ZIMETRICS TECHNOLOGIES PRIVATE LIMITED.

PATTO – PANJIM, GOA

INTERNSHIP REPORT

DOISY DIAS

1909

INTERNSHIP REPORT

Completed by

Doisy Dias | 1909

For the course

MCA Semester VI

At

ZiMetrics Technologies private Limited

Patto – Panjim, Goa

zimetrics

ZIMETRICS TECHNOLOGIES PRIVATE LIMITED

Date: 31/05/2022

<u>TO WHOM IT MAY CONCERN</u>

This is to certify that Miss. Doisy Dias, a student of MCA Goa university roll no- 1909 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 15:45 GMT+5.5)

Sincerely,
For Zimetrics Technologies Pvt Ltd
Ashwini Barve
GM Operations

GOA UNIVERSITY

GOA BUSINESS SCHOOL

CERTIFICATE OF EVALUATION

This is to certify that Ms. Doisy Dias is currently doing her internship at ZiMetrics technologies private limited, Patto – Panjim , Goa in partial fulfilment for the award of the degree in Master of Computer Application.

_____   _____

Examiner 1   Examiner 2

Place: Goa University
Date: 23/06/2021   Dean, Goa Business School

Acknowledgment

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 our internship in ZiMetrics Technologies private limited. 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 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 necessary guidance and support.

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

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 Technology Pvt Ltd.), Yashwanti Patil (HR Manager) and Maseera Shaikh (HR Generalist, ZiMetrics) for helping me out whenever needed.

I would also like to thank Aamod Talauliker, and the whole ZiMetrics family for providing us the resources, guidance, knowledge, and advice on how to go about the issues and work assigned to me. With their support, I got a significant amount of exposure and experience.

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.

Contents

Introduction

This report is a short description of my ongoing internship at ZiMetrics technologies private limited, Patto – Panjim 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 assignments and tasks I worked on, and the other training 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 team I was assigned to, a piece of brief information about the assignments, the modules I built, and the tasks I completed. This report highlights mylearning experience and my contributions to the organization as an intern. However, considering company data security policies, I won't be able to give details of the work I've contributed to. This will describe the knowledge that I gained by 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 hashelped 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, and Data Science. It enables enterprises to harness the vast economic potential and transformational business value of data from systems, machines, connected devices, and human-generated content. ZiMetrics do so by combining, deep industry expertise, pre-engineered solution frameworks and an extensive focus on transformational customer value creation. It is a Confluent & HashiCorp Partner and recognized as the "Top 20 IoT Solutions Providers" by CIO Outlook APAC.

Founded in 2015, ZiMetrics today serves leading global enterprises across Industrial, Oil & Gas, FMCG, MedTech, Internet Advertising, and Retail.

- Research and Innovation at Core
    - Innovation and research are at the center stage of what ZiMetrics do. ZiMetrics is heavily open-source-driven, invest heavily in research across data engineering, decision models, computer vision, and embedded technology. The research initiatives span across both in-house and academia-sponsored research.

- Across Locations & Growing
    - ZiMetrics is headquartered in India, Pune and have regional delivery centres in Goa, Delhi, NCR, and Bangalore (upcoming) in India. The network delivery centres in Canada, British Columbia, and in the USA, Texas.

PROJECT – TELEMATICS DATA PROCESSING(POC)

PROBLEM STATEMENT:

Given Large amount of data 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 help visualize the performance of the vehicles in different areas such as elevation, mountain pass or mountain range.

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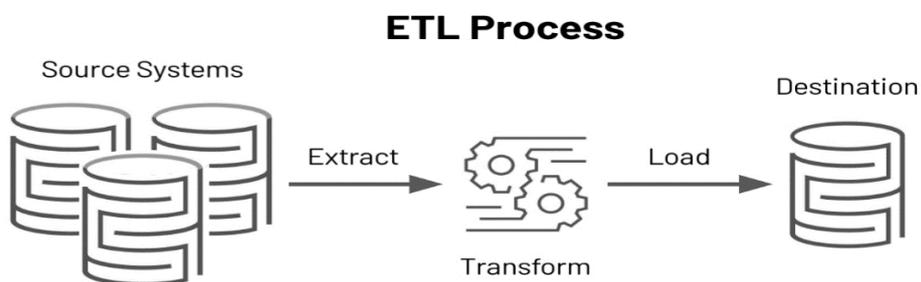
OVERVIEW

The aim of the project 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 pipelines would be scheduled to run at a specific time. The data pipeline integrates 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.

TOOLS AND TECHNOLOGIES USED

- Google Cloud Platform
  - Google Cloud Storage
  - Google Data Fusion
    - Wrangler
    - Studio
  - BigQuery
- Maven
- Git
- Postman
- Google Data Studio

My Contributions

I was assigned to do an end-to-end Dynamic pipeline where in

- Connect to Cloud Storage Bucket
- Load the data in wrangler
- Cleanse the data and transform
- Store it in BigQuery

My detailed contribution is as follows:

- Understand the data.
- In Data Fusion using the GCS plugin had to connect to Google Cloud Storage and get all the files from there and pass to wrangler plugin for enrichment.
- In wrangler plugin parse the files as csv/excel, cleanse the data, transform the data i.e., enrich the data
- And then using the BigQuery plugin store the data in BigQuery data Warehouse.
- Monitor the pipeline for errors, and pipeline failure
- Custom plugin where in adding of csv/excel columns for enriching the data.
- Write queries.

Screenshots
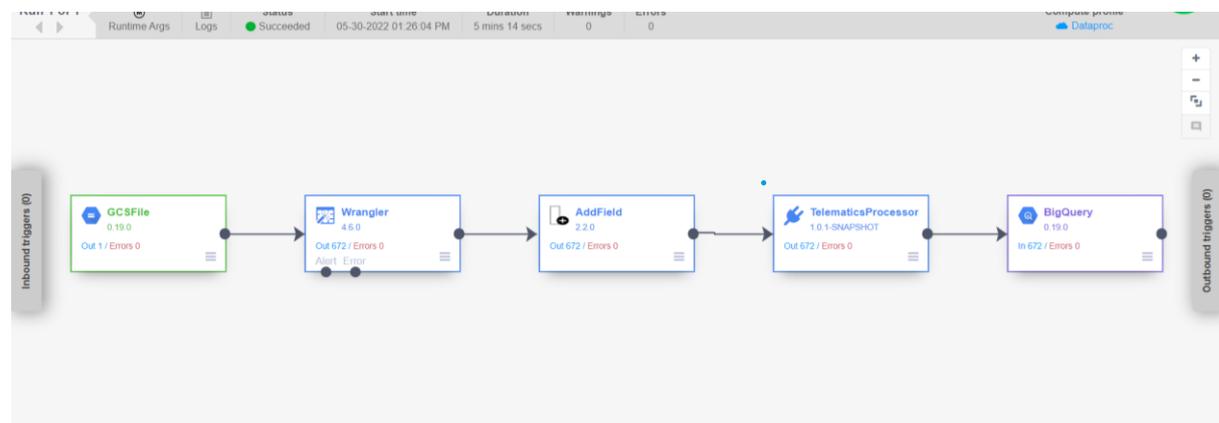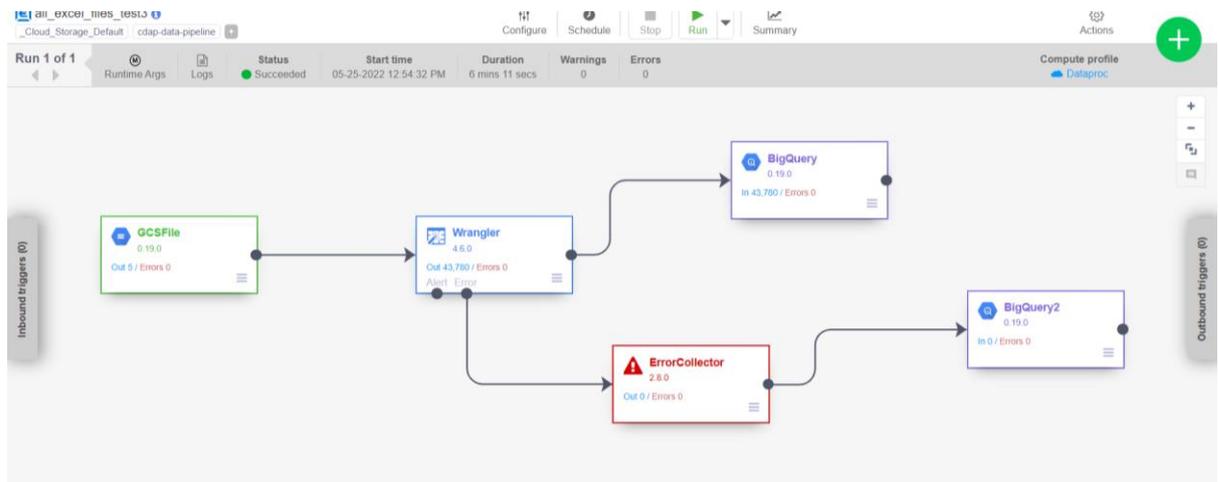
Other Tasks

- Plugin Research

  checking out different plugin and trying them out, which would be used in the pipeline for transforming the data, where in to handle duplicate data, parse batch data i.e., parsing multiple files at runtime, error handling, python plugin to write code to transform data as per our needs.

- Visualizations tools

  Researching on different visualization tools that connects to BigQuery and documenting it.

Course, training and Self-Study

- BigQuery Coursera Course

- Google Cloud Platform Coursera Course

- Data Fusion Coursera Course

- Python

- MySql

- Linux

- Git

- java

Tools and Technologies Used

MySql Shell

MySQL Shell is an advanced client and code editor for MySQL. This document describes the core features of MySQL Shell. In addition to the provided SQL functionality, similar to **mysql**, MySQL Shell provides scripting capabilities for JavaScript and Python and includes APIs for working with MySQL. X DevAPI enables you to work with both relational and document data, see Using MySQL as a Document Store. AdminAPI enables you to work with InnoDB Cluster, InnoDB ClusterSet, and InnoDB ReplicaSet.
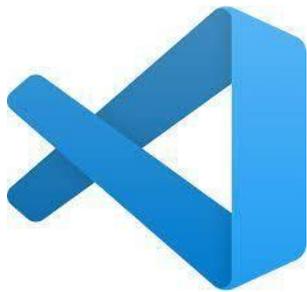
Visual Studio Code

Visual studio Code is a source code editor developed by Microsoft for Windows, Linux and macOS. It includes support for debugging, embedded Git control and GitHub, syntax highlighting code completion, snippets and code refactoring. It is highly customizable, allowing users to change the theme, keyboard shortcuts, prefernces, and install extensions that add additional functionally. The source code is free and released under the permissive MIT License. The compiled binaries are freeware and free for private or commercial use.

Postman

Simplify each step of building an API and streamline collaboration so you

can create better APIs—faster.

PyCharm

PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python programming language. It is developed by the Czech company JetBrains (formerly known as IntelliJ). It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django as well as data science with Anaconda. PyCharm is cross-platform, with Windows, macOS and Linux versions.

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.

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.
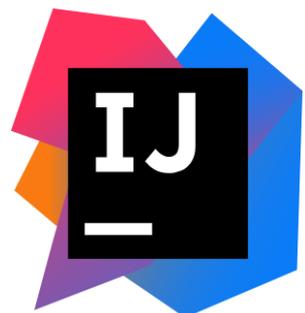
IntelliJ IDEA

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.

Internship Timeline

January 2022

- Week 1

  - Onboarding, installation of different software

  - MySql training

- Week 2

  - Linux training

- Week 3

  - Python training

February 2022

- Week 4

  - Worked on health Care Project

    - Understanding the calcutaions

    - Writing psedocode for calculation

- Week 5

  - Continuing with Python training

  - 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

- o Understanding the GCP

- o Account setup on GCP

- o Watched Coursera videos on GCP, CDF.

- Week 9

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

- Week 10

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

- Week 11

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

- Week 12

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

April 2022

- Week 13

  - o Learnt Google Roads API and Elevation API usage.
  - o Started with task of connecting to google cloud storage and passing data to wrangler.
  - o Learnt cloud storage and its default permissions.
  - o Started with dynamic pipelines/ passing of multiple files at the same time.

- Week 14

- Learnt Service account and key generation.

- Compared BigQuery and Bigtable and its pricing.

- Compared different visualizations tools.

- Week 15
  - Understanding of code
  - Learnt BigQuery and its default permissions.
  - Developed cloud Functions
  - Learnt service account and giving specific roles and permissions.
- Week 16
  - Completed dynamic pipeline
  - Handling errors
  - Handling duplicate data

May 2022

- Week 17
  - Researched on different plugins for error handling, duplicity, transformations
  - Compared plugins to write python code.
- Week 18
  - Started working with python plugin updating.
  - Learnt python plugin code written in java.

- Week 19
  - Researched on excel plugin
  - Started working on excel plugin
  - Parsed files in excel format

- Week 20
  - Deployed pipelines which were in Excel format
  - Started working on custom plugin
  - Attended Java, Git, Maven training

- Week 21

- o   Continued with custom plugin.

My reflections/experiences of internship

My internship at ZiMetrics has been a wonderful and growing experience. Starting from the interviews to date, it helped us in exploring our potentials, and focus on our growth while not letting die the hunger to learn more.

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 an excellent experience. We gathered much knowledge in the classroom, but a hands-on approach has been 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.

We had so much to learn about the company itself, as to what kind of projects it deals with, who are the clients, what are the company traditions, etc. To help us understand the company more, and give an idea of what each team in the company does, every Tuesday we have knowledge sharing where in each team explains their work using presentations.

Overall, I'm glad to be a part of ZiMetrics and looking forward to facing new challenges with the knowledge that I am acquiring 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