

Priyank Mishra

Solutions Architect | AWS, Snowflake and Databricks Data Platforms

Bangalore, Karnataka | +(91)-767-596-5055 | [LinkedIn](#) | mail2priyankmishra@gmail.com

SUMMARY

Certified Solutions Architect with 13+ years designing and delivering scalable AWS, Snowflake and Databricks data platforms. I own end to end architecture: reference architectures, HLD and LLD, dimensional and Data Vault modeling, governance and migration, building reliable, secure and cost aware lakehouse platforms. Strong in modern patterns (Lambda, Kappa, medallion lakehouse, data mesh, event driven), ELT, CDC and streaming, with hands on DevOps and DataOps across CI/CD, IaC, data quality, observability and lineage. Currently building AI ready data foundations and GenAI solution architecture: RAG, LLM applications, vector search, AWS Bedrock and Snowflake Cortex, with MLOps and AI data governance.

CORE SKILLS

Cloud (AWS): Glue, Lambda, EMR, S3, Redshift, Redshift Spectrum, Athena, Kinesis, DMS, SNS, SQS, Step Functions, Lake Formation, Glue Data Catalog, API Gateway, IAM, KMS, CloudWatch, Parameter Store

Data Platforms: Snowflake (Snowpark, Streams, Tasks, Dynamic Tables, Shares, RBAC, masking, performance tuning), Databricks (Delta Lake, Delta Live Tables, Unity Catalog, Auto Loader, Photon, PySpark)

AI / GenAI: LLM and RAG architecture, vector databases and embeddings, AWS Bedrock, Snowflake Cortex, Databricks Mosaic AI, MLflow, MLOps and LLMOps, AI data governance, data quality for AI/ML

Architecture Patterns: Lambda architecture, Kappa architecture, Medallion (bronze, silver, gold), Lakehouse, Data Lake, Data Mesh, Event driven, Microservices, CDC, batch and streaming unification

Data Modeling: Dimensional modeling (Kimball, star and snowflake schema), Data Vault 2.0, Inmon 3NF, Slowly Changing Dimensions

Data Engineering: PySpark, Spark, dbt, Airflow, Dagster, Kafka, Kinesis, Flink, ELT and ETL, incremental loads, schema evolution, partitioning, query and cost performance tuning

DataOps: CI/CD for data, automated testing and data quality gates, pipeline observability, monitoring and alerting, data lineage, orchestration, version controlled data products

DevOps and IaC: Terraform, CloudFormation, Docker, Jenkins, GitHub Actions, CI/CD pipelines

Governance and Quality: Data catalog, lineage, RBAC, PII masking, data contracts, Great Expectations, Deequ, data quality SLAs, observability

Languages and Frameworks: Python, SQL, Scala, Java, Shell; FastAPI, Flask, Django

Architecture Practice: HLD and LLD, ADRs, migration strategy, cost optimization (FinOps), security and access control, presales and estimation, stakeholder management

CERTIFICATIONS

Snowflake: SnowPro Core, Data Engineer, Solution Architect

AWS: Solutions Architect Associate, Developer Associate, Data Engineer Associate

PROFESSIONAL EXPERIENCE

phData | Solutions Architect

02.2024 - Present

Waste Management (Snowflake + AWS) Stack: Snowflake, dbt, Python, AWS (Lambda, S3, Glue, SNS, SQS, Parameter Store), Airflow, Terraform, GitHub Actions, Docker, Great Expectations

- Architected an end to end Snowflake data platform on AWS using dbt, Lambda and Airflow, applying a Lambda style architecture for both batch and real time ingestion, improving processing efficiency and cost.
- Defined the platform reference architecture (HLD and LLD), dimensional data models and a governance model, delivering governed, AI ready self serve data products to analytics and reporting teams.
- Built a DataOps workflow with CI/CD, automated data quality gates (Great Expectations) ensuring trusted input quality for ML and GenAI, dbt tests, schema evolution checks and pipeline observability.
- Led infrastructure automation (IaC) with Terraform and GitHub Actions, provisioning Snowflake roles, S3 buckets and Lambda functions for consistent, secure deployments.
- Orchestrated scalable ELT pipelines with Airflow and dbt incremental models, reducing load time by 60%, and implemented RBAC, masking and lineage governance in Snowflake as the access control foundation for AI workloads.

Enterprise Lakehouse Modernization (Databricks + AWS) Stack: Databricks on AWS, Delta Lake, Delta Live Tables, Unity Catalog, AWS Glue, S3, Auto Loader, PySpark, Airflow, Terraform

- Designed and built a medallion lakehouse on Databricks on AWS using Delta Lake and Delta Live Tables for ACID transactions and incrementally processed, versioned data.

- Automated ingestion from multiple sources into S3 and Delta using AWS Glue and Auto Loader on SLA based schedules, applying partitioning and filter pushdown to cut latency and compute cost.
- Enforced fine grained access, lineage and data quality via Unity Catalog, with validations wired as DataOps gates, and automated jobs with Terraform IaC, CI/CD and observability.

Publicis Sapient | Data Engineering Lead

06.2021 - 02.2024

Lloyds Bank (Databricks + AWS) Stack: Databricks, Delta Lake, AWS (EMR, S3, Kinesis), Kafka, Scala, PySpark, Airflow

- Designed a real time streaming ingestion architecture using Kafka, Kinesis and Spark Structured Streaming on Databricks on AWS, enabling low latency processing of high volume transactional banking data.
- Built Scala and PySpark ELT pipelines into a Delta Lake medallion lakehouse, modeling curated layers for critical business reporting.
- Architected fault tolerant, idempotent pipelines with retries and recovery, ensuring 99.9% uptime, and optimized Spark jobs for cost and performance.
- Operated pipelines with CI/CD, monitoring and alerting (DataOps), fine grained access controls and banking audit compliance, using S3 as a data lake for lineage and historical backfills.

Albertsons (Snowflake + AWS) Stack: Snowflake, dbt, Airflow, Terraform, AWS (S3, Lambda, Glue)

- Led migration of legacy ETL workflows to a scalable Snowflake architecture on AWS, significantly improving performance and maintainability.
- Built modular dbt models with incremental logic, tests and documentation, and orchestrated ELT pipelines with Airflow using retries, monitoring and alerting as a DataOps practice.
- Automated infrastructure provisioning (IaC) with Terraform for Snowflake roles, warehouses, dbt jobs and Airflow, and produced HLD and LLD documentation to align stakeholders.

Pitney Bowes | Data Engineering Lead

04.2017 - 04.2021

Presort (AWS, with Snowflake) Stack: AWS (EMR, Glue, Athena, S3, Glue Data Catalog, Lake Formation, Step Functions, Lambda, IAM, KMS), PySpark, Parquet, Snowflake

- Designed PySpark ETL pipelines on AWS EMR using an S3 data lake with raw, staging and curated zones (medallion layout), converting sources to partitioned, compressed Parquet to cut query cost.
- Orchestrated workflows with AWS Step Functions and Lambda with retries and failure notifications, and implemented Glue ETL with bookmarks for incremental loads and CDC plus Glue Crawlers for schema evolution.
- Served curated datasets to Athena and Snowflake for reporting, and enforced governance and access control via Lake Formation, IAM and KMS encryption.
- Integrated Deequ data quality checks as a DataOps gate and documented assets in the Glue Data Catalog for lineage and discovery, with operational runbooks for on-call and backfill.

Commerce Cloud (AWS, with Snowflake) Stack: AWS (S3, Lambda, API Gateway, CloudWatch, IAM), Python, Flask, Jenkins, Snowflake

- Designed data preparation and curation pipelines on an S3 data lake and served curated datasets to Snowflake for analytics and reporting consumers.
- Developed and containerized Flask APIs to serve real time data services, secured via API Gateway and IAM policies for controlled access.
- Built automated CI/CD pipelines with Jenkins and Lambda triggers for automated data refresh (DevOps and DataOps), with CloudWatch monitoring and alerting for service and pipeline SLAs.

Cognizant Technologies | Senior Data Engineer

07.2016 - 03.2017

Unilever (AWS) Stack: AWS EMR, Redshift, S3, Hadoop, Hive, Sqoop, Oozie

- Led migration of Hadoop Hive pipelines to AWS Redshift, improving query performance and simplifying the reporting architecture.
- Rewrote legacy Hive transformations using optimized SQL and Redshift data models, and tuned Sqoop ingestion with parallelism and compression.
- Operated ETL workflows with Oozie scheduling, retries, monitoring and alerting (DataOps), with version controlled migration scripts.

Hewlett Packard Enterprise | Data Engineer

07.2015 - 07.2016

Avaya Stack: Python, Pandas, Teradata, Tableau

- Built Python and Pandas ETL pipelines over Teradata customer analytics data and Tableau dashboards for sales and support KPIs.

Wipro Infotech | Project Engineer

01.2013 - 06.2015

Kellogg's Stack: Hadoop, Hive, Python, Shell

- Built Hive ETL pipelines on Hadoop for large scale sales data, automating ingestion and validation with Python and Shell.

EDUCATION

B. Tech, Dr. M. G. R. Educational and Research Institute University, 2008 - 2012