



**CS 686:** Special Topics in Big Data

# DFS Project Test Cases

Lecture 11

# File Sizes

---

- To start out, I will generate a couple files:
  - $0.5 * \text{chunk size}$
  - $1.75 * \text{chunk size}$
- A few files ranging from 1 – 10 MB
- One larger file, ~100 MB
- We will note the MD5 checksums of these files before storing them in your DFS
  - Checksums should match when retrieved

# Storage Test Process

---

1. Store the files
2. Use the client **list** command to show where they are located (what storage nodes)
3. We will ssh to these nodes and verify the file's location on disk as well as their sizes
4. Retrieve the files
  - Check to make sure the MD5 checksum is the same

# Disk Space

---

- Your client disk space command also helps during testing
- We'll check the space before and after storing the files

# Fault Tolerance (1)

---

- First step: terminating a storage node
  - **Important:** this does not mean “gracefully” shutting the node down with a command or similar
- Your controller node should react appropriately by printing an error message and starting chunk replication

# Fault Tolerance (2)

---

- Next step: locating the **primary** replica for a chunk, corrupting the chunk on disk, and retrieving the file
- The node with corrupted data should notify the client and repair its copy of the file
- The client should retrieve a good copy of the file

# Heartbeats

---

- We should be able to terminate and restart the controller as well
- Heartbeats will help the controller reconstruct its state
- Not necessarily a fault tolerance test, but will determine if your heartbeats work as intended