TCSS 562: SOFTWARE ENGINEERING FOR CLOUD COMPUTING

Group Presentations IV

Wes J. Lloyd
School of Engineering and Technology
University of Washington – Tacoma

MW 5:50-7:50 PM

OBJECTIVES – 12/9

- **Questions from 12/7**
- **Presentation Questions**
- **Term Project Paper, Term Project Presentation**
- **Group 4 – Paper: *BATCH: Machine Learning Inference Serving on Serverless Platforms with Adaptive Batching**
  David Melanson, Samuel David Adams , Richard Brun*

- **2nd hour:**
- **Group 2 – Paper: *Serverless in the Wild: Characterizing and Optimizing the Serverless Workload at a Large Cloud Provider**
  Enbei Liu, Jingru Zhao*
- **Office Hours / Tutorial questions / Team planning**

December 9, 2020	TCSS562:Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.2

ONLINE DAILY FEEDBACK SURVEY

- **Daily Feedback Quiz in Canvas – Take After Each Class**
- **Extra Credit for completing**

Announcements
Assignments
Discussions
Zoom
Grades
People
Pages
Files
Quizzes
Collaborations
UW Libraries
UW Resources

▼ Upcoming Assignments

- Class Activity 1 – Implicit vs. Explicit Parallelism
  Available until Oct 11 at 11:59pm | **Due** Oct 7 at 7:50pm | -/10 pts

- Tutorial 1 - Linux
  Available until Oct 19 at 11:59pm | **Due** Oct 15 at 11:59pm | -/20 pts

▼ Past Assignments

- TCSS 562 - Online Daily Feedback Survey - 10/5
  Available until Dec 18 at 11:59pm | **Due** Oct 6 at 8:59pm | -/1 pts

- TCSS 562 - Online Daily Feedback Survey - 9/30
  Available until Dec 18 at 11:59pm | **Due** Oct 4 at 8:59pm | -/1 pts

December 9, 2020	TCSS562: Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.3

TCSS 562 - Online Daily Feedback Survey - 10/5

Started: Oct 7 at 1:13am

Quiz Instructions

Question 1　　　　0.5 pts

On a scale of 1 to 10, please classify your perspective on material covered in today's class:

1	2	3	4	5	6	7	8	9	10

Mostly
Review To Me

Equal
New and Review

Mostly
New to Me

Question 2　　　　0.5 pts

Please rate the pace of today's class:

1	2	3	4	5	6	7	8	9	10

Slow

Just Right

Fast

December 9, 2020	TCSS562: Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.4

MATERIAL / PACE

- **Please classify your perspective on material covered in today's class (18 respondents):**
- **1-mostly review, 5-equal new/review, 10-mostly new**
- **Average – 5.89 (↑ - *previous 5.58*)**

- **Please rate the pace of today's class:**
- **1-slow, 5-just right, 10-fast**
- **Average – 5.00 (↓ - *previous 5.29*)**

December 9, 2020	TCSS562: Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.5

FEEDBACK FROM 12/7

- ***Can you please discuss the answers to questions 6c and 12 in Tutorial 7***

December 9, 2020	TCSS562: Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.6

FEEDBACK - 2

- *For our project, we would like to use a Lambda function to read from a csv file that is stored on S3 and insert records to a SQL table created on RDS.*

- *We are facing a problem. To access RDS, the Lambda has to be in its VPC. But if the Lambda is placed in a VPC, it is no longer able to connect to the S3 where the csv file is stored.*

- An S3 VPC Endpoint is needed !

- See page 9 in
  Tutorial 9
  for a description:

Now, while still in VPC, create an "Endpoint" so the VPC can access S3.
Click on the "Endpoints" link in the LEFT-Hand sidebar menu.

9

Click the button "**Create Endpoint**".
Scroll down through the long list and select: "com.amazonaws.us-east-1.s3"
Make sure the default VPC is selected.
When the route tables, appear, select your public and private route tables.
Then create the endpoint, by clicking the button:

[Create endpoint]

December 9, 2020	TCSS562: Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.7

FEEDBACK - 3

- *I would like to update you about our term project issue.*
- *We were able to solve that issue by adding an s3 endpoint at VPC*
- *Adding the endpoint enabled Lambda to communicate with both s3 and VPC at the same time*
- *We don't clearly understand what is actually happening though*
- *VPC already has network gateway enabled, so what is the need of adding VPC endpoint there, and how did it resolve the connectivity issue between Lambda and RDS?*
- Connectivity to an RDS database is only possible from within a VPC – no resource outside the VPC can connect
- Placing your Lambda function in a VPC cuts its internet connection
  - VPC is a private network
  - To allow traffic to reach the Internet a NAT Gateway, Router, or VPC endpoint is required

December 9, 2020	TCSS562: Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.8

TUTORIAL SUMMARY

- Tutorial 7: Sunday Dec 6th @ 11:59p

- Tutorial 8: _Extra Credit_ – Dec 18th @ 11:59p

- Tutorial 9: _Extra Credit_ - Dec 18th @ 11:59p

- Tutorial 10 – No Credit – Posted 11/25

December 9, 2020	TCSS562: Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.9

OBJECTIVES – 12/9

- **Questions from 12/7**
- **Presentation Questions**
- **Term Project Paper, Term Project Presentation**
- Group 4 – _Paper: BATCH: Machine Learning Inference Serving on Serverless Platforms with Adaptive Batching_
  David Melanson, Samuel David Adams , Richard Brun

- **2nd hour:**
- Group 2 – _Paper: Serverless in the Wild: Characterizing and Optimizing the Serverless Workload at a Large Cloud Provider_
  Enbei Liu, Jingru Zhao
- **Office Hours / Tutorial questions / Team planning**

December 9, 2020	TCSS562:Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.10

GROUP PRESENTATION QUESTIONS

- Assignment created as quiz on Canvas
- Only **ONE MEMBER** of each team needs to submit the quiz
- Quiz collects questions for group presentations in one place
- Best to submit all questions at once on/after Wed Dec 11

- Please provide 2 questions for each presentation not occurring on your team's presentation day

- Monday Nov 30 – Quiz for Groups 2, 3, 4, 5, 6, 9, 12
- Wednesday Dec 2 – Quiz for Groups 1, 2, 4, 6, 7, 8, 9, 12
- Monday Dec 7 – Quiz for Groups 1, 2, 3, 4, 5, 7, 8
- Wednesday Dec 9 – Quiz for Groups 1, 3, 5, 6, 7, 8, 9, 12

December 9, 2020	TCSS562: Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.11

GROUP PRESENTATIONS – WEEK 11

Wednesday Dec 9

Slot #1 - **BATCH: Machine Learning Inference Serving on Serverless Platforms with Adaptive Batching**

- *Group 4: David Melanson, Samuel David Adams , Richard Brun*

<MOVED> Slot #2 - Serverless in the Wild: Characterizing and Optimizing the Serverless Workload at a Large Cloud Provider
Group 2: Enbei Liu, Jingru Zhao

December 9, 2020	TCSS562: Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.12

OBJECTIVES – 12/9

- **Questions from 12/7**
- **Presentation Questions**
- **Term Project Paper,** Term Project Presentation
- **Group 4 – Paper:** *BATCH: Machine Learning Inference Serving on Serverless Platforms with Adaptive Batching* *David Melanson, Samuel David Adams , Richard Brun*

- **2nd hour:**
- **Group 2 – Paper:** *Serverless in the Wild: Characterizing and Optimizing the Serverless Workload at a Large Cloud Provider* *Enbei Liu, Jingru Zhao*

- **Office Hours / Tutorial questions / Team planning**

| December 9, 2020 | TCSS562:Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma | L20.13 |

OBJECTIVES – 12/9

- **Questions from 12/7**
- **Presentation Questions**
- **Term Project Paper,** Term Project Presentation
- **Group 4 – Paper:** *BATCH: Machine Learning Inference Serving on Serverless Platforms with Adaptive Batching* *David Melanson, Samuel David Adams , Richard Brun*

- **2nd hour:**
- **Group 2 – Paper:** *Serverless in the Wild: Characterizing and Optimizing the Serverless Workload at a Large Cloud Provider* *Enbei Liu, Jingru Zhao*

- **Office Hours / Tutorial questions / Team planning**

| December 9, 2020 | TCSS562:Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma | L20.14 |

OBJECTIVES – 12/9

- **Questions from 12/7**
- **Presentation Questions**
- **Term Project Paper, Term Project Presentation**
- **Group 4 – <u>Paper:</u> *BATCH: Machine Learning Inference Serving on Serverless Platforms with Adaptive Batching***
  David Melanson, Samuel David Adams , Richard Brun
- **<u>2nd hour:</u>**
- **Group 2 – <u>Paper:</u> *Serverless in the Wild: Characterizing and Optimizing the Serverless Workload at a Large Cloud Provider***
  Enbei Liu, Jingru Zhao
- **Office Hours / Tutorial questions / Team planning**

| December 9, 2020 | TCSS562:Software Engineering for Cloud Computing [Fall 2020]
School of Engineering and Technology, University of Washington - Tacoma | L20.15 |

WE WILL RETURN AT ~7:00PM

OBJECTIVES – 12/9

- **Questions from 12/7**
- **Presentation Questions**
- **Term Project Paper, Term Project Presentation**
- **Group 4 – <u>Paper:</u> *BATCH: Machine Learning Inference Serving on Serverless Platforms with Adaptive Batching* David Melanson, Samuel David Adams , Richard Brun**
- **<u>2nd hour:</u>**
- **Group 2 – <u>Paper:</u> *Serverless in the Wild: Characterizing and Optimizing the Serverless Workload at a Large Cloud Provider* Enbei Liu, Jingru Zhao**
- **Office Hours / Tutorial questions / Team planning**

December 9, 2020	TCSS562:Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.17

QUESTIONS

December 9, 2020	TCSS562: Software Engineering for Cloud Computing [Fall 2020] School of Engineering and Technology, University of Washington - Tacoma	L20.18

QUESTIONS

December 9, 2020

TCSS562: Software Engineering for Cloud Computing [Fall 2020]
School of Engineering and Technology, University of Washington - Tacoma

L20.19

TCSS 562

OFFICE HOURS

PLEASE SAY HELLO

OFFICE HOURS

HAVE STEPPED OUT

WILL RETURN SHORTLY

AREAS OF THE CLOUD

- **Area:** Serverless Computing
  - Function-as-a-Service
  - Container-as-a-Service
- Infrastructure-as-a-Service Cloud
  - Virtual Machines
  - Containers & container clusters (Kubernetes)
- **Perspective:** cloud provider vs. cloud consumer
- **Applications:** tsunami modeling, bioinformatics, environmental modeling
- **Problem:** driven by the area & perspective
  - Common problems: what is the right abstraction? → observability
  - resource contention, resource heterogeneity, provisioning variation, performance variability (delta between min/max performance)