Prototype Big Data Archive in a Public Cloud

Group 56: Pathfinder of Big Data

Zhi Jiang, Isaac T Chan, Zhaoheng Wang

CS 461: Senior Capstone Fall 2016

Oregon State University

Abstract

This document is to present a reviewing and summarizing for progress of our group in the past ten weeks. It includes purpose of this document, brief explanation of what we did in the past 7 weeks, current situation and a retrospective form.

✦

CONTENTS

1 PURPOSE

The purpose of this document is to demonstrate where we are in the project after fall term. We will briefly recap the project purposes and goals to provide context. Then we will provide a week-by-week summary of the project for fall term. This summary will describe the activities, problems, and solutions of each week. Then we will discuss our current position in the progress and describe a high-level plan for spring and winter term. This document will conclude with a retrospectives chart, with positives, deltas, and actions for the last seven weeks.

2 PROGRESS

2.1 Week 3

We had an appointment with our client. Before writing the problem statement, it was important to fully understand the entire problem. Our client clarified the details and we were able to understand more of what was wanted.

The most important issue that we encountered was the database. As a group, we are familiar with SQL databases, but the concept of a NoSQL database was unfamiliar. We needed to evaluate some ways to implement the NoSQL database. As we began writing, we researched NoSQL databases and learned more about them.

Additionally, big data and AWS was unfamiliar to most of our group. Some side research had to be done to catch up.

Another problem that occurred was not accounting for the time the client needed to read through and assess our document. We were lucky that he was able to sign and return it in a timely manner, and we resolved to finish assignments well before deadlines in the future, in order to not have to rush our client.

2.2 Week 4

Due to the extended deadline for the problem statement (which we had already completed), this week somewhat of a transition week.

We received feedback of our first problem statement submission, which had minimal feedback from our instructors. After making the changes, we sent it to our client for another signature and also scheduled an appointment for the next week, in order to discuss our requirements document.

Also, this week was our first week meeting with the TA. He discussed future meeting format, and we asked a lot of questions about the requirements document.

Because there was not a primary assignment due this week, we mainly worked on filling personal gaps in knowledge and planned the requirements document. Zhi reviewed commands on Git. Isaac learned more in-depth in Latex. Zhaoweng researched NoSQL compared to SQL databases and also learned about Latex.

2.3   Week 5

In this week, we have finished the our problem statement document and start to write the requirement document. In this document, there are some requirement details that we didn't know. Therefore, we make an appointment with our client and our TA. After talking with TA, We get more clear visual about the general process for dealing with big data. After checking with our client, we define the requirements that we need.

Actually, the specific requirement part for our project is the main problem we have. We are still not clear with the some parts such as interfaces and functional requirements. As a result, we finish the rough draft of the requirements document and we send it to the client in order to get more feedbacks about the specific requirement parts.

2.4   Week 6

We gained the grade of problem statement but it was unsatisfactory, so we met with Kirsten and discussed our grade. Kirsten given us useful suggestions and comments about the organization of the entire document, and fortunately, we are allowed to do some appropriate adjustment for this document.

Second important thing is that we finished the first draft of requirements document and we received feedback from our client and TA. Our client thinks it is good but in TAs opinion, there still exist some serious issues in our document. Our format of Latex is not correct because it is not IEEE format. On the other hand, we did not complete Specific requirements section in this document. The reason for this issue is actually we did not understand our product completely at that time, and we did not consider more details about this portion. Moreover, we still did not put our schedule in this document.

Benefit of requirements document is we can know what our product really is according to specified each requirement continually. TA told us this document would be evaluation criteria for the final product, in the other word, we need be able to implement all of the requirements we presented in this document.

2.5   Week 7

We completed requirements document on the basis of feedback our TA given in last week. We changed the format of Latex, specified Specific Requirements section, and added a schedule made by Gantt Chart. Moreover, we still improved our document in other places. As we had a deeper understanding of our product, we added two more sections, which are Hardware interface and Software interface. On the other hand, we still added more details for Performance metrics.

We started to prepare for upcoming technology review, but we all encountered some problems while we were researching for this document. Zhi thought there are many concepts he does not know, so it is hard to understand the relationship between these technologies. Isaac found the information we have right now is difficult to envision of the final product thus he worried about our final product could deviate far from the preliminary design. The problem Zhaoheng encountered is he is not familiar with visualization, so he wants to more research about it.

Primary we planned is to talk our questions with TA on account of our bewilderment on technologies of the product. We believed that TA will guide us how to understand these technologies.

2.6   Week 8

We finished technology review document and submitted it in this week. The importance of technology review document is to let us find the best solution for our product. When we compared several alternative solutions, we learned how to consider issues more thoroughly and how distinguished what the product needs are. According to our researches on this document, we still discovered AWS can provide a lot of useful tool for our product.

On the meeting with our TA, we explained more details to him and asked some questions such as the attribute of data storage and differences between data storage and database. The most important thing was TA thinks sample data is helpful us to design our product, so the problem we encountered was we did not gain sample data from our client.

We plan started to do design plan document. The purpose of this document is to expound how to implement technologies we chose in technology review, so which means we need to give an insight into these attributes of each technology.

2.7   Week 9

In this week, we make an appointment with our client to talk about our project. In this meeting, we request the sample data from our client. This sample data is important for us because the our database structure depends on how the sample data is set up. Besides, there are some elements which will relate with the sample data such as whether do we need to clean, parse, or do other process the data before it can be stored in the database. Besides, we also finish revising our problem statement and organize it into more clear way.

The problem for this week is we are still waiting for the sample data from our client. The sample data will give us more clear visual about how database structure should be organized. Besides, it will clearly show the step we need to do in process data part. Another problem is that we are not famliar with the our technologies. Therefore, we decide to do more research for our technologies and once we get the sample data we can start doing our design document as soon as possible.

2.8   Week 10

In this week, we receive our sample data and finish our preliminary design document.In this document, we introduce our technology based on various such as context viewpoint, algorithm viewpoint and information viewpoint. This document gives us more clear visual for our project. Besides, we make a new timeline for our project because we make some changes for our project.

The most difficult problem is about our technologies. We are still in hypothetical terms when we finish our design document.Therefore, we still need more time to familiar with how to implement these technologies. Fortunatly, we receive the sample data from our client but there is one more problem about the sample data. The sample data is storing in csv file type which is already readable by using Excel therefore, we might check again with our client to figure out whether the furturn data type is the same with the sample data type. Besides, we will also check whether the sample data contains all the information we need to store. After that, we plan to implement the technologies base on the sample data.

3   CURRENT SITUATION

From the problem statement document, we understand the purpose of this project and the general process for the project. From the requirements document, we start to narrow down the scope from the problem into requirements and we describe the project from attributes, functionalities and performance from a high-level. From technology review document, we break the requirements of project into small pieces. Each group member choose their three pieces and figure out the optimal choices. Now, we have finished the design document of the project. In this document, each group members introduce their technology part in more detail based on different viewpoints. After finishing this document, we get more clear visual for the project and we plan to figure out the specific step for implement each technology. After that, we would like to make a more specific milestone for each step to start implementing the prototype.

4   RETROSPECTIVE

Positives	Deltas	Actions
Met our client timely	Preparation for meeting was not substantial	Prepare questions in advance to thoroughly cover understanding
Completed problem statement	Organizational problems	Met with Dr. Winters to address problems and resubmitted
Completed first draft of requirements document	Unsure about functional requirements	Met with the client to discuss functional requirements. Resolved the issue with user stories and acceptance tests.
Our client thought the requirements document was good	Our TA did not approve of conversion from Google document to IEEE	Will start other documents in IEEE earlier
Finished the requirements document	Overwhelmed by planning in-depth technologies for the technology review	Met with the TA to look closer at these technologies and gain an understanding
Finished the technology review	Worried that our assumptions can be wrong when we actually see the data and begin implementation	Our TA recommended us to get some sample data to examine
Got the sample data from our client	Still is difficult to envision the prototype before implementation	Examine the data types and how to implement
Finished the preliminary design document	Unsure how well the sample data we received from our client is representative of the actual data we will receive	Meet with our client when we return from break to determine next steps to begin implementation