CS15-319 / 15-619
Cloud Computing

Recitation 10

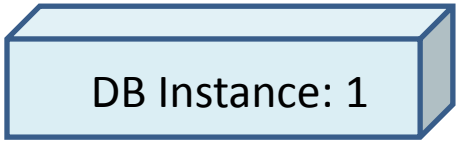
March 25nd and March 27nd, 2014

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

Project 3, Module 3 Reflections

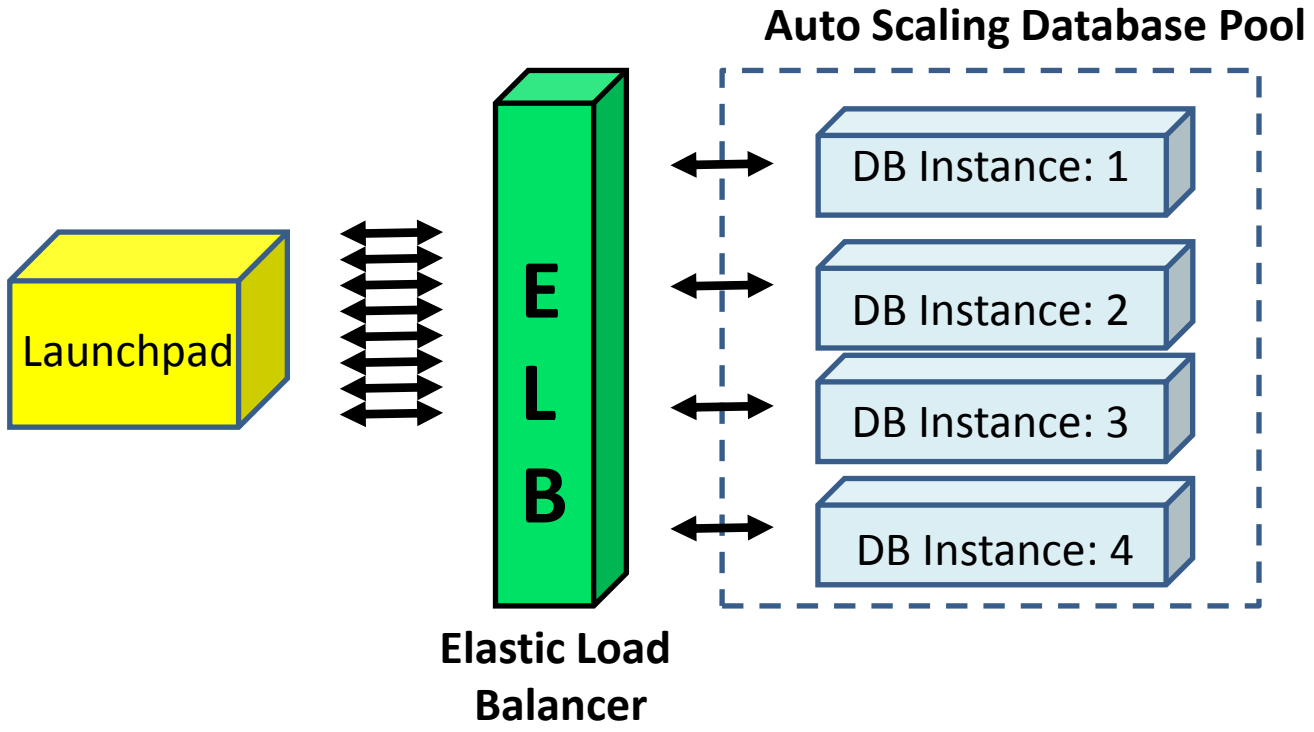
- Cost different between running a fixed number of instances vs autoscaling
- Custom Metrics



  - Queries per Second = Queries / (Uptime1 – Uptime2)
  - Transactions Per Second (TPS)
    - Factor in the two queries used to find the number of queries and uptime.
    - *sysbench* transaction equals to 16 queries
  - Send TPS information to CloudWatch using API

Project 3, Module 3 Reflections

- Horizontal Scaling of Databases



- Horizontal Scaling of SQL Database is not easy to implement.

Project 3, Module 3 Problems

- The **/etc/init.d** directory contains a number of start/stop scripts for various services on your system.
- **/etc/rc.local** file runs after all other init level scripts have run. You can put commands that you want to have issued upon startup.
- **cron:** enable users to execute commands or scripts automatically at a specified time/date.

Module to Read

- UNIT 4: Cloud Storage
  - Module 12: Cloud Storage
  - Module 13: Case Studies: Distributed File Systems
  - Module 14: Case Studies: NoSQL Databases 
  - Module 15: Case Studies: Cloud Object Storage
  - Quiz 4: Cloud Storage

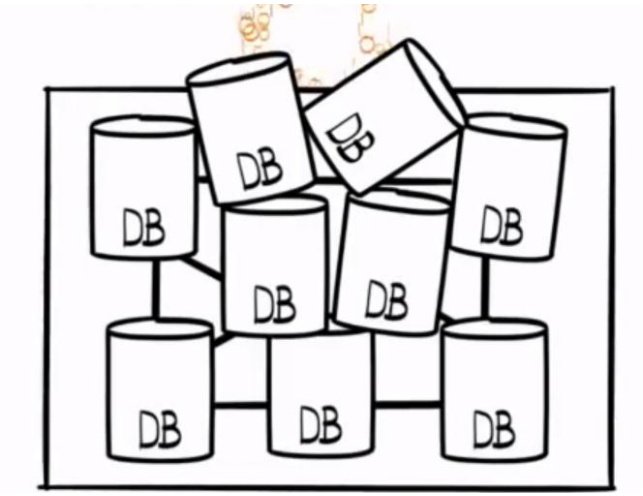
Project 3

- Files vs. Databases
  - File vs. Database
- Vertical Scaling in Databases
  - Vertical Scaling
- Horizontal Scaling in Databases
  - Horizontal Scaling
- Working with NoSQL: DynamoDB / Hbase
  - Amazon DynamoDB
  - DynamoDB vs. HBase

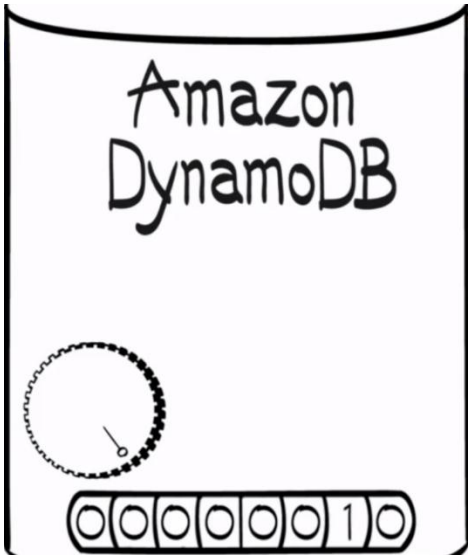
DynamoDB

- Amazon DynamoDB
  - DynamoDB is not a database software, it is a database service.

Relational Database



DynamoDB



Images from http://aws.amazon.com/dynamodb/

DynamoDB

- Amazon DynamoDB
  - NoSQL vs. SQL
    - Flexibility
    - High write performance at very large scale
    - Limited functionalities
  - Items and Attributes vs. Rows and Columns

Items →

Product_ID	Timestamp	Quantity_Sold
1009	2013-4-1	2
2023	2013-4-2	100
1009	2013-5-1	5

  - Solid State Disks (SSDs)

DynamoDB

- DynamoDB Primary Key
  - **Hash Type Primary Key:** the primary key is made of one attribute, a hash attribute.
  - **Hash and Range Type Primary Key:** the primary key is made of two attributes. The first attribute is the hash attribute and the second one is the range attribute.

Hash Key ↓ Range Key ↓

Product_ID	Timestamp	Quantity_Sold
1009	2013-4-1	2
2023	2013-4-2	100
1009	2013-5-1	5

DynamoDB Cost

- Provisioned Throughput Capacity
- Throughput Capacity Parameters
  - **Read capacity units:** The number of reads of items up to 4 KB in size per second.
  - **Write capacity units:** The number of 1 KB writes per second.
- Cost Structure
  - **Write:** $0.0065 per hour for every 10 units of Capacity
  - **Read:** $0.0065 per hour for every 50 units of Capacity
- http://aws.amazon.com/dynamodb/pricing/

DynamoDB Cost

- Example:
- Read: 500 Units
- Write 500 Units
- Each Student's Cost
  - $290 per month
  - **$10 per day**
- **Class Cost: $2,000 per day**
- **Switch to free tier (10-read & 5-write) when done, and terminate DynamoDB after grading.**

DynamoDB Use Case

- Blitz is a fresh approach to to carry out load and performance testing on the cloud.
- Blitz enables you to instantly burst up to 50,000 concurrent users against your app in seconds from multiple points of presence around the world.

From https://www.blitz.io/for/performance

DynamoDB Use Case

- “Eventually we chose Amazon’s DynamoDB as our storage database for this task because it frees us from tedious operations like backups, monitoring, and increasing capacity”.

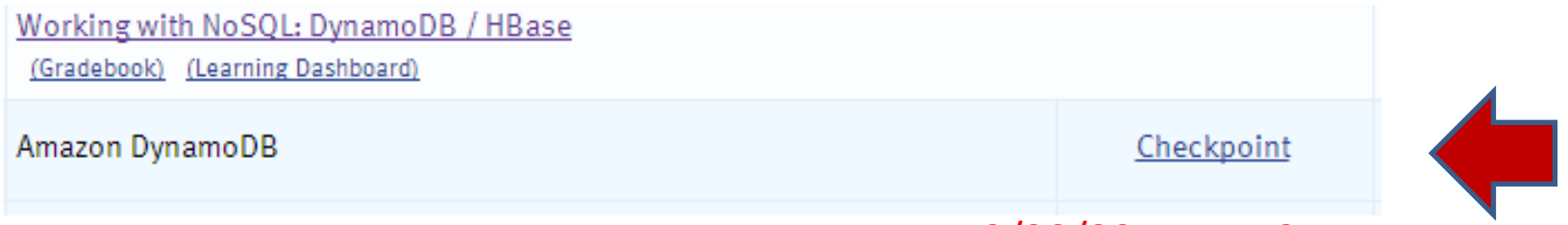
From http://blog.blitz.io/2012/10/how-we-use-dynamodb-and-what-weve-learned-along-the-way/

Review

- After you complete the task in this project module
  - Immediately set the throughput to 10-read / 5-write (which is the free usage tier).
    - This will be enough for grading purposes.
  - After the TAs finish grading, we will notify you by email and then you should terminate your DyanmoDB ASAP.

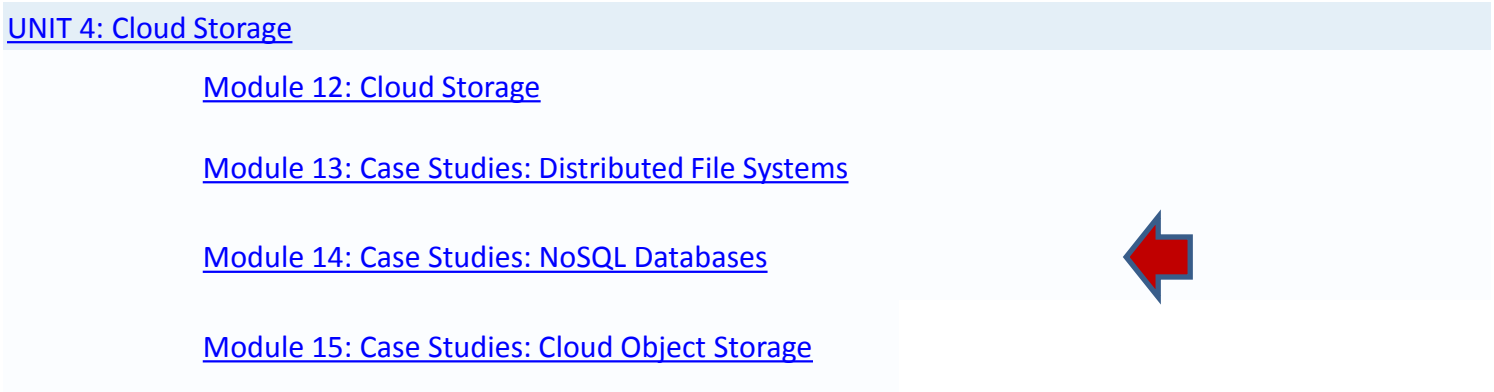
Upcoming Deadlines

- Project 3:



Due 3/30/2014 11:59 PM

- Unit 4:



Demo Outline

- 1. DynamoDB
  - Using Web Console
    - Create DynamoDB Table
    - Check Table Status
  - Using API
    - Put Items into Table
- 2. The Picture Lookup Webpage
  - Using data from DynamoDB

The Dataset

- Caltech256
  - 30,607 images separated into 256 categories
  - Each row represents an image with three fields
    - Category
    - Picture
    - S3URL

The Architecture