Recitation 01/22: Course Projects

Sagar Jha
Kwanghyun Lim
Theo Gkountouvas
About me

- 5th year PhD student in CS (https://www.cs.cornell.edu/~sagarjha/)
- Working with Prof. Ken Birman on Distributed Systems
- Interested in building efficient reliable systems, specially on RDMA networks
- TAed the course in 2018
- Office Hours - Tuesdays 2pm - 3 pm, Thursdays 12pm - 1pm in Rhodes 405
- Meeting by appointment for project discussion
Recitations plan

- I will go over the lectures in more detail, at a slower pace
- I will answer questions about the course material, projects and logistics
- I will discuss solutions for homework/ quizzes
- We will have guest lectures from time to time
- Next recitation will be Prof. Julio Giordano talking about dairy projects
Project Organization (tentative)

- Group Formation and Project Idea (February 7th)
- Project Plan (February 21st)
- Intermediate Report (March 27th)
- Peer Reviews (April 10th)
- Final Report (May 1st)
- Presentation/Poster/Demo (May 4th-May 7th)
  - Exception might be made, since your presence is required.
Project Grading

- Project + homework accounts for ~50% of your total grade.
  - 10% -> Intermediate report
  - 10% -> Peer Reviews
  - 20% -> Final Report
  - 20% -> Poster
  - 20% -> Presentation
  - 20% -> Demo
  - +10% -> Presentation in BOOM (April 29th).

- MEng Project
  - MEng project grade same as the course grade
Group Formation & Project Idea

➢ List of group Members
  ➢ Name, Net ID

➢ If this is an M.Eng. project mention who is taking the M.Eng. credits.

➢ Two paragraphs about the project idea.
  ➢ What are you trying to achieve?
  ➢ Why is it useful?
  ➢ Briefly mention how you are going to do it (input data, analysis, etc.) in one paragraph.
Intermediate & Final Report

Should consist of the following sections:

1. Motivation (idea, why it is useful, etc.)
2. Background (what is the current state?)
3. Design + Implementation
4. Evaluation
5. Conclusions
Project - IoT application

- Input
  - Sensor Data
- Data Analysis
  - Machine Learning/Optimization
- Output
  - Visualization of Data
  - Control Actions
  - Recommendations
  - Alarms
IoT application - Architecture
IoT application - Technical Challenges

1. Data Collection
   - Integrate data from possibly multiple sources
   - Integrate data from third-party sources (Weather)

2. Data Analysis
   - Utilize ML/Optimization tools to analyze data.

3. Scalability
   - Scale up to million/thousands devices

4. Hardware on Site*
   - Tune devices on site to work accordingly

5. Cost-Effective Solutions
Digital agriculture projects

- Pair up with project groups from the dairy management course
- Some ideas include cow tracking, cow pregnancy analysis etc.
- Other projects: Analyze data from greenhouses
- Inter-disciplinary work: Interact with students from other departments
- Impactful projects
Project - Traditional cloud application

- Examples - New key/value store, Storage system, Banking application etc.
- Guarantee availability and consistency across failures
- Manage application membership (sharding/replication/leaves/joins)
- Optimize for performance - High throughput and low end-user latency
- Scalability - Scale with number of user, request rates, internal resources
- Recommended for
  - Students interested in pursuing MS/PhD or a career in Systems
Azure Accounts

➢ Once you form groups, you will receive some Azure credits for your project.
➢ We will have examples using Azure technologies later in the course.
➢ Other Cloud vendors might be used. We will not provide funding or assistance for them.
Examples of projects
My internship project with FarmBeats

A Case Study in Smart Agriculture

Connected Farms
Data collection with sensor deployments, drone flights, and farm equipment

AI-based Advisory
Real-time, actionable insights based on ground conditions combined with remote sensing and weather patterns

Precision Farming
Irrigation, fertilizing, weeding, and spraying applications

Traceability
Use of blockchain to track usage & compliance
My internship project with FarmBeats

Our Solution: HighwayDB + DootMQ

FarmBeats Edge

DootMQ: Serverless Data-Flow Compute System for P.R.I.M.A.E.

- Graph of Functions
  - Heterogenous data types per function
  - Rich functions
  - Spanning edge & cloud

- Queuing & Scheduling
  - Smart function instantiation
  - Tracks availability of new inputs for each function

- Data Flow Processor
  - Efficient resource management
  - Efficient data retrieval and generation

HighwayDB: Integrated, Intelligent Storage System for P.R.I.M.A.E.

- Data Intake & Organization
  - Semantic namespaces
  - Notifications
  - Efficient data placement

- Storage Manager
  - Tiering: Zoned 3D-NAND & 3DR
  - Smart archival
  - Aging: intelligent lossy compression & deletion

- Backup & Upload Service
  - Priority uploading
  - Throttling
  - Stateful & resilient

Other sources: satellite etc

FarmBeats Cloud

Containers...

Batch: Data from a single survey

Data Batches

Stream of batches per drone

Containers...

Data box disk for low-pri

Azure Files/DB/ SMB for high-pri
Image-based authentication for banking

Yu Gu, Gloria Xiao, George Li
Admomo

Saksham Ppreja, Vidhant Maini
Smart Highway Simulation

Cameron Love