DATA STORAGE TECHNOLOGIES & NETWORKS
(CS C446, CS F446 & IS C446)

LECTURE 05 – STORAGE

Data Explosion –Technology Factors

- Personal / Organizational storage is increasing & affordable
    - Personal Multimedia content (songs, digicam/mobile pictures/videos) is increasing
    - Organizational examples:
        - All Course content is online and growing;
        - All questions answered on the web –volume of a sphere?
        - every technology has its website (dhcp.com, snia.org, …)
        - Every town has its website(s)

Data Explosion –Technology Factors

- Mass storage services are feasible
  - Gmail gives 2+x GB, x is monotonically increasing over time
  - Already > 1.4 billion web-mail users and growing
  - Blog sites, social networking sites, custom service sites
    - {orkut, youtube, myspace, facebook, Flickr, linkedIn}, {{naukri, monster, shine} {shaadi, matrimony, paltalk, fropper} …}
- High volume requirements reduce the cost of storage

==> Low cost storage enables high volume applications

==> vicious cycle?

  - Nothing is ever deleted from websites –local or global!
  - Data is the new entropy!!! Tera < Peta < Exa < Zetta < Yotta
  - Exercise: How many zeroes in "Yotta"?

Data & Storage Characteristics

- **Data**
  - May be transactional or stream data
  - But 80% of data is "semi-structured" or "unstructured":
    - X-Ray image does not have any structure
    - A website (in HTML) is semi-structured
  - Is business critical

- **Storage**
  - Must be highly available
    - With redundancy/replication and across non-local networks
  - Must provide high data rates
  - Must support both streaming and transactional access!

Intelligent storage system

- **Four components**
  - Front end
    - Interface btw host & storage system
    - Consists of front end port [with processing logic – transport protocol like SCSI, Fibre channel or iSCSI] & front end controller [route data from & to cache]
  - Cache
    - Semi-conductor memory where data is placed temporarily
  - Back end
    - Interface btw cache and physical disks
    - Consists of back end ports & back end controllers [communicates with the disks when performing reads and writes. Provides additional but limited data storage]
  - Physical disks
    - Connected to back end with either SCSI or Fibre channel.