

# KUNDANA CHERUKURI



## Senior Software Engineer

Email ID: kundanach9@gmail.com

Contact: +1 (571)409-6005

LinkedIn: [www.linkedin.com/in/kundana-cherukuri](https://www.linkedin.com/in/kundana-cherukuri)

GitHub: <https://github.com/Kundanach9>

## Professional Summary:

- Over **8 years** of experience as a **Software Engineer** by delivering scalable **ETL/ELT pipelines** across **AWS, Azure, GCP** with strong expertise in data integration, transformation, **orchestration**, and governance for business-critical systems.
- Specializing in building **big data platforms** using **Apache Spark, Kafka, and Hadoop** to support real-time and batch data processing pipelines with optimized performance and fault tolerance.
- Developed **CI/CD pipelines** with **Terraform, Jenkins, and GitHub Actions** to automate infrastructure provisioning, data workflows, and code deployments across development, QA, and production environments.
- Designed **orchestration workflows** using **Apache Airflow** and **Azure Data Factory**, managing dependency logic, retry strategies, and **SLA monitoring** for high-throughput, low-latency data movement.
- Built and maintained **Delta Lake** and **Snowflake-based lakehouse** architectures to unify streaming and historical analytics with scalable, cost-efficient, and query-optimized data models.
- Engineered event-driven data ingestion with **Kafka Streams** and **AWS Kinesis** for real-time **pipeline execution**, minimizing lag in telemetry and transactional datasets across cloud platforms.
- Proficient in **Python, SQL, and PySpark** for data modeling, transformation, and pipeline modularization; created reusable code libraries to support **cross-functional development** and onboarding.

- Built containerized **microservices and dashboards** using **React.js, Node.js, HTML, and CSS**, integrating them with **backend APIs** and **databases** to support internal operational tools.
- Managed infrastructure as code using **Terraform and CloudFormation**, enforcing security, version control, and scalability across **DevOps layers** for both data and application services.
- Delivered **data quality pipelines** using Great Expectations and **in-DAG validation** checks to ensure schema conformity, missing value detection, and downstream data integrity across workflows.
- Applied data governance policies, including encryption, masking, access control, and auditing to ensure **HIPAA, GDPR, and SOC2 compliance** across cloud-based data lakes and APIs.
- Created **Power BI and Tableau** dashboards from curated data marts to track **KPIs**, user behavior, and system health metrics for finance, healthcare, and customer **analytics domains**.
- Configured observability using **CloudWatch, Prometheus, and Datadog** for real-time insights into pipeline execution, **API performance**, and system-level resource usage.
- Designed and deployed **optimized schemas** in **PostgreSQL, DynamoDB, and MongoDB** to support low-latency access patterns in reporting tools and **full-stack application services**.
- Collaborated with product owners, analysts, and engineering teams across the **SDLC**, mentoring peers and owning solutions across **data engineering, DevOps automation**, and front-end delivery.

**Areas of Expertise:** ETL/ELT Pipelines, CI/CD Automation, Infrastructure as Code, Apache Spark, Kafka, Airflow, Snowflake, RESTful APIs, React.js, Node.js, Docker, Kubernetes, AWS, Azure, SQL/NoSQL, Monitoring & Alerting, Agile/Scrum, Data Modeling, Machine Learning Pipelines, Troubleshooting, User Interface (UI) Design, GraphQL, A/B testing.

## Technical Skills

**Cloud Platforms:** AWS, Azure, GCP, Databricks Lakehouse, Snowflake Native App Framework

**Data Engineering Tools:** Apache Spark, Kafka, Airflow, Hadoop, dbt, AWS Glue, EMR, Dataflow, Databricks, Tableau, Power BI

**Data Warehouses / Lakes:** Snowflake, Redshift, BigQuery, Azure Synapse, Delta Lake, Lake Formation

**ETL/ELT:** Informatica, Talend, Glue Studio, ADF

**Orchestration & Scheduling:** Airflow, Cloud Composer, Azure Data Factory

**Languages:** Python, SQL, PySpark, Scala, Shell, R, JavaScript, TypeScript, HTML, CSS

**Databases:** PostgreSQL, MySQL, Oracle, DynamoDB, MongoDB, Firebase Realtime DB, SQLite

**DevOps & CI/CD:** GitHub Actions, Jenkins, Terraform, Docker, Kubernetes, AWS CodePipeline, Azure DevOps

**Monitoring & Logging:** CloudWatch, Stackdriver, Datadog, Prometheus, Grafana, ELK Stack  
**Version Control & Repos:** Git, GitHub, GitLab, Azure Repos, Bitbucket  
**Web & Full Stack Technologies:** React.js, Node.js, Express.js, REST APIs, Next.js, Bootstrap, Tailwind CSS

## Professional Experience

**Client: Home Depot- Boston**

Role: Senior Data Engineer

January 2025 – Present

### Responsibilities:

- Designed and deployed real-time data pipelines using **Apache Kafka** and **Spark Structured Streaming** to process event-driven datasets, supporting sub-second analytics and incident detection.
- Conducted **performance benchmarking** and **capacity planning** for Spark clusters using **AWS CloudWatch**, **Spark UI**, and dynamic resource scaling strategies.
- Built **modular PySpark components** to standardize batch and streaming ingestion, reducing pipeline development time by 40% and improving reusability across projects.
- Led **Agile sprint planning**, backlog refinement, and **peer code reviews** to ensure velocity and delivery quality in a fast-paced engineering environment.
- Orchestrated complex **ETL workflows** using **Apache Airflow** and **dbt**, transforming data from **Amazon S3** into analytics-ready layers in **Snowflake**, with SLA monitoring and alerting via Slack.
- Implemented an end-to-end **data quality framework** using **Great Expectations**, embedding schema validation and null checks within Airflow DAGs.
- Developed **CI/CD pipelines** with **GitHub Actions** and **Terraform** for deploying data workflows and infrastructure, ensuring consistency across dev, QA, and production.
- Modeled **Snowflake data marts** and **star schemas** optimized for **customer analytics**, leveraging **clustering keys** and **materialized views** to reduce query latency by 60%.
- Collaborated with BI and product teams to define **semantic data layers** and **KPI definitions**, streamlining downstream reporting in **Power BI** and **Tableau**.
- Integrated **monitoring solutions** like **Prometheus** and **Datadog** to track Airflow DAG health, Spark executor metrics, and SLA violations.
- Architected **lakehouse-style ingestion** patterns using **Delta Lake** and **Snowflake** to unify batch and real-time processing for scalable analytical workflows.
- Mentored junior engineers on **data governance**, **best practices**, and **pipeline architecture**, fostering a knowledge-sharing culture aligned with engineering excellence.

- Optimized Snowflake compute costs by implementing **Resource Monitors**, **auto-suspend/auto-resume**, and **multi-cluster warehouses**, achieving a 25% reduction in compute resource consumption.
- Established **row-level security**, **data masking**, and **access control policies** in Snowflake and dbt to ensure fine-grained access management and data compliance with **GDPR** standards.
- Integrated **Apache Atlas** with Airflow and dbt to automatically catalog metadata and data lineage, improving traceability and audit readiness across data pipelines.
- Standardized **unit, integration, and regression testing** using Pytest and dbt tests in CI workflows, increasing deployment confidence and maintaining high code coverage.
- Partnered with platform engineers to evaluate and pilot a transition from legacy batch workflows to a **Databricks Lakehouse** model, benchmarking performance and cost trade-offs.
- Implemented a **blue/green deployment strategy** for data pipelines using feature flags and DAG versioning, enabling safe rollbacks and reducing production downtime risk.

**Client: Citizens Bank- North Carolina**

**Role:** Devops Engineer

October 2023 – November 2024

### **Responsibilities:**

- Implemented CI/CD pipelines using **Jenkins** and **GitHub Actions** for microservices deployments, reducing release failures by 70%.
- Automated AWS infrastructure provisioning with **Terraform** (EC2, VPC, IAM, RDS) using reusable modules and remote state.
- Containerized applications with **Docker** and deployed on **Amazon EKS**; configured **ALB** for secure ingress and blue-green rollouts.
- Developed **Helm charts** to manage templated Kubernetes deployments, enhancing release repeatability across environments.
- Integrated **Prometheus** and **Grafana** to monitor container health, resource metrics, and alert on anomalies.
- Built operational automations in **Bash** and **Python** for log parsing, diagnostics, and incident recovery.
- Centralized logging via **ELK Stack** (Elasticsearch, Logstash, Kibana) for distributed services in Kubernetes.
- Configured AWS **IAM roles** and **security groups** ensuring least-privilege access and compliance.
- Collaborated cross-functionally using **Git branching**, **Jira**, and **Confluence**, fostering code review discipline and sprint alignment.
- Optimized AWS costs by automating EC2 schedules and rightsizing with **Trusted Advisor** and CloudWatch insights.

- Educated peers on cluster architecture and best practices, and mentored junior team members on DevOps tooling.

**Client: Tech Mahindra– Hyderabad, India**

Role: Jr Data Engineer

May 2021 – July 2023

**Responsibilities:**

- Led pipeline capacity planning and performance tuning using **Azure Advisor** and **Azure Monitor**, improving data integration throughput across **development** and **production** environments.
- Migrated legacy ETL pipelines to **Azure Data Factory** and **Azure Synapse Analytics**, utilizing **mapping data flows**, **parameterized pipelines**, and **partitioned tables** to reduce latency by 40%.
- Developed reusable **PySpark** modules in **Azure Synapse Spark Pools** to standardize ingestion and transformation logic for trading analytics and customer profiling use cases.
- Facilitated agile ceremonies, conducted peer code reviews, and maintained high **Python** test coverage using **pytest**, ensuring consistent delivery across sprints.
- Automated third-party API data ingestion using **Python scripts** triggered via **Azure Functions** and orchestrated with **ADF**, with embedded exception handling and data completeness checks.
- Built **CI/CD pipelines** using **GitHub Actions**, **Terraform**, and **Azure DevOps Repos** to automate deployment of ADF workflows and **Azure Functions** across environments.
- Implemented data governance using **Azure Purview** for cataloging, lineage tracking, and sensitivity labeling; monitored data movement with **Azure Monitor Logs** and **Log Analytics Workspaces**.
- Collaborated with data science teams to create **feature stores** and model-ready ETL pipelines, staging preprocessed features in **Azure Synapse dedicated pools**.
- (Hybrid Stack) Supported **Snowflake on Azure Marketplace**, designing **dimension** and **fact tables** for curated analytics layers used in **Power BI** reporting.
- Instrumented **ADF pipeline metrics**, **trigger outcomes**, and custom ingestion logs into **Azure Monitor**, enabling proactive SLA alerting through **alerts**, **dashboards**, and **KQL queries**.

**Client: SAK INFORMATICS- Hyderabad, India**

Role: Jr Data Engineer

April 2019– March 2021

**Responsibilities:**

- Conducted **exploratory data analysis (EDA)** on structured datasets using **Excel** and **Python**, identifying anomalies and trends via pivot tables and statistical summaries.
- Cleaned, transformed, and standardized raw data using **Pandas** and **NumPy**, handling outliers, nulls, and inconsistent formats to improve data readiness.
- Built visualizations using **Matplotlib** and **Seaborn** to communicate insights through histograms, heatmaps, and trend plots.
- Automated **Excel dashboards** using **formulas, Macros, and VBA**, delivering weekly KPIs for sales, operations, and marketing teams.
- Developed **Python scripts** to merge, join, and standardize data from multi-departmental CSV and Excel sources for faster reporting cycles.
- Participated in stakeholder meetings to gather data requirements and translated them into **analytical queries** and visual summaries.
- Utilized Excel functions like **VLOOKUP, INDEX-MATCH, and IF statements** to enhance data extraction accuracy and simplify reporting.
- Created **data dictionaries** and documented transformation pipelines to ensure reproducibility and team alignment.
- Performed **time-series analysis** to track customer behavior and product demand trends using **moving averages** and seasonal breakdowns.
- Supported **A/B testing evaluations**, performing **hypothesis testing** (t-tests) using Python's SciPy library to identify statistically significant differences.

**Client: CodeBees PVT LTD– Hyderabad, India**

**Role:** Software Full-Stack Engineer

February 2017 – March 2019

#### **Responsibilities:**

- Developed a **responsive web application** using **React.js, HTML5, CSS3, and Bootstrap** to manage client onboarding, access roles, and internal dashboards.
- Built **RESTful APIs** using **Node.js** and **Express.js**, integrated with **PostgreSQL** for secure CRUD operations and analytics endpoints.
- Implemented **JWT authentication** and **role-based authorization** to manage user sessions and enforce access control policies.
- Managed state with **Redux**, optimizing React component rendering and ensuring smooth data flow between frontend and backend layers.
- Wrote complex **SQL queries** and **stored procedures** in PostgreSQL to support reporting and advanced filtering features.
- Containerized the application using **Docker**, and created **Docker Compose** configurations for dev and staging environments.

- Integrated **Stripe** for billing and **SendGrid** for transactional email delivery to enhance application capabilities.
- Configured **CI/CD pipelines** using **Jenkins** (or Travis CI) for automated testing, code linting, and deployment to **AWS EC2** instances.
- Developed **unit and integration tests** using **Jest** and **Mocha**, improving code reliability and reducing production defects.
- Used **AWS S3** for secure document storage and generated **pre-signed URLs** for access control.
- Analyzed and improved frontend performance using **Google Lighthouse** and **Chrome DevTools**, optimizing lazy loading and re-renders.
- Worked closely with UX designers to convert wireframes into fully responsive, **mobile-first designs** with accessibility compliance.
- Led Git-based workflows for version control and peer-reviewed code to ensure adherence to best practices.

### **Certifications:**

- Azure Fundamentals (AZ-900)
- Salesforce Certified AI Associate
- Aviatrix Certified Multi-Cloud Networking Associate (ACE-MCNA)
- AWS Certified Data Engineer Associate