Project Requirements

- Run in public cloud

- Difficulty varies for listed candidates - will take difficulty into consideration when grading.

- Will grade based on
  - project difficulty
  - quality and depth of work
  - workload distribution among team members

- Milestones: proposal, demo1, demo2, final demo, report.

- Start early! Discuss with us and Dr. Lu
Project Topics – Cloud Computing

- **IaaS (Infrastructure as a Service)**
  - Real-Time Performance Measurement in
    - Amazon: Xen Virtual Machine and Docker Container
    - Google: Compute Engine and Container Engine

- **PaaS (Platform as a Service)**
  - Real-Time Streaming
    - Amazon: Kinesis
    - Google: Cloud DataFlow
  - Internet-of-Things
  - Real-Time Messaging Service
    - RabbitMQ, NSQ, ZeroMQ, Google Cloud Pub/Sub
Cloud Computing

Cloud computing provides shared pool of configurable computing resource to end users on demand

- Public cloud computing providers allows users to avoid upfront infrastructure costs (e.g., purchasing servers)

Three service models

- IaaS (Infrastructure as a Service): virtual machines, storage, network ...
- PaaS (Platform as a Service): execution runtime, middleware, web server, database, development tool ...
- SaaS (Software as a Service): email, virtual desktop, games ...
Virtualization & IaaS

- Virtualization is widely used in clusters, cloud computing or embedded systems
  - Provides resource isolation, security, high availability
  - Introduces overhead, may result in long latency, extra resource consumption
- IaaS provides configurable **virtual machines** to end users
Amazon IaaS

- Amazon EC2 (Elastic Computing, https://aws.amazon.com/ec2/) provides IaaS to users

- Virtual machines (instance) enabled by Xen
  - HVM and Para-Virtualization
  - Different instances with different features: CPU, memory, disk, networks …

- EC2 container service
  - Easy to launch and scale applications in Docker containers (https://aws.amazon.com/ecs/)
    - Docker: a layer of abstraction and automation of operating-system-level virtualization
  - Containers run on top of VMs
Google IaaS

- Container Engine (https://cloud.google.com/container-engine/)
  - Format: Docker container (https://www.docker.com)
  - Container management: Kubernetes (http://kubernetes.io)
Projects on IaaS

- Real-time performance evaluation for Amazon or/ and Google IaaS

- Run benchmark tests in VMs/containers. Metrics include:
  - Network latency/throughput
  - CPU and memory consumption
  - Disk read/write speed
  - Performance fluctuation
  - Real-time scheduling evaluation[^1]

- Benchmarks:
Platform as a Service

- Cloud providers deliver computing platform, including operating system, programming language execution runtime, middleware, database, web servers …

- With PaaS, application developers can develop and run the specific software solution without cost for configuring underlying hardware and software layers.

Real-Time Streaming

Real-time streaming

- Web, software, sensors generate large amount of data streams
  - Need to be processed/stored continuously and in real-time

Amazon Kinesis (https://aws.amazon.com/kinesis/)

- Create delivery stream
- Send data from your frontend applications to delivery stream via Kinesis Agent or APIs
- Stream is analyzed by Amazon S3 or Redshift
- Results are down-streamed to feed real-time dashboards, monitoring, notification, alarm, or any other interested services
Real-Time Streaming – Contd.

- Google’s Cloud Dataflow (https://cloud.google.com/dataflow/)
  - Dynamic resource provisioning
  - Automated workload rebalancing
  - Unified programming model

https://cloud.google.com/dataflow/
Internet-of-Things

- IoT in public cloud provides bi-directional communication between sensors, actuators, embedded devices, smart applications and cloud

- Amazon IoT (https://aws.amazon.com/iot/)
  - APIs for frontend apps
  - MQTT-based protocol for message delivery
  - Message processing & storage in cloud
Real-Time Messaging Service

- Messaging service provides
  - Message routing, delivery, persistency for applications
  - A typical communication paradigm: publish/subscribe
    - Publishers publish messages of different topics; middleware forwards messages of interest to corresponding subscribers

- RabbitMQ, NSQ, ZeroMQ, Google cloud pub/sub
- Try different configurations
  - Number of publishers/subscribers, number of topics, message size, sending rate, enable/disable persistency, number of brokers
  - Measure latency and locate system bottleneck
Projects on PaaS

Tasks:

- Design and implement a (real-time) application
  - With latency requirement
  - E.g., user uploads messages to cloud via Amazon IoT API. After backend (in cloud) processes the messages, push notifications back to users

- Measure the end-to-end latency (average, WCET, jitter), throughput, CPU/memory …
  - Benchmarks
  - Emulate real-world workload traces

- Analyze experiment results
Steps

1. Choose your favorite topic
2. Form a team
3. Propose a plan
4. Implement
5. Measure and analyze
6. Demo: 1, 2, final
7. Write a technical report
Project 0

- Create & host a static webpage on Amazon or Google cloud
  - Submit **individually**
  - Email the webpage address to me (chong.li@wustl.edu)

- Show names of all team members on your webpage

- Example
  - [https://cloud.google.com/container-engine/docs/docs/quickstart](https://cloud.google.com/container-engine/docs/docs/quickstart)
Pointers

- Apply for $40 credits for Amazon AWS
  - [https://aws.amazon.com/education/awseducate/apply/](https://aws.amazon.com/education/awseducate/apply/)

- Credit programs for the Google Cloud Platform
  - Free trial: 60 Days, $300 credits.
  - Education credit: $50 per student (use your wustl email to apply)
    - [http://goo.gl/gcpedu/9PEI2q](http://goo.gl/gcpedu/9PEI2q)
  - Also, submit your interest via [https://cloud.google.com/edu/](https://cloud.google.com/edu/) for other credit programs

- Resource list for course projects