

Hemasree Randhi

Senior Data Engineer

Mobile: (940) 252-4727 | Mail: hemasree9767@gmail.com | Austin, Tx | LinkedIn: [linkedin.com/in/hemasreerandhi](https://www.linkedin.com/in/hemasreerandhi)

PROFESSIONAL SUMMARY

Senior Data Engineer with over 6 years of experience designing and building scalable cloud data platforms, ETL/ELT pipelines, and enterprise data solutions across telecom and financial services. Strong expertise in Azure Data Factory, Azure Databricks, ADLS Gen2, PySpark, Spark SQL, Python, SQL, Delta Lake, Snowflake, Oracle, Azure Synapse Analytics, and Azure SQL Database. Experienced in building end-to-end data pipelines for high-volume ingestion, transformation, cross-system reconciliation, data quality validation, and analytics-ready data delivery. Hands-on experience developing complex reconciliation frameworks, distributed Spark workloads, incremental processing, and automated data-quality controls across enterprise datasets. Skilled in Spark performance optimization, workflow orchestration, CI/CD, monitoring, UAT, and production deployments across multiple environments. Proven ability to translate complex business and data requirements into reliable, maintainable solutions while collaborating with engineering, QA, and business teams throughout the full development lifecycle.

PROFESSIONAL EXPERIENCE

Senior Data Engineer | Global Logic (Client: T-Mobile), Dallas June 2026 – Present

- Built an Azure-based telecom reconciliation solution using Azure Data Factory, Azure Databricks, PySpark, ADLS Gen2, Snowflake, and Oracle to compare billing, subscriber, cancellation, and return-order data across multiple systems.
- Developed end-to-end ADF and Databricks pipelines to ingest, transform, reconcile, and process data from Snowflake and Oracle, supporting automated cross-system validation and discrepancy detection.
- Created PySpark and Spark SQL reconciliation rules to compare key attributes such as BAN, subscriber ID, cancellation date, and order details, identifying mismatches and data-quality issues between source systems.
- Built matching and fallback logic for incomplete or inconsistent subscriber data using multi-table lookups, null handling, historical date windows, return-order filtering, and order-type standardization.
- Designed Delta Lake-based case management logic to create cases for new discrepancies, update recurring failures, automatically close resolved cases, and maintain a history of processed records.

- Developed reusable Databricks notebooks for data ingestion, transformation, reconciliation rules, Delta processing, exception handling, and unit testing, making the solution easier to maintain and extend.
- Established secure Databricks-to-Snowflake connectivity across DEV, QAT, PPD, and PROD using OAuth/service-account authentication and environment-specific configurations.
- Performed data-quality and reconciliation validation through schema checks, null and duplicate analysis, referential-integrity checks, source-to-target comparisons, and detailed result validation during development and UAT.
- Worked closely with Data Engineers, technical leads, and business teams to clarify reconciliation requirements, troubleshoot data issues, refine business rules, review code, and support UAT.
- Supported releases across DEV, PPD, and PROD using GitLab and Git-based CI/CD workflows, handling feature development, rebasing, merge requests, testing, deployment, and post-deployment validation.

Tech Stack: Azure Data Factory (ADF), Azure Databricks, PySpark, Apache Spark, Spark SQL, Python, SQL, Delta Lake, Snowflake, Oracle, GitLab, Git, CI/CD, OAuth.

Data Engineer | Fiserv, Dallas June 2023 – May 2026

- Designed and developed complex data ingestion and transformation pipelines on Microsoft Azure Data Platform using Azure Data Factory, Azure Databricks, Azure Data Lake Storage Gen2, Azure Synapse Analytics, Azure SQL Database, PySpark, and Spark SQL to process high-volume financial data from settlement systems, merchant portals, and clearinghouses.
- Built scalable batch and near real-time ingestion frameworks to load structured and semi-structured financial datasets from relational systems, APIs, and flat files into ADLS Gen2, improving source onboarding, reliability, and operational consistency.
- Developed distributed transformation workflows in PySpark and Spark SQL to standardize transaction records, perform currency conversions, normalize timestamps, map merchant categories, and generate trusted downstream datasets for reporting and analytics.
- Implemented schema validation, null checks, duplicate detection, referential integrity validation, and reconciliation controls to improve data quality, processing accuracy, and trust in finance and audit reporting.
- Built reconciliation pipelines to compare internal transaction data with clearinghouse settlement records, identify mismatches, and automate exception reporting, replacing manual spreadsheet-based audit workflows.

- Calculated settlement amounts, net fees, and transaction-level revenue impact using advanced SQL programming, curated transformation logic, and Azure-based reporting layers to support business and finance operations.
- Optimized large-scale joins, aggregations, and Spark workloads using partition pruning, caching, data partitioning, and query tuning techniques, improving performance of financial data pipelines and downstream reporting.
- Developed curated reporting datasets and analytics-ready serving layers in Azure Synapse Analytics and Azure SQL Database to support transaction reporting, settlement summaries, fee analysis, and operational finance dashboards.
- Automated workflow orchestration, dependency handling, retry logic, and SLA-based execution using Azure Data Factory, improving resiliency and end-to-end control over enterprise data pipelines.
- Implemented monitoring and alerting using Azure Monitor, operational logs, and ETL health checks to proactively detect pipeline failures, delayed feeds, reconciliation breaks, and data anomalies.
- Partnered with finance teams, analysts, and business stakeholders to translate settlement, reporting, and audit requirements into scalable Azure-based technical solutions, source-to-target mappings, and curated data models.
- Supported data modernization and enterprise reporting initiatives by designing maintainable cloud data pipelines, improving data accessibility, consistency, governance, and business visibility.
- Provided technical guidance on data pipeline design, validation controls, and performance tuning, while contributing to best practices across Azure-based financial data engineering workflows.

Tech Stack: Azure Data Factory, Azure Databricks, Azure Data Lake Storage Gen2, Azure Synapse Analytics, Azure SQL Database, Azure Monitor, PySpark, Spark SQL, Python, SQL, Delta Lake, Terraform, Docker, Jenkins

Junior Data Engineer | Sasken Technologies, India June 2018 – July 2021

- Designed and maintained enterprise ETL/ELT pipelines using Azure Data Factory, Azure Data Lake Storage, PySpark, Spark SQL, Hive, and SQL to ingest, process, and transform large-scale network, billing, and operational datasets for analytics and reporting.
- Built ingestion workflows for CSV, JSON, and Parquet datasets from APIs, FTP sources, and enterprise systems into cloud and big data environments, improving consistency, standardization, and reliability of source data integration.

- Developed transformation pipelines in PySpark to parse and flatten nested JSON logs, standardize timestamps across global regions, and prepare curated datasets for downstream reporting, operational analysis, and business intelligence.
- Applied schema enforcement, datatype checks, null validation, duplicate detection, and record-level filtering to prevent malformed or incomplete records from impacting downstream reporting and analytics.
- Performed data profiling, source-to-target mapping, and transformation rule design to support technical construction of scalable data flows aligned with business and reporting requirements.
- Built quality control and reconciliation frameworks to identify usage anomalies, billing inconsistencies, and invalid records, improving data trust for revenue reporting and operational decision-making.
- Used advanced SQL and Spark-based processing to support large-scale joins, aggregations, filtering, and exception analysis across subscriber usage, billing, and operational data domains.
- Optimized Spark jobs and reporting pipelines using broadcast joins, partition pruning, caching, and workload tuning techniques, reducing runtime and improving data delivery for executive KPI reporting.
- Automated workflow orchestration, retries, dependency management, and scheduled executions using Azure Data Factory and workflow tools to improve operational reliability and batch processing efficiency.
- Supported development of curated reporting datasets for business analysts and reporting teams, enabling visibility into subscriber usage trends, billing behavior, fraud indicators, and churn-related metrics.
- Implemented monitoring, logging, and operational alerting practices to detect failed jobs, abnormal record patterns, delayed processing, and validation issues across production data workflows.
- Worked closely with analysts, reporting teams, and business stakeholders to translate operational and reporting needs into technical solutions, data mappings, and scalable pipeline designs.
- Contributed to architecture discussions, technical documentation, and best-practice adoption for data engineering workflows, while supporting junior team members in debugging, validation, and job troubleshooting.
- Strengthened governance and compliance by supporting secure data handling, controlled access, auditability, and masking of sensitive data across non-production and production environments.

Tech Stack: Azure Data Factory, Azure Data Lake Storage, PySpark, Spark SQL, SQL, Hive, Hadoop, JSON, Parquet, CSV, Python, Azure Monitor, Git, Linux

TECHNICAL SKILLS

Programming Languages: Python, SQL, Spark SQL, Bash, Shell Scripting

Big Data & Distributed Processing: PySpark, Apache Spark, Spark SQL, Azure Databricks, Databricks, Delta Lake, Hive, Hadoop

Microsoft Azure Data Platform: Azure Data Factory, ADLS Gen2, Azure Synapse Analytics, Azure Databricks, Azure SQL Database, Azure Monitor, Azure Storage, Azure Cloud Services

Data Engineering: ETL, ELT, Data Ingestion, Data Transformation, Batch Processing, Near Real-Time Processing, Workflow Orchestration, Data Integration, Incremental Loads, CDC, Merge, Upsert Logic, Metadata-Driven Ingestion

Data Warehousing & Analytics: Data Warehousing, Dimensional Modeling, Fact and Dimension Modeling, Curated Data Layer Development, Data Marts, Reporting Data Models, Advanced SQL Programming, KPI Reporting, Audit Reporting, Operational Reporting

Data Quality & Governance: Schema Validation, Data Profiling, Data Mapping, Source-to-Target Mapping, Null Checks, Duplicate Detection, Referential Integrity Validation, Reconciliation, Audit Logging, Data Quality Frameworks, Data Cataloging

Performance Optimization: Partition Pruning, Caching, Broadcast Joins, Query Optimization, Spark Performance Tuning, Join Optimization, Aggregation Tuning, Partitioning, Compression

Monitoring & Production Support: Azure Monitor, Pipeline Monitoring, ETL Health Checks, Logs-Based Alerting, Job Status Tracking, Incident Response, Troubleshooting, Operational Support

DevOps & Tools: Terraform, Docker, Jenkins, Git, GitHub, Linux

Data Formats: CSV, JSON, XML, Parquet, ORC, Avro

Visualization & Reporting: Power BI, Tableau, Grafana

EDUCATION

- Master of Science in Computer Science - University of Central Oklahoma
- Bachelor of Technology in Computer Science - Andhra University, India