CPS 516: Data-intensive Computing Systems

Shivnath Babu

Grading: Total 50 Points

Note: Deadline dates may be changed by a few days

- Part 0: Starts on Jan 7 and Due on Jan 21
  - 10 points
- Part 1: Starts Feb 1 and Due on Feb 21
  - 10 points
- Part 2: Starts Feb 22 and Due on March 10
  - 10 points
- Part 3: Starts March 11 and Due on April 10
  - 20 points

Project Part 0: Due in 2 Weeks

- For every single system listed in the "Data Platforms Map", give as a list of succinct points:
  - Strengths (with numbered references)
  - Weaknesses (with numbered references)
  - References (can be articles, blog posts, research papers, white papers, your own assessment, …)

- Your own thoughts only. Don't plagiarize. List every source of help. We will enforce honor code strictly.

- Submit on github (md format) into repository given by TA

- Outcomes: (a) Score out of 10; (b) Project leader selection.

Project Parts 1, 2, 3

- Shivnath will work with project leaders to assign one system per project. Will also try to have one mentor per project

- Each student will join one project. Project starts Feb 1

- Part 1: Feb 1 – Feb 21
  - Install system
  - Develop an application workload to exercise the system
  - Run workload and give demo and report

- Part 2: Feb 22 – March 10
  - Identify system logs/metrics and other data that will help you understand deeply how the system is running the workload
  - Collect and send the data to a Kafka/MySQL/ElasticSearch routing and storage system set up by Shivnath. Give demo and report

- Part 3: March 11 to April 10
  - Analyze and visualize the data to bring out some nontrivial aspects of the system related to what we learn in class. Give demo and report