

Priyanka Ponnamm
Data Engineer | Cloud AI Engineer
Email: priyanka240630@gmail.com
Phone: 716-452-7552



- 10+ years of experience with a solid understanding of **Business Requirement Gathering, Business Process Flow, Business Process Modeling, and Business Analysis.**
- I am certified at **Professional Data Engineer** ([Certification Link](#))
- Having good amount of Hands-on experience in **AWS (Amazon Web Services) Cloud**
- Having Knowledge on **AWS S3, Lambda, RDS, Redshift, Glue, EMR, EC2, Athena, IAM, Step Functions, Quick Sight.**
- Leveraged **Apache Iceberg** and **Apache Hudi** to manage large-scale data lakes, enabling efficient querying and versioning for big data workflows.
- Utilized **AWS Database Migration Service (DMS)** to migrate and replicate data across heterogeneous databases seamlessly.
- Built and optimized batch and streaming data pipelines with **AWS Glue, Amazon S3, Amazon Redshift, and Apache Spark.**
- Deployed and managed **Snowflake** data warehousing solutions for scalable analytics across cloud platforms.
- Proficient in **Live Video Production**, including operating video switchers and managing content flow between multiple camera inputs, presentation decks, and display systems.
- **Managed professional audio systems for production**, including digital audio consoles, wireless microphone systems, and mixing for both live audience and recording feeds.
- **Automated the deployment and management of cloud infrastructure** using **Terraform**, enabling consistent, repeatable, and scalable infrastructure provisioning.
- Monitored and tracked pipeline performance using **Amazon CloudWatch** and **AWS Lambda** for real-time insights and alerts.
- Collaborated with security teams to enforce data compliance and implemented safeguarding measures using **IAM** and **AWS KMS.**
- Utilized **AWS Bedrock** to build and deploy generative AI models, enabling enhanced capabilities for natural language processing, text generation, and machine learning applications.
- Leveraged **PySpark** and **SQL** to implement ETL pipelines, ensuring seamless data integration, transformation, and loading across diverse platforms.
- Designed and implemented **CI/CD pipelines** using **GitLab** and **Jenkins**, automating build, test, and deployment processes for faster and more reliable software delivery.
- Proficient in deploying and managing **Snowflake data warehousing solutions on AWS and GCP.**
- Having Knowledge in GCP services like **Cloud Storage, BigQuery, Cloud Composer, Dataflow, DataProc, Cloud Functions, Vertex AI, Document AI, AlloyDB, Cloud Monitoring, Cloud Logging, Google Kubernetes Engine (GKE), DLP, Pub/Sub.**
- Proficient in building batch and streaming data pipelines using **big data technologies and GCP cloud services.**
- Utilized **Gemini models like Gemini 1.5 Pro and Gemini 1.5 Flash** to enhance machine learning workflows, improving accuracy and efficiency in tasks like **text classification, translation, and semantic understanding.**

Priyanka Ponnamm
Data Engineer | Cloud AI Engineer
Email: priyanka240630@gmail.com
Phone: 716-452-7552

Technical Skills

- **Generative AI & LLM Orchestration:** Agno, LangChain, LlamaIndex, Haystack, CrewAI, AutoGen, PromptLayer.
- **Large Language Models (LLMs):** Gemini (1.5 Pro/Flash), GPT-4/3.5, Amazon Titan, Claude (Anthropic), Llama 2/3, Falcon, Mixtral, DeepSeek, Cohere, BLOOM.
- **Vector Databases & Retrieval:** FAISS, ChromaDB, Pinecone, Weaviate, Milvus, OpenSearch, SentenceTransformers, Hugging Face Embeddings.
- **Cloud Platforms & AI Services:** AWS (Bedrock, SageMaker, Kendra, Lex), GCP (Vertex AI, Document AI, Model Garden).
- **Data Engineering & Big Data:** PySpark, Apache Spark (**Java/Scala UDFs**), Spark SQL, Apache Iceberg, Apache Hudi, Hadoop (HDFS, Hive, Kafka), dbt.
- **Data Warehousing:** Amazon Redshift, Google BigQuery, Snowflake, AlloyDB.
- **Scripting & Languages:** Python, SQL, Java (Core, Spring Boot), Shell, R, Scala API.
- **Infrastructure & DevOps:** Terraform, AWS CloudFormation, Docker, Kubernetes (AWS EKS, GKE, Rancher), Helm Charts, Jenkins, GitLab CI/CD.
- **Orchestration & Workflow:** Apache Airflow, Google Cloud Composer, AWS Step Functions.
- **Data Analytics & Visualization:** Amazon QuickSight, Power BI, Tableau, Looker, Google Data Studio, Matplotlib, Seaborn.
- **Data Science Packages:** NumPy, Pandas, Scikit-learn, TensorFlow, PyTorch, Spacy, Beautiful Soup, Selenium.

Professional Experience:

Role: Cloud AI Engineer

Feb 2024 - Present

Client: JPMorgan Chase

About Project: As a Data Enable team, our responsibility is to build and maintain batch and stream data pipelines, analyze data, and create comprehensive reports and interactive dashboards. We leverage AWS services such as S3, Lambda, Redshift, and QuickSight to ensure efficient data processing and insightful analysis. Additionally, we focus on optimizing data workflows and enhancing data accessibility for informed decision-making across the organization.

Responsibilities:

- Utilized Agile methodology for software development, encompassing **problem definition, requirement gathering, iterative development, business modeling, and communication with technical teams.**
- Designed and implemented ETL workflows using **Talend** to extract, transform, and load large datasets from multiple sources into data warehouses, improving data availability and accuracy for analytics.
- **Architected agentic workflows** using the **Agno framework** to orchestrate multi-model LLM interactions, enabling autonomous decision-making in complex data transformation tasks.
- Built notification systems using **AWS SNS**, integrating with other AWS services to send **alerts, updates, and operational data to end-users or systems.**

Priyanka Ponnamp
Data Engineer | Cloud AI Engineer
Email: priyanka240630@gmail.com
Phone: 716-452-7552

- Optimized **AWS SQS message retention, dead-letter queues, and visibility timeouts** to improve system reliability and ensure high availability of services.
- Designed and deployed RESTful APIs using **Java Spring Boot** to serve machine learning inference endpoints, ensuring secure and scalable access to GenAI models.
- Used **terraform** to deploy **AWS services**, developed terraform code for **modules and resources**.
- **Developed specialized AI Agents** with Agno to automate the monitoring of **CI/CD pipelines**, reducing manual intervention for deployment bottleneck identification by 40%.
- Optimized memory management and garbage collection tuning for **Spark** jobs running on **Java/Scala** runtimes (JVM) to handle terabyte-scale datasets on AWS EMR.
- Designed and implemented **ETL workflows** using **Talend** to extract, transform, and load large datasets, including financial data from **SAP ERP systems**, into data warehouses, improving data availability and accuracy for analytics.
- Collaborated with ML engineers and data scientists to build AI-powered solutions, including conversational chatbots, **RAG** pipelines, and embedding workflows using **AWS GenAI** tools such as **Amazon Bedrock, Titan models, SageMaker JumpStart**, and **SageMaker Pipelines**.
- Designed **Retrieval-Augmented Generation (RAG)** frameworks combining **Amazon Kendra** for semantic search with **Bedrock-hosted** large language models (LLMs) for dynamic response generation.
- Integrated **Splunk** into **GenAI** document retrieval pipelines to monitor vector ingestion, API latency, and semantic search performance, enabling root-cause analysis of **LLM** response delays at different stages of vector lookup.
- Engineered and managed scalable data pipelines that power **RAG-based conversational AI**, delivering real-time contextual responses enriched with external knowledge.
- Built chatbot solutions combining **Amazon Lex**, **Comprehend**, and **Bedrock** to deliver intelligent, multilingual, and context-aware conversational experiences.
- Developed embedding pipelines and vector feature stores using **SageMaker with FAISS**, optimizing semantic search performance for structured and unstructured datasets.
- Architected and implemented data pipelines supporting **GenAI** and **ML platforms**, enabling seamless integration of generative models into business workflows.
- Designed and implemented **comprehensive reports and interactive dashboards** utilizing **AWS S3, Lambda, Athena, and QuickSight** to visualize metrics, usage patterns, trends, and user behaviors.
- Leveraged **AWS DMS (Database Migration Service)** for seamless migration of on-premises databases to **AWS RDS** and **Amazon Redshift**, ensuring data accuracy and consistency.
- Integrated **AWS ECS** with **AWS RDS** and **S3** for seamless data storage and retrieval, ensuring high availability and optimized resource utilization.
- Developed data processing solutions using **PySpark** and **Spark SQL** to enable faster testing, transformation, and analysis of large datasets.
- Built Spark applications to **consume streaming data from Apache Kafka**, supporting real-time processing workflows.
- Utilized **Spark and Spark SQL** to read Parquet files and dynamically create Hive tables using the **Scala API** for optimized storage and querying.
- Designed and deployed **containerized applications** using **Amazon ECS**, ensuring scalability and efficient resource utilization.

Priyanka Ponnamm
Data Engineer | Cloud AI Engineer
Email: priyanka240630@gmail.com
Phone: 716-452-7552



Role: Data Engineer
Client: Bank of America

Sep 2021 - Feb 2023

About Project: As an Enterprise Data Enable Analytics team member, I create batch and streaming data pipelines using cloud services GCP. I utilize GCP services like BigQuery, Vertex AI, and AlloyDB for data processing and management. I build monitoring solutions to track pipeline and application performance using tools like Google Cloud Monitoring and Logging. My expertise includes optimizing data models in Power BI and writing complex SQL queries for data analysis.

Responsibilities:

- Collaborated on building reliable big data infrastructure using **Spark and container services**.
- Developed and optimized data processing pipelines in **Spark** for analytics.
- Deployed and managed **Snowflake data** warehousing solutions on **GCP Cloud**. Designed optimized data models in Snowflake for scalability.
- Utilized GCP services like **BigQuery, Dataflow, Pub/Sub, and Cloud Storage** for data processing. Integrated GCP services for seamless data storage, retrieval, and real-time processing.
- Proficient in **Databricks** for data engineering tasks and workflow optimization.
- Implemented data transformations and models using DBT for structured datasets.
- Used **Terraform** for deploying GCP services and infrastructure.
- Implemented **CI/CD** pipelines for automated deployment processes.
- Collaborated with security teams for data policy compliance and retention models. Implemented data safeguarding measures using **GCP's Data Loss Prevention (DLP)**.
- Centralized and analyzed logs using **Google Cloud Logging**.
- Implemented monitoring solutions for GCP services using **CloudWatch** and relevant tools.
- Extensive knowledge in **Vector Embeddings and Generative AI Data Engineer Services, including Document AI, Vector AI, Cloud Storage, and AlloyDB**.
- Utilized GCP services like **Document AI** with different parsers for extracting data from PDF files.
- Leveraged **Document AI**'s pre-trained models for automatic text extraction, classification, and entity recognition, enabling rapid processing of large volumes of documents.
- Used **Vector AI** with **embedding models** to generate embedding vectors for the extracted text.
- Integrated **Vertex AI** with other GCP services like **BigQuery** and **Cloud Storage** for seamless data processing and analysis, creating powerful AI-driven solutions for business intelligence.
- Integrated **AlloyDB** for loading embedding vectors for faster retrieval and data processing.
- Having Good knowledge in **Llama Index and Gemini AI** for prompt engineering.
- Used **Cloud functions, Cloud pub-sub and BigQuery** to build streamline dashboard to monitor services.
- Created jobs using **Cloud composer (Airflow DAG)** to migrate data from Data Lake (cloud storage) to transform it using **DataProc and ingest in BigQuery** for further analysis.
- Built batch and streaming jobs using **GCP Services like BigQuery, Pub/Sub, DataProc, Dataflow, CloudRun, Compute engine and Cloud Composer**.

Priyanka Ponnamp
Data Engineer | Cloud AI Engineer
Email: priyanka240630@gmail.com
Phone: 716-452-7552



Role: Big Data Engineer

Sep 2017 - Aug 2021

Client: S&P Global Market Intelligence

About Project: As a Platform Enhancement team, we are responsible to build self-service applications using cloud services. We also built self-services batch and streaming pipelines.

Responsibilities:

- Collaborated with business, analytical teams, and data scientist to improve efficiency, increase the applicability of predictive models, and help translate ad-hoc analyses into scalable data delivery solutions.
- Collaborated with DevOps team to integrate innovations and algorithms into a production system.
- Worked with the DevOps team to create and manage deployment workflows for all scripts and code.
- Develop and maintained scalable data pipelines that will ingest, transform, and distribute data streams and batches within the **AWS S3** and **Snowflake** using **AWS Step Function**, **AWS Lambda**, **AWS Kinesis**, **AWS Glue** and **AWS EMR**.
- Created batch pipelines using **AWS S3**, **AWS Lambda**, **AWS Glue**, **AWS EMR**, **AWS Athena**, **AWS RedShift**, **AWS RDS** etc.,
- Used **AWS** services like **AWS S3**, **AWS RDS** and **AWS Redshift** for storage.
- Created **Apache Airflow** dags to ingest data from sources like API's, Servers, and Databases to transform using **PySpark** in **Glue** and **EMR** and loaded data into Data Warehouse like **AWS RedShift**.
- Created Data pipelines using **AWS services** like **S3**, **Glue**, **EMR**, **Lambda**, **Athena**, **IAM**, etc.,
- Created reports and dashboards that provide information on metrics, usage, trends, and behaviors using **AWS services** like **S3**, **Lambda**, **Athena** and **QuickSight**.
- Orchestrated pipelines and dataflow using **Apache Airflow** and **Step Function**.
- Created reports and dashboards using **AWS services** like **Lambda**, **Glue**, **Step Function** and **QuickSight**.
- Created monitoring service using **AWS CloudWatch**, **AWS Lambda**, **AWS Glue**, **AWS Step Function**, **Grafana** and **ElasticSearch**.
- Created Airflow dags to extract, transform and load data into Data Warehouse.
- Developed and deployed **Kubernetes** pods to extract, transform and load data.
- Used **Docker** and **Kubernetes** for Data Pipelines and ETL Pipelines.
- Used **Hadoop** ecosystems like **Apache HDFS**, **Apache Spark**, **Apache Hive**, **Apache Airflow** and **Apache Kafka**
- Facilitated the developed and deployed proof-of-concept machine learning systems.

Priyanka Ponnamm
Data Engineer | Cloud AI Engineer
Email: priyanka240630@gmail.com
Phone: 716-452-7552



Role: Data Analyst

Jan 2016 - Aug 2017

Client: HP

About Project: As a Data Analyst, I supported enterprise-wide data initiatives by analyzing large datasets, creating dashboards, and providing insights that helped guide business decisions. I worked closely with business stakeholders, data engineers, and data scientists to ensure data accuracy, improve data pipelines, and promote a data-driven culture across the organization.

Responsibilities:

- Gathered, cleaned, and transformed raw data from multiple sources (databases, APIs, flat files) for analysis.
- Developed and automated reports, dashboards, and visualizations using tools like **Power BI / Tableau / Google Data Studio / Looker**.
- Performed exploratory data analysis (EDA) to identify patterns, trends, and anomalies.
- Wrote complex SQL queries to extract, manipulate, and analyze large datasets from relational databases (e.g., **PostgreSQL, MySQL, Redshift, BigQuery**).
- Collaborated with business users to define KPIs and metrics for performance tracking.
- Built predictive and descriptive models in collaboration with data science teams.
- Conducted A/B testing and statistical analysis to support marketing and product decisions.
- Automated recurring data tasks and reports using **Python, R, or Excel macros**.
- Validated and ensured the **quality, accuracy, and consistency of data**.
- Documented data definitions, business rules, and data lineage.
- Partnered with data engineering teams to design and optimize **data pipelines**.
- Provided **ad-hoc analysis** and insights to senior leadership on demand.
- Trained business users on how to interpret **dashboards** and **reports**.
- Defined and implemented **data governance** practices for the project.
- Worked closely with **stakeholders** to translate business needs into technical data requirements.
- Integrated external data sources (e.g., third-party market data) into internal **reporting frameworks**.
- Created presentations to summarize analytical findings and recommendations.
- Identified process improvement opportunities through data insights and recommended solutions.
- Supported data migration and data cleansing efforts during system upgrades or new system implementations.
- Ensured compliance with data privacy and security standards (**GDPR, CCPA**) during data handling and analysis.

Priyanka Ponnamp
Data Engineer | Cloud AI Engineer
Email: priyanka240630@gmail.com
Phone: 716-452-7552



Role: SQL Developer

Aug 2014 - Dec 2015

Client: Nationwide

About Project: Worked as part of the Business Intelligence and Data Warehousing team to design, develop, and optimize enterprise reporting solutions. The project focused on building efficient ETL processes, dimensional data models, and interactive dashboards to support decision-making. Delivered robust SSIS, SSAS, and SSRS solutions to enhance data accessibility and analytics across the organization.

Responsibilities:

- Experience in developing complex **stored procedures**, efficient triggers, required functions, creating indexes and indexed views for performance.
- Excellent Experience in monitoring **SQL Server Performance tuning in SQL Server**
- Expert in designing **ETL data flows using SSIS**, creating mappings/workflows to extract data from **SQL Server** and Data Migration and Transformation from Access/Excel Sheets using **SQL Server SSIS**.
- Efficient in Dimensional **Data Modeling for Data Mart design**, identifying Facts and Dimensions, and developing fact tables, and dimension tables using Slowly Changing Dimensions (SCD).
- Experience in **Error and Event Handling**: Precedence Constraints, Break Points, Check Points, Logging.
- Experienced in **Building Cubes and Dimensions** with different Architectures and Data Sources for Business Intelligence and writing MDX Scripting.
- Thorough knowledge of **Features, Structure, Attributes, Hierarchies, Star and Snowflake Schemas of Data Marts**.
- Good working knowledge on **Developing SSAS Cubes, Aggregation, KPIs, Measures**, Partitioning Cube, **Data Mining Models and Deploying and Processing SSAS objects**.
- Experience in creating **Ad hoc reports** and reports with complex formulas and querying the database for Business Intelligence.
- Flexible, enthusiastic, and project-oriented team player with excellent written, verbal communication and leadership skills to develop creative solutions for challenging client needs.
- Expertise in developing Parameterized, Chart, Graph, Linked, Dashboard, Scorecards, Report on **SSAS Cube** using Drill-down, Drill-through and Cascading reports using SSRS.

Education:

- Bachelors in Electronics from Osmania University – May 2014
- Masters in Information Technology & Management from St Francis College, Brooklyn, NY – May 2025