Project Topics

Spring 2014
CSci 8980
Project Topics

• I have listed here some ideas for project topics.

• Your project need to be confined to these topics

• Ultimately what pick to do should be based on:
  – Your genuine interest
  – You background and qualification
  – Desire to learn deeper in that topic

• Define the scope and milestones in a realistic and reasonable manner.
  – You should be able to achieve at least some minimal goals
  – If you plan to work alone, then make sure that the scope is properly defined.
1. Reliable Group Communication

- Implement causal and atomic broadcast
- Include group membership management
- Include support for fault-tolerance and reliable delivery
- Measure performance and scalability
- Appropriate for a group of 3
2. Impact of Failures on MapReduce Applications

- Study the impact of failures on performance of MapReduce applications
- Failure injection at different stages to evaluate how performance is affected
- Appropriate for a group of 2
3. Performance Evaluation of Key-Value Data Stores

Hbase, Cassandra, MongoDB, CouchDB etc

• Benchmarking of for a range of operations.
• Determine how scaling out affects performance
• Impact of failures on performance
• Appropriate for a group of 2
4. Performance Evaluation of HBase

- Dynamic and automatic scaling of HBase applications.
- Continuously measure load and performance, and then scale out when it crosses some bound.
- Appropriate for a group of 2
5. Implementation Paxos Protocol

• Implement Paxos with dynamic group configuration
• Use this protocol for building a resilient service
• Measure performance under process crashes and recovery
• Appropriate for a group of 2
6. Fault tolerant data storage in DHT Systems

• Develop and integrate a suitable data replication management and fault-tolerance protocol in either:
  – Chord or Pastry
  – FreePastry is an open-source DHT available but does not have support for fault-tolerance
  – How to support range-queries in hash-based systems

• Appropriate for a group of 2 students
7. Investigation of Trigger Mechanisms in HBase

- Study how triggers can be used at application
- Possible use for alerts, notifications, event stream processing
- What is the performance impact of adding triggers.
8. Impact of Replication and Failures in Hbase/Cassandra Applications

• Study the impact of failures on performance of HBase or Cassandra applications
  – Inject failures of region servers

• Can we define difference replication levels for different tables at the HDFS level and tune performance?
9. Scalable Architectures for Publish-Subscribe

• Explore publish-subscribe architectures using cloud data storage and computing models.
10. Management of Spatial and Temporal Data in Cloud Storage Systems

• Investigation of efficient techniques for managing spatial and temporal data on key-value based storage systems