R & D Project – Compressed Analytics

Student: Morten Lyng Rosenquist – 201706031

Supervisor: Daiel E. Lucani Rötter

Date: 16/12 – 2021

Semester: Spring 2022

Project Statement

How can the random access capabilities of Generalized Data Deduplication (GDD) be utilized to carry out analytics on compressed IoT data? How is the database to be structured to ensure efficient and accurate queries?

Objectives

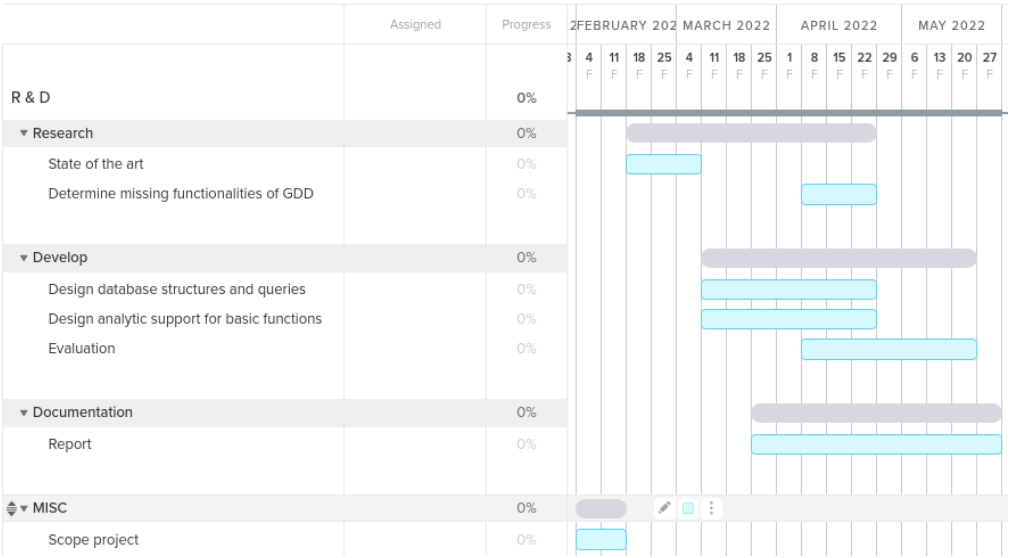
The goal of the project is to develop and design a database to perform analytics on Generalized Data Deduplication [1] compressed IoT data. Existing compressor and queries developed by the NETX group at Aarhus University may be utilized. Reasonable transformations for GDD must be investigated in terms of IoT data. Furthermore, the performance of the database and queries must be analyzed and evaluated.

Work Plan

What is to be done:

- Research state-of-the-art for GDD, databases, analytics, and compression.
- Study and design of database structure and queries, starting by basic query types (e.g., value, range).
- Study and design analytic support for basic functions, e.g., average, standard deviation.
- Determine missing primitives/functions for GDD in the context of analytics and queries. (Ideally, provide an early implementation of these missing primitives)
- Evaluate the performance for specific tasks.

A days work is to be used weekly by the student and a biweekly meeting will be in place with the supervisor. An initial plan can be seen on the gantt chart.



References

[1] R. Vestergaard, D. E. Lucani, Q. Zhang, “Generalized deduplication: Lossless compression for large amounts of small IoT data,” in European Wireless Conference. VDE, 2019, pp. 1–5.