

# On the Data Explosion and Storage Networking



Dongwon Lee

Assistant Professor  
IST / Penn State

April 21, 2006



## Storage Networking



- Storage Networking is ideal for:
  - High availability
  - High reliability
  - Maximizing storage assets
  - Better I/O performance
  - Centralized storage management
- Their promise to IT was:
  - “Just tell us how much disk space you need, and we’ll give you ideal performance...”
- Do we need large disk space? ⇔ Do we have large data?

Source: [“Storage Bricks for Data Mining”](#), J. Gray and R. Johnson, TechReady, 2006

# How Much Data?

- Kilo < Mega < Giga < Tera < Peta < Exa < Zetta < Yotta < ... (X  $10^3$  per jump)
- How big are these?
  - 1 Gigabyte: A pickup truck filled with books
  - 10 Terabytes: The print collections of the U.S. Library of Congress
  - 200 Petabytes: All printed material
  - 5 Exabytes: All words ever spoken by human beings

Source: <http://www.sims.berkeley.edu:8000/research/projects/how-much-info/>

3

# Data Growth (est. 2003)

- New stored information grew about 30% a year between 1999 and 2002
  - IM generates 5 billion messages a day ~ 750 Gigabytes
  - WWW contains ~ 175 Terabytes
  - More than all the money in the world ~ 1 Petabyte
  - Telephone calls worldwide (on both landlines and mobile phones) contain ~ 17.3 Exabytes

Source: <http://www.sims.berkeley.edu:8000/research/projects/how-much-info/>

4

## Eg: Sensor Data

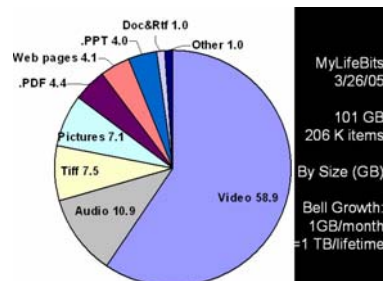
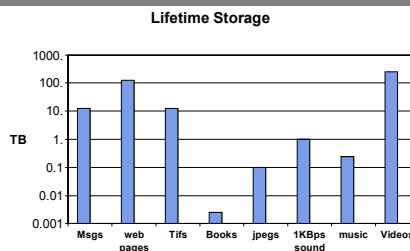
- Earth Observation
  - 15 Petabytes by 2007
- Airplane Engines
  - 1 Gigabyte sensor data/flight
  - 100,000 engine hours/day
  - 30 Petabytes/year
- Medical Images & Information + Health Monitoring
  - Potential 1 GB/patient/year
  - 1 Exabyte/year

Source: [The Information Avalanche](#), J. Gray, Adobe, San Jose, CA, 2006

5

## Eg: MyLifeBits Project

- Gordon Bell @ Microsoft has scanned a lifetime's worth of:
  - articles, books, CDs, letters, memos, papers, photos, pictures, presentations, home movies, and voice recordings
- So far (3/26/05):
  - Only 101GB / 206 K items
  - 1GB/month = 1TB/lifetime



MyLifeBits  
3/26/05  
101 GB  
206 K items  
By Size (GB)  
Bell Growth:  
1GB/month  
= 1 TB/lifetime

Source: <http://research.microsoft.com/barc/MediaPresence/MyLifeBits.aspx>

6

# So, What Does This Mean?

- In near future
  - Data explodes into Tera, Peta, Exa, and more
  - Data will be spitted, replicated, and distributed
  - Special-purpose super-ultra network for data sharing is needed
- New Species:
  - Even an individual will need SAN and NAS
    - ⇔ Personal Storage Networking
  - Will be integrated into DBMS
    - ⇔ DBMS-AN and NA-DBMS