METU, Department Of Computer Engineering

Graduation Project
Proposal Form

Important Notes

A project could be proposed by (i) a student group, (ii) a company, or (iii) a faculty member of the department by filling in this form and submitting it to 49x-proposal@ceng.metu.edu.tr by e-mail. For a project proposal, there might be a sponsoring company supporting the project and providing some form(s) of resources for the project.

Each project will be carried out by a group of 4 students over the course of 9 months, which amounts to 36 man*months. It is very important that your project's workload is around 36 man*months. Please make sure that you have a rough justification about the workload of the project.

If your proposal might contain a patentable idea or any type of intellectual property, please first make sure to follow appropriate steps (apply for a patent, etc.) before sending your idea to us. Once this form is received from you, the instructor(s) and the department has no responsibility regarding to intellectual properties of your project/idea.

All sources and documentation developed for this course are assumed to be public domain (GPL, CC or similar license) by default. If you need any exception for license and disclosure of project work, please specify this in detail in IP section of the form.

Please note that source codes, documents and issue tracking should be kept in department servers. No restrictions can be requested for limiting faculty and assistants access to student work.

Project Information

Title

Streaming Data Analytics Platform

Target

Public [X] Restricted [ ]

Proposer Information

Name(s)	*Dr. Pınar KARAGÖZ*
E-Mail(s)	*karagoz@ceng.metu.edu.tr*

IP (Intellectual Property) Information

IPR is co-owned by the proposer and the project group(s).

Project Description and Background Information

Description

The need for streaming data processing is increasing in business life. Especially in telecommunications and smart grid environment, streaming data processing projects are on the increase. IoT solutions that can handle streaming data sent from sensors. Applications like Uber are also using such technologies.

In this project, the aim is to develop a software for streaming data analytics. The input to the software is a set of streaming data resources. The software will provide a set of functionalities including statistical analysis, such as counting, aggregating and predictive analytics/data mining functionalities on the streaming data.

Similar Products/Projects

There are several solutions available on the web, here are two samples, one from Turkey and one international

Turkcell Curio
https://curioweb.turkcell.com.tr/home.html

MS Azure Stream Analytics
https://azure.microsoft.com/en-us/services/stream-analytics/

Justification of the proposal

The need for streaming data processing expertise is increasing in business life. Especially in telecommunications and smart grid environment, streaming data processing projects are on the increase. Applications like Uber are also using such technologies. With this project, I aim to provide means for gaining experience for streaming data processing technologies, especially for streaming data analytics.

Contributions, Innovation and Originality Aspects of the Project

There are open source stream processing solutions available. The project will make use of them. There are also commercial tools that claim to have analytics functionality. However, how well they answer the need is unclear.

The contribution of the project is basically due to analytics capability. The idea is to provide a set of functionalities common for different types of streaming data, and to design/develop an expansible environment with user defined functionalities.

The project can initiate further research on developing novel streaming data mining/predictive analytics algorithms.

Technical Aspects of the Project

In very broad terms, the software can be considered as of two layers

At the client site/ level, the software will involve input (streaming data resource) definition and output (result of the applied functionalities) presentation. The output will be possible in the form of a dashboard in the form of charts, graphics, time series presentation etc.
This level should also have another interface that support defining new functionalities such as a specific prediction algorithm or a specific complex event to be traced.

The nature/type of the data stream may vary, but within the scope of the project, it is planned to focus on numeric (such as financial or IoT data streams) and text (such as microblog posts). Processing streaming image data may have interesting applications on surveillance systems such as keeping track of number of certain objects or prediction on the basis of what has been tracked.

The server site will involve several modules including data access, analytics functionalities, storage of the results and querying.

Targeted Output, Targeted User/Domain Profile

The end-product will be a web/desktop software, possible also having a mobile device app, involving a server module.

The success of the project will be measured through as to whether the following abilities are fulfilled

- *ability to process at least one data stream*
- *ability to provide a basic set of statistical and predictive functionalities*
- *ability to support developing new functionalities*

(Provide information about the users / user groups and/or domain that will utilize the product.)
There are many domains that can use this software such as tel-co, smart grid/energy, IoT, …

Project Development Environment

The project will involve using open source (probably Apache) stream processing solutions (such as Kafka, Storm/Heron, maybe Spark)

External Support

There is no specific hardware requirement for the project. The project will involve using open source (probably Apache) stream processing solutions (such as Kafka, Storm/Heron, maybe Spark). Such software generally calls for a multicore server.

References

Turkcell Curio
https://curioweb.turkcell.com.tr/home.html

MS Azure Stream Analytics
https://azure.microsoft.com/en-us/services/stream-analytics/