CLOUD COMPUTING CONCEPTS

with Indranil Gupta (Indy)

DISTRIBUTED FILE SYSTEMS

Lecture A

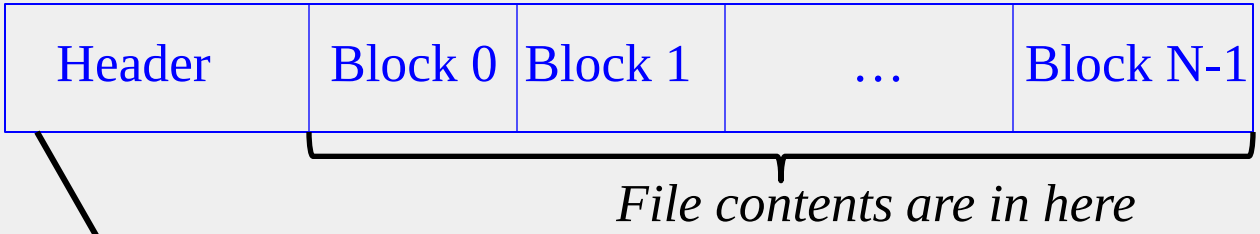
FILE SYSTEM ABSTRACTION

File System

- Contains files and directories (folders)
- Higher level of abstraction
  - Prevents users and processes from dealing with disk blocks and memory blocks

File Contents

- Typical File



  - Timestamps: creation, read, write, header
  - File type, e.g., .c, .java
  - Ownership, e.g., edison
  - Access Control List: who can access this file and in what mode
  - Reference Count: Number of directories containing this file
    - May be > 1 (hard linking of files)
    - When 0, can delete file

What about Directories?

- They’re just files!
- With their “data” containing
  - The meta-information about files the directory contains
  - Pointers (on disk) to those files

Unix File System: Opening and Closing Files

- Uses notion of *file descriptors*
  - Handle for a process to access a file
- Each process: Needs to open a file before reading/writing file
  - OS creates an internal datastructure for a file descriptor, returns handle
- *filedes*=open(*name*, *mode*)
  - mode = access mode, e.g., r, w, x
- *filedes*=creat(*name*, *mode*)
  - Create the file, return the file descriptor
- close(*filedes*)

Unix File System: Reading and Writing

- *status*=read(*filedes*, *buffer*, *num_bytes*)
  - File descriptor maintains a *read-write pointer* pointing to an offset within file
  - read() reads *num_bytes* starting from that pointer (into buffer), and *automatically advances pointer by num_bytes*
- *status*=write(*filedes*, *buffer*, *num_bytes*)
  - Writes from buffer into file at position pointer
  - Automatically advances pointer by *num_bytes*
- *pos*=lseek(*filedes*, *offset*, *whence*)
  - Moves read-write pointer to position offset within file
  - *whence* says whether offset absolute or relative (relative to current pointer)

Unix File System: Control Operations

- *status*=link(*old_link*, *new_link*)
  - Creates a new link at second arg to the file at first arg
  - Old_link and new_link are Unix-style names, e.g., "/usr/edison/my_invention"
  - Increments reference count of file
  - Known as a "hard link"
    - Vs. "Symbolic/Soft linking" which creates another file pointing to this file; does not change reference count
- *status*=unlink(*old_link*)
  - Decrements reference count
  - If count=0, can delete file
- *status*=stat/fstat(*file_name*, *buffer*)
  - Get attributes (header) of file into *buffer*

Distributed File Systems (DFS)

- Files are stored on a server machine
  - Client machine does RPCs to server to perform operations on file

Desirable Properties from a DFS

- Transparency: client accesses DFS files as if it were accessing local (say, Unix) files
  - Same API as local files, i.e., client code doesn't change
  - Need to make location, replication, etc. invisible to client
- Support concurrent clients
  - Multiple client processes reading/writing the file concurrently
- Replication: for fault-tolerance

Concurrent Accesses in DFS

- One-copy update semantics: when file is replicated, its contents, as visible to clients, are no different from when the file has exactly 1 replica
- At most once operation vs. At least once operation
  - Choose carefully
  - At most once, e.g., append operations cannot be repeated
  - *Idempotent* operations have no side effects when repeated: they can use at least once semantics, e.g., read at absolute position in file

Security in DFS

- Authentication
  - Verify that a given user is who they claim to be
- Authorization
  - After a user is authenticated, verify that the file they're trying to access
  - Two popular flavors
  - **Access Control Lists (ACLs)** = per file, list of allowed users and access allowed to each
  - **Capability Lists** = per user, list of files allowed to access and type of access allowed
    - Could split it up into capabilities, each for a different (user,file)

Let's Build a DFS!

- We'll call it our "Vanilla DFS"
- Vanilla DFS runs on a server, and at multiple clients
- Vanilla DFS consists of three types of processes
  - *Flat file service*: at server
  - *Directory service*: at server, talks to (i.e., "client of") Flat file service
  - *Client service*: at client, talks to Directory service and Flat file service

Vanilla DFS: Flat File Service API

- Read(*file_id*, *buffer*, *position*, *num_bytes*)
  - Reads *num_bytes* from absolute *position* in file *file_id* into *buffer*
    - *File_id* is *not* a file descriptor, it's a unique id of that file
  - No automatic read-write pointer!
    - Why not? Need operation to be *idempotent* (at least once semantics)
  - No file descriptors!
    - Why not? Need servers to be *stateless*: easier to recover after failures (no state to restore!)
  - In contrast, Unix file system operations are neither idempotent nor stateless

Vanilla DFS: Flat File Service API (2)

- write(*file_id*, *buffer*, *position*, *num_bytes*)
  - Similar to read
- create/delete(*file_id*)
- get_attributes/set_attributes(*file_id*, *buffer*)

Vanilla DFS: Directory Service API

- *file_id* = lookup(*dir*, *file_name*)
  - file_id can then be used to access file via Flat file service
- add_name(*dir*, *file_name*)
  - Increments reference count
- un_name(*dir*, *file_name*)
  - Decrements reference count; if =0, can delete
- *list*=get_names(*dir*, *pattern*)
  - Like ls –al or dir, followed by grep or find

Can We Build a Real DFS Already?

- Next: Two popular distributed file systems
  - NFS and AFS