

JAY ZILPILWAR

DATA ENGINEER| EXPERIENCE - 5+ YEARS

✉ jayzilpilwar@gmail.com

☎ 9657023743

📍 Pune

EDUCATION

Vishwakarma Institute Of Information Technology,Pune
B.E. Computer Engineering 2019
CGPA: 8.04

Government Polytechnic, Yavatmal
Diploma Computer Engineering 2016
Percentage - 84.31%

SUMMARY

- Experienced Data Engineer with expertise in designing and developing data pipelines and infrastructure.
- Skilled in building scalable and high-performance data processing systems using technologies Cloud Tech like GCP, AWS, Hadoop.
- Strong programming skills in Python, Java, and SQL for data manipulation, processing, and analysis.
- Expertise in data modeling, ETL development, and data integration using tools like GCP BigQuery , DataFlow , DataStage.
- Hands-on experience with various data storage and processing technologies such as HDFS, Hive, PySpark , CloudSQL and Redshift.
- Familiarity with DevOps tools and practices like Jenkins and Gitlab for building, deploying, and managing data pipelines.

SKILLS

PROGRAMMING LANGUAGE: Python, SQL, Shell Scripting, SparkSQL
CLOUD TECHNOLOGY: GCP, AWS
DATABASE: MySQL, Oracle, MongoDB, Teradata, PostgreSQL, Hive, Hadoop
ETL TOOLS: Debezium, Fivetran, DataStage, Informatica
GRAPH DATABASE: Neo4j
OTHERS TOOLS: Jenkins, GitLab, Terraform

EMPLOYMENT

PERSISTENT SYSTEM LIMITED Lead Software Engineer

Tech Stack - GCP BigQuery , Dataflow, PubSub ,Cloud Function | Composer | Github | Postman | Oracle | SSIS
• Developed complex and logical legacy logic inside BigQuery Stored Proc.
• Created DAG to orchestrate Stored Proc in BQ.
• Developed CICD workflows for Cloud Function,Cloud Run,BigQuery,Data Fusion,Dataflow using Git Actions

Pune
June 2023 to Current

KPMG ASSURANCE AND CONSULTING SERVICE LLP Cloud Data Engineer

Client : Ports and Logistics industry

Pune
Oct. 2021 to May 2023

- Technology: GCP | BigQuery| Composer | DataFlow | PubSub| Debezium | AWS | RDS | Lambda | EC2 | S3 | Python | Oracle | Git
- Led Oracle, MySQL, AWS to GCP data migration project, transferring over 1TB of data using Google DataFlow and Golden Gate.
- Analyzing AWS EC2 python script and then creating BQ compatible query to meet ETL requirements.
- Analyzing AWS Lambda function and maintaining same replica in GCP using Cloud Function which schedule via Cloud Scheduler and Pub/Sub.
- Developed CDC dataflow pipeline using Debezium an open source tool with PubSub, DataFlow and BigQuery. Giving data at every 2-3 min on final BigQuery layer.
- Developed complex and logical merge logic inside Dataflow to reduce trigger dependency on BigQuery Stored Proc.
- Developed end to end CDC data pipeline using GCP DataStream tool.
- Developed Sql Server, MySQL to BigQuery pipeline and orchestrate via Cloud Composer.
- Developed BigQuery to SQL Server pipeline and orchestrate via Cloud Composer.
- Developed File ingestion pipeline using Cloud Function and Storage and load final data to BigQuery.
- Developed API for logistics service on existing data using GCP API Gateway.
- Developed backend data pipelines for Power BI dashboards.

PLUTO7 Data Engineer

- Technology: GCP | BigQuery| Cloud Function | MongoDB | Data Modelling | Python | Git | Looker
- Developing Customer Data Platform(CDP) data model for one of the client and data load inside the BigQuery.
- Build data pipeline between MongoDB,MySQL to BigQuery using Fivetran tool
- Cloud Function used to trigger file based event and perform ETL using python code and load data in BigQuery then used Data Studio for Dash boarding.

Bangalore
Apr. 2021 to Aug. 2021

DATAMETICA SOLUTIONS PRIVATE LIMITED Engineer I

Client : Macy's

Pune
July 2019 to Mar. 2021

- Technology: GCP | Airflow | BigQuery| DataStage | Python | Hadoop | Sqoop | Hive| Git | Control-M
- Developed automatic script which can extract code from Hadoop cluster using Sqoop utility and load into GCS bucket then moving to GCP BigQuery bq load.
- Developed on Airflow DAG's to scheduled and perform ETL Logic.
- Worked on File ingestion Script to load data into BigQuery.
- Understanding DataStage ETL logic and create BQ compatible SQL query for ETL migration.
- Worked on building data pipeline -using BigQuery ,Python script , Storage, Airflow and schedule them using Control M.

Client :- IBX

- Technology: GCP | DataStage | Informatica | Python | Java | SQL Server | SSIS | Neo4j | Hadoop | Hive | Spark SQL | Git
- Developed automatic script which can extract code from client database and load into GCS bucket smoothly.
- Finding insights from database warehouse and graph database using Neo4j and Big Query.
- Load data into Data Studio for interactive reports for Managers Decision and for client better understanding of their database warehouse patterns.
- Developed and Configure Mailing System with shell script so that every activity is notified and track status convey to every team member.
- Working and knowledge on Teradata and Netezza databases.
- Worked on Data Extraction from DataStage XML and DSX files.
- Worked on Data Processing in Hadoop-Hive,BigQuery,Data-proc.
- Finding workload of DataStage and Informatica through Python code.
- Experience in writing and creating spark Sql queries and handling workload

AWARDS

SUPER TEAM AWARD | KPMG INDIA | MARCH 2023

- Developed, Designed and Delivered data migration project for one of largest Ports and Logistics Client.

ABOVE AND BEYOND AWARD | KPMG INDIA | DEC 2022

- Develop and deliver quality work for backend data pipelines for dashboard within time.

SPOT AWARD | KPMG INDIA | JUNE 2022

- Develop and deploy successfully on priority dashboard with end to end backend data pipelines

SPOT AWARD | DATAMETICA SOLUTION PVT. LTD. | MARCH 2020

- Developed streamline python code to parse the DataStage XMLs to identify correct data lineage.

CERTICATIONS

- GOOGLE PROFESSIONAL DATA ENGINEER CERTIFICATION
- ASTRONOMER CERTIFICATION FOR APACHE AIRFLOW FUNDAMENTALS
- APACHE BEAM | A HANDS-ON COURSE TO BUILD BIG DATA PIPELINES