Passive Wireless Monitoring System

Final Project Report

Date: May 8, 2019

Prepared by: Justin Plassmeyer, Hammad Hanif, Rebecca Yim, Mansour Faragallah

Table of Contents

1.0 Executive Summary 3

2.0 Objective 3

3.0 Challenge/Issues 3

4.0 Approach 3-4

5.0 Architecture/Data Model 4-5

6.0 Verification Approach and Results 5-6

7.0 Schedule Summary 7-11

8.0 Deliverables 12

9.0 Customer Satisfaction 12-13

10.0 Lessons Learned 13

11.0 Project Costs 14

**1.0 Executive Summary**

Team Olympus created an end to end passive monitoring system to collect environmental data and provided a time series analysis of said data. We will be using the time series analysis to provide a long-term trend analysis of the collected data. We used a time series database with CRAN-R to develop the analysis. The purpose of this project was to research and deepen our understanding of time series data in order to extrapolate meaningful data from our analysis.

**2.0 Objectives**

1. Evaluate and select a Time Series Database.
2. Integrate collected data into Time Series Database.
3. Integrate Time Series Database with the CRAN-R Space-time library.
4. Provide a document explaining the pros and cons of different databases, reason behind our decision, and value of the approach.

**3.0 Challenges/Issues**

Team Olympus explored many new concepts while working on building the Passive Wireless Monitoring System, Challenges and Issues were encountered while implementing these concepts.

1. Setting up the Time Series database.
   1. Configuring the underlying components such as Hadoop, Hbase and Zookeeper.
2. Connecting the sensors to database.
3. Formatting the data pulled from the database in a format that R accepts.

4.0 **Approach**

Task 1: We created a document explaining our choice of Time Series Database and how it works. We also explained usability and what time series analysis are used for. In that document, we also discussed the pros and cons of each TSDB we found and compared them with each other.

Task 2: We researched the CRAN-R library and learned how to use required functions to run our analysis.

Task 3: We installed the Linux partition on our machine and installed OpenTSDB and its underlying components to hold the collected data from sensor nodes.

Task 4: Once we received the sensor nodes, we researched on how to set up those sensor nodes and we connected them to our TSDB.

Task 5: We used R with CRAN-R library in order to perform our time series analysis.

**5.0 Architecture/Data Model**

After researching we decided to go with a database-centric Architecture. This type of architecture made sense for our project as our central hub of information was our database. Also, since we were not making a user interface with the system, there was no need for a more complex architecture, as this one satisfied all of our needs. After we decided on the architecture to use we moved onto setting up the database. The architecture of our project is displayed in the figure 5.1.

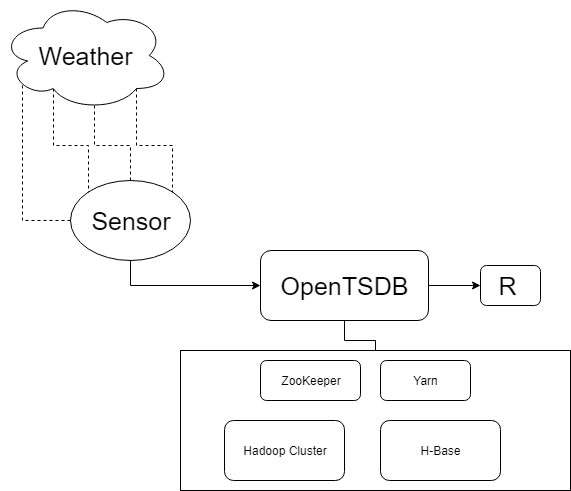


Figure 5.1 The overall architecture of our project.

The data model is unique as it is specific to that of a NoSQL Time Series Database. Data in a TSDB has a few core properties, something signifying who the data point belongs to. In Figure 5.2, “What” is normally the value being stored, and “when” is the timestamp of when the data was collected. This timestamp attribute is one of the main reasons why it is called a time series database. Figure 5.2 represents all the attributes of a single data point.

{

**"Metric"**:"d1.lsm.accel.z", ----> Who

**"Timestamp"**:1551471760, ----> When

**"value"**:-9.909276592249998572, ----> What

**"tags"**:{ ----> Who

**"sensor"**:"LSM9DS1", ----> who

**"Device"**:"1" ----> Who

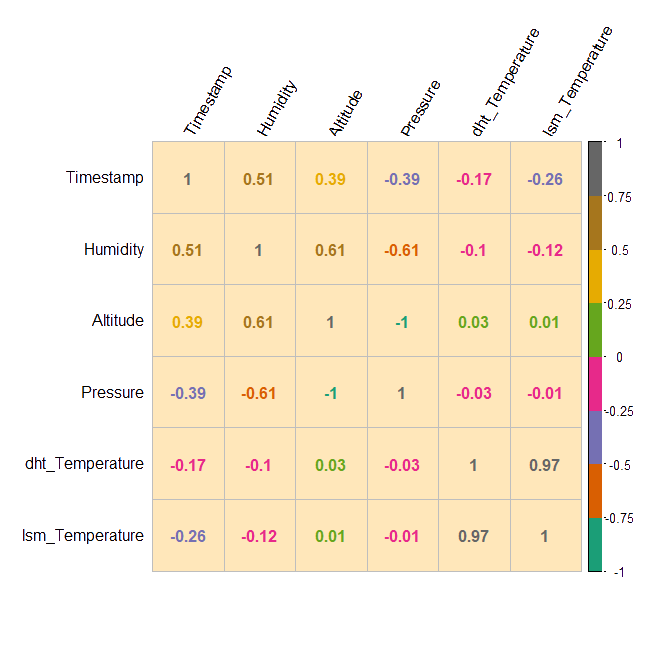
}

Figure 5.2 Example of data sent to database.

**6.0 Verification Approach and Results**

We verified our data by comparing the data collected between the two sensors that were provided. We could also verify our analysis by confirming that the results matched up with what is scientifically true. We also cross-validated our analysis between the results of the two sensor nodes.

In order to better understand the data and find interesting results, several different types of time series analysis were performed. The first time series analysis to be performed was detection of any correlation between the different data fields. Several strong correlations were found in the data; however, the strongest correlation found was between pressure and altitude. They had a correlation of -1 which means they have a perfect negative correlation. This was a surprising result, as the sensors nodes were stationary throughout the duration of the data collection. The correlations are illustrated in Figure 6.1.

Figure 6.1 Correlation between data fields.

The decomposition of data was then performed to better visualize any trend of the data. Time series decomposition is a mathematical procedure which transforms a time series into three different time series: trend, seasonality, and “random noise”. The trend time series illustrates any underlying trends of the metric. Seasonality displays the patterns that repeat with a fixed period of time. And “random noise” is the residuals of the time series after the trends and seasonality has been extracted. However, in order to decompose the data, it’s necessary to detect the period of the seasonality first. Fourier Transforms in R is one popular way to find the period of seasonality in time series data. The idea of it is to decompose a signal into all the possible frequencies that comprise it, and the frequency with the highest power represents the greatest periodicity. The decomposition of the temperature data we collected is illustrated in Figure 6.2.

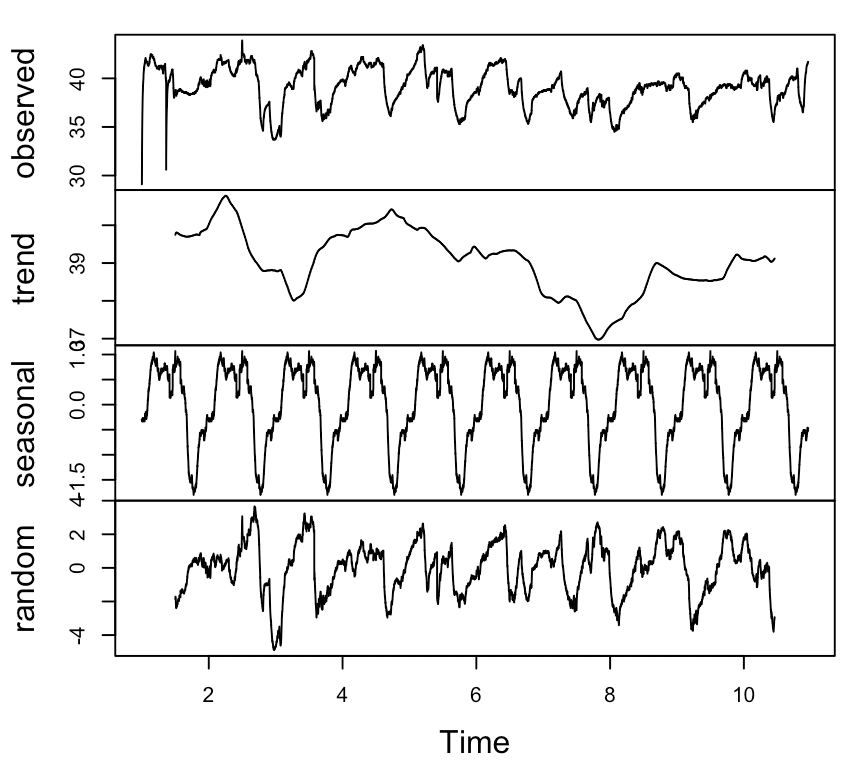


Figure 6.2 Decomposition of temperature data.

The last time series analysis performed was data prediction. There were two approaches to predict the next range of data. The first approach, displayed in the figure below, used CRAN-R’s forecast library which created a range of values where the future data would fall in.

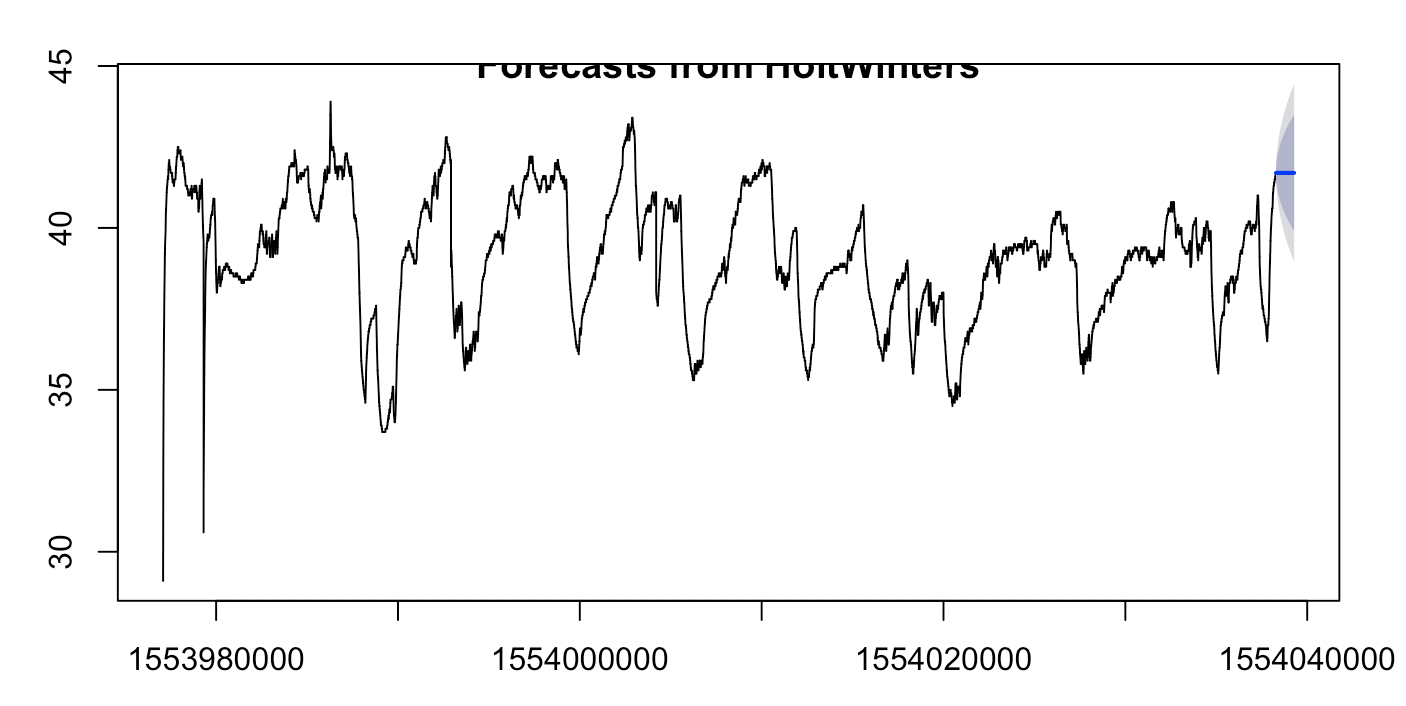


Figure 6.3 Data prediction using approach one.

**7.0 Schedule Summary**

Project Schedule Fall 2018

In the Fall, the team focused on selecting a database, researching, and setting up the database.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Task ID | Activity | Duration | Wk1 | Wk2 | Wk3 | Wk4 | Wk5 | Wk6 | Wk7 | Wk8 | Wk9 | Wk10 | Wk11 | Wk12 | Wk13 | Wk14 | Wk 15 |
| 1 | Define what a Time Series Database is | 1 wk |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Research the different available TSDB | 2 wk |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Conduct further research on chosen TSDB | 1 wk |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Research the CRAN-R Library | 2wk |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Research provided Sensor Module | 2 wk |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Setting up the TSDB | 10 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Setting up the Sensor module | 2 wk |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | Connect Sensor Module to the TSDB | 4 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 | Use R to code for analysis of TS data | 6 wk |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 1 | Customer Reporting "Quad-Pack" | 7 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 2 | Weekly Activity | 30 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 3 | interim Report and lessons learned Document | 5 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 4 | Poster Paper | 3 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 5 | Final report and Team Presentation | 5 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 6 | Product deliverables | 30 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 7 | Preliminary Design Package |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 8 | Final Design Package | 5 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 9 | Database Design Documents | 6 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Table 7.1 Project schedule for Fall 2018 is broken up by week with the assigned tasks for that week highlighted.

Project Schedule Spring 2019

In the Spring, the team focused on implementing the code and analysis, and testing.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Task ID | Activity | Duration | Wk  16 | Wk17 | Wk18 | Wk19 | Wk20 | Wk21 | Wk22 | Wk23 | Wk24 | Wk  25 | Wk26 | Wk27 | Wk28 | Wk29 | Wk 30 |
| 1 | Define what a Time Series Database is |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Research the different available TSDB |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Conduct further research on chosen TSDB |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Research the CRAN-R Library |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Research provided Sensor Module |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Setting up the TSDB |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Setting up the Sensor module |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | Connect Sensor Module to the TSDB | 2 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 | Use R to code for analysis of TS data | 12 wk |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 1 | Customer Reporting "Quad-Pack" | 7 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 2 | Weekly Activity | 30 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 3 | interim Report and lessons learned Document | 5 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 4 | Poster Paper | 3 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 5 | Final report and Team Presentation | 5 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 6 | Product deliverables | 30 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 7 | Preliminary Design Package | 4 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 8 | Final Design Package | 5 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CLIN 9 | Database Design Documents | 6 wks |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Table 7.2 Project schedule for Spring 2019 is broken up by week with the assigned tasks for that week highlighted.

**8.0 Deliverables**

The following contact line items were delivered to the customer throughout the project.

CLIN-1: Customer Reporting “Quad-Pack” - We will generate a monthly report for you including an executive summary, consisting of key accomplishments, schedule, and cost status, as well as the project status and customer satisfaction survey.

CLIN-2: Weekly Activity - Our team leader will, on a weekly basis, send to our professor, an activity sheet that includes the weekly accomplishments, issues, labor hours, and expenses.

CLIN-3: Interim Report and Lessons Learned Document - A interim report will be generated on the status of the project after the first semester.

CLIN-4: Poster Paper - We will create a poster containing key components and other major findings.

CLIN-5: Final Report and Team Presentation

CLIN-6: Deliverable-1 - Proposal - Proposal will be delivered to General Dynamics by early-October.

CLIN-7: Deliverable-2 - Document Report - By the end of October, the team will deliver a document that explains which time series database the team has chosen and why. It will also list the pros and cons of all the databases considered and the value of the approach.

CLIN-8: Deliverable-3 - Progress Report - Report that will be delivered at the beginning of each month. It will provide details regarding the set-up of the database, data from the sensor nodes, and the progress made thus far.

CLIN-9: Final Design Package - The final product will be delivered to the customer at the end of the project.

**9.0 Customer Satisfaction**

General Dynamics has provided us with positive feedback throughout the project. John Kohler has recognized our progress working with Skip and was happy to see us on track despite the constraints of the effort. Skip also mentioned that we displayed agility overcoming and a number of technological constraints and issues and achieved our target. Our last customer satisfaction survey is display below in Figure 9.1.

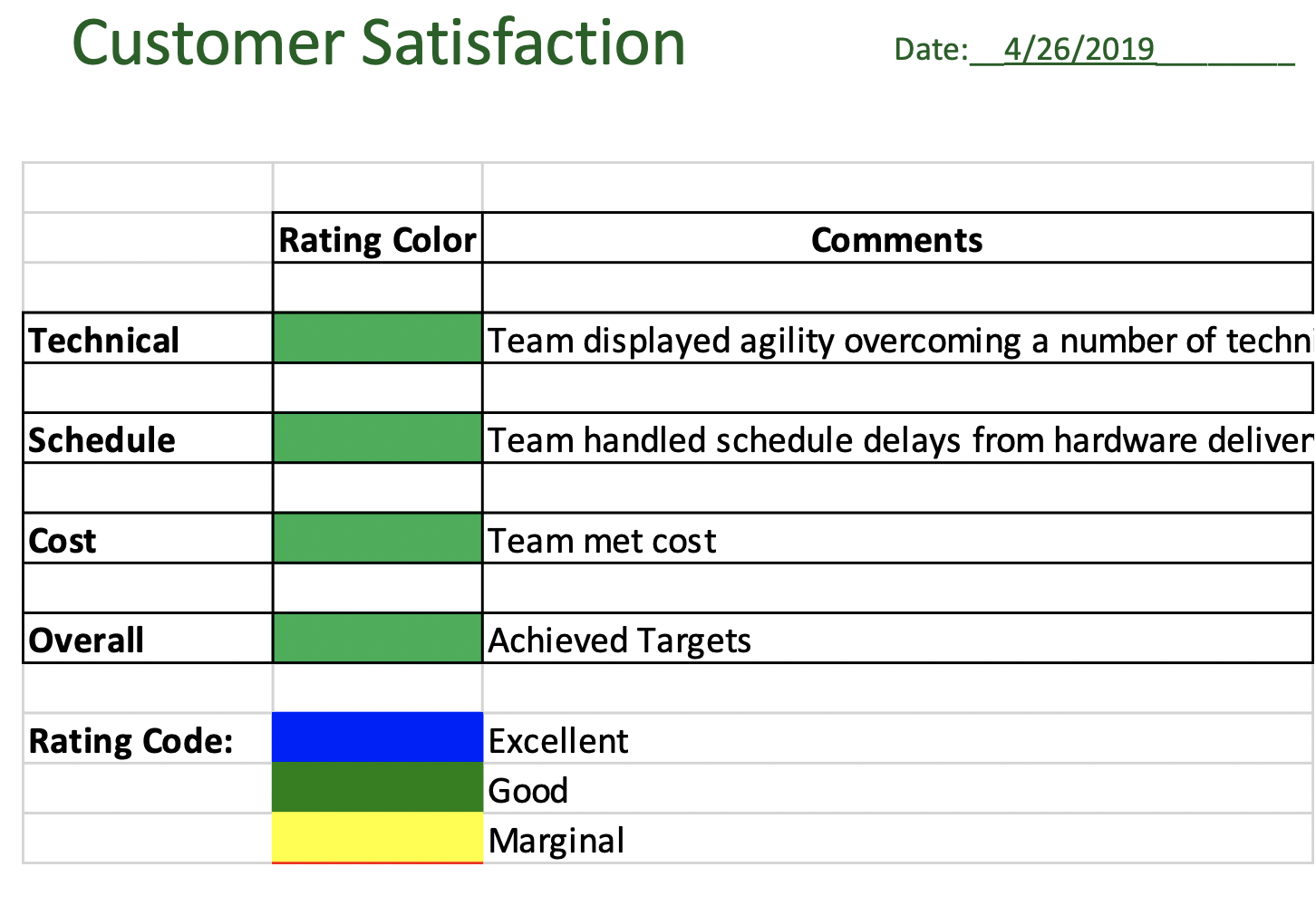
****

Figure 9.1 Customer Satisfaction Survey in April.

**10.0 Lessons Learned**

1. Installed and configured Linux operating system and components.
2. How to set up OpenTSDB and its underlying components.
3. Write Shell Scripts to send data from nodes to database.
4. Deeper understanding of what a Time Series Database is.
5. How to use R to perform time series analysis.

* Became familiar with the different strategies on how to perform analysis on Time Series database.
* Learned how to find trends using CRAN-R library.
* Learned how to predict the future using the collected data.
* Learned how to find correlations between different data fields.

**11.0 Project Costs**

The only costs associated with the project were labor costs. Figure 11.1 portrays the hours applied per week on average for the entirety of the project.

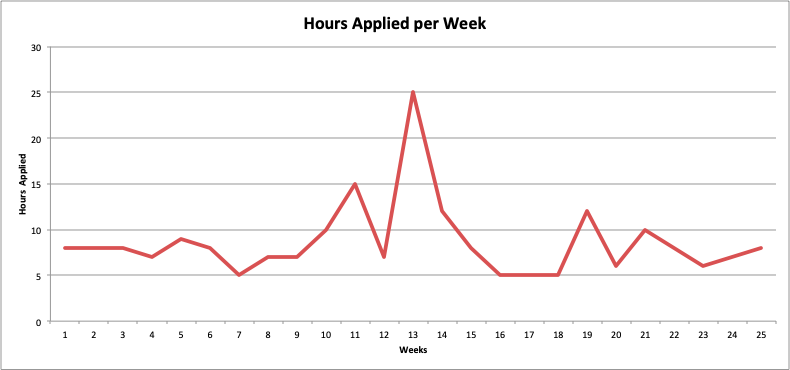


Figure 11.1 Hours applied per week.