Development Process and Project Plan
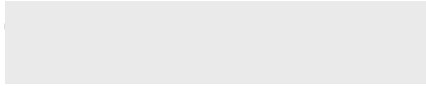
Table of Contents

Project Plan Guidelines

This project plan document is meant to be a living document that is meant to artifact the overall history and thought processes of Team _____. This project plan document will be useful in creating a development cycle and process going forward with the development of this project. While this document can be used to give an overview of the project that _____ is tasked with, this will also deal with various factors that the team involves itself with. This involves scheduling, risks, metrics, and technical process.

Overview

Team _____ is currently assigned to creating an extension to the BeakerX application. BeakerX is an open-source collection of kernels and extensions to the Jupyter interactive computing environment. Jupyter is an open-source notebook application that enables users to share and create code, equations, and visualizations with one another.

BeakerX's current JSON-based data translation will be overhauled in order to utilize the Apache Arrow format. Apache Arrow is an open-source project for common cross-language, in memory, data formats. By providing this in-memory data format as a cross-language common data representation, the time and computation power required to convert one language to another is greatly reduced. This will allow data scientists to allow their dataset sizes within Jupyter to grow without as much of a concern for high overhead.

Increased focus will be given towards Python (for utilization with the Pandas framework), JVM (for utilization with both Groovy and Scala with Apache Spark), and Javascript (for utilization with the d3.js or ArrowJS framework).

To demonstrate the improved performance of the Apache Arrow implementation, example notebooks will be created in Jupyter in order to produce a performance comparison against the original JSON-based implementation.

Goals and Scope

Conversion from JSON translation to Apache Arrow - BeakerX's existing JSON-based data translation will be replaced with an Apache Arrow based implementation. The Apache Arrow format will be able to be used by Python (for utilization with the Pandas framework), JVM (for utilization with both Groovy and Scala with Apache Spark), and Javascript (for utilization with the d3.js or ArrowJS framework).

Benchmarking - The new implementation will be able to transfer data from different languages within Jupyter by using BeakerX faster and with potentially much larger datasets than the original JSON-based implementation. Benchmarking will include performance metrics related to the timing, resource usage, and scalability.

Example Notebooks - Example notebooks will be created to demonstrate the new implementation.

Developer and User Documentation - Documentation describing the new components and features of the new Apache Arrow based implementation.

Open-Source - Interact and contribute to the open-source community throughout development.

Deliverables

Project Website - Website that stores the information about the project that ⬚ is in charge of. This website will give an overview and other project planning deliverables.

Application - The ⬚ feature promised to Two Sigma.

Risk Management

Please see 'Risk Management' document.

Scheduling & Estimates

A scrum based methodology will be used.

[Insert Chart Here]

Estimations will be addressed in various stories detailed in the Trello board that ⬚ is involved in. These will be measured in the amount of hours that each task may require.

Measurements & Metrics

Metrics required for scheduling and estimating our performance and progress will be divided up between progress and effort metrics. Progress metrics will consist of a slippage chart in order to track our ability to meet estimates and deadlines, and an earn value chart for comparing our planned values to our earned values and costs in terms of functionality and time spent on the project. Metrics required for tracking effort will consist of estimated accuracy of our predictions and measured effort by time of activity.

Technical Process

Sprint 0 - Prior to the project kickoff, this sprint will be dedicated towards elicitation of requirements and setting up tools for group communication and collaboration. We will be using tools such as Trello and Slack as a means of communication, as well as a story tracker board for delegating tasks and planning our development process going forward.

Sprints - Sprints will be divided up into two week intervals. The first week of each sprint will be dedicated to research, planning, and discussing our progress and status with the project sponsor. Each sprint will focus on a varied amount of stories detailed and pointed by based on the input received from the sponsor. The second week in each sprint will be dedicated towards developing and fulfilling each story detailed.

Sprint Review Meetings - Sprint review meetings will be held at the end of each sprint, the day before the meeting with the project sponsor. Each meeting will go over what was completed in the ongoing sprint, and the completed stories will be discussed with the project sponsor during meetings. This will also be the meeting used to discuss various questions or concerns for the project sponsor.

Story Grooming Discussions - These discussions will be take place after the project sponsor meeting. They will provide   the opportunity to reflect on the current direction of the project and to allocate upcoming development time towards higher priority user stories based on project sponsor feedback.

Daily Standup -   members will briefly go over everything they have accomplished since the last meeting, as well as what they plan on working on next. Members should state any known or foreseeable problems with their current task and ask for assistance.