**Dr B R Ambedkar National Institute of Technology, Jalandhar**



**INDUSTRIAL WEEKLY DIARY**

**Data Analyst Internship**

**at**

****

**For the award of the**

**Degree of Bachelor in Technology**

**in**

**Department of Computer Science & Engineering**

(Session: 2018-2022)

***Topic : Azure Data Platform for COVID Data Analysis for Health Provider Industry***

**Institute Supervisor: Submitted by:**

Dr. Urvashi Sanyamdeep Singh

(Assistant Professor) 18103083

Dept. of CSE B.Tech. CSE (Final year)

**WEEKLY REPORT (13th January 2022 – 10th July 2022)**

*Worked on Azure cloud services*

## Week 1(13 January 2022 - 15 January 2022):

* It was the introduction week where different sessions and orientations were conducted with different levels of management in the company
* Gave us information about the different departments, work culture, the company’s core values etc., and told us whom to contact in case of any help required.
* Had a formal discussion about scope of the team and different verticals involved in it.
* Session on various practices and policies that need to be followed while working with company.

## Week 2 (17 January 2022 - 21 January 2022):

* Laptop configurations required for the company work were done during this week.
* Setup of laptop and installation of required softwares was done.
* We were given company specific training that is necessary for learning the company ethics, values, and their work culture like the Linkedin general courses.

## Week 3 and Week 4 (24 January 2022 – 4 February 2022):

* Basic Linkedin learning courses that were given need to be completed.
* Continue with Linkedin learning and consultation courses.

## Week 5 (7 February 2022 - 11 February 2022):

* Studied about azure cloud services provided and how it is required to cope with coming technologies.
* We have also learned the major topics of cloud computing which includes:

• Cloud Models - Public Cloud, Private Cloud and Hybrid Cloud

• Consumption Based Model

• Cloud Services – IAAS, PAAS, SAAS

* learned basic of Azure portal and different services offered by Azure:

• Azure compute and networking services

• Azure storage service

• Azure identity and security services

• Azure Management and Governance

* Assessment

## Week 6 (14 February 2022 – 18 February 2022):

* Basic teachings of DBMS.
* Studied about how data is collected and which azure services can be used for data analysis
* Case studies were analysed to choose service can be used under which condition.
* Fundamentals of database concept in a cloud environment. In this we have learned topics like:

• OLTP and OLAP

• Batch and Streaming data

• Azure services for relational data

• Azure services for non-relational data.

* Assessment

## Week 7 (21 February 2022 – 25 February 2022):

* We learnt about the basis of python programming which includes

• Python – Overview

• Basic Syntax and Variable Types

• Loops in Python

• Lists, Tuples and Dictionary

• Functions in Python

• Python File Handling.

* Assessment

## Week 8 (28 February 2022 – 4 March 2022):

* We learnt about the Azure Data Factory which is a a fully managed, serverless hybrid data integration service that provides a data integration and transformation layer that works across your digital transformation initiatives.
* Enables integrators and data engineers to drive business and IT led Analytics/BI.
* Studied how to build pipelines, set triggers, how to copy data automatically online to azure data storage, cleaning data, conversion of data formats like parquet, json, csv.
* ETL and ELT processing.
* Hands on were performed on racing data set.
* Assessment

## Week 9 (7 March 2022 - 11 March 2022):

* Learnt SQL Programming.
* Studied about queries required to be performed for anlaysis of data as a data analyst.
* Introduction to T-SQL.
* Assessment

## Week 10 (14 March 2022 – 18 March 2022):

* We were introduced to Apache Spark
* Apache Spark is a parallel processing framework that supports in-memory processing to boost the performance of big-data analytic applications.
* How to handle big data, how clusters can be used for processing of data.
* How to perform analysis using pyspark.
* Assessment

## Week 11 and Week 12 (21 March 2022 – 1 April 2022):

* We were introduced to Azure Synapse Analytics

## Azure Synapse Analytics is a limitless analytics service that brings together data integration, enterprise data warehousing and big data analytics.

## Studied about building pipelines, using sql queries to query data

## Building visualizations.

## Performed hands on

## Assessment

## Week 13 (4 April 2022 – 8 April 2022):

* We were introduced to Azure Databricks
* Azure Databricks is optimized for Azure and tightly integrated with Azure Data Lake Storage, Azure Data Factory, Azure Machine Learning, Azure Synapse Analytics, Power BI and other Azure services to store all your data on a simple, open lakehouse and unify all your analytics and AI workloads.
* Studied about storage mounting with databricks.
* Flexibility of using sql, python, scala, R with databricks.
* Creating visualizations, sharing dashboards.
* Connecting power BI

## Week 14 (11 April 2022 – 15 April 2022):

* Azure log analytics
* Learnt how to query data collected from monitor insights from on premises data and from azure services used.
* Performed hands on by implementing a pipeline and performing queries also by connecting with databricks
* Assessment

## Week 15 (18 April 2022 – 22April 2022):

* Azure Devops
* Studied about CI/CD (continuous integration and continuous deployment).
* Performed hand on to implement continuous tracking, integration and delivery
* Introduce to Forage platform
* Consulting 101 Program

## Week 16 (25 April 2022 – 5 May 2022):

* Break and time given for completion of certifications given by the company.
* **AZ-900(AZURE DATA FUNDAMENTALS CERTIFICATE)**
* Successfully passed az-900 online proctored exam by Pearson VUE securing more than 70% marks

## Week 17 and Week 18 (9 May 2022 – 20 May 2022):

* 2-weeks break for college end semester exam and major project evaluation.

## Week 19 and Week 20 (23 May 2022 – 31 May 2022):

* Two more certifications completed
* **DP-900 (DATA FUNDAMENTALS CERTIFICATE)**
* Successfully cleared dp-900 online proctored exam conducted by Pearson VUE by securing more than 70% marks.
* **DP-203 (DATA ENGINEER ASSOCIATE)**
* Successfully cleared dp-203 online proctored exam conducted by Pearson VUE by securing more than 70% marks.

## Week 21,22,23 (1 June 2022 – 24 June 2022):

* Capstone Project where we need to apply all the learnings learned from training and certifications
* Building azure data platform for COVID data analysis for health provider industry.
* The company want to have a dedicated team which can analyze data collected by ECDC Epidemic Intelligence to keep an eye on worldwide COVID trends such as reported cases, mortality rate, hospitalization & ICU admission rates & current occupancy, testing volume and country response which can act as feedback to improve their existing healthcare processes and patient monitoring system to respond optimally towards COVID.
* We had to come up with an automated solution rather than manual updation using the technologies we had learnt.

## Week 24 and 25 (27June 2022 – 8 July 2022):

## Forage platform tasks and 101 consultation program.

* **Gathering insights through research and client interviews**(scenario given ---Zomato sales needs to be increased).
* **Creation of an execution plan**
* Come up with solutions to the identified problem
* **Becoming the trusted advisor through client engagement**
* Setting up a presentation with ceo and team members of company to propose the solution.
* **Build your personal brand and networking**
* **Issue Prioritization**
* **Presenting with impact to the client**
* Giving the updates about progress of project and which new features can be added to attract customer attention.