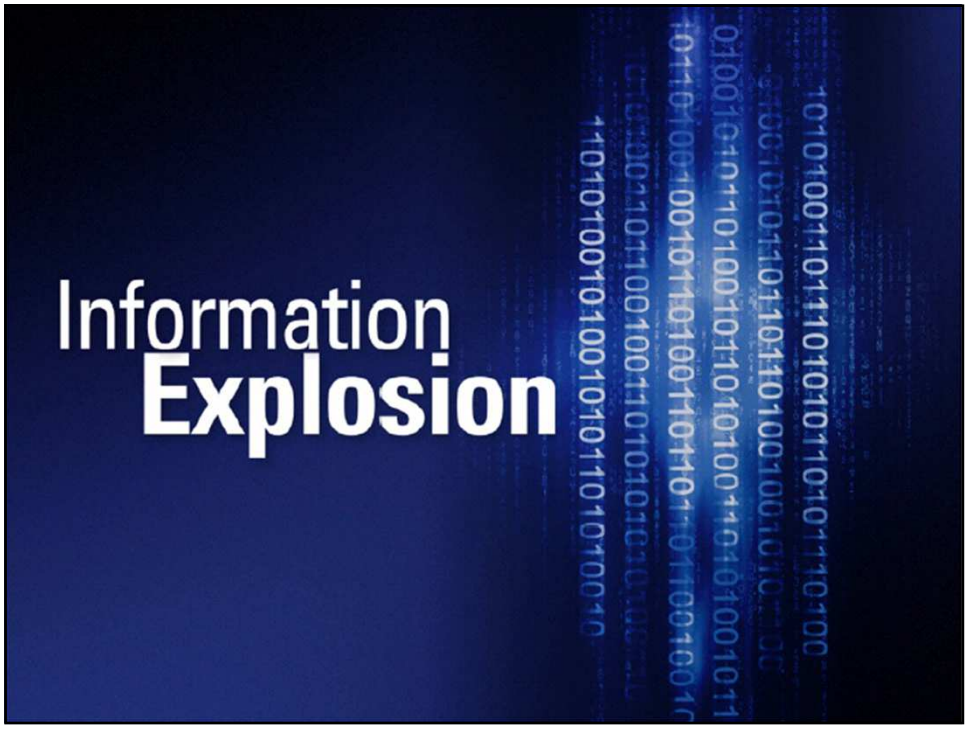


Multimedia = Big Databases => Big Data

Object	Typical Size
High-definition television	200 MB/sec.
Feature-length high-resolution movie	6-7 GB
Video	1 GB/hour
Radiological image	40-60 MB
Color image	20-40 MB
Voice/Audio	2-10 MB
Text	30-40 KB/page
Check image	45 KB
Small image	30-40KB




TV & Movies
Maps,
Travel planning
Spotting trends for
business intelligence
Music
Calendars,
Scheduling
Presence
Unified
communication
Memories
Improving
personal
productivity
Education
Digital
Lifestyle
Digital
Workstyle
Games
Insights and
structured
workflow
Team
collaboration
Productivity
Finding
information
Buying/Selling
Optimizing
supply chains
Communications
Digital Life


Information
Explosion

How much information?

Google "how much information" Search Advanced Search Preferences

Web Show options... Results 1 - 10 of about 569,000 for "how much information". (0.17 seconds)

How Much Information?
Attempts to measure **how much information** is produced in the world each year by looking at media and other variables. Produced by the School of Information ...
www2.sims.berkeley.edu/research/projects/how-much-info/ - Similar pages

UCSD: Global Information Industry Center
Jun 3, 2008 ... To answer these questions and others, a new **How Much Information**? (**HMI**) research program is underway. . . continue > ...
hmi.ucsd.edu/ - 8k - Cached - Similar pages

How much information is too much in cyberspace? - web - Technology
May 18, 2009 ... When does sharing about one's personal life cross the line and become too much information? - The Age.
www.theage.com.au/news/technology/web/**how-much-information**-is-too-much-in-cyberspace/2009/05/18/1242498689035.html - Similar pages

How Much Information Is There In the World?
How much information is there in the world? This paper makes various estimates and compares the answers with the estimates of disk and tape sales, ...
www.lesk.com/mlesk/ksg97/ksg.html - 20k - Cached - Similar pages

How Much Information? 2003

Summary (Exec Summary) **Stored Information** (Paper | Film | Magnetic | Optical) **Information Flows** (Broadcast | Telephony | Internet) **Wrap-up** (Thanks | Printable (PDF))

HOW MUCH INFORMATION 2003?

Senior researchers: Peter Lyman and Hal R. Varian

Project coordinator: Kirsten Swearingen

Researchers: Peter Charles, Nathan Good, Laheem Lamar Jordan, Joyojeet Pal

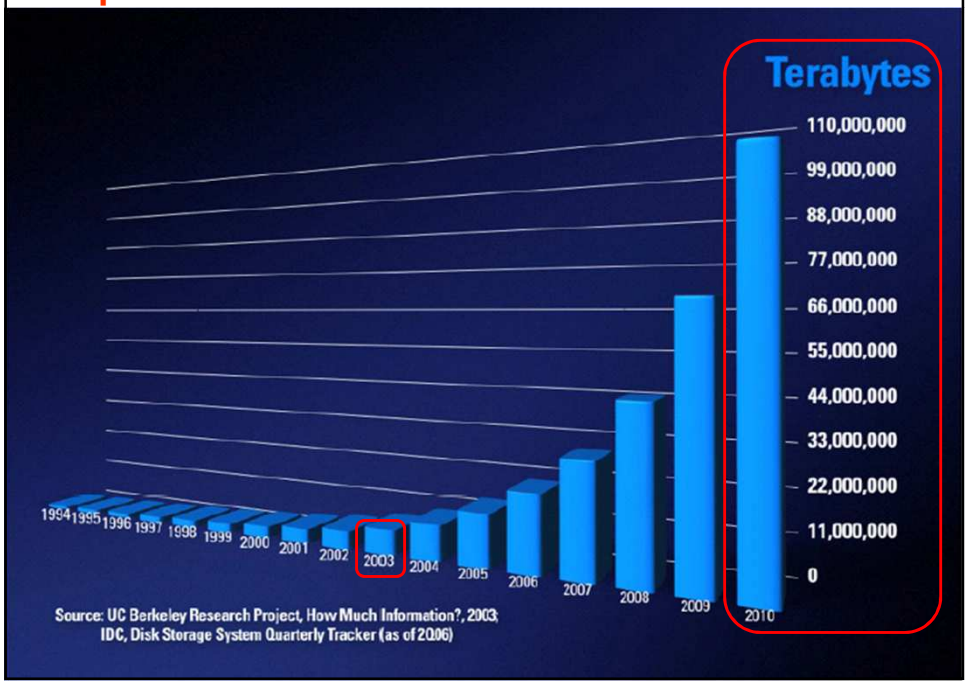
This study is an attempt to estimate how much new information is created each year. Newly created information is distributed in four storage media ? print, film, magnetic, and optical ? and seen or heard in four information flows ? telephone, radio and TV, and the Internet. This study of information storage and flows analyzes the year 2002 in order to estimate the annual size of the stock of new information contained in storage media, and heard or seen each year in information flows. Where reliable data was available we have compared the 2002 findings to those of our 2000 study (which used 1999 data) in order to identify trends ? recognizing that 1999-2002 were years of relatively low economic activity. The 2000 study is located on the Web at http://www.sims.berkeley.edu/how-much-info/. Note that this ? the 2003 study ? has revised certain of the 1999 estimates when we have found new and better data sources.

Table 1.2: Worldwide production of original information, if stored digitally, in terabytes circa 2002. Upper estimates assume information is digitally scanned, lower estimates assume digital content has been compressed.

Storage Medium	2002 Terabytes Upper Estimate	2002 Terabytes Lower Estimate	1999-2000 Upper Estimate	1999-2000 Lower Estimate	% Change Upper Estimates
Paper	1,634	327	1,200	240	36%
Film	420,254	76,69	431,690	58,209	-3%
Magnetic	5187130	3,416,230	2,779,760	2,073,760	87%
Optical	103	51	81	29	28%
TOTAL:	5,609,121	3,416,281	3,212,731	2,132,238	74.5%

Source: How much information 2003 http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/

A perfect storm



What is Large?

Abbreviation	Term	Amount
KB	Kilobyte	1,024bytes
MB	Megabyte	1,024KB
GB	Gigabyte	1,024MB
TB	Terabyte	1,024GB
PB	Petabyte	1,024TB
EB	Exabyte	1,024PB
ZB	Zettabyte	1,024EB
YB	Yottabyte	1,024ZB

EMC: How Much Info?

How tall is 1 TB?

1,450 feet

23 x Sears Tower

1 Page: 80 bytes/line x 60 lines
x 2 sides = 9,600 bytes
~ 10K

1 TB = 1,000,000,000 KB

= 100,000,000 pages

1 Brick = 500 sheets

= 200,000 bricks

* Each brick is 2” thick

= 400,000 inches

= 33,333 feet

Page 17

Data, Data Everywhere ….

- Not all the data created equal
- The value of data changes over time
- Is data an Asset, Liability, or Distraction?

Are you data rich & information poor?

In the past 50+ years, we learned how to "store" information.
Today and in the future, we need to figure out how to "find" information.

DATABASE TRENDS AND APPLICATIONS
The Journal of Information Integration and Management
August 2008
TRENDS AND APPLICATIONS

Data Discovery is Next Evolutionary Step in Data Integration

By Robert Eve

"In the struggle for survival, the fittest win out at the expense of their rivals because they succeed in adapting themselves best to their environment." – Charles Darwin, The Origin of Species

Data Integration (DI) technology, (specifically, extract, transform, and load (ETL) middleware), when combined with an intermediate data store such as a warehouse or mart, has played a key role in advancing business intelligence (BI) and performance management since the mid-1990s. Virtualized DI evolved from these technologies in the mid-2000s. Alternatively known as virtual data federation or enterprise information integration (EII), virtual DI eliminates the intermediate data store by leveraging high-performance query techniques that let the consuming application pull data directly from the source, in real time.

The next evolutionary DI step is currently in a nascent stage. Data discovery revolutionizes how business professionals can leverage enterprises' ever-expanding data assets, thus changing the competitive dynamic with its speed and simplicity.

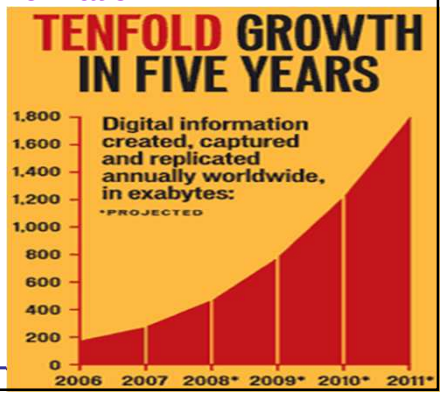
Drivers of the DI Evolution: Data Volume and Source Complexity

Recently, IDC estimated the rate of compound enterprise data growth to reach nearly 60 percent annually. In other words, enterprises will likely have 10 times today's data by 2013, and 100 times by 2018.

Concurrent with this growth has been the rapid expansion of data complexity. Data can be structured in rows and columns within transactions systems. Data can be unstructured in documents stored on desktops. Recent

Data Growth

- The recent growth of information sources such as blogs, social networks, news aggregators, microblogs like Twitter, instant messaging and e-mail has been exponential.
- The good … There's so much more information available.
- The bad ….. Too much information available.
- The ugly …. Can I trust the information?
  Can I find the right information?
- According to market research firm IDC, by 2011 the digital universe will be 10 times the size it was in 2006.

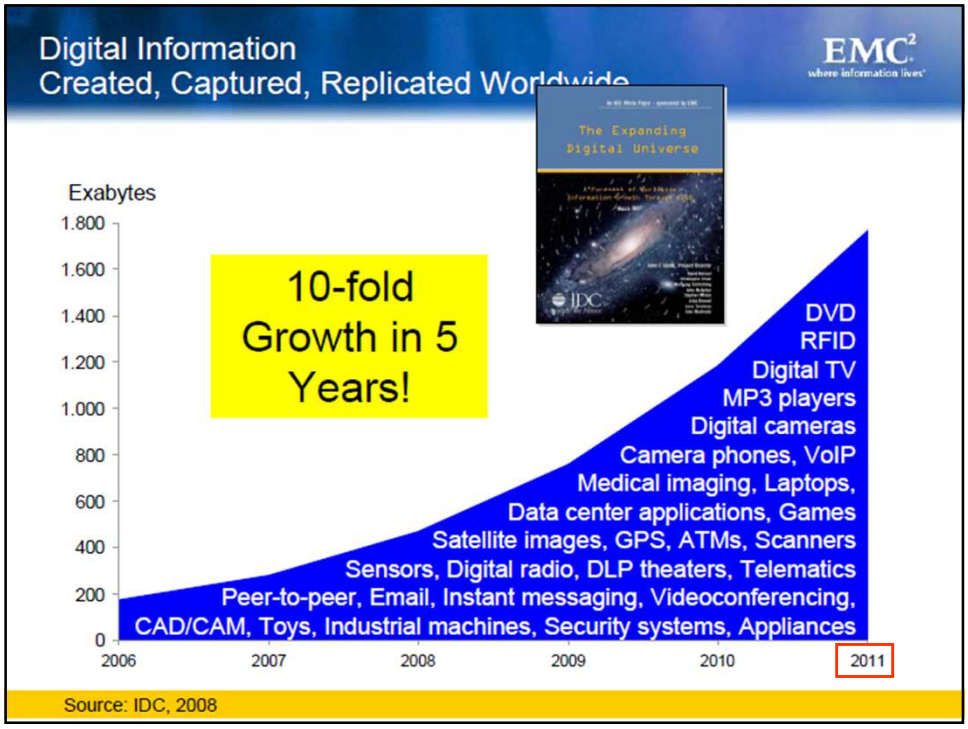


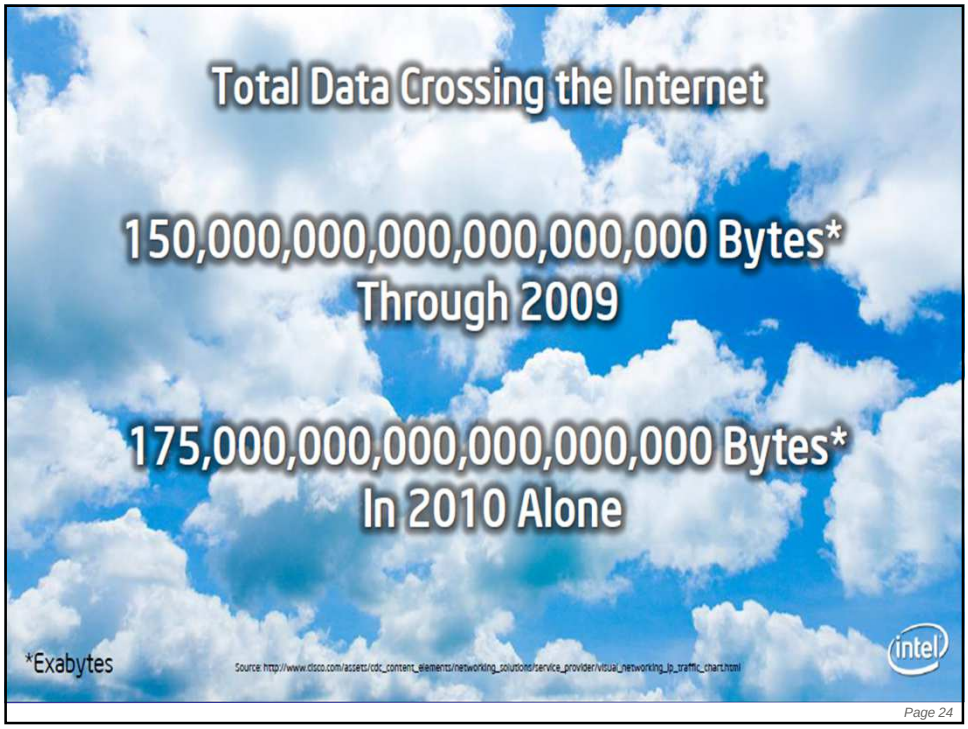
- **Today you have $1**
- **With 60% annual growth**
- **What's the return in 5 years?**
- **What's the return in 10 years?**

2014

2015 - $1.6
2016 - $2.56
2017 - $4.09
2018 - $6.55
2019 - $10.48

2020 - $16.7
2021 - $26.8
2022 - $42.9
2023 - $68.7
2024 - $110.0


Digital Information
Created, Captured, Replicated Worldwide
EMC²
where information lives
The Expanding Digital Universe
IDC
Exabytes
1.800
1.600
1.400
1.200
1.000
800
600
400
200
0
10-fold Growth in 5 Years!
DVD
RFID
Digital TV
MP3 players
Digital cameras
Camera phones, VoIP
Medical imaging, Laptops,
Data center applications, Games
Satellite images, GPS, ATMs, Scanners
Sensors, Digital radio, DLP theaters, Telematics
Peer-to-peer, Email, Instant messaging, Videoconferencing,
CAD/CAM, Toys, Industrial machines, Security systems, Appliances
2006
2007
2008
2009
2010
2011
Source: IDC, 2008


Total Data Crossing the Internet
150,000,000,000,000,000,000 Bytes*
Through 2009
175,000,000,000,000,000,000 Bytes*
In 2010 Alone
*Exabytes
Source: http://www.cisco.com/assets/cdc_content_elements/networking_solutions/service_provider/visual_networking_ip_traffic_chart.html
intel
Page 24

Are You Ready?

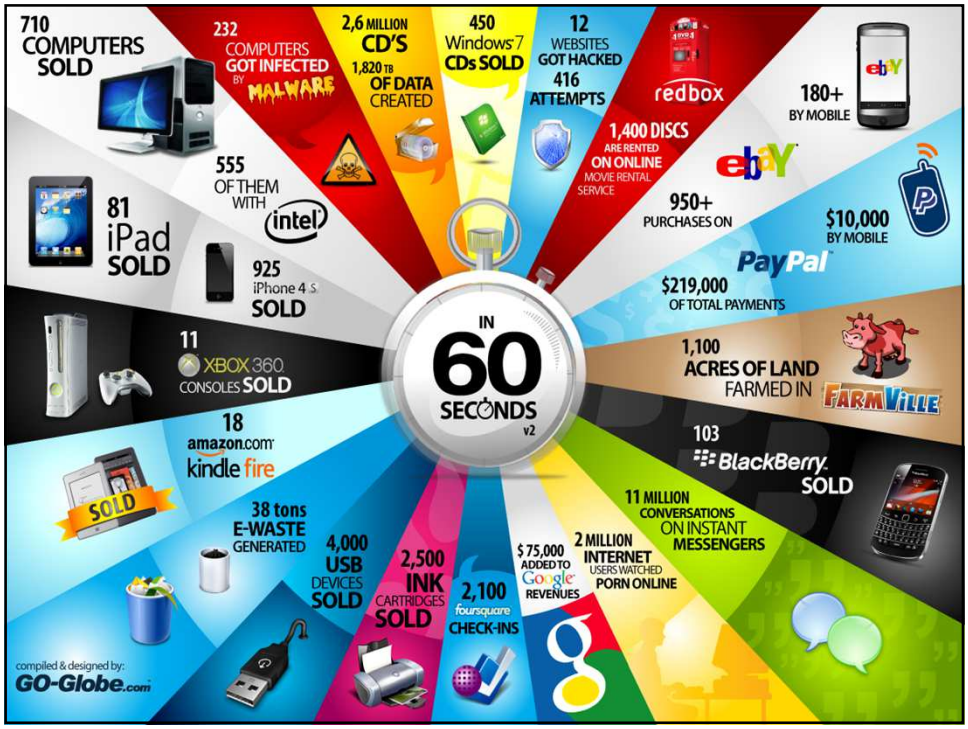
Page 25

How much data do you generate?

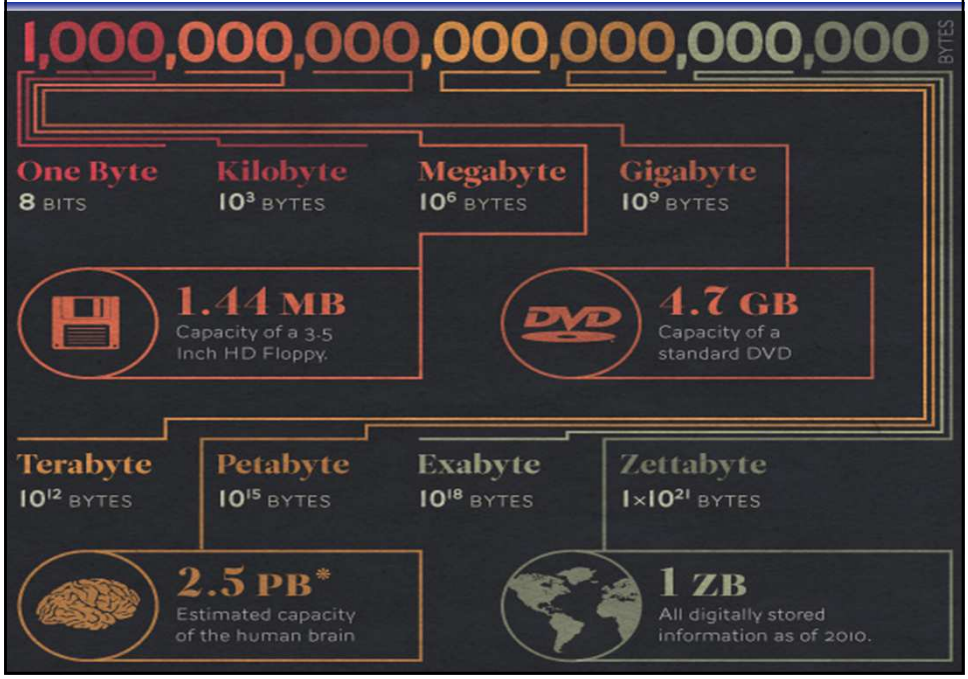
How do you generate?

- E-mail
- Text
- Tweet
- Phone
- Photo
- Audio
- Video
- Document
- Download/replicate
- Transaction (buy, sell, pay, subscribe/register)

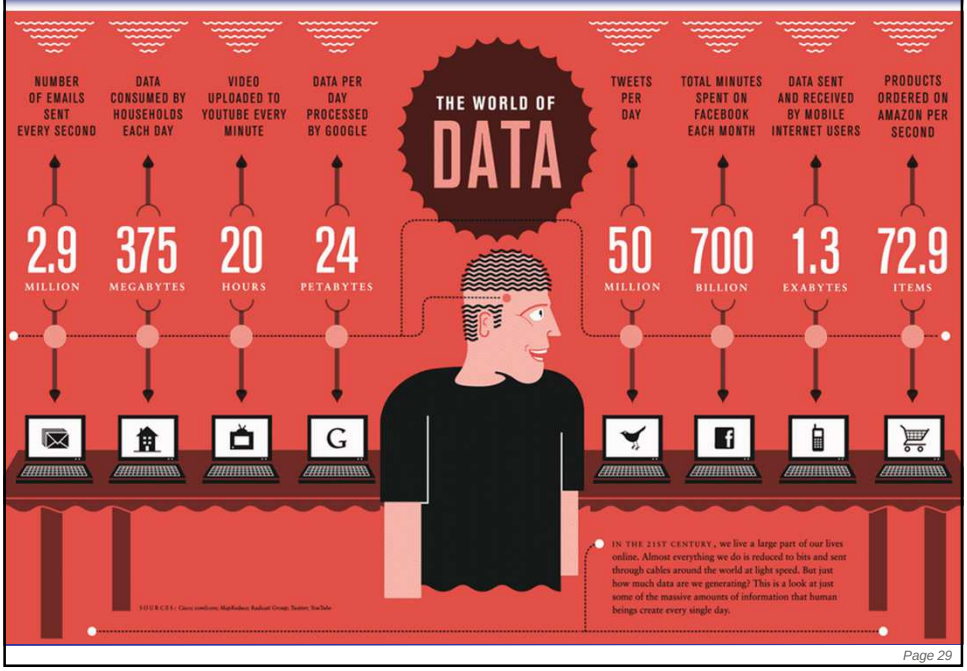
Page 26


710 COMPUTERS SOLD
232 COMPUTERS GOT INFECTED BY MALWARE
2,6 MILLION CD'S
1,820 TB OF DATA CREATED
450 Windows 7 CDs SOLD
12 WEBSITES GOT HACKED
416 ATTEMPTS
redbox
180+ BY MOBILE
1,400 DISCS ARE RENTED ON ONLINE MOVIE RENTAL SERVICE
555 OF THEM WITH intel
81 iPad SOLD
925 iPhone 4 S SOLD
950+ PURCHASES ON
$10,000 BY MOBILE
PayPal
$219,000 OF TOTAL PAYMENTS
IN 60 SECONDS v2
11 XBOX 360 CONSOLES SOLD
1,100 ACRES OF LAND FARMED IN FARMVILLE
18 amazon.com kindle fire
SOLD
103 BlackBerry. SOLD
11 MILLION CONVERSATIONS ON INSTANT MESSENGERS
38 tons E-WASTE GENERATED
4,000 USB DEVICES SOLD
2,500 INK CARTRIDGES SOLD
2,100 foursquare CHECK-INS
$75,000 ADDED TO Google REVENUES
2 MILLION INTERNET USERS WATCHED PORN ONLINE
compiled & designed by: GO-Globe.com

What is a ZB?


1,000,000,000,000,000,000,000 BYTES
One Byte
8 BITS
Kilobyte
10³ BYTES
Megabyte
10⁶ BYTES
Gigabyte
10⁹ BYTES
1.44 MB
Capacity of a 3.5 Inch HD Floppy.
DVD
4.7 GB
Capacity of a standard DVD
Terabyte
10¹² BYTES
Petabyte
10¹⁵ BYTES
Exabyte
10¹⁸ BYTES
Zettabyte
1×10²¹ BYTES
2.5 PB*
Estimated capacity of the human brain
1 ZB
All digitally stored information as of 2010.

The World of Data



Page 29

How much data?

1.2 zettabytes (ZB)

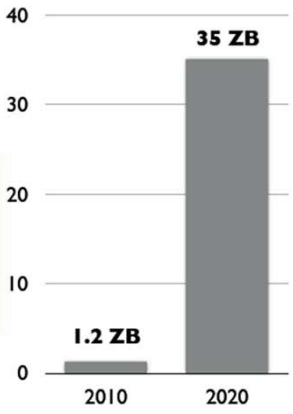
The amount of digital data that was produced worldwide in 2010.

(Measured in typewritten pages, this would be enough to cover the earth 30 times over.)

IDC *Digital Universe Study*, April 2010

In one decade the amount of digital data produced worldwide is projected to grow 30 times larger.

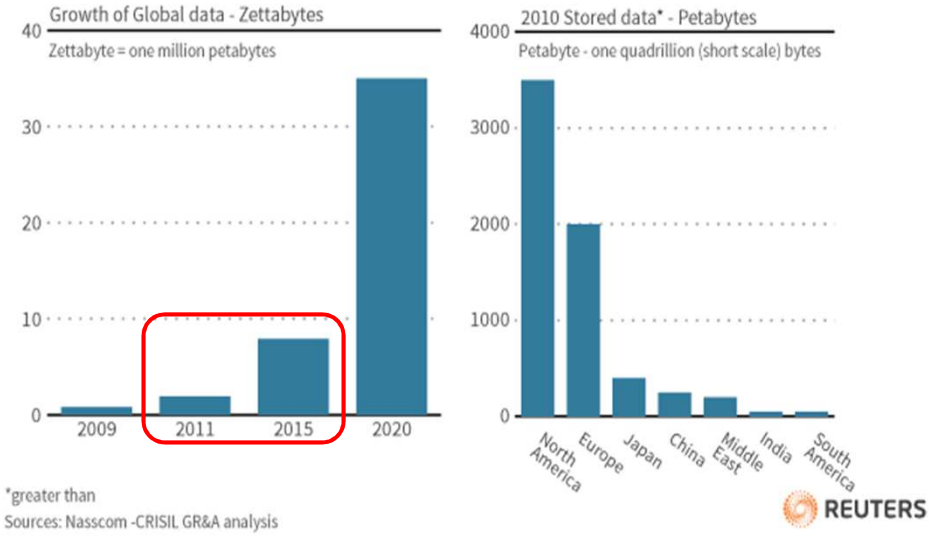
IDC Digital Universe Study, April 2010



Page 30

Big data growth

Big data market is estimated to grow 45% annually to reach $25 billion by 2015

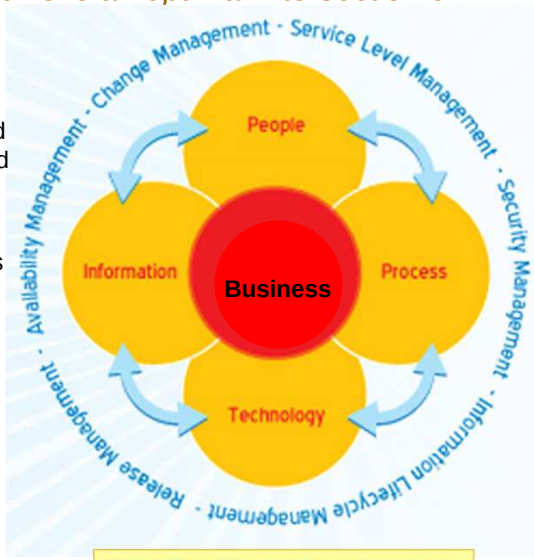


Reuters graphic/Catherine Trevethan 05/10/12

Business = PEOPLE…PROCESS…TECHNOLOGY...INFORMATION

Business thrives when you achieve an optimal intersection of all four building blocks.

- **PEOPLE**
  - Clearly defined roles and responsibilities provide a sound decision-making framework and solid personnel infrastructure
- **PROCESS**
  - Implement and enforce policies by integrating information, technologies, and people with automated processes
- **TECHNOLOGY**
  - Products and technologies automate and integrate tasks and processes
- **INFORMATION**
  - Standardized, federated accessible information



Run IT like a business

Beyond the basics

- **Traditional maintenance:** backups, recovery, performance monitoring, tuning, space management, upgrading, etc.
- **More focused on specialty areas:** compliance, applications, business intelligence, high availability, virtualization, cloud, big data, etc.

There are two predominant trends that are changing the face of database administration:

Go beyond the administration basics and strive to learn one or more of these specialty areas.

- **Increased functionality:**
  - The functionality (and often the complexity) of the software expands significantly with each new release.
  - DBAs are constantly faced with a myriad of new features and technologies to learn.
  - DBAs must welcome these changes and embrace any opportunity to learn something new.
- **Increased automation:**
  - The simple, administrative and repeatable tasks of the traditional DBA are becoming more automated as database vendors continually try to create and market the "self-managing" or "self-healing" database.
  - Toolsets like Oracle's OEM are becoming more robust and new features like self-managing tablespaces, undo segments, automated backups, etc. are making physical database maintenance more automatic.

Some specialized areas you may want to focus on …

- **Compliance management**
  - Sarbanes-Oxley, auditing, security, etc.
  - Compliancy issues will continue to dominate public businesses and are even now spreading into the private sector.
- **Best practice frameworks**
  - IT Service Management, ITIL v3, ISO 20000, etc.
  - ITIL is becoming the most widely accepted framework. ITIL focuses on IT service delivery and support including the help desk, incident management, change management, problem management and service-level management.
- **ERP application skills**
  - Oracle, PeopleSoft, SAP, JD Edwards, etc.
  - If your organization is using an ERP package, take advantage and learn it. You must be able to support the application from a software or technical standpoint.
- **Data management**
  - Data modeling, archiving, data mining, retention strategies, consolidation, etc.
  - The amount of data in organizations is growing rapidly as is the business dependency on it. Get involved in any development or data-related strategies in your organization.
- **Hardware and storage**
  - Disk farms, storage area networks (SANs), grid computing, virtualization, etc.
  - Stay on top of new technologies in hardware and storage.
- **High-availability solutions**
  - Standby, failover, clustering, RAC (Real Application Clusters), etc.
  - High availability is an absolute must for growing organizations. These skills also require an advanced understanding of the operating system and disk structures.
- **The next big thing**
  - The key to staying on top of technology is to read about it! The secret to success is to find out where people are going, and get there first!

Your Knowledge

- So ask yourself: Do I have all the specific knowledge that may be required for a specific task whether that's handling a server, a storage device, a database, or an application?
- If not, and if you care in **"Raising the Bar"** in IT industry, then understand that the time has come to for you to focus on improving your knowledge, your skill set and increasing the value you bring to the organization.

- It is no longer about just being a DBA.
- It is no longer about just having a diploma.
- It is no longer about a certification with your name.
- The company is looking for real knowledge, proven experience and professional quality service.

Database: Past, Present, and Future

- **Databases are at the heart of many commercial systems.**
- **Databases serve a couple of purposes.**
  - They compensate for the fact that we have terrible memories - forget something - just look it up
  - They compensate for the fact that we can analyze only small pieces of information at a time -- because they group data for us, aggregate it much faster than any human could.
  - They allow us to share facts and eventually knowledge - for example with Credit databases -- such as TRW all creditors can lookup and arrive at a credit risk for you - they know right away if you are likely to default on a loan
- **Past:** Hierarchical, Network
- **Present:** Relational, Object databases, Object-Relational, Spatial Database, OLAP database, NoSQL databases.
- **Future:**
  - More Intelligent, Robust, Automation (Self- …..)
  - More Choices - Any Place (Cloud, Mobile)
  - More Complex Data Type (Social, Big Data)

Data/Database Trends

- Database -> Data Warehouse (DW)
- Data Analysis -> Data Mining (DM) -> Business Intelligence (BI)
- Data Appliance
  - Oracle: ExaData
  - IBM: PureData
- Cloud Storage
  - iCloud, Dropbox, Google Drive, Skydrive
- Data as a Service
  - DaaS, SaaS, DR as a Service, Archive as a Service, etc.
  - AWS S3, Redshift, Glacier
- Big Data
  - Data Variety: QR Code, RFID, Sensor, Click Stream, Video, etc.
  - Real-time DW (Real-time analysis): Hadoop
- Data Virtualization / Integration (Structured + Unstructured)

It's All About Control of the Data

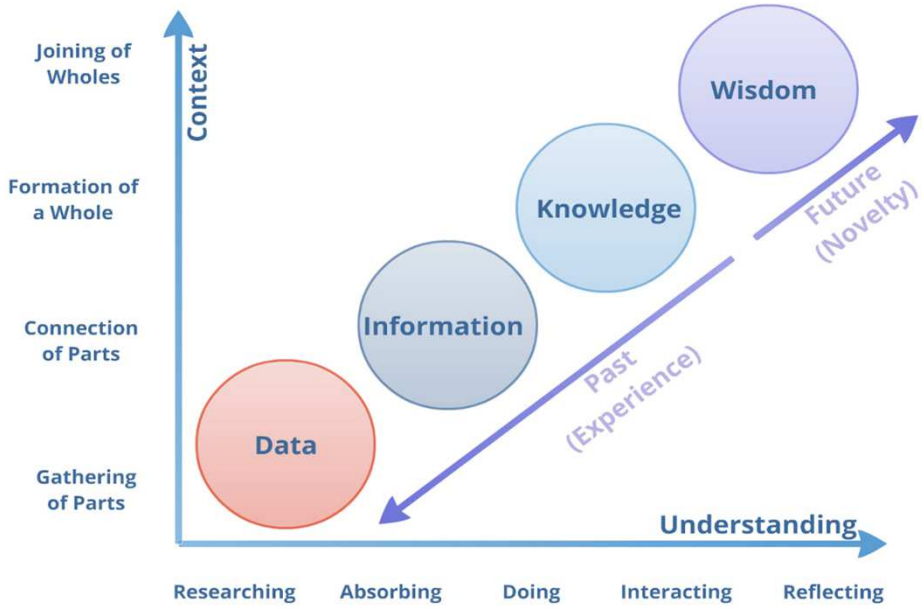
If yesterday is an archive and today is a real-time view, then tomorrow is an idea that encompasses both history and present-day experience.

- How data enters a system, where it resides, how it is processed, and who can access and manage it, as well as who can store and archive data. That's where the real power is.
- Those who know how to control both the archival and current views are most often the ones who come up with significant new ideas and promote business progress.
- These critically important technology trends include cloud services and systems; data centers that use less electricity; the larger-than-life workloads and storage capacities we call "big data"; the increasing use of automation in systems of all kinds; the integration of business intelligence into just about everything; and the ever-growing volume of stored data in all its formats.

eWEEK.COM

http://www.eweek.com/c/a/Data-Storage/IT-2012-Its-All-About-Control-of-the-Data-579216/

Data -> Database -> Information -> Insight



What's the value of information?

Value	Information	Level	Compensation
Efficiency	Facts	Worker	$50,000
Effectiveness	Knowledge	Professional (understand the context)	$100,000
Innovation	Wisdom/vision (Insight)	Executive (make better decisions)	$500,000+

know ….

What happened

What is happening

What will happen

How to make it (not) happen