

+91-70201 61630
Pune, Maharashtra
akashpahilwan53@gmail.com

Akash Pahilwan

Data Engineer

GitHub: akashpahilwan
LinkedIn: akashpahilwan53

EDUCATION

B.E. in Mechanical Engineering, *Sinhgad College of Engineering, Pune.* MAY 2019
Higher Secondary Education, *PVG's Mukhtangan English School and Jr. College.* JUN 2015
Secondary Education, *Vidya Niketan English Medium School.* JUN 2013

SKILLS

Technical Skills and Tools Azure Databricks, Azure data factory, Airflow, DBT, Python, Pyspark (Dataframes and SQL), Excel, Power BI, Alation, basics of Soda(DQ framework)

Communication English(fluent speaker), Marathi (fluent speaker), Hindi (reading and writing)

TECHNICAL EXPERIENCE

Consultant: Data Engineer SEPT 2022 — PRESENT
Cervello Bangalore, India

- Currently working on creating and maintaining ETL pipeline for data transfer from traditional datamarts(MS SQL) to Salesforce using Azure Data Factory maintaining the parent-child relationship
- Requirement gathering for any new object/field and collaborating with client technical and Salesforce team.
- Create de-normalized tables and implement SCD Type1(For capturing any change in data) and CDC(for delta loads) as per the requirement from the Client.
- Optimize existing ETL pipelines for effective utilization of resources. Optimized the run-time of the few existing views by almost 50%. Debugging and troubleshooting for any failures in ADF pipelines.
- Participated in multiple internal initiatives such as Aiflow and DBT POC, Alation initial implementation and gave presentations on these on COE calls for personal and professional growth.

Senior Systems Engineer / Daimler Project FEB 2021 — SEPT 2022
Infosys Pune, Maharashtra

- Working primarily on Checkpoint Firewall R80.30, VmWare NSX-T Distributed Firewall, F5 Loadbalancer and DNS System
- Creation/Modification/Deletion of Firewall and NAT Rules
- Managing and troubleshooting Incidents related to Firewall Clearances
- Creation/Modification/Deletion of LB VIP IPs and certificates on LB VIPs(F5 Load-balancer)

Accelerate Candidate / Accelerate Tasks 1: Data Engineer APR 2021 — JUN 2022
Infosys Pune, Maharashtra

- Creating scripts using **SQOOP** utility to ingest data into **HDFS** or directly into **tables** as per the requirement(On Premise)
- Creating/modifying **hive** scripts for data transformation as per the requirement(On Premise)
- Experience in working with different file formats such as **CSV, JSON, Parquet** using **PySpark**
- Creating **ETL** scripts using **pyspark** in Azure data bricks to load data from raw folder in **Azure data lake gen2** and store data frame as **parquet file/delta table** as per requirement into processed folder
- Creating **external** tables and **temp** views for **data analysts** to access the data.
- Creating **pipelines** in **Azure data factory** for data ingestion using metadata entries and creating linked Services to schedule run of **pyspark notebooks** on **daily** or **weekly** basis as required.

PROJECTS

DataFlowX: Automated ETL Solution with Azure, Snowflake, and DBT.

Personal project Pune, Maharashtra

- Designed an end-to-end **ETL solution** using **Azure Data Factory** for data extraction and ingestion, **Snowflake** for Data Warehousing, **DBT** (Data Build Tool) for data transformation, and **Airflow** for orchestrating and scheduling services.
- Implemented an **Azure Data Factory pipeline** to automate daily data transfers from source ADLS (Azure Data Lake Storage) to target ADLS, **dynamically** passing source data object names and dates.
- Established **external** tables in Snowflake during the initial setup, leveraging DBT model **dependencies** to facilitate data transformations efficiently.
- Developed comprehensive and reusable test cases for DBT models, ensuring **data quality** and integrity, and utilized DBT's **incremental** model for efficient data loading into the target tables.
- Implemented a versatile **logging mechanism** to handle different types of errors, including **data-related** issues triggering **email** notifications and **functional** programming errors in Airflow, which notify through both **Slack** and **email**.

Formula1 Racing Project using Azure Databricks, Delta Lake and Azure Data Factory

Personal project

Pune, Maharashtra

- Building a solution architecture for a data engineering solution using **Azure Databricks, Azure Data Lake Gen2, Azure Data Factory** and **Power BI**(for visualization)
- Mounting **Azure Storage** in **Databricks** using secrets stored in **Azure Key Vault**
- **PySpark**: Ingestion of CSV, simple and complex JSON files into the **data lake** as **parquet** files/ tables, using Transformations on dataframes such as **Filter, Join, Simple Aggregations, GroupBy, Window functions** etc. and Creating **local** and **temporary views**
- **Spark SQL**: Creating **databases, tables** and **views**, using Transformations such as **Filter, Join, Simple Aggregations, GroupBy, Window functions** etc. and Creating **local** and **temporary views**
- Read, Write, Update, Delete and Merge to **delta lake** using both **PySpark** as well as SQL and Converting **Parquet files to Delta files**
- Creating pipelines in **Azure data factory** to execute **Databricks** notebooks
- Designing **robust** pipelines to deal with **unexpected** scenarios such as **missing** files
- **Scheduling** the pipelines using data factory **triggers** to execute at **regular intervals**

CERTIFICATES AND ACHIEVEMENTS

- | | |
|---|-------------|
| • Alation Certified Professional Consultant | 2022 — 2027 |
| • DA100: Microsoft Certified Data Analyst Associate | 2021 — 2023 |
| • Infosys Certified Power BI Advanced Developer | |
| • Infosys Certified Python Programmer | |