

Research Topic Term Paper Project Description (group project)

1 Overview

For this two-part assignment, each student group will research a topic related to this course and present that topic as a class presentation and as a written paper.

2 Purpose

The assignment purpose is to have students know some important topics in detail, through research and presentation. Such familiarity is important in job interviews, as well as computing work. The topics are complimentary to the textbook topics. Please ensure that your research goes beyond that presented in the class discussion and readings.

3 How to complete the topic term project

1. Send three topic choices to the instructor, who will confirm one
2. Research the topic
3. Create a PowerPoint presentation
 - 3.1. Present the PowerPoint presentation, according to the scheduled list of topics (see the syllabus)
 - 3.2. Upload your presentation to the course web site
4. Create a term paper for the topic
 - 4.1. Upload the Word document term paper on the web site, at the end of the term (see the syllabus)

4 How to research the topic

Your class peers are the intended audience for this article. They will want to know the details of the method you describe, as well as “why do I care?” ☺

In describing your topic, please consider the following aspects:

- Consider the common format of: what is it, what problem it solves, how its used in industry, why it's valuable.
- Describe both theory and computer tool support.
- Describe implications for practitioners, i.e., does the method really help?
- Include at least **four** academic references (peer-reviewed articles) in your research.
- Do not summarize (substantially) the course materials; assume them as background and add new materials.
- Relate the topic material to the course materials

4.1 Literature Review (for the references)

Ensure that you have some **peer reviewed** articles. These articles are found on scholar.google.com or using our library.gsu.edu and check the box labeled “Scholarly (Peer Reviewed) Journals”.

Search for articles using keywords at these sites.

1. Scan the web
 - a. www.google.com
2. Scan the web using scholar search engines
 - a. <http://scholar.google.com/>
 - i. Set the [Google Scholar Preferences](#) to

1. Show library access links for Georgia State University
2. Show links to import citations into EndNote
- b. <http://academic.live.com/>
3. Scan using library databases (@GSU)
 - a. <http://www.galileo.usg.edu>
 - b. In particular, the following databases
 - i. [ABI/INFORM Complete](#)
 - ii. [ACM Digital Library](#)
 - iii. [IEEE Xplore](#)

4.2 How the topic presentation is evaluated

Presentations are typically about 15 PowerPoint slides. The following rubric is used to evaluate the topic presentation.

Criteria	required	Points
Within time limit	15 minutes	
Academic peer-reviewed references	4	12
Introduction		5
Intermediate topics		5
Conclusion		5
Well presented		10
Informative		10
No more than 20% quotes, videos, etc		
Total		100

4.3 How the topic term paper is evaluated

Ensure that your article:

- Style, grammar, and spell checked (in Word)
- Minimum of **1,500 words**; about 3 pages (single spaced)
 - figures, tables, and references do not included in page count
 - Liberal use of quotations are allowed; however, the **sources must be referenced**.
No more than 20 percent of the paper word count can be quotations
- References to articles using a standard format (e.g., Chicago); see EndNote.

The following rubric is used to evaluate the term paper artifact.

Criteria	required	Points
Minimum words	1,500	x/1,500
Academic peer-reviewed references	4	12
Introduction		5
Intermediate headings/sections		5
Conclusion		5
Well written		10
Informative		10
No more than 20% quotes		
Total		100

5 Research Topics

Determine your preference for [three \(3\)](#) presentation topics. Email your choices early! First comes, first served. You may also define your own research topic, but obtain approval from the instructor.

Consider these overlapping categories of research topics:

5.1 Business case study

Review one or more cases where an organization used elements of the Hadoop ecosystem (e.g., MapReduce, Pig, Spark, etc.). Ideally, select one application within one organization and provide a detailed description. However, if you do not have access to sufficient information, then you can do more than one application or more than one organization.

In your report, where possible, present:

- **the context before and after the use of Hadoop (or some kind of Big Data technology)** —especially some description of the new capability derived from Hadoop
- the Hadoop components, processing, and any details available
- any issues that were overcome to enable the use of Hadoop
- generalize the case, to suggest what kinds of organization or problems may likewise benefit from a Hadoop solution

[Organizations that benefit from Hadoop](#) include Google, Yahoo!, Facebook, Netflix, NSA, Spotify, Amazon, Expedia, LinkedIn, and many others.

5.2 Technical case study

Review how to use a Hadoop (Big Data) component or write a Hadoop application. These are typically blog posts, such as found on the [Cloudera Engineering Blog](#) or on a GitHub.com site. To qualify, the component

- cannot be one that we have covered in class
- must be substantial, such that you can present its technical details and write a few pages about it
- cannot be the same as the application that you extend for a programming assignment deliverable

In your report:

- Give a detailed **How To** explaining of how to use important features, as well as some associated theory of the problem it solves
- Summarize results of your own person experience of using the component (not just book or blog descriptions)

References

- This report is likely to only have book and blog references (no academic references are required)

Examples of Big Data components include Pig, Hive, HBase, Spark streaming, Sqoop, Oozie, ZooKeeper, Ambari, Mahout, Grafana, or specialized extensions such as provided by Netflix, Spotify, etc.

Example of important Spark related API include TensorFlow, Keras, Pipelines, Kafka, Apache Storm.

A simple way to find a important technology is to go to Google, Amazon, or Microsoft and search for their newest cloud technology offering. Be sure to check the date of the technology, as their newer offerings are often intended to replace existing offerings.

5.3 Container case study

Review how to use containers in the Big Data context. Docker is introduced in this course. So, the presentation must go beyond an introduction of Docker, or more likely, present Docker or Kubernetes related technologies and strategies. These include:

- Docker details: Docker compose, services, stacks
- Design of service stacks: microservices architecture
- Container management: Swarm, Kubernetes, Amazon ECS

- Resource management: Mesos, Yarn

Note that many of these topics have GitHub code demos, video demonstrations, and some have live online demonstrations, which can be presented in class.

5.4 Best big data company to invest in?

Review one company or compare and contrast companies that specifically develop a Big Data technology. This is like the technical case study above. However, look to see if a company's value is mostly derived from a Big Data technology. Some company (or subsidiaries) that are developing important Big Data technologies include:

- Data management
 - Snowflake
 - Teradata
 - Amazon's Redshift Spectrum (part of Amazon, so may be used as a comparison)
 - Azure Synapse
 - Firebolt
- Data services
 - Cloudera (startup) vs Google, Amazon, Microsoft, etc.
- Prediction
 - Palantir
- Data acquisition, monitoring
 - Splunk
 - Datadog

When analyzing the company, determine what is unique about their technology and why it will propel the company to profits, especially compared to their competitors.