

# Introduction to IoT: Autumn 2019

Exercise set: 5

Due on 9th October 2019 by 16:00.

**Instructions:** All course participants are requested to submit their exercise solutions (in English) electronically to the instructors Agustin Zuniga (agustin.zuniga at helsinki.fi) and Prof. Petteri Nurmi (petteri.nurmi at cs.helsinki.fi) by the due date. Use the following subject in your email: *IoT-week[#]-[last name\_first name]-[student number]*, (i.e. *IoT-week4-Zuniga-Agustin-12345*)

Your submission have to contain no more than **five (5)** single-spaced and numbered pages. Use font type Arial or its equivalent with size no smaller than 10 points. Include the exercise set number, your full name and student Id in the upper right corner of the first page.

In all the exercises, do not just give the answer, but also the derivation how you obtained it. Participants are encouraged to review course material to answer the problems and in some cases write computer programs to derive solutions.

**Learning objective:** In this set of exercises you will understand better the data management layer in *IoT* contexts. The tasks will help you to be more familiar regarding the basic requirements for data management, data representation, storage and retrieval in IoT applications.

## Task 1 (2 pts.)

Answer the following questions regarding the queries in the bottom:

1. What does the query do?
2. Is this an event-based, temporal aggregate or lifetime query? Refers to Lecture 9.

*Queries:*

1. ON EVENT fire-detect(loc):  
SELECT time, event.loc, MAX(temperature)  
FROM sensors  
WHERE event.loc\_id == 'critical'  
SAMPLE PERIOD 600s FOR 3600s
2. SELECT WINAVG(pressure,10s,2s), WINMAX(pressure,10s,2s)  
FROM sensors  
WHERE altitude > 1000  
SAMPLE PERIOD 2s

## Task 2 (2 pts.)

An IoT application requires the following operations to retrieve data from sensors:

- Accessing every sensor's features.
- Performing data aggregation by grouping the tuples.
- Implementing indexes for boosting future operations.
- Performing multiple queries simultaneously.

Which query data management system(s) would you use and why? Refers to Lecture 9.

### Task 3 (2 pts.)

Assume there are 3,000 temperature sensors deployed distributed throughout Helsinki Area.

- How would you parallelize the tasks to get the GPS coordinates of the location with the highest average temperature in Helsinki every hour? What is the job and what would the tasks be? How should the query be distributed along different nodes?
- Perform the query to calculate the number of times that sensor at GPS coordinates (0, -78.467834) collected data during the last 10 minutes. Consider that query is retrieved for one week every one hour.

### Task 5 (4 pts.)

A car manufacturer has added IoT features on its vehicles. Specifically, its cars use IoT-based intelligence to improve driving quality and safety while driving. One of the functions includes periodically collecting data while the vehicle is being operated. Drivers can access this data locally to know the current driving conditions including speed, location and traffic. Additionally, data is sent to the back-end of the manufacturer to evaluate the performance of the components in the vehicle, to assess they could be improved in upcoming models, and to inform the driver of potential faults in the car.

- How data management would be performed and where? Draw a free form diagram illustrating the whole data management pipeline.
- What type of jobs and tasks can you identify? Where would you execute the jobs and tasks? Explain what would you perform in the cloud, gateway and locally.
- (Bonus): How would you implement the task using nearby vehicles instead of remote infrastructure?

Justify your answers to get the full points. Refers to Lecture 10: IoT and Big Data.

### Bonus task 1 (4 pts.)

Write a simple pseudocode for (a) *Antelope: Max heap* and (b) *TinyDB's SRT*. Refers to slides 26 and 32 of Lecture 9.

### Bonus task 2 (6 pts.)

Select two different IoT applications, one that uses warehousing model and the other that uses distributed model.

- Describe the application and justify why it uses such model.
- Which dimensionality would you use in your application? Enumerate the features you would use and their function.
- Describe how the 5'Vs fit in your application?