

EDUCATION

Master of Science in Computer Science, University of Illinois Chicago	Aug 2024 - Dec 2025
Bachelor of Engineering in Computer Science, RV College of Engineering	Aug 2016 - May 2020

TECHNICAL SKILLS

Programming Languages: Java, SQL, Python, Scala, C++, Bash/Shell scripting

Data Engineering: ETL/ELT Data pipeline, Data modeling, Schema evolution, Data integration, Data Validation, Data Quality Data orchestration, Data Lakes, Big Data Processing, Data Governance, Data Catalog, Visualization, Security

Distributed Systems: Apache Spark, Hadoop MapReduce, Apache Flink, Apache Kafka, Hive, Airflow, gRPC

Technologies & Tools: AWS, GCP, Azure, UNIX, Git, GitHub, CI/CD, Agile, Docker, Kubernetes, Terraform, Tableau, dbt

CERTIFICATIONS

AWS Certified Data Engineer – Associate Jul 2025

Skills: AWS EC2, Lambda, Step Functions, S3, EBS, Glue, EMR, ECS, RDS, Athena, EventBridge, IAM, KMS, CloudWatch

EXPERIENCE

Fivetran (Offer automated ELT data pipelines across cloud platforms) Jan 2020 - Aug 2024
Senior Software Engineer Bengaluru, India

- Redesigned and built a new data writer for the **BigQuery** data warehouse destination, aligning with SQL-based data writers, eliminating maintenance overhead by 90%, and streamlining the implementation of new features
- Enhanced the efficiency of the **Warehouse Data Writer** pipeline by 30% through the implementation of multi-threading with workers, enabling concurrent file append and processing of split files
- Optimized the BigQuery data writer using **data partitioning and clustering**, reducing customer BigQuery bills by 90%. This hackathon prize-winning idea showcased significant performance and cost-efficiency improvements
- Engineered support for JSON type in **Redshift**, streamlining data migration and improving query performance by 25%

Software Engineer 2

- Designed and developed a high-performance new **DynamoDB** database source connector, achieving an impressive 15x increase in incremental data retrieval speed compared to the previous version of the connector
- Architected a new incremental sync strategy for NoSQL **MongoDB** database using Change streams, improved the data extraction performance by 5x speed increase and reduced latency during data imports
- Implemented support for new connector **Azure CosmosDB** for MongoDB API, enabling efficient data connectivity

Software Engineer

- Authored the **hackathon project**, Isolated Endpoint Sync, now widely adopted in **500+ connectors**, adapted by over 10 teams, and integrated into key frameworks such as Priority Sync, Coil, and multithreading frameworks
- Implemented a new **ETL** connector for ADP using **REST API** with supported ERD and documentation

Software Engineering Intern

- Conducted performance benchmarking of the end-to-end ETL/ELT data pipeline using the **Snowflake** warehouse
- Implemented **Webhooks** for the Recharge connector which increased the extract speed by 10-fold

PROJECT

AWS vs GCP Data Pipeline: Comparative Analysis of Real-Time Data Streaming Jan 2025

- Designed scalable real-time data pipelines on AWS and GCP to compare performance and cost-efficiency
- Leveraged AWS (EC2, Kinesis, Lambda, DynamoDB) and GCP (Compute Engine, Pub/Sub, Cloud Functions, Bigtable)
- Conducted benchmarking and sustainability analysis focusing on throughput, latency, and resource utilization

Massively Distributed Parallel LLM System using Apache Spark, Hadoop, AWS EMR, and Bedrock Aug 2024

- Built a scalable LLM pipeline using Apache Hadoop and Spark on AWS EMR, processing large datasets stored in S3
- Deployed the final model using AWS Bedrock and serverless architecture with Lambda, API Gateway, and Akka HTTP
- Enabled real-time client interaction with a distributed microservices architecture optimized for low-latency inference