CS15-319 / 15-619
Cloud Computing

Recitation 14

November 26th, 2013

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 Tuesday
  - Tuesday **12/03/2013 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

This Week's Goal

Example:

this

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

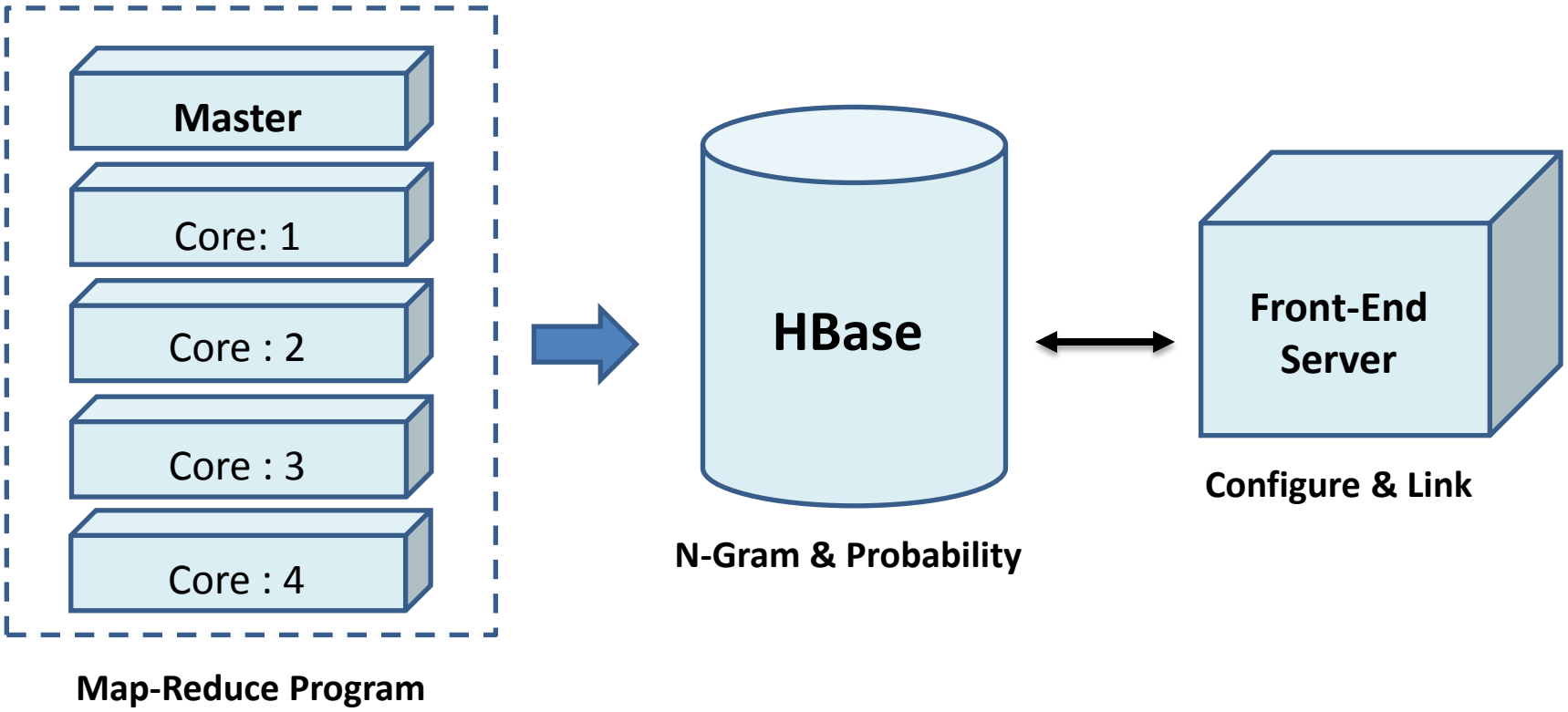
This Week's Goal

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

#	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

This Week's Goal

Store and index the words and their probabilities to use in an application



Upcoming Deadlines

- Project 4:

Project 4		
Input Text Predictor: Language Model and User Interface		
Language Model Generation	Checkpoint	11:59PM 12/6/2013

- 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		
Quiz 5: Distributed Programming and Analytics Engines for the Cloud	Checkpoint	Available Now Due 12/3/13 11:59 PM