YAMINI CHENNA

Chicago, IL | 815-649-2125 | chennayamini802@gmail.com | https://www.linkedin.com/in/yamini-chenna/

SUMMARY

Skilled and performance-driven Data Engineer with over 4 years of experience building end-to-end data solutions across cloud, big data, and enterprise-scale platforms. Proven ability to design scalable ETL pipelines, implement modular data lake architectures, and automate data workflows in AWS, Azure, and Hadoop ecosystems. Expertise in Python, PySpark, SQL, and cloud-native tools like AWS Glue, Redshift, Azure Data Factory, Databricks, and Snowflake. Demonstrated success in processing over 1B+ records/month, enabling self-service BI, and ensuring data accuracy, governance, and security across highly regulated environments in finance and insurance sectors.

SKILLS

Programming Languages: Python, SQL, Scala, Shell Scripting, PL/SQL
Big Data Technologies: Snowflake, Spark, Databricks, DBT, Kafka, Delta Lake, Hadoop (HDFS, Hive)
Cloud Platforms: AWS (S3, Redshift, Lambda, EMR, Glue), Azure (Data Factory, Synapse, Data Lake)
Databases: Snowflake, PostgreSQL, MySQL, Oracle, DynamoDB
ETL & BI Tools: DBT, Python/Pandas, Airflow, Tableau, Power BI
DevOps & CI/CD: Git, Docker, Kubernetes, Terraform, Azure DevOps, Jenkins, GitHub Actions
Integration Tools: REST APIs, JSON, CSV, JDBC, FTP, CloudWatch, Splunk

WORK EXPERIENCE

BMO Bank – Chicago, USA

Data Engineer ***August 2024 - Present***

Description: Emulated a real-world data engineering role at BMO Bank by designing and deploying scalable AWS-based data pipelines to support financial reporting, executive dashboards, and regulatory compliance. Delivered robust batch ETL solutions, Streamlined infrastructure, and ensured enterprise-grade security and governance.

- Engineered and deployed 20+ scalable ETL pipelines using AWS Glue (PySpark), Lambda, and Redshift, processing over 500M+ records/month from ERP, CRM, and external APIs.
- Orchestrated data workflows with Apache Airflow (MWAA) and AWS Step Functions, reducing pipeline failures by 30% and improving SLA compliance by 40%.
- Assembled a modular data lake architecture on Amazon S3, integrating Glue Catalog, Athena, and partitioned data zones (raw, processed, curated) to streamline analytics and reduce costs.
- Streamlined infrastructure provisioning using AWS CDK and CloudFormation, accelerating deployment cycles by 60% and maintaining consistency across dev, QA, and prod environments.
- Authored 15+ reusable Python modules for data validation, enrichment, and transformation using Pandas and NumPy, cutting processing time by 35%.
- Engineered high-performance Redshift models and SQL scripts, enabling portfolio trend reports, cash flow analysis, and financial KPIs across departments.
- Integrated data from 10+ disparate sources (API, CSV, JSON, legacy RDBMS) using Glue and Informatica, ensuring 99.9% data accuracy in downstream systems.
- Implemented enterprise-grade data security with IAM, S3 encryption (KMS), and data masking policies aligned with SOX, BCBS 239, and internal audit standards.
- Enabled self-service BI via Amazon QuickSight and Power BI, providing access to financial reports and operational dashboards for business users.

Tools & Technologies: AWS Glue (PySpark), Amazon Redshift, S3, AWS Lambda, Step Functions, MWAA (Airflow), AWS CDK, CloudFormation, Python, SQL, Pandas, NumPy, Informatica, Athena, Tableau, QuickSight, IAM, KMS

Northern Trust – Chicago, USA

Data Engineer ***May 2023 – June 2024***

Description: Simulated an enterprise data engineering role by architecting a modern Azure-based Data Mesh platform inspired by Northern Trust. Designed and deployed production-grade pipelines, automated infrastructure, and ensured enterprise-level security, enabling analytics at scale for financial insights and regulatory compliance.

- Built and managed 15+ scalable data pipelines using Azure Data Factory and Databricks, processing over 400M+ records/month from APIs, Excel files, and SQL sources into ADLS Gen2 with structured bronze, silver, and gold layers.
- Crafted 12+ PySpark notebooks in Azure Databricks for transformation, feature engineering, and enrichment, reducing data processing latency by 45% for analytics teams.
- Provisioned infrastructure using Terraform and automated CI/CD with Azure DevOps, decreasing environment setup time by 60% and ensuring consistent deployments across dev, test, and prod.
- Integrated Snowflake with Azure for warehousing and analytical modeling, delivering 10+ BI-ready datasets with sub-second query performance and supporting executive dashboards.
- Acted as Databricks workspace admin, configuring 5 production workspaces, managing cluster utilization, enforcing access controls with Azure Key Vault, and enabling detailed audit logging via Azure Monitor.
- Enabled self-service analytics using Power BI, publishing 5+ dashboards connected to structured Snowflake models, which reduced ad hoc reporting requests by 40% and improved time-to-insight.

Tools & Technologies: Azure Data Factory, Azure Databricks (PySpark), ADLS Gen2, Snowflake, Azure Synapse Analytics, Terraform, Azure DevOps, Power BI, Azure Monitor, Key Vault, Python, SQL

Aviva India – Mumbai, India

Data Engineer ***July 2020 - June 2022***

Description: Modeled a real-time enterprise data engineering project at Aviva India focused on building a Big Data platform for scalable claims processing, customer insights, and compliance reporting. Leveraged open-source technologies within the Hadoop ecosystem to design robust data pipelines, automate workflows, and enable efficient analytics across millions of insurance records.

- Constructed and implemented a scalable data lake on HDFS, ingesting and managing 100M+ policy and claims records/month across raw, refined, and refined layers for regulatory and business analytics.
- Developed 20+ PySpark transformation jobs to clean, enrich, and validate insurance data, improving data processing efficiency by 60% and supporting downstream risk and underwriting analysis.
- Created 15+ Hive tables over structured HDFS data with optimized partitioning and bucketing, enabling sub-3 second SQL query response times for analytics and compliance reporting.
- Systematized daily data workflows using Apache Airflow, orchestrating 10+ pipelines with SLA tracking, reducing manual intervention by 70% and increasing reliability of reporting.
- Delivered 5+ Power BI dashboards by connecting Hive datasets via ODBC, supporting visual insights into claim approval rates, policy lapse trends, and fraud detection patterns.

Tools & Technologies: Hadoop, HDFS, Apache Spark (PySpark), Apache Hive, Apache Airflow, Apache NiFi, Python, SQL, Power BI, Git

EDUCATION

Illinois Institute of Technology, Chicago, IL **August 2022 - May 2024**

M.S. Information Technology