CMPE282 – Cloud Services

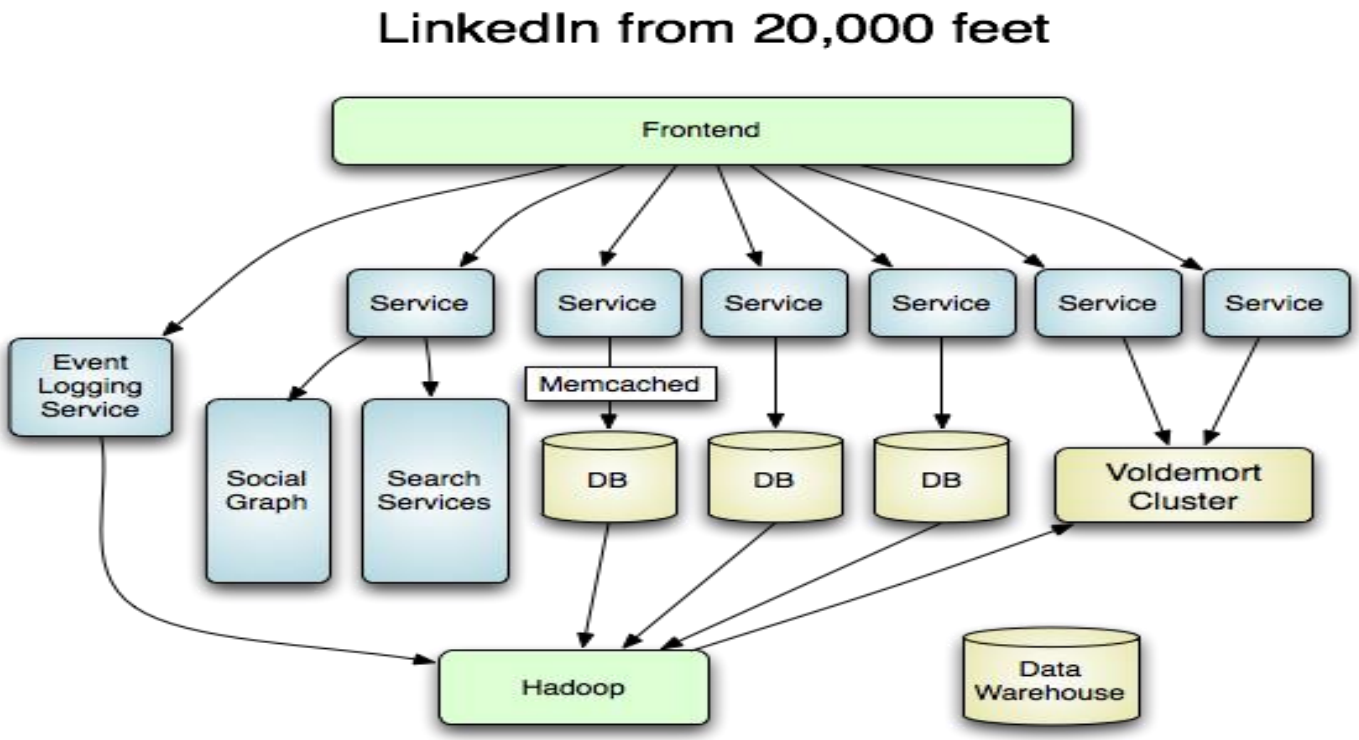
Project

Due date: December 1 2014

Midterm Check: Your 1 page midterm document is due by Nov 5th

Team-Based Class Project (up to 6 people)

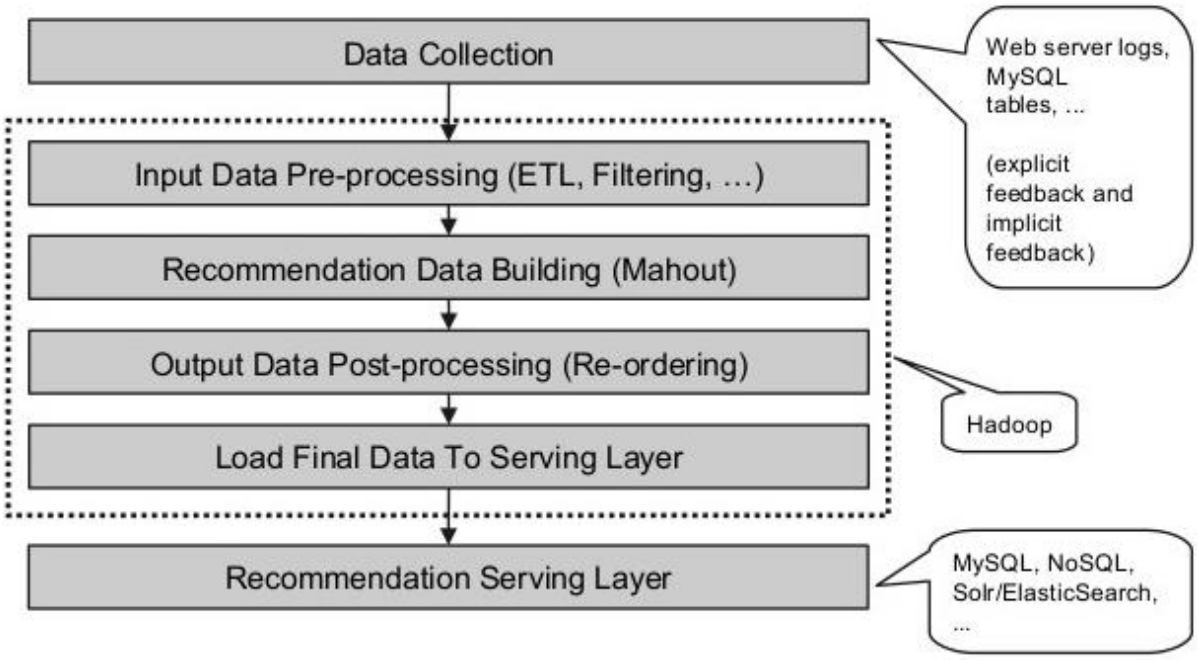
In this project, **you will design and implement an item based recommendation** (using REST) that contains data analysis (Using Hadoop and Mahout). You will build a working prototype system.



The goal of the project is to create a digital music recommendation engine for Amazon.

Project descriptions:

Architecture:

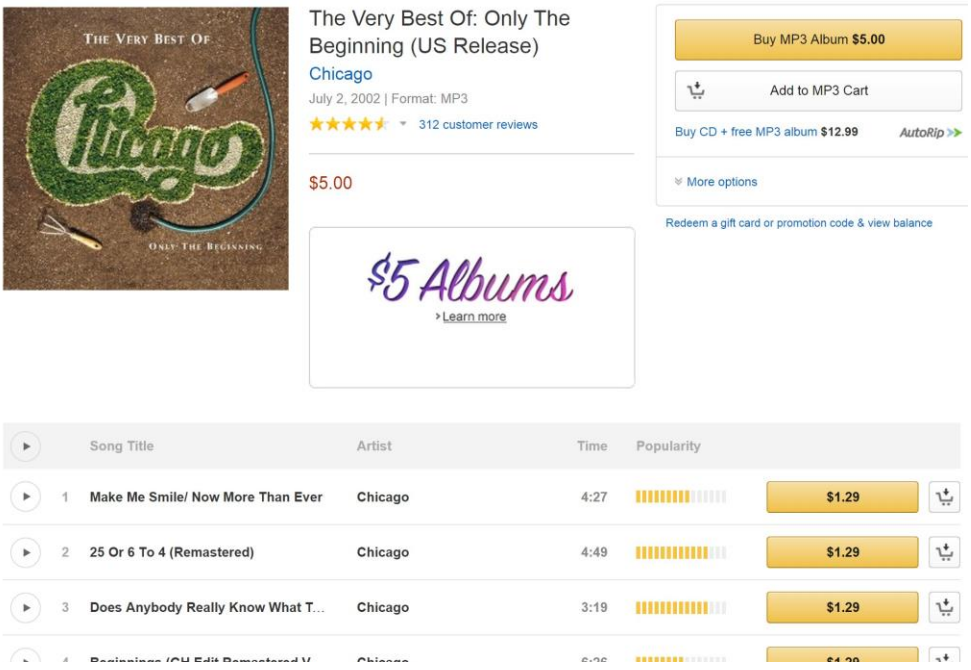


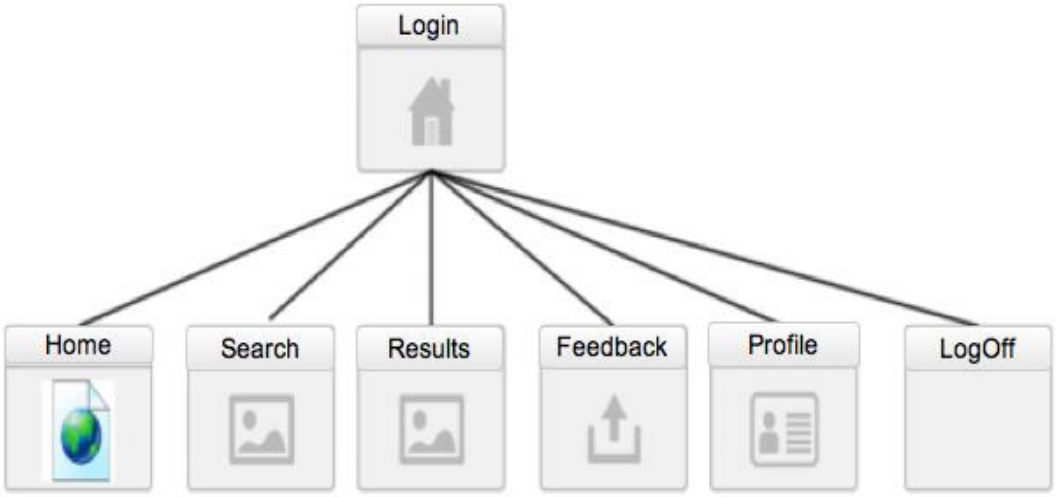
Your team must implement login, input your information, and search for the recommendation on Yahoo data-set: **http://webscope.sandbox.yahoo.com/catalog.php?datatype=c.** The required data-set is C15.

Recommendation database are built using Mahout.

1. Convert your Lab1 to include digital music. Run them in AWS.

Example Screenshot for the Project Deliverable:



2. 

Recommendation (Run it on AWS)

1. Make Top-N recommendation based on items liked by a user
2. Build collaborative filtering similar to Amazon's product recommendation. Construct co-occurrence matrix and find the most co-occurred items as a recommendation.
3. Build user based recommendation Build personal recommendation and represent other users as feature vectors and calculates CooccurrenceCountSimilarity, LogLikelihoodSimilarity, CosineSimilarity, PearsonCorrelationSimiliarity, and EuclideanDistanceSimilarity (K-mean) using Mahout. Make a recommendation based on user data. Final output should be looked like userID, (itemid, Score), (itemID, Score). …
4. Make your own recommendation and build personalized recommendation described in the lecture notes.

Algorithm

```
for every item i that u has no preference for yet
  for every item j that u has a preference for
    compute a similarity s between i and j
    add u's preference for j, weighted by s, to a running average
return the top items, ranked by weighted average
```

Figure 1: Mahout's Item Recommendation Algorithm

Project deliverables: (submit them to Canvas)

Midterm document (5%)

A design document, architecture and functions, API design, recommendation algorithm, analysis mapping and high-level specification, deployment, provision, reporting. (20%)

A project demo and program source codes (20%)

Deliverables Required:

- Submissions shall include source codes in zipped format.
- Project directory must include the group ID/Names
- project report into one word format
- Do not submit binaries, class files, or supporting libraries
- Project report
    - Introduction: state your goals, purpose of system,
    - System Design: Describe your chosen system design
    - Architecture: describe architectural components
    - Algorithms: Describe algorithms used. Why did you use these algorithms?
    - Results: Screen image captures of each run
    - Analysis: What was the recommendations? Does it make sense? Analyze results.
    - Lessons Learned
    - Conclusion
- Presentation slides based on project report

Extra Credit:

You can get Extra Credits(5pts) if you use implement the above algorithm using Spark, SparkML and show performance graph. Explain what you see.