Analyzing San Francisco Fire Department Service Responses

Project Plan

Version 1.0

Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Version** | **Description** | **Author** |
| 05/07/2023 | 1.0 | Project design and implementation, documentation | Aishwarya Joshi,  Dharani Aningi,  Sai Keerthana Kattige,  Supriya Kamble,  Vrushali Menthe |

Table of Contents

1. Introduction 4

1.1 Purpose of this document 4

1.2 Intended Audience 4

1.3 Scope 4

1.4 Definitions and acronyms 4

1.4.1 Definitions 4

1.4.2 Acronyms and abbreviations 5

1.5 References 5

2. Background and Objectives 5

3. Organization 5

*3.1* Project group 5

3.2 Customer 6

4. Development process 6

*5.* Deliverables 6

6. Project risks 6

7. Communication 7

7.1 Canvas 7

7.2 Git 7

8. Project plan 7

8.1 Time schedule 7

8.1.1 Remarks 7

8.2 Test plan 9

8.2.1 Testing Remarks 10

9. References 10

# Introduction

## Purpose of this document

The purpose of this document is to provide a detailed project description of the application called Analyzing San Francisco Fire Department Responses which is designed to help people buy better and more quality items. This document includes details about organization, roles, deliverables, project risks, time plans and financial plans.

## Intended Audience

This document shall be used in all phases of the project as a guideline. Intended audiences of this project are all project stakeholders:

* Professor Andrew H. Bond
* Dharani Aningi
* Aishwarya Joshi, Sai Keerthana Kattige, Supriya Kamble and Vrushali Menthe

## Scope

This document defines the project plan of the Analyzing San Francisco Fire Department Responses application. The overview includes objectives of the project, organization of the project team, development process that is going to be used during the project, assessment of possible risks, communication used between project stakeholders and project plan that includes time schedule and activity plan.

## Definitions and acronyms

### **Definitions**

|  |  |
| --- | --- |
| **Keyword** | **Definitions** |
| Project Name | Analyzing San Francisco Fire Department Responses |
| Project Supervisor | Professor Andrew H. Bond |
| Project Leader | Dharani Aningi |
| Team Member | Aishwarya Milind Joshi, Sai Keerthana Kattige, Supriya Kamble and Vrushali Menthe |
| Milestone | 05/07/2023 |
| Git | <https://github.com/Dharani2805/BigDataProject_228> |
| Scrum | An iterative and incremental agile software development method for managing software projects and product or application development |
| Kunagi | Web-based tool for integrated agile project management and collaboration based on Scrum |
| Scrum sprint | The basic unit of development in Scrum |
| Scrum master | Dharani Aningi |
| Product owner | Supriya Kamble |

### **Acronyms and abbreviations**

|  |  |
| --- | --- |
| **Acronym or**  **abbreviation** | **Definitions** |
| AWS | Amazon Web Services |
| ETL | Extract, Transform, Load |
| IAM | Identity & Access Management |
| S3 | Simple Storage Service |

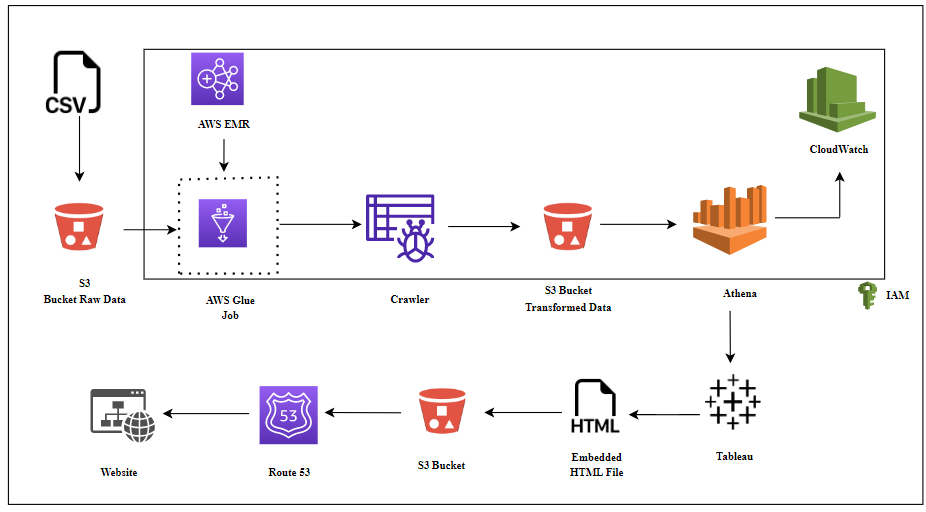
## References

1. [https://data.sfgov.org/Public-Safety/Fire-Department-Calls-for-Service/nuek-vuh3](about:blank)
2. <http://www.scrum.org/>
3. <http://kunagi.org/>

# Background and Objectives

We notice uncertain events happening around every corner every day such as fire, medical emergency, accidents etc. In such situations, usually the initial response is given by the Fire Department. In this project, we are going to analyze San Francisco Fire Department service calls and their responses. The project's objective is to pinpoint areas for improvement and assess how well the fire department's current operating procedures are by looking at trends and patterns in the demand for emergency services. To accomplish this, the project will make use of a cloud-based system, specifically the AWS cloud platform, which offers effective data utilization in addition to other advantages like dependability, security, and scalability. For additional analysis and visualization purposes, Tableau will also be used in the project.

# Architecture and High-Level Design

****

The team started by collecting the raw data from the DataSF website and loading it into the S3 input bucket. AWS Glue Studio was utilized to transform the raw data. To begin, a workflow consisting of an AWS Glue job and crawler is created. AWS Glue job was created to transform some column values to make them consistent and rename columns for interoperability. After the transformations were executed, the transformed data was crawled using a Glue Crawler and saved in S3 output bucket. CloudWatch was used for logging the activities in the job and crawler. Athena was employed to query the data and investigate its schema. Tableau Desktop was chosen to create visualizations as it is free and useful to produce interactive dashboards. Tableau was linked to Athena, and the data source was extracted to create visualizations and interactive dashboards. The dashboard was subsequently shared on Tableau Public, and an embed code was generated, which was hosted on the S3 bucket using Route 53. Throughout the project, IAM policies were implemented to guarantee secure access to the used AWS resources.

# Organization

# Project group

|  |  |  |
| --- | --- | --- |
| **Name** | **Initials** | **Responsibility (roles)** |
| Dharani Aningi | DA | Data Transformation, System Design, Documentation |
| Aishwarya Joshi | AJ | Data Visualization, Documentation |
| Sai Keerthana Kattige | SKK | Data Loading, Documentation |
| Supriya Kamble | SK | Data Processing, System Design, Documentation |
| Vrushali Menthe | VM | Data Visualization, Documentation |

## 

## Customer

The target customers are listed below:

1. Fire Departments
2. Emergency Responders
3. Government Agencies
4. Insurance Companies
5. NGOs
6. Public

# Development process

The project will use AWS Services like AWS S3 for storing the input and output file, AWS Glue Studio for creating ETL job, Glue Crawler and Glue Workflow, AWS CloudWatch for generating and analyzing logs, AWS Route 53 for creating a hosted zone using AWS S3 to host a website, AWS Certificate Manager for provisioning the SSL/TLS certificate for the website. The project also uses Tableau to create visualizations and understanding the pattern of the data. HTML script has been used create a website and host using S3 bucket and Route 53 services of AWS.

# Deliverables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| To | Output | Planned week | Promised week | Days +/- | Delivered week | Notes |
|  | Download data from DataSF website | 2/22/2023 | 2/22/2023 |  | 2/22/2023 |  |
|  | Load data into S3 | 3/29/2023 | 3/29/2023 |  | 3/29/2023 |  |
|  | Data Transformation | 4/12/2023 | 4/12/2023 | +7 | 4/19/2023 |  |
|  | Querying data in Athena | 4/30/2023 | 4/30/2023 |  | 4/30/2023 |  |
|  | Data Visualizations | 4/30/2023 | 4/30/2023 | +3 | 5/3/2023 |  |
|  | Static Website Creation | 5/1/2023 | 5/1/2023 | +3 | 5/4/2023 |  |
|  | Project Report & PPT | 5/2/2023 | 5/2/2023 | +5 | 5/7/2023 |  |

# Project risks

|  |  |  |
| --- | --- | --- |
| **Possibility** | **Risk** | **Preventive action** |
| Cost of used AWS services | Cost | Deletion of files and resources used. |

# Communication

We had weekly meetings over zoom calls for project discussions and implementation activities. Along with this, we have also taken out time to meet in-person for project related activities.

## Collaboration

1. Defining project objectives and goals.

2. Assigning tasks and responsibilities to team members.

3. Communication with team members on a frequent basis to track progress and provide feedback.

4. Coordination of tasks and activities to ensure completion on time and within budget.

## Git

All source code and finished documentation will be uploaded to GitHub repository.

Repository URL: <https://github.com/Dharani2805/BigDataProject_228>

# Project plan

## Time schedule

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Id** | **Milestone**  **Description** | **Responsible Dept./Initials** | **Finished week** | **Forecast Week** | **+/-** | **Actual** | **Metr.** | **Rem.** |
| 1 | Topic finalization, Finding Data Set | AJ, DA, SK, SKK, VM | 2/22/2023 |  |  | 2/22/2023 |  |  |
| 2 | Loading the data into S3 | AJ, SKK | 3/29/2023 |  |  | 3/29/2023 |  |  |
| 3 | Data Transformation | DA, SK | 4/12/2023 |  | +7 | 4/19/2023 |  |  |
| 4 | Querying in Athena | SKK, VM | 4/30/2023 |  |  | 4/30/2023 |  |  |
| 5 | Data Visualizations using Tableau | SK, AJ | 4/30/2023 |  | +3 | 5/3/2023 |  |  |
| 6 | Static Website hosting in S3 | DA, SK, VM | 5/1/2023 |  | +3 | 5/4/2023 |  |  |

### **Remarks**

|  |  |
| --- | --- |
| **Remark Id** | **Description** |
|  |  |
|  |  |
|  |  |
|  |  |

## Test plan

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test No.** | 001 | **Phase:** | 1 | **Authors:** | Aishwarya Joshi, Sai Keerthana Kattige | Date: 04/12/2023 |
| **Test Category:** | | System Test | | |  |  |
| **Software Product:** | | S3, Glue ETL Job | | | |  |
| **Test Title:** | | Connection between S3 and Glue Job | | | | |
| **Test Purpose:** | | Read input file from S3 via Glue Job | | | | |
| **Test Setup:** | | Script is created and executed | | | | |
| **Prerequisites:** | | Input file availability in S3 | | | | |
| **Procedure:** | | Reading file from S3 using a PySpark script and generate output file in S3 bucket | | | | |
| **Checks:** | | Data table availability in Athena for query | | | | |
| **Expected Results:** | | Output files is S3 | | | | |
| **Result:** | | Output file is generated in S3 bucket | | | | |
| **Reason for Failure:** | | N/A | | | | |
| **Remarks:** | | N/A | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test No.** | 002 | **Phase:** | 2 | **Author:** | Vrushali Menthe | Date: 04/17/2023 |
| **Test Category:** | | **System Test** | | |  |  |
| **Software Product:** | | Glue Workflow | | | |  |
| **Test Title:** | | Execute Glue Workflow | | | | |
| **Test Purpose:** | | To execute Glue job and Crawler under one Workflow | | | | |
| **Test Setup:** | | Glue workflow is created | | | | |
| **Prerequisites:** | | Glue job and Crawler are created | | | | |
| **Procedure:** | | Execute Glue Workflow | | | | |
| **Checks:** | | NA | | | | |
| **Expected Results:** | | Successful execution of Glue workflow | | | | |
| **Result:** | | Glue Workflow executed successfully | | | | |
| **Reason for Failure:** | | N/A | | | | |
| **Remarks:** | | NA | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test No.** | 003 | **Phase:** | 3 | **Author:** | Supriya Kamble | Date: 05/01/2023 |
| **Test Category:** | | System Test | | |  |  |
| **Software Product:** | | Tableau Desktop | | | |  |
| **Test Title:** | | Connection between Tableau Desktop and AWS Athena | | | | |
| **Test Purpose:** | | To connect AWS Athena to Tableau Desktop to create visualizations | | | | |
| **Test Setup:** | | Access Key ID and Secret Access Key | | | | |
| **Prerequisites:** | | Data availability tables in Athena | | | | |
| **Procedure:** | | Configuring settings for Athena and Tableau | | | | |
| **Checks:** | | NA | | | | |
| **Expected Results:** | | Connection successful between Athena and Tableau | | | | |
| **Result:** | | Athena is connected to Tableau successfully and data is available for visualizations | | | | |
| **Reason for Failure:** | | N/A | | | | |
| **Remarks:** | | N/A | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test No.** | 004 | **Phase:** | 4 | **Author:** | Dharani Aningi | Date: 05/02/2023 |
| **Test Category:** | | **System Test** | | |  |  |
| **Software Product:** | | Tableau Desktop, Tableau Public | | | |  |
| **Test Title:** | | Connect Tableau Desktop to Tableau Public | | | | |
| **Test Purpose:** | | To get embed code for dashboard in Tableau Public and host it using S3 | | | | |
| **Test Setup:** | | Publish dashboard to Tableau Public | | | | |
| **Prerequisites:** | | Dashboard is created | | | | |
| **Procedure:** | | Use Publish Workbook feature from Tableau Desktop | | | | |
| **Checks:** | | NA | | | | |
| **Expected Results:** | | Dashboard visible in Tableau Public | | | | |
| **Result:** | | Dashboard published in Tableau Public | | | | |
| **Reason for Failure:** | | N/A | | | | |
| **Remarks:** | | We used extracted data from Athena to connect to Tableau Public | | | | |

### **Testing Remarks**

|  |  |
| --- | --- |
| **Remark Id** | **Description** |
|  |  |
|  |  |
|  |  |
|  |  |

# References

<https://help.tableau.com/current/pro/desktop/en-us/examples_amazonathena.htm>

<https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00154>

<https://doi.org/10.1007/s10694-019-00870-4>

<https://www.springfield.il.us/Docs/SpringfieldILFireAnalysis012722FinalReport.pdf>

<https://docs.aws.amazon.com/AmazonS3/latest/userguide/Welcome.html>

<https://docs.aws.amazon.com/IAM/latest/UserGuide/introduction.html>

<https://docs.aws.amazon.com/athena/latest/ug/what-is.html>

<https://docs.aws.amazon.com/glue/latest/dg/what-is-glue.html>

<http://noobcoders.link.s3-website-us-east-1.amazonaws.com>