ZIMETRICS TECHNOLOGIES PRIVATE LIMITED

PANJIM, GOA

INTERNSHIP REPORT

TANVI ULHAS DESAI

1906

INTERNSHIP REPORT

Submitted By:

TANVI ULHAS DESAI

1906

Under the Guidance of:

Omkar Prabhu
(Center Head, ZiMetrics)

Aditya Kunkolienkar
(Sr. System Analyst, ZiMetrics)

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 Miss. Tanvi Desai, a student of MCA Goa university roll no- 1906.is currently undergoing long internship program at Zimetrics Technologies Pvt Ltd.(10th Jan 2022- Till date)

During the period of her internship program, she is working on Developing ETL for Telematics solutions as an engineering intern.

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

As per the campus drive conducted on 8th November 2021, she 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 UNIVERISTY

GOA BUSINESS SCHOOL

CERTIFICATE OF EVALUATION

This is to certify that **Ms. Tanvi Ulhas Desai** is currently doing her internship at **ZiMetrics Technology Pvt. Ltd., Patto, Panjim, Goa,** in partial fulfillment for the award of the degree in Master of Computer Application.

_____ _____

Examiner 1 Examiner 2

Place: Goa University _____

Date: 25th June 2020 Dean, Goa Business School

Acknowledgement:

I am privileged to have done my internship in ZiMetrics Technologies Private Limited, Panjim, Goa. I got a great chance for 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 Kumar Verma (founder & President of Engineering at ZiMetrics Technologies) for allowing me to be an intern at ZiMetrics Technologies Private Limited.

I would like to thank Mr. Omkar Prabhu (Center Head – Goa RDC.) for giving me the opportunity to intern at ZiMetrics Technologies Private Limited and for all the guidance and support.

My sincerest gratitude to Mr. Aditya Kunkolienkar for being supportive and guiding me throughout the project.

I would like to extend my gratitude to Ms. Swati Patil (Head – HR ZiMetrics Technology Pvt Ltd.), Mr. Shantanu Waghmare (Associate Manager – HR at ZiMetrics ZiMetrics Technology Pvt Ltd.), Yashwanti Patil (HR Manager) and Maseera Shaikh (HR Generalist, ZiMetrics) for helping me out whenever needed.

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

Finally, I would also like to thank my family and friends who were always ready to help me and guide me in all aspects of life.

Table of Contents:

Introduction:

This report is a short description of my full-time on-site internship at Zimetrics, Goa.
I joined as an Intern at Zimetrics on 10th January 2022 and have been here since then. This report contains necessary information about the organization, the projects and mini projects I worked on and the other tasks and trainings I completed in this internship period.
In the chapters that will follow, I will talk about the company, the work here, the culture, etc. Then, I shall elaborate on the project I worked on, a brief information about the projects, the modules I built and the tasks I completed in those modules.
This report highlights my learning experience and my contributions 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, Big Data Analytics, Cloud Computing and Data Science. Its vision is to enable enterprises to harness the vast economic potential and transformational business value of machine data and connected devices. ZiMetrics do so by combining, deep industry expertise, pre-engineered solution frameworks and an extensive focus on transformational customer value creation.

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: IOT, Machine Data, Big Data & Data Science.

ZiMetrics is built on the founding principles of Integrity, Intellect & Ingenuity. These form the core principles around which Our value system, our engineering expertise, and our organizational behaviour are nurtured.

ZiMetrics is headquartered in India, Pune. We have regional delivery centres in Goa, Delhi, NCR, and Bangalore (upcoming) in India. Our network delivery centres in Canada, British Columbia, and in the USA, Texas.

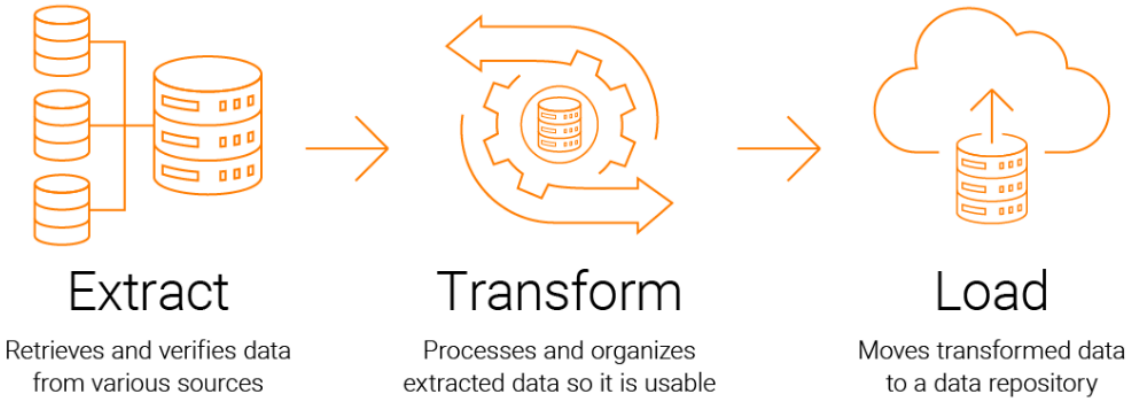
ZiMetrics is a Confluent & HashiCorp Partner and recognized as the "Top 20 IoT Solutions Providers" by CIO Outlook APAC.

PROJECT – TELEMATICS DATA PROCESSING (POC)

PROBLEM STATEMENT:

Given in large number of Electronic Vehicles being tracked continuously in a specific time gap (after every 10 min), the goal is to clean, transform, enrich, integrate and visualize to see the performance of the vehicles in different regions such as in plains, mountain pass or in mountain range.

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



OVERVIEW:

This project aims to perform ETL operations on the Electronic Vehicles (EV) data using Google Cloud Platform. The Cloud Data Fusion by Google Cloud is a fully managed, code-free data integration service that helps users efficiently build and manage ETL/ELT data pipelines. The data pipelines would be scheduled to run at specific time. The pipeline would integrate the data stored in different files inside Google Cloud Storage bucket, cleanse, transform, and enrich the data using various available plugins in Cloud Data Fusion and finally store the processed data in Google BigQuery.

To check the performance of vehicles in different regions visualization is performed using Google Data Studio. The Google Roads API and Google Elevation API are used to enrich the data.

PLATFORM, TOOLS & TECHNOLOGIES USED:

- Google Cloud Platform
  - Cloud Data Fusion
  - Google Cloud Storage
  - Big Query
  - Compute Engine

- Google Data Studio
- Postman
- Git
- Maven

MY CONTRIBUTIONS:

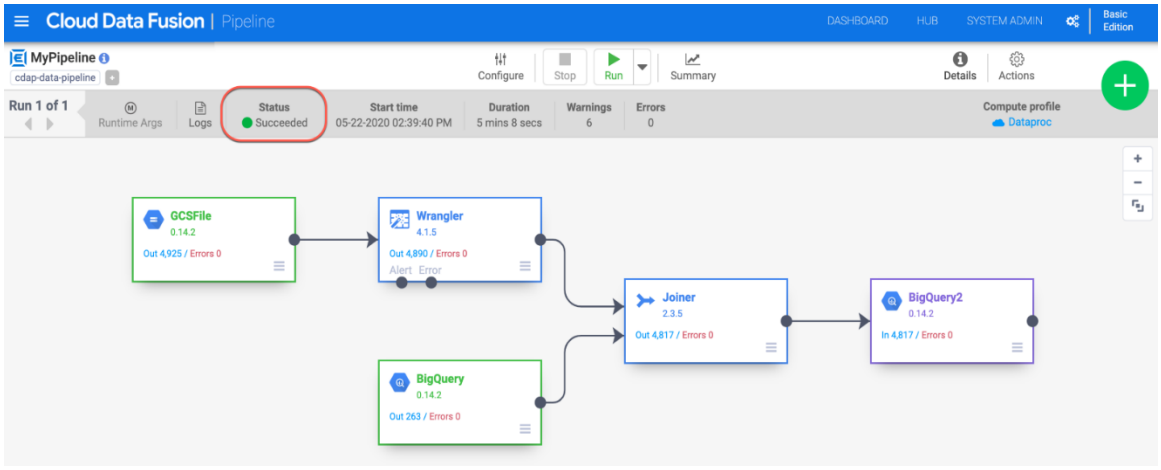
I was assigned the following tasks for building the end-to-end dynamic pipeline in Cloud Data Fusion:

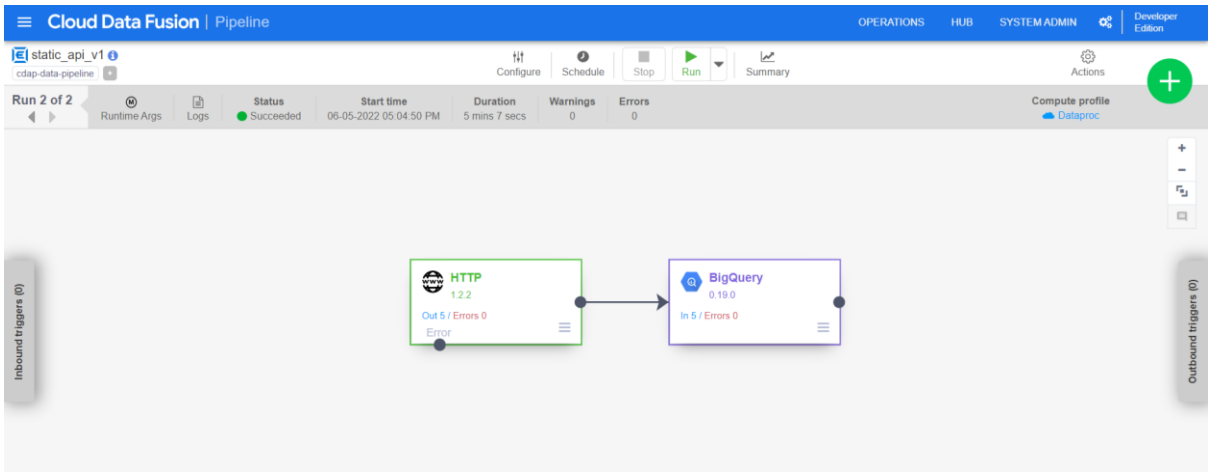
- Push the data to Google Cloud Storage.
- Loading multiple files in pipeline and converting its data format.
- Custom plugin creation to enrich data.

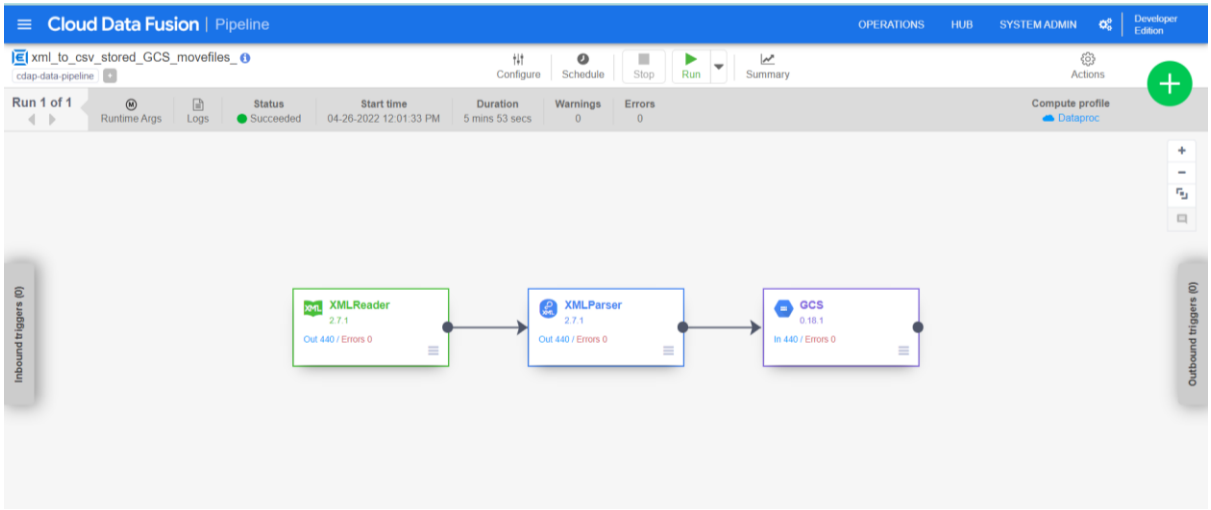
My detailed contributions are as follows:

- Understanding various services of Google Cloud Platform.
- I had to push the received data to Google Cloud Storage from my local machine by writing python script. It required the Service account key generation. The data received had to be converted from excel to csv and then push to the Google Cloud Storage and then archive the pushed files.
- The data received by calling one Sample API In the VM instance (created inside Compute Engine) had to be pushed to Cloud Storage in python. This required creation of Service account and giving proper roles to that account. After pushing the files, they were moved to different folder.
- Uploading multiple XML files all together in pipeline and converting all the files to CSV using the appropriate plugins available in Cloud Data Fusion.
- Understanding the client data thoroughly.
- Understanding the usage of python plugin that is used to enrich the data. Setting up the native mode and path to libraries.
- Creating custom plugins in java that were used to enrich the fields.

SCREENSHOTS:







Other tasks related to project:

- Integrating multiple data sources and storing the data in BigQuery warehouse.
- Calling static / dynamic API using Http plugin and store the data in BigQuery.
- Transformation applied in Wrangler plugin on a sample dataset and stored in BigQuery.
- BiqQuery and BigTable comparison and pricing.
- Understanding the Google Roads API and Google Elevation API.
- Understanding of IAM Roles and Service account.
- Creating custom plugin in Java using CDAP plugin development.

Courses, Training & Self learning:

- MySQL
- Python
- Linux
- Java
- Git
- BigQuery
- Cloud Data Fusion
- Google Cloud Platform

Platforms, Tools & Technologies Used:

Google Cloud Platform:

Google Cloud is a suite of public cloud computing services offered by Google. The platform includes a range of hosted services for compute, storage and application development that run on Google hardware. Google Cloud services can be accessed by software developers, cloud administrators and other enterprise IT professionals over the public internet or through a dedicated network connection.

Google Cloud offers services for compute, storage, networking, big data, machine learning and IoT, as well as cloud management, security and developer tools.

Cloud Data Fusion:

Cloud Data Fusion is a fully managed, cloud-native, enterprise data integration service for quickly building and managing data pipelines.

Cloud Data Fusion provides a graphical interface to help increase time efficiency and reduce complexity and allows business users, developers, and data scientists to easily and reliably build scalable data integration solutions to cleanse, prepare, blend, transfer, and transform data without having to wrestle with infrastructure.

Google Cloud Storage:

Cloud Storage is a RESTful service for storing and accessing your data on Google's infrastructure. The service combines the performance and scalability of Google's cloud with advanced security and sharing capabilities.

BigQuery:

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. BigQuery's serverless architecture lets you use SQL queries to answer your organization's biggest questions with zero infrastructure management. BigQuery's scalable, distributed analysis engine lets you query terabytes in seconds and petabytes in minutes.

Compute Engine:

Compute Engine is a computing and hosting service that lets you create and run virtual machines on Google infrastructure. Compute Engine offers scale, performance, and value that lets you easily launch large compute clusters on Google's infrastructure. There are no upfront investments, and you can run thousands of virtual CPUs on a system that offers quick, consistent performance.

Google Data Studio:

Data Studio is Google's reporting solution for power users who want to go beyond the data and dashboards of Google Analytics. The data widgets in Data Studio are notable for their variety, customization options, live data and interactive controls (such as column sorting and table pagination). Data sources include Google products (Analytics, AdWords, Search Console, Sheets, YouTube, etc.), database connectors, file upload, and "community" connectors to popular marketing services.

PyCharm:

PyCharm is a dedicated Python Integrated Development Environment (IDE) providing a wide range of essential tools for Python developers, tightly integrated to create a convenient environment for productive Python, web, and data science development.

PyCharm provides smart code completion, code inspections, on-the-fly error highlighting and quick-fixes, along with automated code refactoring's and rich navigation capabilities.

Visual Studio Code:

Visual Studio Code is a lightweight but powerful source code editor which is available for Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other languages (such as C++, C#, Java, Python, PHP, Go) and runtimes (such as .NET and Unity).

Visual Studio Code combines the simplicity of a source code editor with powerful developer tooling, like IntelliSense code completion and debugging.

IntelliJ IDEA:

IntelliJ IDEA is an intelligent, context-aware IDE for working with Java and other JVM languages like Kotlin, Scala, and Groovy on all sorts of applications. Additionally, IntelliJ IDEA Ultimate can help you develop full-stack web applications, support for JavaScript and related technologies, and advanced support for popular frameworks like Spring, Spring Boot, Jakarta EE, Micronaut, Quarkus, Helidon.

Moreover, you can extend IntelliJ IDEA with free plugins developed by JetBrains, allowing you to work with other programming languages, including Go, Python, SQL, Ruby, and PHP.

Postman:

Postman is an API platform for building and using APIs. Postman simplifies each step of the API lifecycle and streamlines collaboration so you can create better APIs—faster.

Postman is an application used for API testing. It is an HTTP client that tests HTTP requests, utilizing a graphical user interface, through which we obtain different types of responses that need to be subsequently validated.

Maven:

Maven is a popular open-source build tool developed by the Apache Group to build, publish, and deploy several projects at once for better project management. The tool provides allows developers to build and document the lifecycle framework.

Maven is written in Java and is used to build projects written in C#, Scala, Ruby, etc. Based on the Project Object Model (POM), this tool has made the lives of Java developers easier while developing reports, checks build and testing automation setups.

Git:

Git is a free and open-source distributed version control system designed to handle everything from small to very large projects with speed and efficiency.

Internship Timeline: (10th Jan – till date)

January 2022:

- Week 1:
  - Onboarding and installation of software's.
  - Started with MySQL training

- Week 2:
  - Linux training

- Week 3:
  - Python training

February 2022:

- Week 4:
  - Health Care project understanding
  - Understanding the calculations
  - Writing pseudocode for calculations

- Week 5:
  - Continued with python training
  - Java Self learning.
  - Understanding Flask framework

- Week 6:
  - Working with Postman tool and consuming API
  - Research on Alooma Tool

- Week 7:
  - Continued with Alooma Tool
  - Introduced to Google Cloud Platform (GCP)

March 2022:

- Week 8:
  - Understanding the GCP
  - Account setup on GCP
  - Watched Coursera videos on GCP, CDF.

- Week 9:
  - Continued with GCP and CDF basics.

    - Understanding basic features in GCP like creating projects, adding IAM roles, service accounts creation, clusters, Cloud shell.
    - Understanding the billing structure in GCP.
    - Running basic pipelines

- Week 10:
    - Went through Cloud Data Fusion documentation and videos.
    - CDF instance creation, used studio to create data pipelines.
    - Usage of different plugins like Google Cloud Storage, Wrangler, BigQuery.
    - Started learning BigQuery

- Week 11:
    - Integrated with 2 datasets and ran pipelines.
    - Applied custom transformations in wrangler and wrote queries in BigQuery.

- Week 12:
    - Searched for datasets, understood data, applied transformations in wrangler and built a data pipeline and saved data to BigQuery.
    - Worked with calling static API in data pipeline.

April 2022:

- Week 13:
    - Learnt Google Roads API and Elevation API usage.
    - Started with task of pushing data to Cloud Storage.
    - Learnt cloud storage and its default permissions

- Week 14:
    - Learnt Service account and key generation.
    - Completed the task of pushing data to cloud storage.
    - Compared BigQuery and Bigtable and its pricing.

- Week 15:
    - Started with task of pushing data to GCS bucket from VM instance.
    - Learnt service account and giving specific roles and permissions.

- Week 16:
    - Completed the task of data pushing from VM instance.
    - Started with conversion of Xml files to CSV in the data pipeline.
    - Worked with plugins to convert xml to CSV.

May 2022:

- Week 17:
    - Completed the task of converting files to CSV in data pipelines.
    - Started in brief with Custom plugin creation, working in CDAP development.

- Week 18:
    - Started working with python plugin updating.
    - Learnt python plugin code written in java.

- Week 19:
    - Tried setting up native mode, setting path to python libraries in python plugin in order to use that plugin and run pipelines
    - Started with creating custom plugins in Java.

- Week 20:
    - Attended Java, Git, Maven training
    - Continued with custom plugin.

June 2022:

- Week 21:
    - Continued with custom plugin, enriching fields.

My reflections/experiences of internship:

My internship at ZiMetrics has been an amazing and growing experience. It helped us in exploring our potentials, and focus on our growth.

I was fortunate to receive training as part of my internship. As I reflect on all we have learned, I realize what this has been superb experience. This hands-on approach was invaluable, it has served as a beneficial ending to my formal education. Skills such as multitasking, communicating, learning to deal with diversity, and dealing with deadlines are different when you are working for someone else.

It is amazing to see how people from different regions stay as one family and work together. To help us understand the company more, and give an idea of what each team in the company does, we had meetings on every Friday of each month. This was often followed by recreational activities sessions, which kept me engaged and made it more interesting.

Overall, I'm very happy to be a part of ZiMetrics and looking forward to facing new challenges with the knowledge that I have acquired here.

References:

- https://www.zimetrics.com/
- https://cloud.google.com/
- https://cloud.google.com/data-fusion
- https://cloud.google.com/bigquery
- https://datastudio.google.com/overview