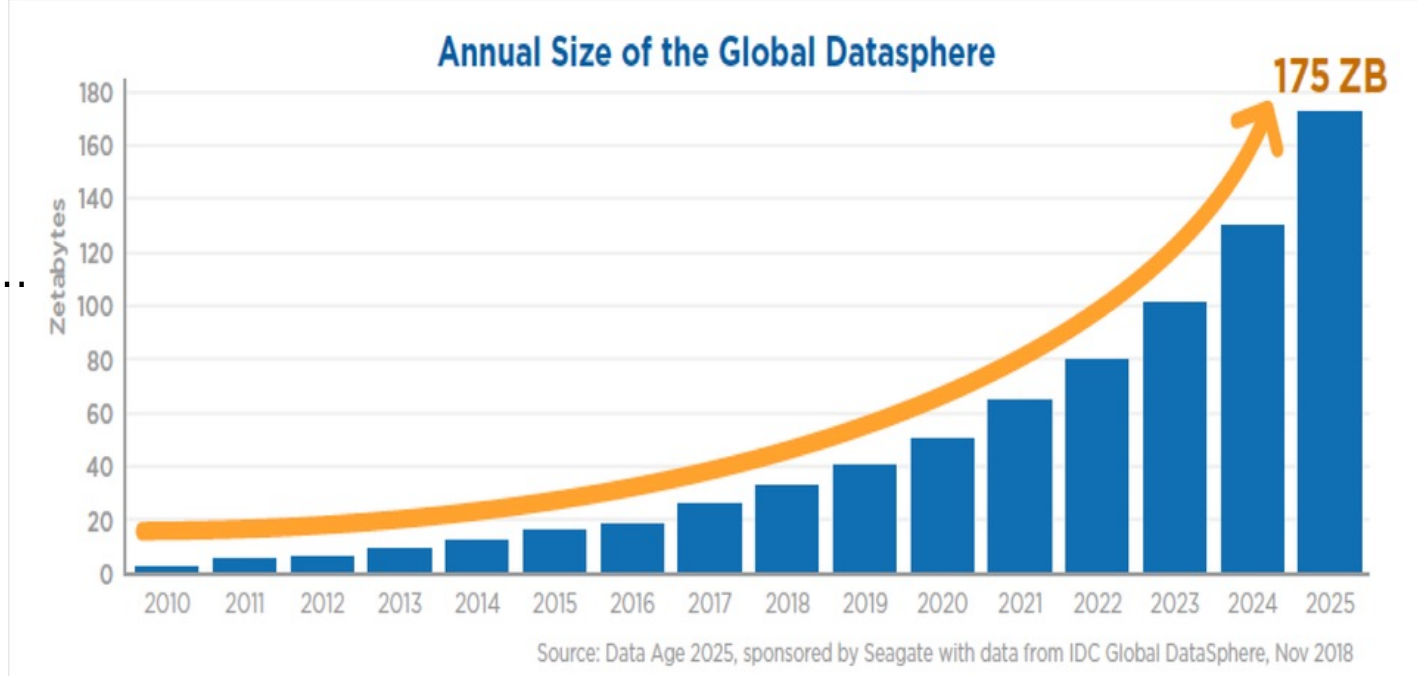
This week

- **Storage in the Cloud**
  - **Big Data**
  - **NFS**
  - **Distributed File Systems (databases next week)**
- **AWS Storage services (ACF Module 7)**
  - Amazon Elastic Block Store (Amazon EBS)
    - Plus some extra notes on Instance Storage
  - Amazon Elastic File System (Amazon EFS)
  - Amazon Simple Storage Service (Amazon S3)
  - Amazon Glacier

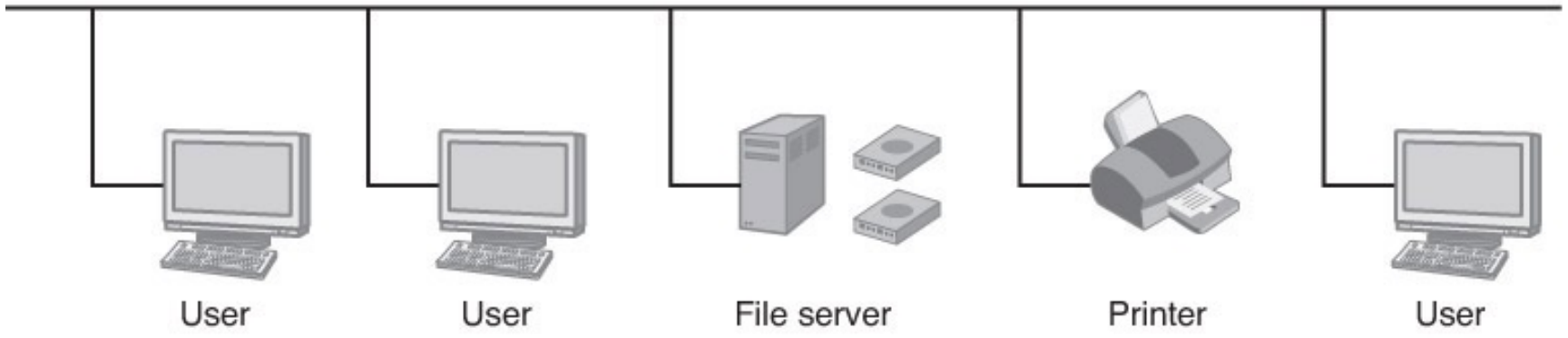
Big Data – Centre of the Universe

- Drivers
  - Internet commerce, Mobile, Social media, IoT and sensors, Science (e.g. biology, astronomy, meteorology), Health, Spooks, …
- Data sizes
  - KB (10^3), MB (10^6), GB (10^9), TB (10^{12}), PB(10^{15}), exabyte (EB, 10^{18}), zettabyte (ZB, 10^{21}), yottabyte (YB,10^{24})
  - Everyday: Facebook 10T, Twitter 7T, Youtube 4.5T+
- 4Vs: volume, variety, velocity, and veracity



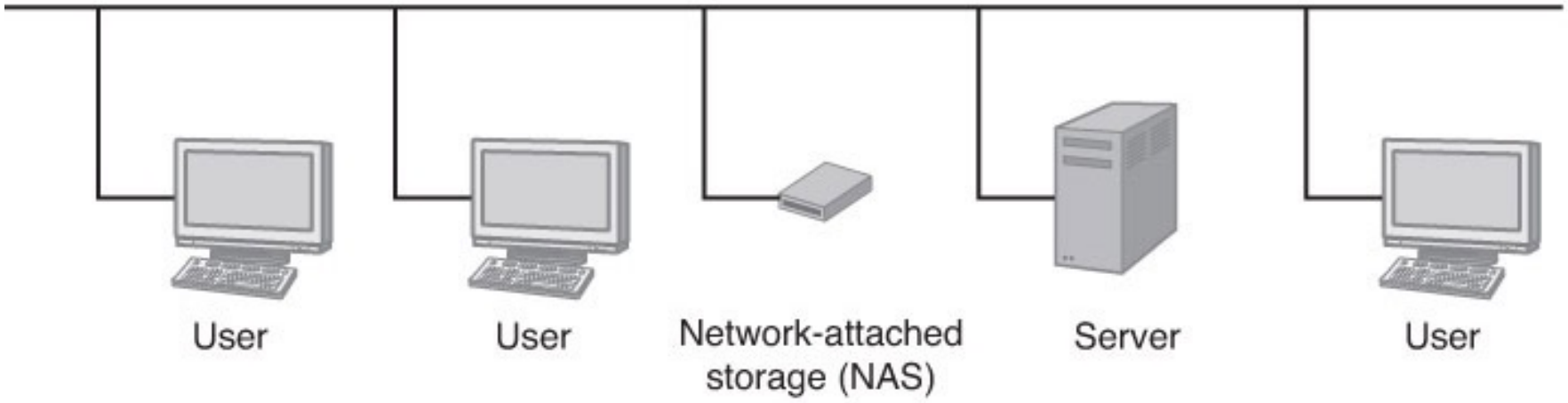
Network Storage Began with File Servers

Years ago, local-area networks used special servers, called file servers, to support file sharing, file replication, and storage for large files.



Network-Attached Storage (NAS)

Plug directly into the network.



Advantages of NAS

- **Reliability**: A NAS device typically provides advanced data striping across multiple volumes within the device. If one (or more) volumes fail, the data striping would maintain the data and allow reconstruction of the file contents.
- **Performance**: Because the NAS device did not run a complete operating system, the hardware had less system overhead, which allowed it to outperform a file server.
- **Compatibility**: NAS devices normally support common file systems, which, in turn, make them fully compatible with common operating systems.
- **Ease of performing backups**: NAS devices are commonly used for backup devices. Within a home, for example, all devices can easily access and back up files to a NAS device.

Cloud-Based Storage

- Cloud-based data storage is the next step in the evolution of NAS devices.
- Across the web (the cloud), many providers offer data storage that resides in the cloud.
- Data may be accessible as follows:
  - Through a web browser interface
  - Through a mounted disk drive
  - Through a set of API (application program interface) calls

Advantages of Cloud-Based Storage

- **Scalability**: Most cloud-based data storage providers let users scale their storage capacity (up or down) to align with their storage needs.
- **Pay for use**: With most cloud-based data storage facilities, users pay only for the storage (within a range) that they need.
- **Reliability**: Many cloud-based data storage facilities provide transparent data replication.
- **Ease of access**: Most cloud-based data storage facilities support web-based access to files from any place, at any time, using a variety of devices.
- **Ease of use**: Many cloud-based data storage solutions let users map a drive letter to the remote file storage area and then access the files through the use of a logical drive.

Disadvantages of Cloud-Based Storage

- **Performance**: Because the cloud-based disk storage devices are accessed over the Internet, they will never be as fast as local drives.
- **Security**: Some users will never feel comfortable with their data in the cloud.
- **Data orphans**: Users may abandon data in cloud storage facilities, leaving confidential private or company data at risk.

Cloud-Based Block Storage

- In the simplest sense, a block of data storage is a fixed-sized sequence of bits. The size of the block normally corresponds to an underlying unit of storage on the **cloud-based block storage device**.
- Some applications work with very large blocks of data, the format of which has meaning only to the application itself—meaning that the data may not map well to storage within a file system or database.

File Systems

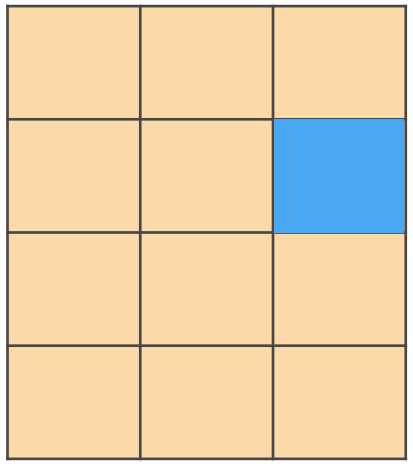
- Operating systems exist to allow users to run programs and to store and retrieve data (files) from one user session to the next.
- Within the operating system, special software, called the file system, oversees the storage and retrieval of files to and from a disk.
- When you copy a file, delete a file, or create and move files between folders, the file system is performing the work.
- Initially, file systems allowed users to manipulate only local files that reside on one of the PC's disk drives.
- As networks became more prevalent, so too did network operating systems, which allow users and programs to manipulate files residing on a device across the network.

Real World: Hadoop Distributed File System

- Apache Hadoop is an open source project, the goal of which is to support reliable, scalable distributed computing.
- Part of the project includes the Hadoop Distributed File System (HDFS), a Java-based file system that is well suited for cloud-based storage.
- HDFS is designed to be highly fault tolerant and robust to maintain operation in the event of a device failure.

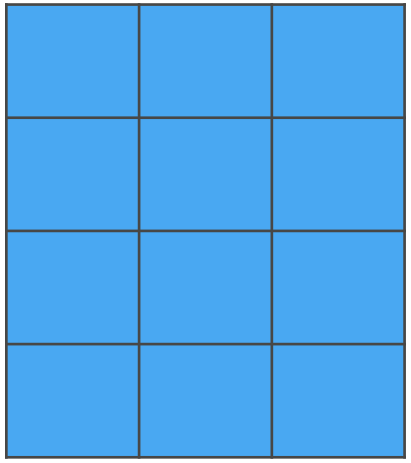
AWS Storage Options: Block vs. Object Storage

What if you want to **change** one character in a 1-GB file?



Block Storage

Change one block (piece of the file) that contains the character



Object Storage

Entire file must be updated

This week

- Storage in the Cloud
  - Big Data
  - NFS
  - Distributed File Systems (databases next week)
- **AWS Storage services (ACF Module 7)**
  - Amazon Elastic Block Store (Amazon EBS)
    - Plus some extra notes on Instance Storage
  - Amazon Elastic File System (Amazon EFS)
  - Amazon Simple Storage Service (Amazon S3)
  - Amazon Glacier