PART I

-Group member: Magno Alessandro

Requirements

I have selected IoT as my domain and the application I plan to consider concerns the monitoring of classroom temperatures within the various departments of the University of Genoa. It is a data analytics application with aggregate operations, like average or plotting graphs. The main entities are the buildings, the rooms, the devices (e.g. raspberry pi or Arduino) equipped with sensors, which acquire some data of the surrounding environment such as temperature, humidity, pressure. There are three relations: the first one connects a building and the corresponding rooms, the second one connects the rooms with the sensing device, the last one refers to the data acquisition through the devices. A typical workload is given a day showing the temperatures acquired in the various rooms of the department of Valetta Puggia. Since devices acquire data at a certain frequency, the application is write-intensive and it is common for metrics to be read in batches for aggregate operations. The system requirements are the following: partitioning, replication, high availability are needed because I want guaranteed data delivery (no losses during transportation of data from the device to the data analytic system), linear scalability, performance and low latency. Another requirement is eventual consistency. The dataset is synthetic, it will be generated with python script and will have the fields described above in the application: department, locale, device, acquired data (temperature, humidity, pressure). The NoSql system that I will be use is Cassandra because supports the requirements described above, it is able to manage fault tolerance, it is horizontally scalable, performant and can be integrated with popular analytics platforms as Hadoop and Spark.