15-319 / 15-619 Cloud Computing

Recitation 2

January 19 & 21, 2016

Accessing the Course

- Open Learning Initiative (OLI) Course
  - Access via Blackboard
- http://theproject.zone
  - AWS Account Setup
  - Azure Account Setup
  - Updated your TPZ profile with AWS & Azure info
  - Primers
  - Project 0
- Piazza Access

You should have access to

- ✓Open Learning Initiative Course
  - ✓Access via Blackboard
- ✓http://theproject.zone
  - ✓AWS Account Setup
  - ✓Azure Account Setup
  - ✓Updated your TPZ profile with AWS & Azure info
  - ✓Primers
  - ✓Project 0
- ✓Piazza Access

Reach out to the instructors over Piazza if you have trouble accessing any of the systems above.

Amazon Web Services (AWS) Account - 1

- Students who are just joining us or who have not completed the AWS Account Setup:

  === ONLY IF YOU HAVEN'T DONE SO ALREADY ===

  - Log on to https://theproject.zone and make sure you follow the instructions in the Account Setup Primer
  - Wait to receive Consolidated Billing Request email from Amazon
    - Manual process, waiting time varies
  - When you receive the linking email, click the link to verify the linked billing
    - Many students have not clicked on the link yet!
      - Check your **SPAM** folder
    - You wont be able to complete the projects.

Amazon Web Services (AWS) Account - 2

- ALL STUDENTS:
  - If you have created an AWS account and have not received notification that it was linked to our account
    - Your credit card on file may get charged!
    - We CANNOT reimburse you!!!
    - It is your responsibility to verify that your account is linked to us correctly.
    - Contact us if there are any issues related to your account.
  - Check your spam box
    - Let us know ASAP so we can resend the link request.

Amazon Web Services (AWS) Account - 3

- Changing your AWS Linked Account
  - You must contact us if you make any changes to your AWS linked account or change your profile on TPZ
    - Otherwise you will not be able to work with our AMIs.
    - It is your responsibility to verify that your account is linked to us correctly.
    - Contact us if there are any issues related to your account.

Microsoft Azure

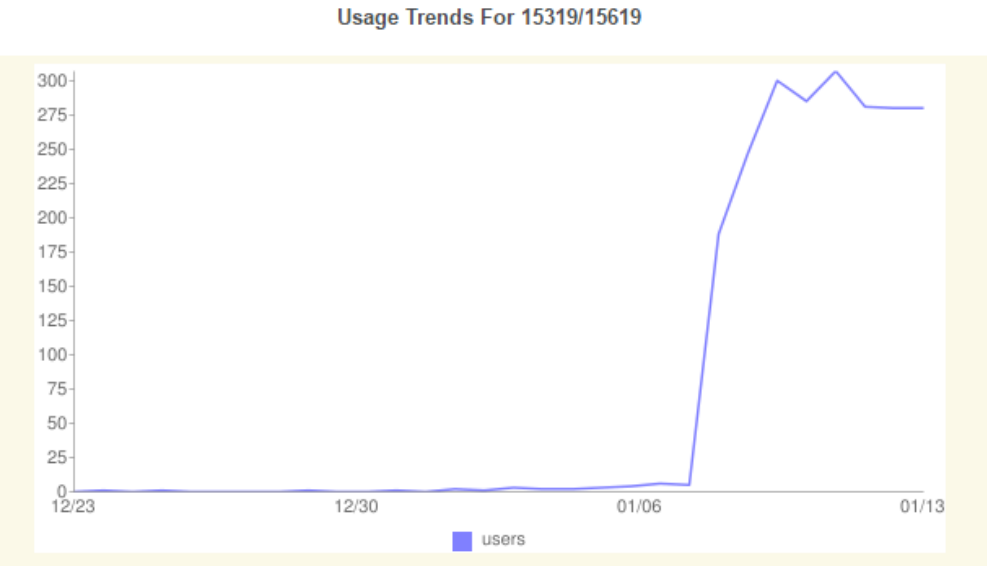
- Experimental Platform for S16
  - Working closely with Microsoft to iron our bugs
- Please contact us if you have trouble signing up
  - Your token is valid for $100 per month for 6 months
- Account Linking Issues
  - We don’t link your account for billing like we do for AWS
  - Purely monitoring and data collection
- If you couldn’t link your account
  - Just make sure you can sign in and provision VMs

Piazza

- Suggestions for using Piazza
  - Discussion forum, contribute questions and answers
  - Read the Piazza Post Guidelines (@9) before asking
  - Read Piazza questions & answers carefully to avoid duplicate ones
  - Give us context and as much information as possible!
  - Don’t ask a public question about a quiz question
  - Try to ask a public question if possible
  - Provide your andrewID if you think we need it to debug
- TA office hours are posted on Piazza and Google calendar

First Week Piazza Volume

- 400+ Piazza Posts, many of which were:
  - duplicates,
  - or covered in primers or
  - solved with a quick google search



Special Thanks To:
Brian Ho (43 answers)
Jun Zhou (14 answers)

Helped many fellow students this week.

Piazza Examples:

"I tried doing xyz as per the instructions, but I got an error, what to do?"


Unclear on what the student did


No additional information on the error condition


No indication of any work the student did to find out more about the error


No information on what the student did to try to fix it

"I tried doing xyz as per the instructions, step number 7, but I got a specific error (error code).

I looked through various resources and on google for this error condition, and tried the following fixes: however, I seem to be stuck. Here's a dump of the error. Can someone help me out?


Clear indication of the exact point where student failed


Additional information on the error (screenshots or text dump)


Indication of the approaches the student took to fix the problem


Solutions or fixes that the student tried

Be clear and concise. Help Us so that we can Help You!

Reflecting on Last Week

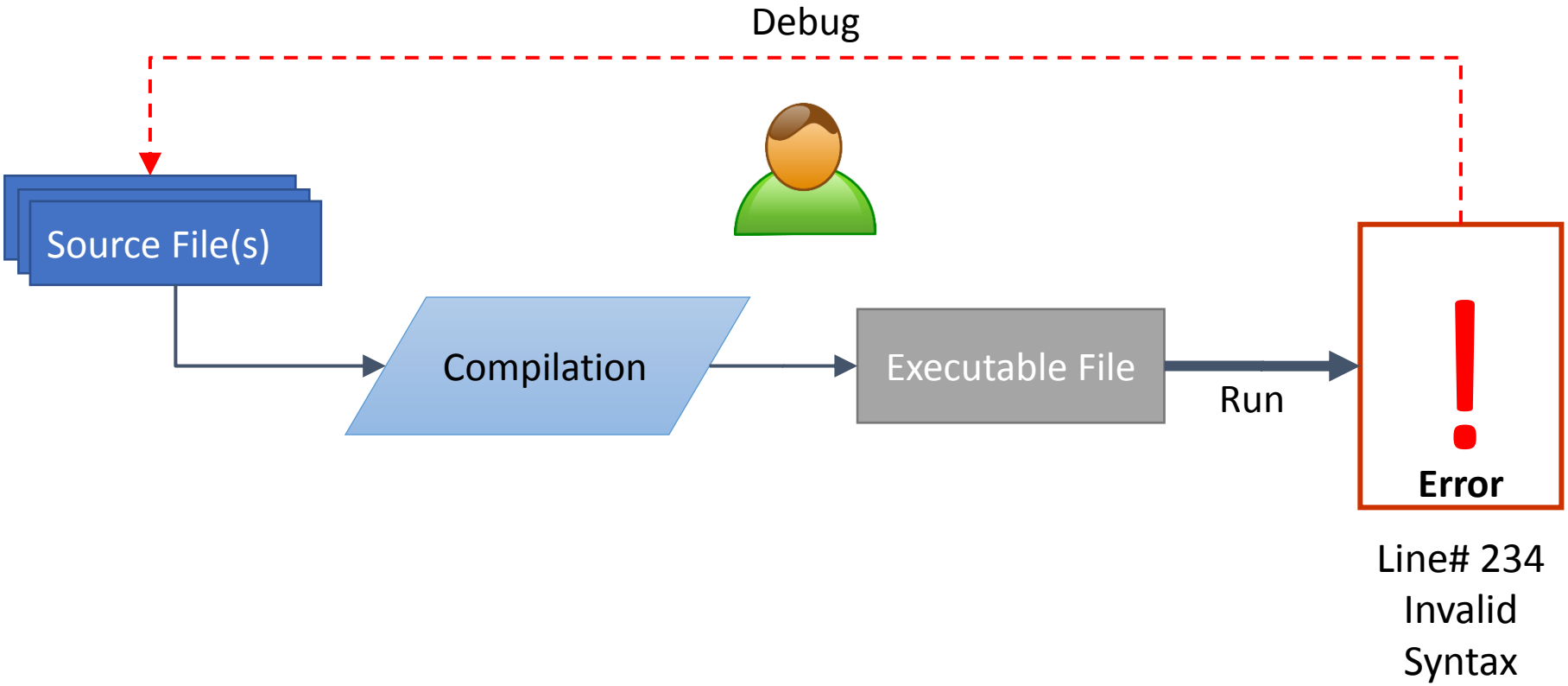
- You should have learned
  - AWS
    - Launching, connecting to and terminating EC2 instances
    - Installing and running software on an EC2 instance
    - How EC2 spot instances work
    - Using S3 to store and retrieve files
  - Azure
    - Launching, connecting to and terminating Azure VMs
    - Installing & running software on an Azure VM
  - Basic Linux/SSH skills
  - Running a web server, testing to access the server over a browser
    - Setting up the web server software.
    - Opening up the required ports.
  - Using AWS and Azure APIs to launch instances

Skill Building in This Course

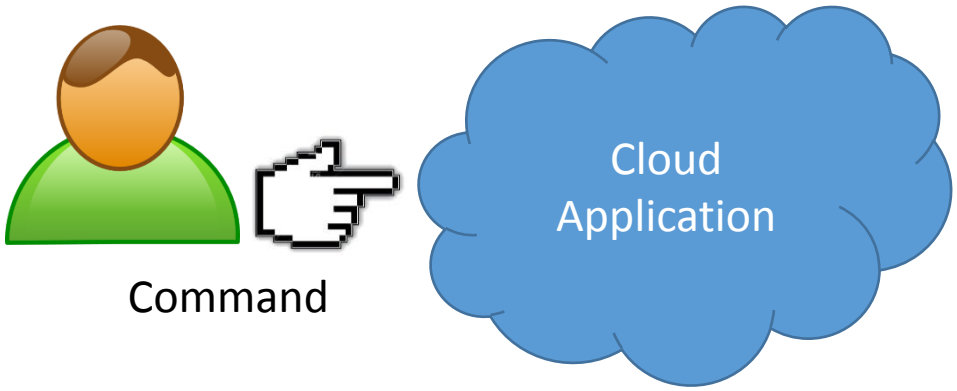
- Complete the Primers
  - Practice working in the Linux shell
  - Understanding AWS/Azure
    - provisioning resources, connecting to VMs, playing around, …
- Important skill in this course!
  - Reduce your debugging time if you know the environment
  - You can bring down the time taken to complete a project from 10 hours to 5 hours per week

Typical Programming Workflow:

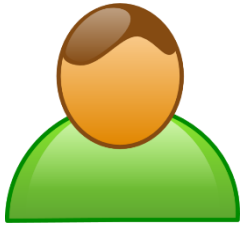
- For most courses:



In the Cloud



In the Cloud



In the Cloud

Y U NO WORK?!?

How do I even begin to fix this?

In the Cloud

Y U NO WORK?!?

How do I even begin to fix this?

AWS Instances

Load Balancers

!

Error

Databases

Front-End Services

Suggested Error Debugging Workflow

What information can I get about the error?

- Read error messages, Look through logs, other information

How can I isolate the source of the problem?

- What component seems to have the problem?

What remedial action can I take?

- The error messages and other information should have clues.
- Configuration changes, command parameters

Am I Still Stuck?

- Google, Piazza, TA Office Hours (In that order!)

Programming Experience Expected

- **Strong proficiency** in at least one of the following, with some fair comprehension of the others:
  - **Java**
  - **Python**
  - **Bash**

- Java is **required** to complete parts of Projects 2, 3 and 4.
- GraphLab (a bonus project) uses a bit of C++ in Project 4 (at the end)
- Use the time now to brush up
- Do not fear bash/python scripting, it will make your life easier!

Completing Projects in this Course

- Provision AWS or Azure Resources
  - Use the AMIs/VHDs we provide for the project
  - Tag all instances!
- Monitor your cost
  - Calculate costs before you provision!
- Complete tasks for each project module
  - Each project module has several sections unlocked by AssessMe
- Submit your work
  - Pledge of integrity
  - Results in scoreboard
- Terminate all resources when you have verified your score and kept a copy of your work

Tagging

- Tag *all* tag-able resources on AWS
  - Before you make a resource request, read the docs/specifications to find out if tagging is supported
  - Apply tags during resource provisioning
  - We need tags to track usage, a grade penalty will be applied automatically if you do not tag!
- Tagging Format
  - `Project: <Project#>`
  - Project#: 0, 1.1, 1.2….etc.
  - Information will always be present in the project instructions

Budgets and Penalties

- No proper tags ➔ 10% grade penalty
- Budget
  - For P1.1, each student's budget is $5
  - Exceeding Budget ➔ 10% project penalty
  - Exceeding Budget x 2 ➔ 100% project penalty (no score)
  - You can see Cost and Penalties in TPZ.
- We will enforce these penalties automatically starting from Project 1.1

Academic Integrity Violation

- Plagiarism → the lowest penalty is 200% & potential dismissal
  - Other students, previous students, Internet (e.g. Stackoverflow)
  - Do not work on code together
  - This is about you struggling with something and learning
  - Penalty for cheating is SEVERE – don't do it!
  - Ask us if you are unsure

How to Work on a Budget

- P1.1 Budget → $5
- You are only allowed to use t1.micro
  - $0.02 per hour (on demand)
  - Total time you have: 250 hours of t1.micro
- Other Costs to consider:
  - EBS is $0.1 per GB/month
  - t1.micro is configured with 30 GB EBS by default.
  - Data transfer costs (minimal)
- **Note**: Free Tier does not apply to any of the linked accounts!

Deadlines!

- **Hard** Deadlines
  - No late days, no extensions
  - Start early!
  - Plan your activities, interviews and other commitments around the deadlines.
  - **No exceptions!**
- Project modules due every Sundays at Midnight ET
- Quizzes are typically due on Fridays
- Team projects are typically due on Wednesdays

Deadlines!

- **Project deadlines**
  - **On TheProject.Zone**
- **Quiz deadlines**
  - **On OLI**

What's due this week:

- On OLI: Quiz 1

Cloud Computing			
Assignment			**Status**
UNIT 1: Introduction to Cloud Computing			
Module 1: Cloud Computing Overview (Gradebook) (Learning Dashboard)			
Module 2: Economics, Benefits, Risks, Challenges and Solutions (Gradebook) (Learning Dashboard)			
Quiz 1: Introduction to Cloud Computing		Checkpoint	Available 1/22/16 12:01 AM Due 1/22/16 11:59 PM

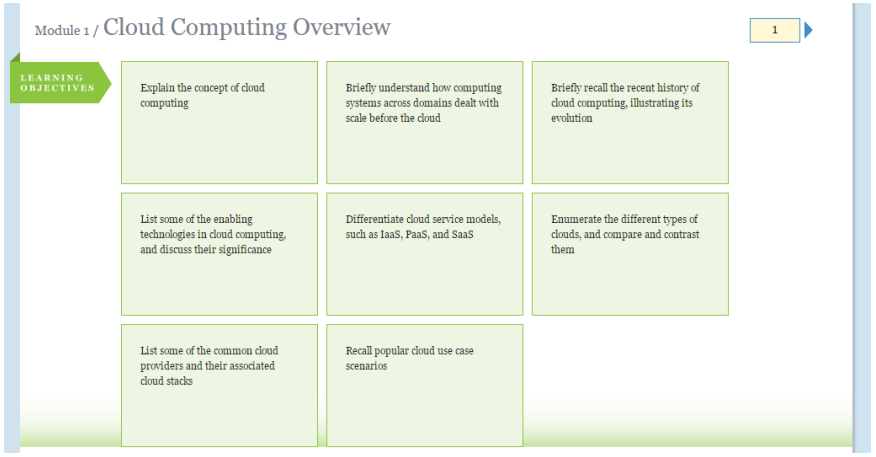
- On http://theproject.zone: Project 1.1

Project 1 Big Data Analytics

Module	Open time	Deadline
Sequential Programming Ongoing 6 days 10 hours left	01/18/2016 00:01 -0500	01/24/2016 23:59 -0500

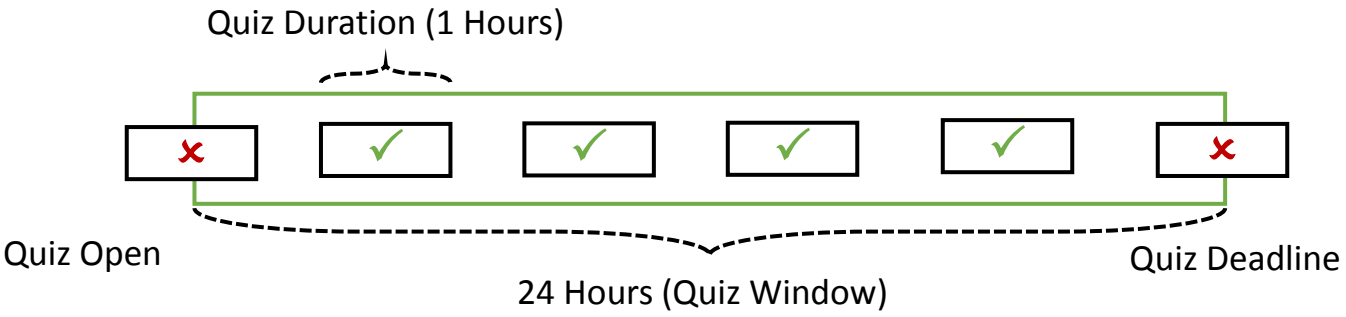
Quiz 1 Preparation

- Tests your understanding in Modules 1 and 2
  - Cloud computing fundamentals, service models, economics, SLAs, security
  - Use the activities in each page for practice.
  - You will be tested on you ability to perform the stated learning objectives on OLI:



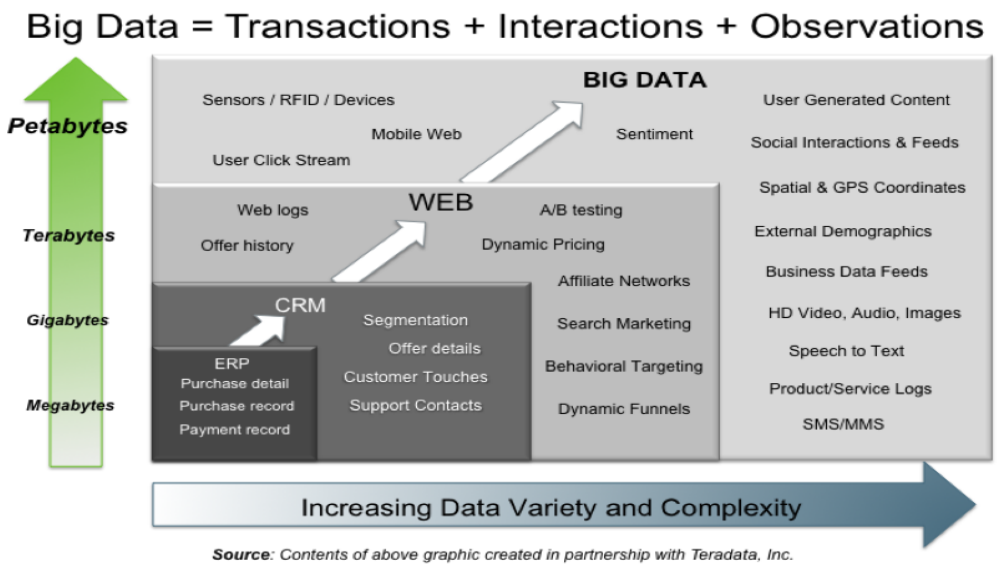
Quiz 1 Logistics

- Quiz 1 will be open for 24 hours, Friday, Jan 22
  - Quiz 1 becomes available on **Jan 22, 00:01 AM EST.**
  - Deadline for submission is **Jan 22, 11:59 PM EST.**
  - Once open, you have **60 min** to complete the quiz.
  - You may not start the quiz after the deadline has passed.
  - Every 15 minutes you will be prompted to save.
  - **Maintain <u>your own timer</u> from when you start the quiz.**
  - **Click submit before deadline passes. No Exceptions!**



Project 1 Motivation: Big Data

- What is Big Data?
  - It is high volume, high velocity, and/or high variety information assets.
  - There is a lot of value in analysis of big data for organizations

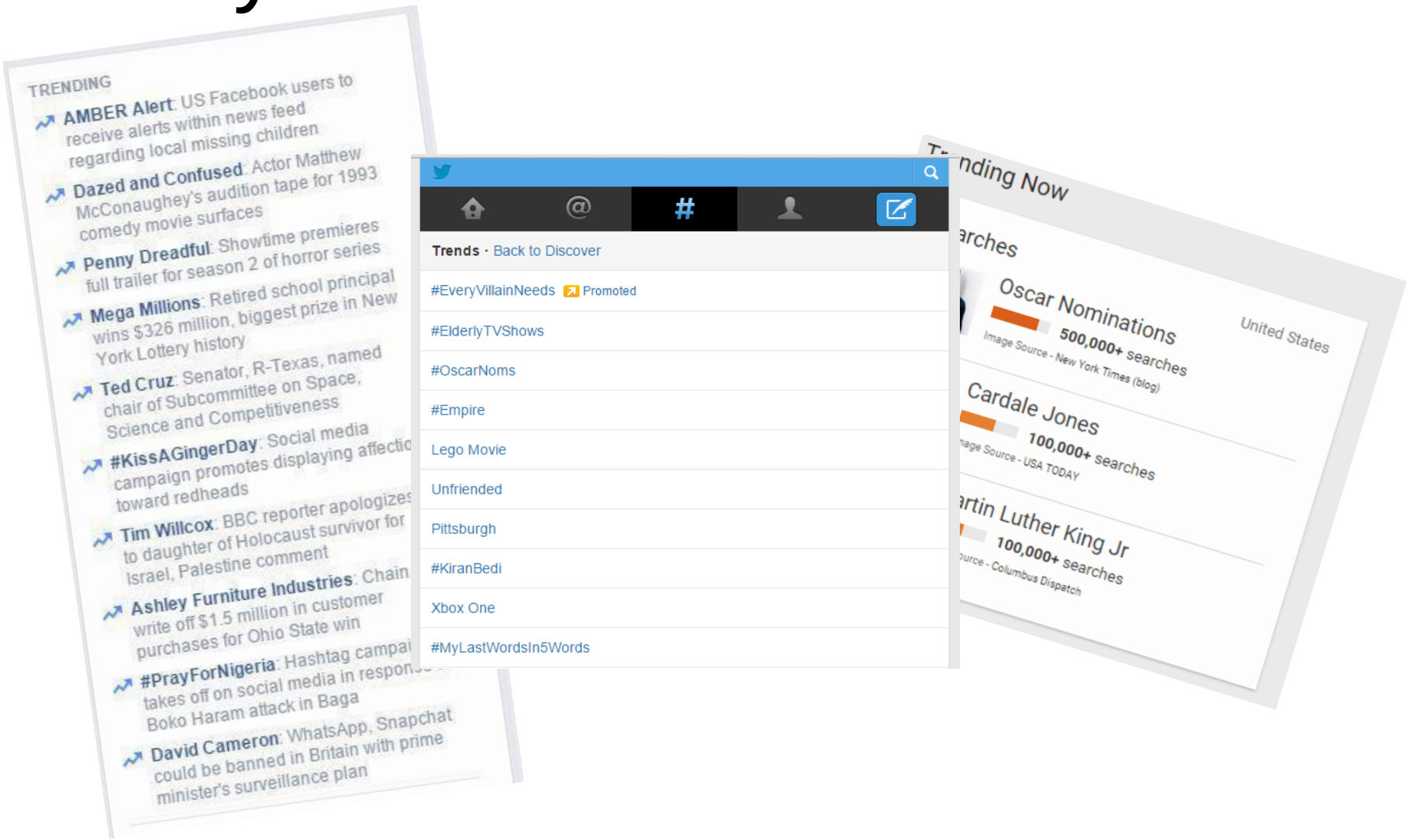


Source: Contents of above graphic created in partnership with Teradata, Inc.

Use Cases: Big Data Analysis

- Online retailers are analyzing consumer spending habits to learn trends and offer personalized marketing campaigns and offers to individual customers.
- Companies such as Time Warner, Comcast etc. are using big data to track media consumption habits of their subscribers and trends to provide value-added information to advertisers and customers.

Trending Topics are Everywhere!



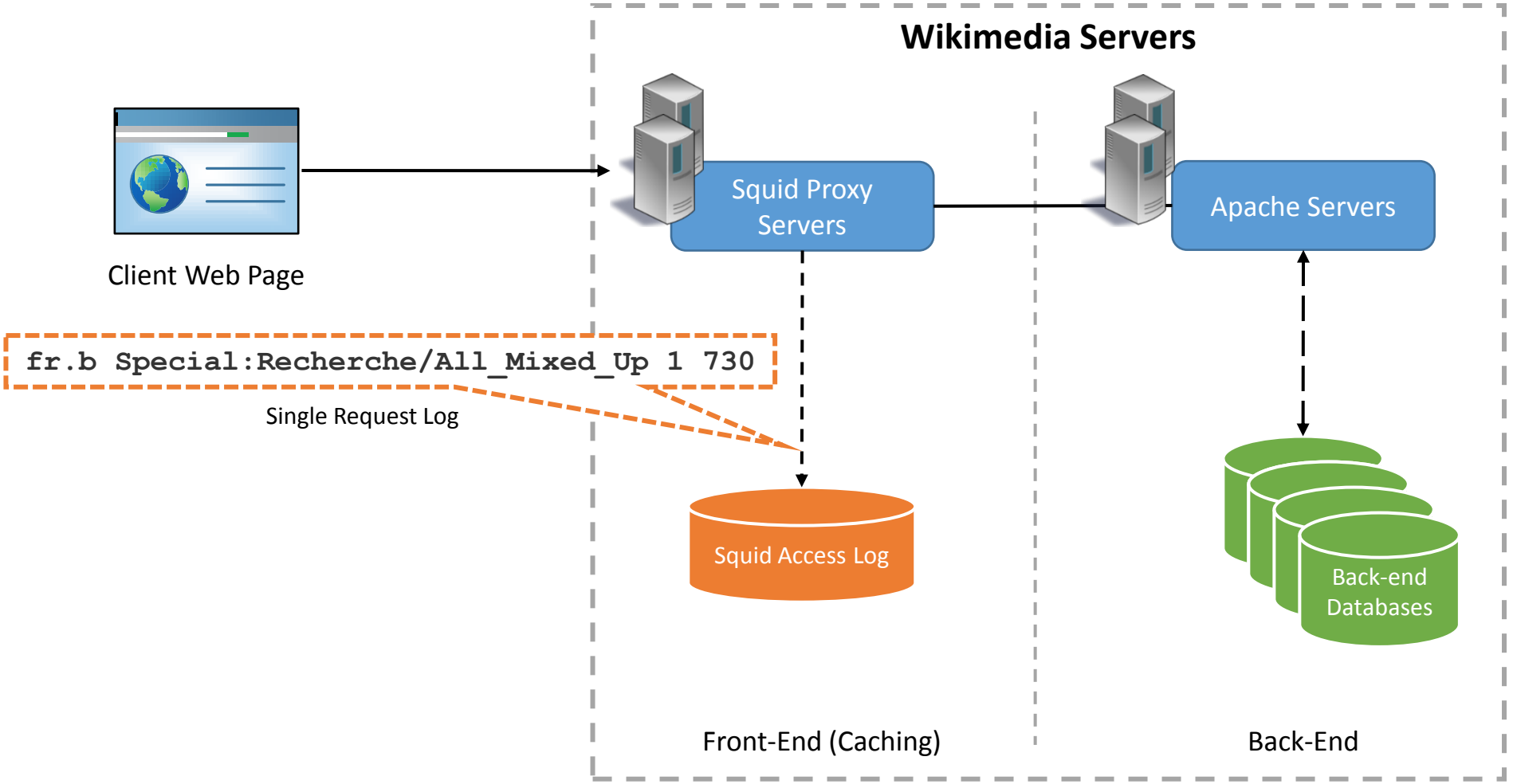
Why Trending Topics?

- Identify trends and viral content
- Maximize ad placement opportunities
- Search Engine Optimization (SEO)
- And more....

Project 1

- Identify Trending Topics on Wikipedia
  - Use the hourly page-view statistics dataset
- Project 1.1: (This Week)
  - Find trends from a single hour of data.
- Project 1.2: (Next Week)
  - Find trends for an entire month.

Wikipedia page requests



The Dataset

- Data set
  - Wikimedia raw page views data
  - One File Per Hour
- Format:
- `<project name> <page title> <number of accesses> <total data returned>`

<Language>.<ProjectName>
en.b = English Wikibooks
fr.wm = French Wikipedia Mobile

Parse and Filter

- We are only interested in English Wikipedia pages
- Filter out the rest
  - Use the filtering rules specified
  - Remove non-titles (without uppercase)
  - Remove special pages
  - Filter out images and other files
  - If ambiguous or a special case – keep the line
- This dataset is raw, real-world
  - Not always clean
  - Use your own discretion
- Sort the pages by number of page-views

Project 1.1 Workflow

- Launch EC2 instance with a special AMI
  - Recommend using the APIs to do this so you get practice before Project 2.
- Download the required dataset
- Write the code to parse, filter and sort
- Complete and run the script
  - `/home/ubuntu/Project1_1/runner.sh`
  - Answer 9 questions of increasing difficulty by providing the commands inside runner.sh
- Submit your code for grading
  - Complete the `references` file
  - Execute `submitter.sh` to submit your code

Grading of your Projects

- Code submissions are auto-graded
- Scores will be made available on http://theproject.zone
  - Instantly updated on the scoreboard for each project
- We will grade all the code (both auto and manually)
  - Be sure to make your code readable
    - Preface each function with a header that describes what it does
    - Use whitespace well
      - Indent when using loops or conditional statements
    - Use descriptive variable and function names
    - For more detail, please refer to Google's Style Guide
- If your code is not well documented and is not readable, we will deduct points
  - Documentation shows us that you know what your code does!
  - The idea is also NOT to comment every line of code

Demo