RAMAKRISHNA KARNATI

Azure Cloud & DevOps Specialist | Data Engineer

Hyderabad, India ramukarnati2001@gmail.com +91 7207000030
ramakrishna-karnati-899066170 RamuK2001

Summary

Results-driven **Azure Data Engineer** with **3** years of experience in **ETL Development, Data Engineering, and Azure DevOps Automation**. Adept at designing and implementing Azure-based ETL solutions, developing **CI/CD** pipelines, and driving cost-efficient cloud operations. Expertise in **Azure Databricks (ADB), Azure Data Factory (ADF), DevOps, ETL,** and **Data Warehousing**. Key achievements include:

- Designed and developed ETL pipelines in ADF migrating **terabytes** of data from on-premise to Azure, achieving a **99%** success rate.
- Developed **PySpark**-based data processing solutions handling **gigabytes** of data daily across multiple formats.
- Automated **Vacuum** and **Optimize** operations on ADB tables, enhancing storage efficiency and boosting table performance by **40%**, and optimized query performance through improved code design and utilizing advanced techniques like **partitioning** and **liquid clustering**.
- **85%** reduction in manual infrastructure setup time using **ARM templates**, **Terraform**, and **Azure DevOps** pipelines.
- Implemented a **selective deployment** feature in the **CI/CD** pipeline, enabling targeted artifact deployment for increased flexibility and reduced deployment time.
- Automated code quality checks by integrating **SonarQube** and **ARMTTK** with **Azure DevOps**, reducing manual review effort by **80%**.
- Automated the weekly cost, ADF pipeline performance, and Databricks clusters usage metrics reports extraction using **Azure PowerShell modules**, and **Azure Automation Account**, reducing manual effort by **95%**.
- Built real-time **Power BI** dashboards for SLA tracking and pipeline monitoring, improving performance insights.

Skills

Programming Languages:	Python, PySpark, Java, C, C++, PowerShell, Unix, SQL
Cloud Tools:	Azure Data Factory, Azure Databricks, Azure Data Lake Storage, Logic App, Automation Account, Purview, ARM Templates Deployment, Terraform, Azure DevOps (YAML Pipelines, Repos, Boards), Microsoft Fabric (Warehouse, Lakehouse, Data Engineering, Data Science)
Big Data:	Hadoop, Scala, HDFS, Pig, HQL, Hive, Kafka, Spark
Other Tools/Technologies:	Power BI, AWS (EC2 & S3), HTML, CSS, JavaScript, Spring Framework, Docker, Kubernetes, SonarQube, Eclipse, JPA, Hibernate, Maven
Soft Skills:	Quick learning ability, Problem solving, Critical thinking, Attention to detail, Adaptability, Communication, Googling

Experience

SPECIALIST PROGRAMMER **Aug 2022 – Present**
INFOSYS, *Hyderabad, India*

Project 1: Azure DevOps and Data Migration

Infrastructure Setup and Automation:

- Configured a comprehensive Azure environment, including Storage Accounts (ADLS), Databricks, Data Factory, and Virtual Machines, with end-to-end private networking using **ARM templates** deployed through **Azure DevOps pipelines**, reducing the environment setup time to just **around 20 minutes, an 85% reduction** in manual configuration time.

- Integrated ADF & ADB with **Azure DevOps Repos** to enable efficient version control and seamless collaboration.
- Implemented **CI/CD** for artifacts such as ADB notebooks, workflows, ADF pipelines, datasets, etc., using Repos and YAML-based pipelines in Azure DevOps, capable of deploying around 150 artifacts within 5 minutes.
- Automated the setup and configuration of Azure Databricks clusters, pools, etc., and **Microsoft Purview** sources, scans, etc., using **REST APIs** through Azure DevOps pipelines, enhancing deployment efficiency by **70%** and streamlining data governance processes.
- Documented the infrastructure setup, CI/CD processes, branching strategy, and other project-related documentation in Azure Wiki, creating a comprehensive reference for implementation and project management.

ETL Development and Data Processing:

- Designed and developed ETL pipelines in ADF, integrating ADB notebooks and ADLS, to orchestrate & automate the migration of **terabytes** of data from various on-premise sources like SAP, SFTP, Salesforce, Database servers, etc., to the cloud, **achieving a 99% success rate.**
- Designed ADF pipelines with detailed logging and **HTML**-formatted email alert mechanisms to promptly notify stakeholders in case of failures, ensuring efficient monitoring and quick issue resolution.
- Developed PySpark code to read and process **gigabytes** of data daily from various file formats, including CSV, JSON, XLSX, XML, and PARQUET, ensuring efficient and accurate data handling to meet business requirements.
- Created **Azure Logic App** workflows to fetch data from SAP source using RFC parameters and land it in ADLS, then orchestrated the movement of data to downstream applications via ADF pipelines and performed data processing using PySpark in ADB notebooks.
- Developed an **event-based triggered** pipeline that automatically loads data into respective tables whenever a file lands in the designated ADLS location, reducing manual intervention by **99%** and processing over **20 files per day**.
- Automated & implemented regular **Vacuum and Optimize** operations on ADB tables to manage storage efficiently and boost read/write performance on tables by **40%**.
- Optimized cost and resource utilization of Databricks clusters by **30%** by improving the code design and leveraging advanced methods like **partitioning** and **liquid clustering**.
- Gained experience in writing advanced SQL queries with **Common Table Expressions (CTEs)**, window expressions, and complex conditions to efficiently process data.
- Worked with managed and external tables within a **Unity Catalog**, utilizing a custom metastore to effectively manage and govern data across various environments in ADB.

Reporting and Support:

- Created an SLA matrix report in **Power BI** to visualize and track incidents and service requests from the ServiceNow platform.
- Developed a centralized Power BI dashboard to monitor production pipeline status, utilizing logs from each pipeline to provide real-time insights and improve performance tracking.
- Handled CI/CD deployments to production for changes, bug fixes, or enhancements during the monthly scheduled production releases, in accordance with client policies.
- Developed **PowerShell** scripts to generate weekly cost and ADF pipeline performance reports using Azure PowerShell modules and Azure Automation Account, **reducing manual effort by 95%**.
- Played a key role throughout the project lifecycle, including the initial infrastructure design, ETL development, and post-production support following go-live.
- Responsible for **troubleshooting or debugging** failures in production ADF pipelines or infrastructure, identifying root causes, and deploying fixes as needed.
- Responsible for analyzing weekly cost reports in Azure to identify deviations, investigate root causes, and provide detailed reports with actionable insights for cost optimization.

Project 2: Data Warehousing and Data Engineering

- Developed workflows in ADB to migrate **millions** of records of data daily from an on-premise database server to the Delta Lake tables in ADB using **JDBC connection**.
- Scheduled the workflows using **CRON** syntax to align with business requirements, ensuring timely and automated data migration and processing.
- Configured the workflows with **job clusters** to enhance job isolation and improve cost efficiency and resource utilization by eliminating the minimum 10-minute inactivity period required for all-purpose clusters before auto-

termination.

- Provided support in troubleshooting and bug fixing throughout the project development phase, successfully **resolving 2 critical issues** related to **Row-Level Security (RLS)** and **concurrency issues**.
- Managed CI/CD deployments to production for workflows and notebooks in ADB, ensuring successful deployments and smooth transitions throughout the project lifecycle, leading up to the on-time project go-live.
- Leveraged tags in Azure Databricks (ADB) resources, such as clusters, to effectively segregate costs by application, enabling better cost tracking and reporting for each business unit.

Project 3: Data Modelling and Data Engineering

- Collaborated in business requirement gathering and walk-through sessions to understand client expectations, while also assisting in project estimations, planning, timelines, and resource allocation.
- Configured **Azure OpenAI** service and developed a **custom model** in ADB, behind client network policies.
- Designed and implemented category classification logic for over **2.5 billion** records using three approaches—UPC lookup, keyword-based, and LLM-based—achieving over **90% accuracy**.
- Designed a cost-optimization strategy for LLM-based categorization by leveraging a subset of data (**around 0.2% of 2.5 billion records**) to classify the majority, significantly reducing computational expenses.
- Developed ADB workflows to seamlessly integrate the three approaches into a unified process for categorizing both **historic and incremental data**.

SENIOR ANALYST - INTERN **Feb 2022 – May 2022**
CAPGEMINI, *Hyderabad, India*

Case Study: Student Attendance Management System

- Worked with a team of 5 members to develop a **Java-based Spring application** for student attendance management using **GitHub** for version control and seamless collaboration.
- Utilized **PostgreSQL** as the DBMS and integrated the **JDBC API** into the application to perform efficient read/write operations on the database.
- Developed **JUnit** test cases to ensure robust functionality, achieving over 90% test coverage.
- Leveraged the **SonarQube** tool for code quality validation and **resolved 100%** of code smells and issues.
- Containerized the entire application using **Docker** and deployed it on an **AWS EC2 instance**, leveraging **Kubernetes** for orchestration and scalability.

Awards & Achievements

- Earned a formal appreciation (individual) from the client for resolving a critical incident during a crucial month-end business closure outside the scope of the project and my usual work hours.
- Received formal client appreciation (individual) for delivering exceptional support on a separate project, particularly for seamlessly automating CI/CD deployments.
- Consistently recognized with "**Rise Insta**" awards for exceptional performance every quarter.
- Awarded with the "**Business Ninja Award**" in recognition of remarkable contributions and outstanding achievements.
- Achieved the "**Outstanding**" rating twice in year-end performance appraisals, reflecting exceptional contributions.

Certifications

- Databricks Certified: Data Engineer Professional
- Microsoft Certified: Fabric Analytics Engineer Associate (DP-600)
- Microsoft Certified: Azure Fundamentals (AZ-900)
- Microsoft Certified: Azure Data Fundamentals (DP-900)

Education

B. TECH. in Electronics and Communication Engineering (ECE) **Jun 2018 - May 2022**
VARDHAMAN COLLEGE OF ENGINEERING, *Hyderabad, India*

- **Project:** Autonomous Obstacle Avoiding Robot