CS 686: Special Topics in Big Data Protocol Buffers & Project Hints

Lecture 5

Today's Agenda

- Pointers on paper evaluations
- A quick tour of Protocol Buffers, project hints

CS 686: Big Data

Paper discussion

9/1/17

Today's Agenda

- Pointers on paper evaluations
- A quick tour of Protocol Buffers, project hints

CS 686: Big Data

Paper discussion

9/1/17

Paper Evaluations

- Look great so far!
- A note on reading the papers: don't worry if you don't understand 100% of the paper
 - Some (most) papers leave out details
 - If something seems contradictory or unsubstantiated, bring it up in our discussions!
- You're certainly not locked in to answering every question in the template

CS 686: Big Data

Today's Agenda

- Pointers on paper evaluations
- A quick tour of Protocol Buffers, project hints

CS 686: Big Data

Paper discussion

9/1/17

9/1/17

Protocol Buffers

- There's a protobuf example repo linked from the assignment spec now
- Using them tends to be a bit verbose, but they are a good alternative to Java serialization
- We'll discuss messaging/serialization/etc. in general next week

Protobuf Workflow

- Add the library jar to your project
- Create .proto files that define your communications
- Compile them with **protoc** to Java classes
 - This is where you could also generate Python, Scala, C#, etc.

CS 686: Big Data

 Use the classes to encapsulate your communications

A Few Hints (1/2)

- You may want to put all your message types in one (or two) .proto files
- An outer class is generated for each file, with each message you define represented by inner classes
 - my_proto.proto → MyProto.java
- Since your components will send/receive multiple message types, you also need a message Wrapper
 - Gives each message a unique ID

CS 686: Big Data

A Few Hints (2/2)

9/1/17

/1/17

- The wrapper allows the library to figure out what kind of message you're receiving
 - Test with hasBlahMsg() method
- Final note: logging
 - Use a logging framework / println() to make it obvious what your system is doing

Today's Agenda

- Pointers on paper evaluations
- A quick tour of Protocol Buffers, project hints
- Paper discussion

Some of Your Tweets

- "By using HDFS, Yahoo! can provide more reliable service base on HDFS file content replication strategy and network topology."
- "Hadoop Distributed File System: Redundancy to the rescue"

CS 686: Big Data

- "HDFS has answers to most of the questions posed by DFS, but security remains a concern!!"
- "Achieve great things with junk computers and lose 2/3 of your storage capacity with HDFS"

CS 686: Big Data

Discussion (1/2)

9/1/17

- We'll split into groups
- Discuss the paper in general within your group
- Next, your job is to analyze a specific portion of the paper
 - Diagram the component(s) and concepts from your section(s)
 - Take note of your group's critical analysis and questions
 - Put your group members' names on it!

Discussion (2/2)

- Each group presents their primary findings
- Then we'll merge back and discuss as a class

CS 686: Big Data

Groups

9/1/17

9/1/17

- Group 1 Sections II-A, II-B
- Group 2 Sections II-C, II-D
- Group 3 Sections II-E, II-F, II-G

- Group 4 Section III-A
- Group 5 Sections III-B, III-C