

CONTACT DETAILS

+91 - 8618837023

patneshirish@gmail.com

Pune, Maharashtra

ABOUT ME

Detail-oriented Data Engineer and DevOps Engineer with over **5+** years of experience delivering scalable data solutions for enterprise clients. Proven ability to design and optimize ETL pipelines, working hands-on with modern data platforms including **Databricks**, **Snowflake**, **AWS** and **Azure**. Strong background in end-to-end data engineering—covering data ingestion, transformation, quality assurance, and stakeholder collaboration. Passionate about building clean, efficient data systems that drive business insights and long-term value.

EDUCATION

Engineering, Instrumentation

Government College of Engineering Jalgaon.

Post Graduate Diploma in Data Analytics and Machine Learning Imarticus Learning Pvt Ltd

SQL RDBMS, Python, EDA, Machine Learning with python, NLP, Statistics for business, Data Visualization

SKILLS

Programming & Scripting

- Languages: Python, SQL, Shell Scripting
- Code Quality Tools: Pylint, flake8 (for static code analysis), Black Formatter (for code formatting)

Databases

- SQL: SQL Server, Oracle, PostgreSQL, Amazon Aurora (RDS)
- NoSQL: MongoDB, DocumentDB, DynamoDB

Big Data & Cloud Platforms

- Big Data Tools: PySpark, Kafka (Confluent Kafka), AWS Glue, AWS EMR, Snowflake, Databricks.
- Cloud Platforms: AWS (Lambda, API Gateway, Fargate, ECS, EventBridge, S3, Lake Formation, Step Functions, ECR), Azure (ADF, Azure AD, Azure functions, Azure Databricks)
- ETL & Workflow Orchestration: Apache Airflow

Data Engineering & Integration

- Data Pipelines: End-to-end design and migration using PySpark, AWS Glue, and Kafka, Databricks, SSIS, Talend.
- Integration Technologies: WebSocket, Webhook, REST and FAST APIs

Data Validation & Testing

- Automated ETL testing using Pytest and pydemeq
- Regression and performance testing for APIs and datasets

Analytics & Visualization

- Visualization Tools: Power BI, Looker, Tableau
- Monitoring & Observability: Dynatrace

Infrastructure Automation & CI/CD

- Tools: Terraform, CloudFormation, GitHub Action, Concourse CI/CD pipelines
- Automated deployment of workflows and end to end infrastructure provisioning

Containerization

- Tools: Docker, AWS Fargate, AWS Lambda Function

Data Governance & Security

- Services: AWS Lake Formation, Databricks Unity Catalog, Delta Sharing
- Data cataloging, access control, and secure data sharing across environments.

Shirish Patne

DATA ENGINEER

PROFESSIONAL EXPERIENCE

DATA ENGINEER

Cloud Migration Expertise:

- Led the end-to-end migration of critical data pipelines and infrastructure to AWS, utilizing Kafka, Lambda, and API Gateway for real-time processing.
- Developed Python-based data validation scripts and automated regression testing for DocumentDB and MongoDB, ensuring data integrity post-migration.
- Optimized workflows by configuring Kafka sink connectors to trigger Lambda, improving resource efficiency and reducing costs.

Advanced Data Migration & Data Modeling

- Led the migration of enterprise-level data from legacy systems (SSIS, Talend) to Snowflake, developing end-to-end data pipelines and provisioning infrastructure using AWS services like Lambda, Step Functions, and Glue.
- Migrated key data from application like Jira, OneStream, Workday, PRS (Patient Renewal System), and Steam (CPA finance data).
- Designed and implemented multiple dynamic table queries in Snowflake for fact, dimension, and aggregation table.
- Built and managed bronze, silver, and gold Delta tables to structure and optimize data workflows in Databricks.
- Developed and maintained several backend data workflows and automation pipelines within the Databricks platform.
- Migrated a Streamlit application from Posit Connect to the Databricks Lakehouse, ensuring smooth integration with the new backend.

Data Engineering & ETL:

- Designed scalable ETL pipelines using PySpark, AWS Glue, and Apache Airflow, enabling efficient data processing.
- Built REST and Fast APIs with API Gateway, load balancer and Containerized Lambda for CRUD operations, enhancing database accessibility and usability.
- Mapped disparate data sources into unified frameworks, improving enterprise data governance.

Infrastructure Automation & DevOps:

- Automated infrastructure provisioning with Terraform, including API Gateway, Lambda, and IAM roles, streamlining deployments.
- Implemented CI/CD pipelines via Concourse and GitHub actions for automated Docker image creation and deployment to Amazon ECR, and automated workflow deployment on databricks.
- Designed scalable multiple CI/CD pipeline using artifacts like webhook, s3 trigger, Git.

Testing & Optimization:

- Developed a Pytest framework for ETL validation, cutting execution time from 1 hour to 10 minutes.
- Automated testing pipelines with Event Bridge and Lambda, generating reports stored in S3 for real-time monitoring.
- Conducted regression and performance testing for APIs, Kafka topics, and databases to ensure system reliability.

Big Data & Real-Time Processing:

- Utilized Confluent Kafka and WebSocket's for real-time data validation, ensuring end-to-end data accuracy.
- Automated scheduling with Apache Airflow DAGs, reducing operational overhead and enhancing pipeline reliability.

CERTIFICATIONS & AWARDS

- ❖ [Databricks Certified Data Engineer Associate](#)
- ❖ [AWS Partner: Sales Accreditation - Credly](#)
- ❖ [AWS Partner: Technical Accredited - Credly](#)
- ❖ [AWS Partner: Cloud Economics Essentials - Credly](#)
- ❖ [Snowflake Technical Sales Professional](#)