

Project Topics (Broad outline) for IDS630E

1. Distributed Log System

- a. Logs collected at different server
- b. API to log events with arbitrary info, send/receive of messages
- c. Different types of query: type, events in time range (others, plus combinations with AND and OR)
- d. Forming process traces from distributed logs

2. Distributed Storage

- a. Web based interface to offer storage service for files
- b. Files to be stored at different servers
- c. Fast search to find a file
- d. Should replicate for fault tolerance, handle writes to replicas

3. Distributed Chat Server

- a. Central server only to maintain group information
- b. Multiple groups, changes possible at anytime.
- c. Chat between group members is totally distributed (no central server)
- d. Ordering to be maintained
- e. Strategies for handling group join/leave, that happens on central server but needs to be known by all members
- f. Voice chat if possible

4. P2P File Sharing

- a. Replicated Peers to store location of files and answers file search queries
- b. Normal node registers/searches in peers
- c. Replication among peers
- d. Files shared directly between nodes storing
- e. Known super-peer to find peer addresses
- f. New peer selection if number of peers is down too much
- g. Incentives/penalties to foster collaboration

5. Online Travel Agency

- a. Web based interface for travel planning
- b. Agency handles multiple hotels and multiple airlines spread over multiple cities
- c. User can ask for only hotels or only airtickets or both
- d. Trips can span over cities
- e. A trip is confirmed if all sectors are confirmed
- f. Provision to cancel

6. Online Retail Store

- a. Sells goods from multiple vendors (access to/from multiple database)
- b. Replicated servers to store and search who sells what and to log transactions
- c. Web based interface, load balanced to servers

7. Implement RPC framework

8. Comparison of Checkpointing/Recovery Protocols

- a. Take 3-4 algorithms (asynchronous and synchronous) and implement them (decide on algorithms based on talk with TA)
- b. Compare their performances under different parameter changes

9. Dynamic Load Balancing for Heterogenous Hadoop clusters

10. Twitter Sentiment Analysis using Map and Reduce

