

CS15-319 / 15-619

Cloud Computing

Recitation 14

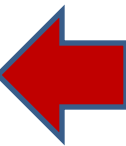
November 25th, 2014

Announcements

- Encounter a general bug:
 - Post on Piazza
- Encounter a grading bug:
 - Post Privately on Piazza
- Don't ask if my answer is correct
- Don't post code on Piazza
- Search before posting
- Post feedback on OLI

Module to Read

- UNIT 5: Distributed Programming and Analytics Engines for the Cloud
 - Module 16: Introduction to Distributed Programming for the Cloud
 - Module 17: Distributed Analytics Engines for the Cloud: MapReduce
 - Module 18: Distributed Analytics Engines for the Cloud: Pregel
 - Module 19: Distributed Analytics Engines for the Cloud: GraphLab



Quiz 5

- Quiz 5 Due Next Wednesday
 - Wednesday **12/03/2014 11:59PM** Pittsburgh
 - Late submissions are NOT accepted
- Timed
 - 180 minutes once started
 - Remember to click SUBMIT before the deadline

Project 4, Module 2 Reflections

Construct an n-gram model of the corpus

- An n-gram is a phrase with n contiguous words
- For example a set of 1,2,3,4,5-grams with counts:

#	Example	Count
1	this	1000
2	this is	500
3	this is a	125
4	this is a cloud	60
5	this is a cloud computing	20

Statistical Language Model (SLM)

- Provide a mechanism to solve common natural language processing problems
- Examples: speech recognition, machine translation and intelligent input method
- SLM estimates the probability of a word given the previous phrases and the N-gram count
- N-gram model is one of the most popular mechanisms to generate an SLM today

This Week's Goal

Build a statistical language model that contains the probability of a word appearing after a phrase

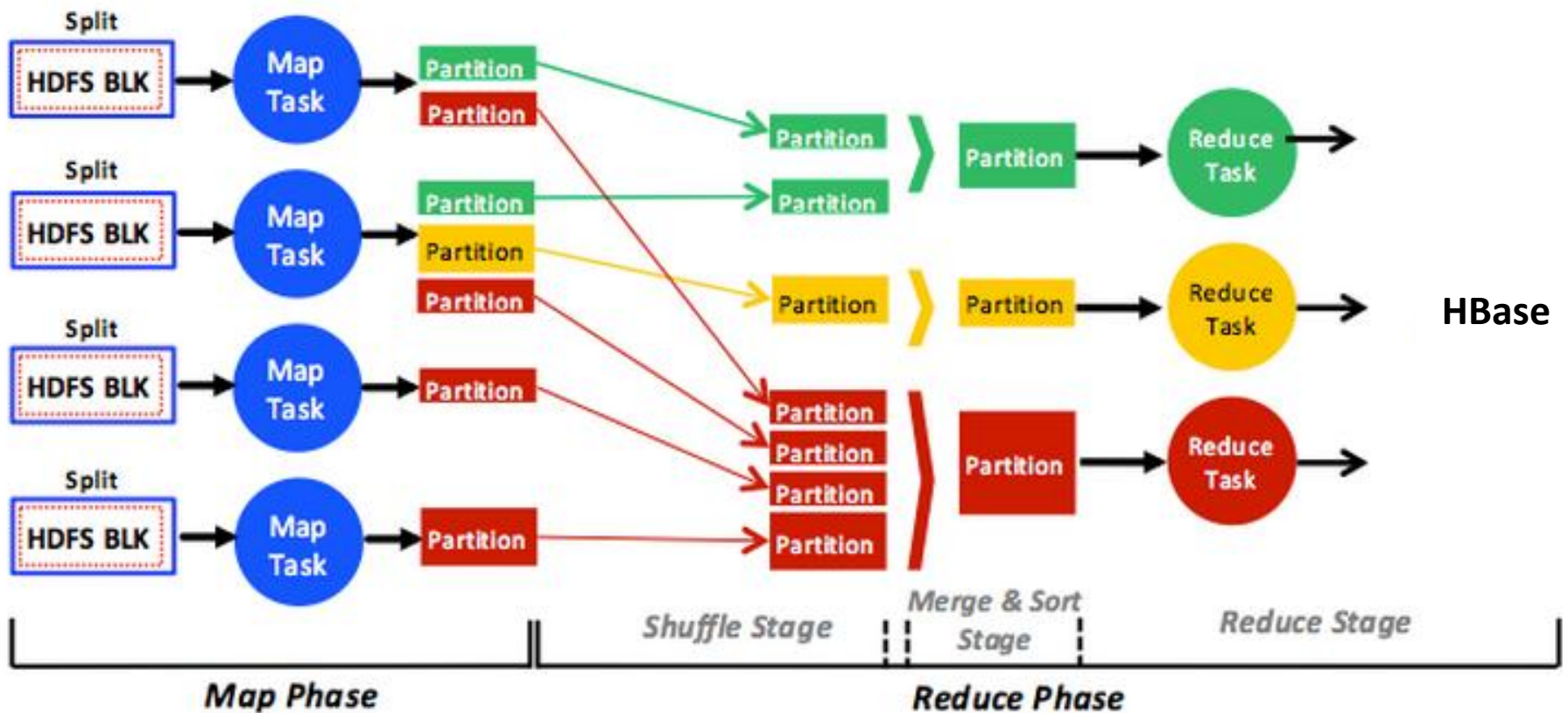
Options	Count	Probability
this was	150	0.15
this is	500	0.50
this day	250	0.25
this kiss	25	0.03
this boy	75	0.08



Options	Count	Probability
this is	500	0.50
this day	250	0.25
this was	150	0.15
this boy	75	0.08
this kiss	25	0.03

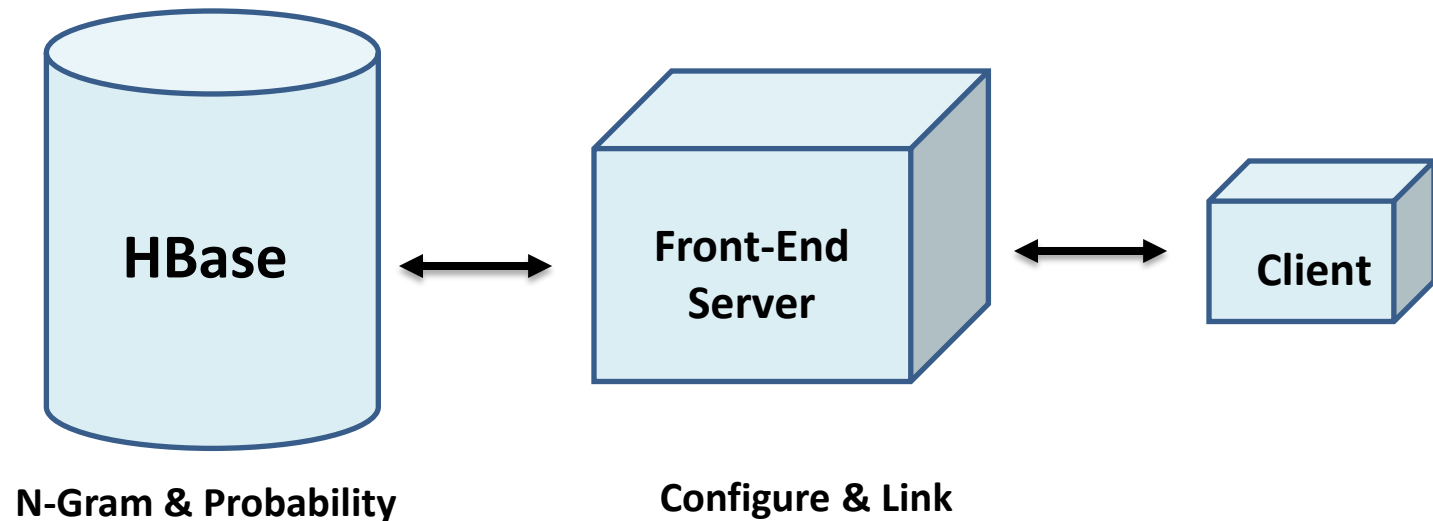
Project 4 Module 3

- Read the input (N-gram) from HDFS and write the output (post-SLM) to HBase



Project 4 Module 3

- Connect HBase with the PHP-based front end server to provide a functional web service.



Upcoming Deadlines

- Unit 5

[UNIT 5: Distributed Programming and Analytics Engines for the Cloud](#)

[Module 16: Introduction to Distributed Programming for the Cloud](#)

[Module 17: Distributed Analytics Engines for the Cloud: MapReduce](#)

[Module 18: Distributed Analytics Engines for the Cloud: Pregel](#)

[Module 19: Distributed Analytics Engines for the Cloud: GraphLab](#)

Wednesday 12/3

Quiz 5: Distributed Programming and Analytics Engines for the Cloud

[Checkpoint](#)

[Available Now](#)

[Due 12/3/14 11:59 PM](#)



- Project 4.3

[Project 4](#)

[Input Text Predictor: Language Model and User Interface](#)

Friday 12/5

Language Model Generation

[Checkpoint](#)

11:59PM

12/5/2014

