INTERNSHIP REPORT

Kamlesh Premanand Naik

Roll No: 1937

Report of Internship done at Zimetrics Technologies Private Ltd

Submitted by

Kamlesh Naik

1937

Under the Guidance of

Mr. Omkar Prabhu

(Center Head – Goa RDC. At ZiMetrics Technologies)

Mr. Aditya Kunkolienkar

(Senior System Analyst at ZiMetrics Technologies)

Internship Certificate

501, Lunkad Sky Station,
Viman Nagar, Pune –411014
Contact No.: 020-41230949
Web: www.zimetrics.com

ZIMETRICS TECHNOLOGIES PRIVATE LIMITED

Date: 31/05/2022

TO WHOM IT MAY CONCERN

This is to certify that Mr. Kamlesh Naik, a student of MCA Goa university roll no- 1937.is currently undergoing long internship program at Zimetrics Technologies Pvt Ltd.(10th Jan 2022- Till date)

During the period of his internship program, he is working on

Developing ETL for Telematics solutions as an engineering intern.

His sincere efforts and dedication towards work are greatly appreciated. He is exhibiting overall very good conduct, flexibility and professionalism during this period.

As per the campus drive conducted on 8th November 2021, he will be working with us as a full-time employee from 9th August 2022 onwards.

AshwiniBarve
AshwiniBarve (May 31, 2022 15:58 GMT+5.5)

Sincerely,
For Zimetrics Technologies Pvt Ltd
Ashwini Barve
GM Operations

CIN: U72900PN2015PTC153852, GSTIN 27AAACZ8110B1ZO
Register office: B-3, Konark Campus,Viman Nagar,Pune. 411014

GOA UNIVERSITY

GOA BUSINESS SCHOOL

CERTIFICATE OF EVALUATION

This is to certify that **Mr. Kamlesh Naik** has been evaluated for the project work titled **"Report of Internship done at ZiMetrics Technologies Private Limited"** undertaken at **ZiMetrics Technologies Private Limited, Pune** in partial fulfilment for the award of thedegree in Master of Computer Applications.

Examiner 1

Examiner 2

Place: Goa University

Date:

Dean, Goa Business School

Acknowledgement

Internship in a company is a golden opportunity for learning and self-development, especially with professionals who have tremendous knowledge of all the aspects of the technology.

I am privileged to have done my internship at ZiMetrics Technologies Private ltd. I got a great chance of learning, and professional development, and growth. The internship wouldn't be complete without expressing my gratitude to all the people who made it possible.

I would like to thank Mr. Vikas K Verma (Founder & President of Engineering, ZiMetrics Technologies) for giving me the opportunity to work as an intern at ZiMetrics Technologies Private Ltd.

I would like to thank Mr. Omkar Prabhu (Center Head – Goa RDC) for giving me the opportunity to intern at ZiMetrics Technologies Private Ltd, for making me feel comfortable in the new environment and also to guide me along the way.

I would like to extend my gratitude to Mr. Aditya Kunkolienkar (Senior System Analyst at ZiMetrics Technologies) for his support and guidance in the project.

I would like to thank Ms. Swati Patil (Head – HR Zimetrics Technologies Private Ltd), Ms. Ashwini Barve (GM - Operations), Mr Shantanu Waghmare (Associate Manager – Human Resource at ZiMetrics Technologies Private ltd), Yashwanti Patil (HR Manager) and Maseera Shaikh (HR Generalist, ZiMetrics) for helping me out whenever needed.

I would like to thank Ms. Ankita Raul (Office Admin, ZiMetrics) for always being kind, friendly and supportive towards me.

I thank M.S. Dayanand (Dean, Goa Business School, Goa University), Mr. Ramdas Karmali (Professor. and TPO, MCA, Goa Business School, Goa University), Mr. Jarret Stevan Anthony Fernandes (Assistant Professor, MCA, Goa Business School, Goa University) and all the faculty of MCA, Goa University for their constant encouragement and support during the project work.

I would like to thank my parents, family and friends for their blessings and support, and always being there with me.

Last but not least, I would like to express my gratitude towards all my colleagues at Zimetrics, especially Ashwin Kolgaonkar and Nityanand Waingankar for being kind, friendly and helpful in nature.

Table of Contents

INTRODUCTION

This report includes a short description of my full-time internship at ZiMetrics Technologies Private Limited.

I joined as an intern in ZiMetrics on 10th January 2022. This report contains all the necessary information about the company, the project I have worked on, the trainings I havereceived, and some other tasks that I completed in this internship period.

In the following sections I shall include information about the company, the work and culture over here. I shall also include details of the project I worked on, a brief description of the project, the modules I built and the tasks I completed in those modules.

This report emphasizes my learning experience and contribution to the organization as an intern. This will describe the knowledge that I gained by successfully completing the tasks that were assigned to me.

I'll also be talking about the tools and technologies that were used followed by my internship timeline. I shall conclude by sharing my experience and how it has helped me to grow, both, on the personal and professional front.

COMPANY PROFILE

ZiMetrics is a niche technology provider and solutions enabler for IoT, Machine data, BigData analytics, and Data Science.

ZiMetrics is a Confluent & HashiCorp Partner and recognized as the "Top 20 IoT SolutionsProviders" by CIO Outlook APAC.

Founded in 2015, ZiMetrics today serves leading global enterprises across Industrial, Oil & Gas, FMCG, MedTech, Internet Advertising, and Retail. Learn of our expertise and transformational value creation stories in Digital Media Analytics, Retail Analytics, Marketing Analytics, Customer Data Analytics, Video Analytics, Big Data, IOT, Machine Vision, Natural language Processing, Machine Learning, and Deep Learning.

RESEARCH AND INNOVATION AT CORE

Innovation and research are at the center stage for ZiMetrics. ZiMetrics is heavily open-source-driven and invests heavily in research across data engineering, decisionmodels, computer vision, and embedded technology.

ACROSS LOCATIONS & GROWING

ZiMetrics is headquartered in India, Pune. ZiMetrics has regional delivery centers in Goa, Delhi, NCR, and Bangalore (upcoming) in India. ZiMetrics has network delivery centers in Canada, British Columbia, and in the USA which helps in delivering a seamless cross-time zone network execution experience.

The fundamentals and aims of ZiMetrics are as follows:

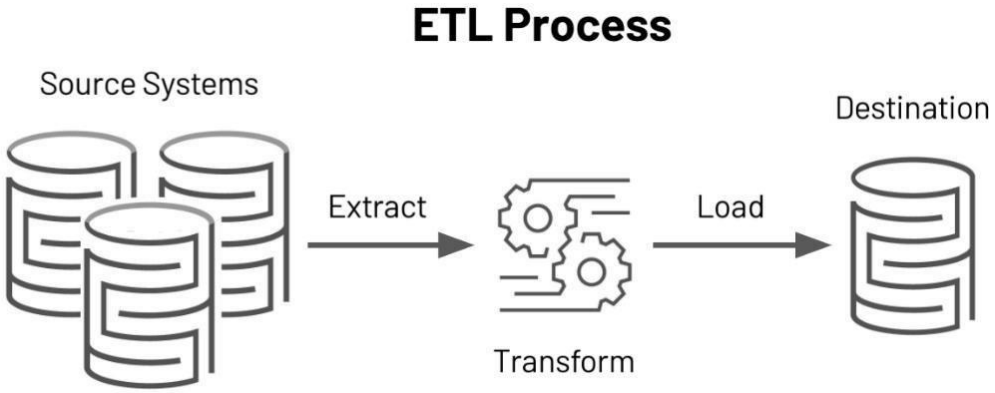
- One that combines data sciences, domain disciplines, and data technologies intocrafts that create transformational customer value.
- One that puts the customer's interests first.
- And one that would be recognized as one of the great places to work.

PROJECT – TELEMATICS DATA PROCESSING

PROBLEM STATEMENT

Given large amounts of data of EV vehicles being tracked continuously with a specific time gap (every 10 mins), the aim is to clean, transform, enrich and integrate all the data to analyze the performance of the vehicles in different areas such as elevations, mountain passor mountain range.

Also, to create visualizations from the data to gain better understanding of the vehicle performance.



OVERVIEW

The project aim is to perform ETL operations on the given data using Google Cloud Platform. The Google Data Fusion platform is a service that helps users efficiently build and manage ETL/ELT data pipelines. The data pipeline would be scheduled to run at a specific time. The data pipeline would integrate data stored in different files within a bucket on the Google Cloud Storage, cleanse, transform and enrich the data and finally store it in Google BigQuery. The Google Roads API and Google Elevation API is used to enrich the data.

Platforms, Tools and Technologies used

- Google Cloud Platform
    - Compute Engine
    - Google Cloud Storage
    - Google Data Fusion
        - Wrangler
        - Studio

    - BigQuery
- Maven
- Git
- Postman
- Google Data Studio

My Contribution

I was assigned to the Telematics Data Processing project wherein my responsibilities were:

- Exploring the Alooma Platform.
- Exploring Google Data Fusion, create and run data pipelines using Data Fusion on different data sources and different data destinations.
- Creating a VM instance on GCP.
- Testing APIs with query parameters in Post Man
- Understanding Google Roads API and Google Elevation API.
- Developing custom plugin to be used in Data Fusion.

More in Detail

Exploring Google Data Fusion:

- Creating and managing Data Fusion Instance.
- Connecting to different data sources such as REST API, files stored with GoogleCloud Storage and data from BigQuery.
- Integrating data from these different data sources, transforming and enriching the data using built-in plugins in Data Fusion.
- Developing custom plugin to be used inside Data fusion and storing the data toBigQuery.
- Triggering one data pipeline to run after another.

Working with Google APIs and creating Custom Plugin in GCP

My task was to call the snap to roads API and elevations API on the raw data (once the data got into source plugin) by passing latitude and longitude field values. These field values are sent by the sensors that are installed to the vehicle, and are stored in raw data file.

The reason we call the snap to roads API and elevations API is to get the accurate latitude, longitude values. Sometimes the latitude, longitude values inside raw data (data from Vehicle) are pointing to the side of the roads or away from the actual path, so to overcome this

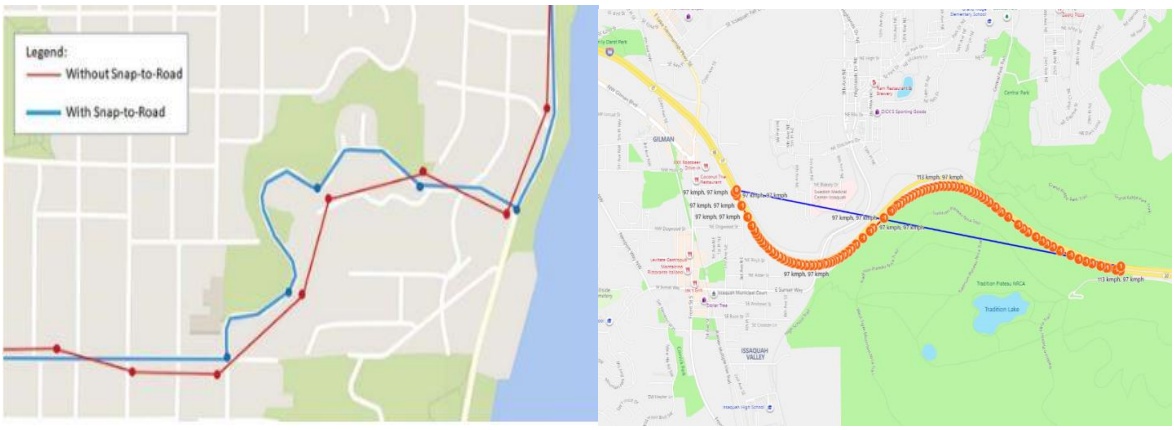
problem we pass these values to the Snap to Roads API and it returns us the accurate latitude, longitude values in the form of lat_path and lon_path.

As we need this data for calculations like getting number of trips the vehicle covered, finding shortest distance between two coordinates, calculating the elevation.

The challenge was to pass different latitude and longitude field values (for all rows) to API call. As this feature to call API with field values as parameters is not supported in google cloud platform, and there is no plugin to do such multi row transformation.

To do such transformation and to call API for each row of data, i created a custom plugin in google cloud platform. The custom plugins are build using Java programming language.

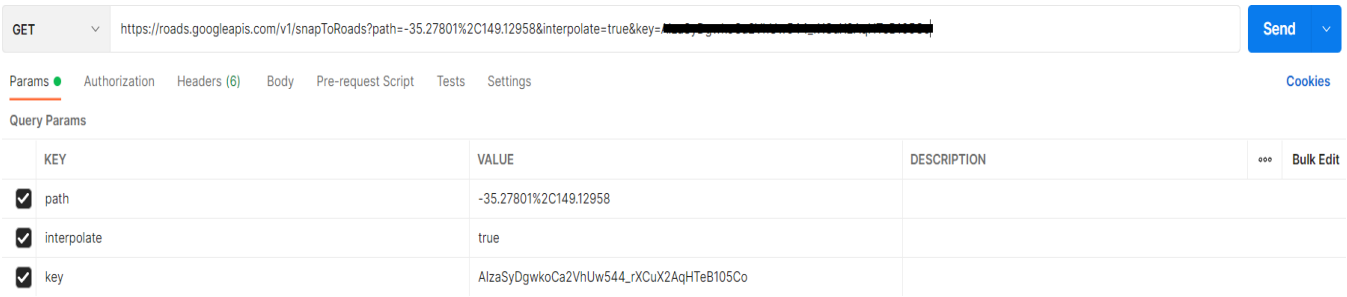
Google Cloud Platform (GCP) provides a set of plugins that are used for data transformation and to create pipelines in google cloud data fusion. Creating custom plugin not just helped us in calling APIs with field values for all rows of data, but we were also able to do batch transformation on the data (As per project requirement).



The Elevation API provides elevation data for all locations on the surface of the earth, including depth locations on the ocean floor (which return negative values).

Elevation Data helps in getting information of steep roads which are very useful for calculations and further data transformations.

Sending request to Snap to Roads API with field values



JSON response of Google Snap to Roads API and Elevation API:

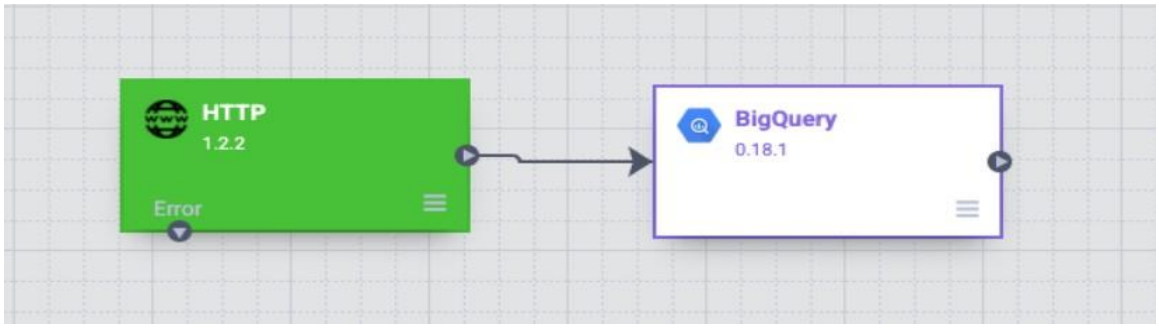

```
{
    "snappedPoints": [
        {
            "location": {
                "latitude": -35.278004899930188,
                "longitude": 149.129531998742
            },
            "originalIndex": 0,
            "placeId": "ChIJr_xl0GdNFmsRsUtUbW7qABM"
        }
    ]
}
```

```
{
   "results" : [
      {
         "elevation" : 567.687744140625,
         "location" : {
            "lat" : -35.27801,
            "lng" : 149.12958
         },
         "resolution" : 38.17580795288086
      }
   ],
   "status" : "OK"
}
```

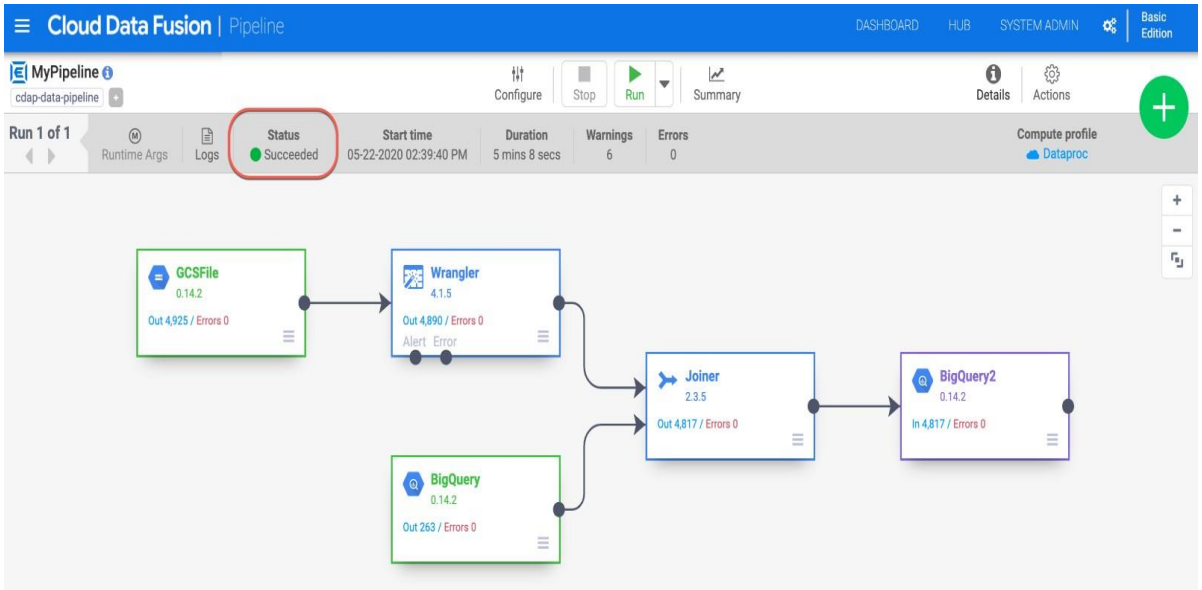

Once the API's are called on Raw data by Passing latitude and longitude, we get the API response in JSON format. The latitude, longitude values are fetched from this response and further stored in two new columns named lat_path and Lon_path in enriched file.

Screenshots

Static API Batch Pipeline:



Integrating multiple data sources and storing the data in BigQuery warehouse:



OTHER TASKS

CONSUMING REST API:

- Consuming REST API in XML format and to get specific data using xpaths.
- Learning flask framework (basic) to call REST API.

TOOLS AND TECHNOLOGIES USED:

- Postman
- Flask

CERTIFICATIONS AND TRAININGS COMPLETED UNDER INTERNSHIP AND SELF STUDY:

COURSERA COURSE:

- Exploring and preparing your data with BigQuery

OTHER SELF-STUDY:

- Learning python
- Learning MySQL
- Learning UNIX commands and shell scripting

TRAININGS COMPLETED UNDER INTERNSHIP:

- Java
- Git

PLATFORMS, TOOLS AND TECHNOLOGIES USED

CLOUD DATA FUSION:

- Cloud Data Fusion is a fully managed, cloud-native, enterprise data integration service for quickly building and managing data pipelines.
- The Cloud Data Fusion web UI allows you to build scalable data integration solutions to clean, prepare, blend, transfer, and transform data, without having to manage the infrastructure.
- Features:
    - **Google Cloud-native:** features like security, reliability, scalability provided by Google.
    - **Delivers hybrid infrastructure:** Cloud data fusion is built using an open-source project called CDAP which ensures data pipeline portability
    - **Multi-cloud integration**
    - **Code free environment**
    - **Seamless operations**: provided by Restful APIs, pipeline state base triggers,logs, metrics and various monitoring dashboards.

BIGQUERY:

- BigQuery is a fully-managed, serverless data warehouse that enables scalableanalysis over petabytes of data.
- BigQuery is a fully managed enterprise data warehouse that helps you manage and analyze your data with built-in features like machine learning, geospatial analysis, and business intelligence.

CLOUD STORAGE:

- Cloud Storage is a service for storing your objects(files) in Google Cloud. An object is an immutable piece of data consisting of a file of any format. You store objects in containers called buckets.

GOOGLE DATA STUDIO:

- Google Data Studio is an online tool for converting data into customizableinformative reports and dashboards.

GIT:

- Git is a distributed version-control system for tracking changes in source code during software development. It is designed for coordinating work among programmers, but it can be used to track changes in any set of files.

MAVEN:

- Maven is a build automation tool used primarily for Java projects. Maven can also beused to build and manage projects written in C#, Ruby, Scala, and other languages.

MYSQL:

- MySQL is an open-source relational database management system.

POSTMAN:

- Postman simplifies each step of building an API and streamline collaboration so you can create better APIs—faster.

PROJECT TIMELINE/PROJECT DIARY

JANUARY 2022:

- Week 1:

  Completed MySQL self-learning training.

- Week 2:

  Completed Python self-learning training.

- Week 3:

  Completed Linux self-learning training.

FEBRUARY 2022:

- Week 1:

  Learnt python flask basics to make REST API calls.

  Learnt to consume REST APIs using the Postman tool.

- Week 2:

  Revised Java Concepts.

  Prepared a short documentation for a medical project.

- Week 3:

  Research on an ETL platform named Alooma.

  Started with a course on coursera to explore and prepare data withBigQuery.

- Week 4:

  Started learning Google Cloud Platform.

MARCH 2022:

- Week 1:

  Understanding of different components of GCP and how data processing works

- Week 2:

  Building static pipelines in GCP and processing a sample CSV file

- Week 3:

  Building Dynamic pipelines in GCP and storing the output in CSV file in GCP bucket

- Week 4:

  Exploring how to store transformed data in pipeline to BigQuery

April 2022:

- Week 1:

  Got familiar with the project statement and what the project is all about.

  Got assigned to a task where I need to create a pipeline that will call APIs (roads API and Elevation API) on each row of the file.

- Week 2:

  Worked with HTTP plugin where we can call the APIs but with only static parameters in API URL.

- Week 3:

  Started working on Python Plugin that is used to write python Programs in GCP pipeline and this program gets executed during pipeline execution

- Week 4:

  Started working on Custom Plugin creation to perform the API call operation on the data.

  Started building custom plugin from scratch, for this we used Maven and the Custom Plugin was wrote in Java Programing language.

May 2022

- Week 1:

  First tried working with a sample data with custom plugin to check whether the custom plugin works

- Week 2:

  Started working on Java code to do API call on each row of data for this I used the java library OkHttp which helps in calling API.

- Week 3:

Worked with testing of the plugin to check whether the API call is done successfully, Here I handled NULL pointer exception and also rows that had invalid data.

- Week 4:

  The API call was successfully handled in custom plugin and then the API response that was in JSON format was returned back and was added as new fields (lat_path and lon_path) to the enriched file.

June 2022

- Week 1:

  Started working on Batch Processing of Data for API call, so that we can call the API only once for a batch of rows.

REFLECTIONS/ EXPERIENCES OF INTERNSHIP

My experience of working as an intern at ZiMetrics has been wonderful. I started working on Google Cloud Platform which was new to me. Got introduced to API concept and how APIs help in communication between applications. It was great learning new technologies. I found myself growing after joining the company.

The work environment at ZiMetrics is very amazing. There is freedom given to interact, or seek help from seniors. The efforts and dedication that every employee puts at ZiMetrics makes you motivated towards your work.

Apart from work, there are Recreational activities being conducted once every week where all employees come together and play some indoor games. Once a month, a meeting is conducted wherein all members talk about their work progress.

You get appreciated for the work you do successfully which motivates you to do your work and also get help whenever needed. Also, if you get stuck to any task, seniors always help you in finding the solution.

Overall, I am happy to be a part of ZiMetrics. My best wishes to everyone here.

REFERENCES

- https://www.javatpoint.com/mysql-tutorial
- https://www.javatpoint.com/linux-tutorial
- https://www.javatpoint.com/python-tutorial
- https://cloud.google.com/compute/docs/instances/creating-instance-with-custom-machine-type#pricing
- https://cloud.google.com/compute/docs/ip-addresses/reserve-static-external-ip-addr ess
- https://cloud.google.com/data-fusion/docs/concepts
- https://datastudio.google.com/overview
- https://www.youtube.com/watch?v=kehG0CJw2wo
- https://medium.com/cdapio/getting-started-with-cdap-plugin-development-bcd21cc7ae66
- https://cdap.atlassian.net/wiki/spaces/DOCS/pages/480313897/Developing+Plugins+Guide