15-619
Cloud Computing

15-619 Project Overview

Recitation 8
Mar 4\textsuperscript{nd} and Mar 6\textsuperscript{th}, 2013
619 Project Overview

• For students who enrolled in 15-619
  – Since 12 units, an extra project worth 3-units.

• A team project for a team no more than 3 students.
  – The team registration was due on Sun, Mar 2\textsuperscript{nd}.
  – Contact us immediately if you have not formed a team yet.
What is the project about

– A client has approached your company and several other companies to compete on a project to build a web service for Twitter data analysis.

– Your company needs to build a web service that meets the requirements of the customer.

– It will be a bidding process, meaning you will compete with other companies.

  • Other companies = other 15-619 groups.

• The link for the project description can be found on Piazza

  – This is an active document, keep checking for updates.
619 Project Overview

• What you are given:
  – A Twitter dataset stored on S3
  – Customer requirements
    • Business Requirement (Query types)
    • Minimum Performance Requirement (Throughput)
  – Cost Constraints
    • If you overrun your budget you will receive a disqualification penalty
    • Design and plan before you execute
619 Project Overview

• What you are asked to do
  – Build a web service
    • Preprocess the data and store them
    • Process queries from customer and send a response
  – Explore solutions to win the bid
    • Front End and Back End

Details are articulated in the project writeup.
619 Project Evaluation

- Three phases
  - Each phase includes development, evaluation, exploration, optimization.

- Live test period
  - A comparison of cost, throughput between all submitted web services
Tag Your Instances

• Tag all your instances as required!

• Why Tagging?
  – To categorize instances for different purposes
    • Production / Development / Testing...
    • OLI Projects / 15-619 Project
  – To have a clearer billing report!!
    • For cost control and monitoring purposes in this class
619 Project Phase 1

• What you have
  – A small dataset (0.8 million tweets, 2.3 GB)
  – Two simple queries
    • Q1 Heartbeat
    • Q2 database query on userid and tweet_time

• What you have to do
  – Front end system development
  – Two back end databases
    • MySQL
    • Hbase
619 Project Phase 1

• What we expect you to learn
  – Hands-on experience of front end system
  – Working experience in HBase and MySQL
  – Design of database schema
  – A solid skill set for Phase 2 & Phase 3
  – Cost really matters!!!

• Start Really Early!!

• A good start leads to the final success!
Q & A