

Shashank Puli

Data Engineer | Portfolio: Shashank

Email: Pulishashank1@gmail.com [LinkedIn: linkedin.com/in/Shashank](https://www.linkedin.com/in/Shashank) Mobile: +1 (484) 487-1200 Place: United States

SUMMARY

Experienced Data Engineer with 4+ years of experience in designing, developing, and optimizing scalable ETL pipelines across AWS, GCP, and on-premises platforms. Engineered real-time data workflows and automated orchestration using Airflow and Kubernetes, improving pipeline efficiency by 20%. Optimized Big Data storage and processing with Spark, HBase, and Snowflake, enhancing query performance. Developed and maintained interactive dashboards in Power BI and Tableau to accelerate operational insights. Proficient in Python, SQL, Java, and Shell scripting for high-volume data processing and integration.

TECHNICAL SKILLS

- **Programming & Scripting Languages:** Java, Python, Shell Scripting
- **Cloud Platforms & Services:** AWS, S3, EMR, Redshift, Lambda, Athena, Glue, Kinesis, Data Lake, GCP, BigQuery, Composer, Azure, Azure Data Lake, ADF, Databricks, Azure SQL
- **Web Technologies:** HTML, CSS, JSON
- **Frameworks & Data Access:** Hibernate, JDBC
- **Databases & Data Storage:** Oracle, MySQL, SQL Server, PostgreSQL, HBase, Cassandra, MongoDB, Snowflake
- **Big Data & Distributed Computing:** Hadoop, HDFS, MapReduce, Hive, Pig, HCatalog, HBase, Sqoop, Impala, Yarn, Kerberos, Cloudera Manager, Spark, PySpark, Kafka, Flume
- **ETL & Data Integration Tools:** Informatica, Talend, SSIS, Airflow
- **Visualization & Reporting Tools:** Tableau, Power BI
- **Web & Application Servers:** Apache Tomcat, WebLogic, WebSphere
- **Version Control & Build Tools:** Git, Maven, SBT, CBT, GitHub
- **Development IDEs & Methodologies:** Eclipse, Dreamweaver, Visual Studio, SharePoint
- **Operating Systems:** Windows, Unix, Linux
- **Database Management & Query Tools:** TOAD, SQL Developer, Azure Data Studio, SSMS, Teradata SQL Assistant
- **API & Web Testing Tools:** SoapUI

PROFESSIONAL EXPERIENCE

Crowe LLP

[AWS Data Engineer](#)

August 2024 – July 2025

- Designed and deployed scalable ETL pipelines on AWS using Lambda, S3, and Python, enabling efficient ingestion and transformation of large financial datasets across many client projects.
- Developed real-time data integration workflows with Kubernetes and Docker, reducing processing latency by 30% across many client reporting streams.
- Implemented IAM roles and CloudWatch triggers for automated Python scripts, enhancing system security and operational uptime.
- Optimized HBase and Redshift data storage and query performance using partitioning and indexing, decreasing query runtime.
- Developed Apache Spark jobs for large-scale data processing on AWS EMR, improving batch processing speed by 40% and enabling faster financial reporting.

Orange S.A.

[GCP Data Engineer](#)

March 2021 – June 2023

- Engineered scalable ETL pipelines in BigQuery and Dataflow, handling large volumes of telecom and customer data for analytics and reporting, ensuring seamless data ingestion and transformation.

- Established Airflow DAGs for orchestrating GCP data workflows, ensuring pipeline reliability and seamless integration with Cloud Storage and Pub/Sub.
- Executed Data Quality checks and monitoring using Cloud Monitoring and Stackdriver, reducing data discrepancies across customer analytics reports by 40%.
- Designed and maintained Snowflake data warehouses, optimizing SQL queries and clustering strategies to accelerate analytics for customer behavior and network performance insights.
- Automated GCP resource provisioning with Terraform and Python, cutting infrastructure deployment time by 50% while ensuring compliance with internal security standards.

Carborundum Universal
Data Engineer

February 2020 – March 2021

- Optimized ETL pipelines to ingest, clean, and transform large-scale manufacturing and supply chain data using Python, SQL, and Airflow, improving data processing efficiency.
- Constructed and maintained real-time dashboards in Power BI to monitor production KPIs, reducing reporting latency by 40% for management decisions.
- Applied data integration workflows using Spark and Hadoop, enabling seamless consolidation of operational data across many facilities.
- Migrated legacy datasets to Snowflake, enhancing data accessibility and query performance by 35% for analytics teams.
- Implemented data validation and anomaly detection processes using Python scripts and SQL, ensuring data accuracy in production and inventory reports.

EDUCATION

Master of Science in Computer Science
University of South Florida

January 2023 – December 2025

CERTIFICATIONS

- Attained AWS Certified Data Engineer Associate.
- Achieved AWS Certified Cloud Practitioner.
- Certified Microsoft Azure Data Engineer.
- Completed Power BI Data Analyst.