





Pratik Nandankar

AWS Data Engineer

 pratik.nandankar1@gmail.com

 +91-8446611969

 www.linkedin.com/in/pratik-nandankar

 Pune, India

Summary

Experienced AWS data engineer with over **9 years** of expertise in constructing and managing AWS Data Lakes, developing Data Architectures, implementing Data Analytics and Data Warehousing, as well as devising Cloud Migration Strategies and Data Integration Solutions. Proficient in collaborating with multidisciplinary teams across diverse locations to empower clients in leveraging cloud computing and data analytics technologies across **Media**, **BFSI**, and **Human Capital** sectors. Skilled in contributing to RFP/Proposals, Sales Pitch, and POC developments. Committed to staying at the forefront of industry trends and continuously enhancing data engineering strategies.

Skills And Strengths

- **Cloud Platform:** - Amazon Web Services (AWS)
- **Programming:** - Python, PySpark, Spark APIs, Pandas, Unix Shell Scripting, SQL & PLSQL
- **Big Data Ecosystem:** - Apache Spark, Hive, Apache Hadoop, HDFS
- **AWS Cloud:** - AWS Glue, EMR, Athena, Lambda, Step Functions, SageMaker, DataSync, AppFlow, MWAA (Airflow) etc.
- **Storage:** - Amazon S3, Amazon Glacier, Amazon EFS and Free NAS.
- **Monitoring:** - AWS CloudWatch, CloudTrail, AWS Guard Duty, New Relic etc.
- **Databases:** - Redshift, Dynamo DB, Oracle, PostgreSQL & MS SQL
- **Data:** - Data Lake, Delta Lake, Data Warehousing, Data Modeling, Data Governance, Data Security
- **Others:** - Databricks, Apache Airflow, Informatica PC, Docker, Git, AWS CDK, AWS Code Pipeline etc.
- **Strengths:** - Proactive Communication, Problem Solving, Critical Thinking, Customer Service

Professional Experience

Deloitte

Jan 2022 - Present

Lead AWS Data Engineer (Big Data Consultant)

Data Driven Business Value Realization Platform (Media and Entertainment Domain)

- Spearheaded the engineering and architecture of the analytics application migration to an **AWS data lake**, serving as a centralized repository for analytical data for a prominent media and entertainment client. Leveraged AWS serverless and analytics offerings to optimize efficiency.
- **Designed the AWS data lake architecture** for developing key use cases such as BARC (Broadcast Audience Research Council) - Viewership analytics reporting, Advertiser 360 analytics, and channel performance reporting for ad sales business.
- **Transformed** SQL Server stored procedures using Spark-SQL and PySpark transformations on raw data, facilitating ingestion into the AWS Data Lake in S3 Hive format across prepared and transformed layers.
- **Engineered** end-to-end ETL pipelines to seamlessly load data from diverse sources including BARC – Yumi, SQL Server, SAP, SharePoint, Salesforce, and Rights Line into AWS S3, Dynamo DB and Redshift using AWS AppFlow, AWS DMS, AWS DataSync and PySpark/Python with AWS Glue and EMR.
- **Innovated** a flexible AWS Step Function for table migration tasks leveraging Glue, Lambda, and S3, significantly reducing development time for each pipeline by **65%**.
- Demonstrated proactive engagement in resolving defects and bugs related to data loading from various source systems. Collaborated with cross-functional teams utilizing SQL and PySpark for analysis, root cause identification, and implementation of solutions, ensuring consistent data quality.
- **Implemented optimization techniques** in AWS Glue jobs using PySpark and Spark SQL to enhance performance.
- **Implemented end-to-end Machine Learning Operations (MLOps) pipelines** for Channel FPC (Fix Point Chart) and Content Plagiarism model using AWS SageMaker, ensuring seamless integration from data preprocessing via Apache Airflow through to model deployment.
- Utilized **Docker** extensively to containerize machine learning models and applications, ensuring consistent behavior across development, testing, and production environments.

- **Collaborated** closely with cross-functional teams including data scientists and DevOps professionals to streamline development and deployment processes.
- **Developed** AWS CI/CD pipelines using AWS CDK for seamless code deployment into AWS production account.
- **Orchestrated automation** by creating AWS Lambda scripts for the automation of **150+** Tableau dashboard refreshes using Python and Tableau APIs, **eliminating 100% manual intervention** in application maintenance activities.
- Created HLD/LLD and maintained documentation of data processes, workflows and troubleshooting procedures ensuring efficient knowledge transfer and team collaboration.
- **Leading** a team of 7 members and trained them on application functionalities and overall project needs.
- Contributed to developing the **Gen-AI proposal** for Ad-sales business, resulting in the addition of over **INR 1 Cr** in revenue to the on-going project.

Tata Consultancy Services Ltd

June 2015 – Dec 2021

AWS Data Engineer

Corporate VPC Finance Data Lake Migration to AWS (Finance domain)

- The project focuses on data lake development using scaled approach to migrate Finance Data Lake to AWS platform.
- Designed AWS data pipeline using Databricks and Apache Spark, data migration task using AWS DMS and Data Sync.
- Developed PySpark scripts for the creation of Data Marts featuring Monthly, Quarterly, and Yearly transformations for payroll data across **50+** countries, enabling aggregated reporting.
- Transformed Oracle SQL procedures into Spark SQL, worked with Hive delta tables and various data formats like Text, CSV, and Parquet files.
- Implemented optimizations for a gross aggregation task, slashing its runtime from 3 hours to a mere 10 minutes within the AWS data lake environment. This achievement translated to annual savings of **USD 90,000** for the client.
- Demonstrated proficiency in documentation, test case writing, and conducting functional and manual testing.

AWS Cloud Engineer (Nearshore Lead)

Corporate ETL Workload Migration to AWS (Human Capital domain)

- Collaborated closely with technical architects to craft a comprehensive AWS solution for migrating the ETL tool, Informatica, onto an AWS-hosted platform supporting over **30** applications. Above migration involved detailed planning around first time setups, parallel runs, and cutover tasks.
- Contributed within the engineering team to outline the data and network security approach by configuring VPC settings, implementing user management and control through IAM, and managing multi-AZ deployment for RDS.
- Worked cross functionally with business stakeholders along with IT and application users to analyze, recommend, design, and implement solutions for SQL based reporting.

ETL Developer

Global Payroll Data-Warehouse (Human Capital domain)

- Applied extensive expertise in data warehouse design, incorporating STAR and SNOWFLAKE schemas.
- Designed and implemented intricate mappings, mapplets, and workflows using Informatica.
- Developed Unix/Python shell scripts for data validation and parsing API responses.
- Implemented an API-driven approach to retrieve banking-related data, mitigating Personally Identifiable Information (PII) exposure in flat files.
- Optimized process flow design, resulting in a significant reduction of workflow execution time by **75-80%**.
- Contributed to the creation of HLD/LLD documents and SOP creations for various processes.

Certifications

- AWS Certified Solutions Architect – Associate
- Lean Six Sigma Green Belt

Academic Qualifications

- Bachelor of Engineering (B.E) in Electronics and Telecommunication from Yeshwantrao Chavan College of Engineering (YCCE), Nagpur with 87.11 percentage (9.52 CGPA) – Gold Medalist, 2015
- Class 12th H.S.C. from Maharashtra State Board, Nagpur with 89 percentage, 2011
- Class 10th S.S.C. from Maharashtra State Board, Nagpur with 91.84 percentage, 2009