

## **Integrated Data Layer (IDL) Project**

### **College Office of Research Enterprise (CORE)**

#### **Data Warehouse for IDL research project:**

The Information and Communications Technologies Research group (ICTRG) undertook a project to create an interactive, integrated data layer for the wealth of IoT data that has been streaming in from several IoT sensors belonging to different applications. The first stage of the IDL project is to create a data layer that collects data from different data sources, including IoT data and offer access to this data to their rightful users. The second stage of the IDL is to create a data warehouse using the collected data. The data warehouse will serve on demand specific sets of data for its users as well as the regular historical data contained in the IDL.

To perform this latter stage of the IDL project, we need to perform thorough literature review research in data analytics and data warehousing to identify the latest and best practices and tools in these fields.

The scope of this research task involves conducting a comprehensive literature review of all the available open-source data warehousing technology stacks and products and data analytics tools that can be readily plugged into such data warehouse stacks.

Part of the research requirements is the highlighting of the pros and cons of the selected products and technologies and listing alternatives to mitigate problems with certain products if any.

The following is a breakdown of the research steps and objectives:

- 1- Familiarization with the IDL project, its architecture and how a data warehouse fits in the project.
- 2- Literature review of data warehousing
- 3- Literature review of data analytics and how they relate to data warehousing
- 4- Researching available data warehousing software and technologies with focus on open source.
- 5- Researching data analytics tools that can be integrated with the warehousing software and technologies, again with more focus on open source.
- 6- Propose one or more solutions based on the research results.
- 7- Prepare diagrams of the stacks of technology to build each proposed solution with some ranking in terms of its being open-source, partly open-source, or proprietary, the overall review of the solution by customers, and any disadvantages or limitations.

- 8- It would be good if you could experiment with the solution that you think best fits our need and use case and show some results or a demo. We can provide you with the sample sensor data.
- 9- Write a comprehensive report that details the findings from the literature reviews and the proposed solutions.

The final deliverables from this research task are the comprehensive report of the literature review and the proposed solution stacks. In the final research presentation, students should present a design of what they think is the best combination of the data warehousing and data analytics products and tools to help implement the data warehouse component of the IDL. You will also give a presentation at the end of the research to show your work and present the results.