SUDHEEKSHITHA MALEPATI
DATA ANALYST

PROFESSIONAL SUMMARY:

- Over 7 years of experience in designing healthcare-focused data pipelines, leveraging Snowflake and dbt to build scalable, testable analytics workflows. Adept at working with clinical and EHR/EMR datasets to develop insights, KPIs, and dashboards supporting regulatory and operational decisions.
- Extensive experience supporting securities and investment workflows, including portfolio construction, asset allocation modeling, and performance reporting for advisory clients.
- Experienced in Microsoft Azure-based data engineering and analytics workflows using Azure Data Factory, Databricks, SSIS, and MS Power BI for dashboarding and insight delivery.
- Delivered advanced analytics solutions for healthcare operations and clinical workflows, building KPIs, patient insights, and compliance dashboards using Snowflake, SQL, and Power BI.
- Proficient in designing scalable data pipelines using Apache Spark, Hadoop, and MapReduce for processing large volumes of structured and unstructured data.
- Expertise in building data pipelines with Talend, Informatica, SSIS, Azure Data Factory (ADF), and dbt (data build tool) for modular, testable ETL and data integration in Snowflake-based environments.
- Hands-on experience with Databricks for large-scale data processing and machine learning model development using PySpark.
- Advanced proficiency in Python, Pandas, NumPy, and scikit-learn for data manipulation, statistical analysis, and building data science models.
- Expertise in HDFS for distributed storage and Apache Hive for building and managing data warehouses to support BI and reporting solutions.
- Skilled in data ingestion, extraction, and transformation using Apache Sqoop, Kafka, Kafka Streams, and Kafka Connect for real-time data streaming and batch processing.
- Proficient in Spark SQL, Presto, and SQL for querying large datasets and optimizing query performance in distributed environments.
- Led healthcare supply chain analytics projects focused on inventory optimization, shrinkage reduction, and logistics efficiency by leveraging Snowflake data models and Tableau dashboards to deliver actionable clinical and operational insights.
- Extensive experience with containerization and orchestration using Docker, Kubernetes, and AWS Fargate to build and deploy cloud-native applications.
- Led cloud migration initiatives involving legacy systems to cloud-native architectures, including schema mapping and workload optimization across AWS and Azure.
- Experienced in visualizing healthcare data using Tableau, Qlik Sense and Power BI to build interactive dashboards and convey business insights to stakeholders.
- Solid experience with database technologies, including MySQL, MongoDB, Oracle, SQL Server, and CosmosDB for reliable data storage solutions.
- Excellent organizational skills, multitasking ability, and communication capabilities; comfortable working in agile environments and with dot-com clients. Experience managing large box-style datasets and structured file deliveries for regulated industries.

TECHNICAL SKILLS:

Cloud Platforms: AWS (Glue, Redshift, Lake Formation, EMR, S3, MSK, IAM), Azure
Big Data Technologies: Databricks, Apache Spark, Hadoop, MapReduce, HDFS, Apache Hive, Apache Sqoop, Apache Airflow, Apache Beam, Apache Kafka, Kafka Streams, Kafka Connect, Talend, Presto
Data Engineering & Processing: SQL (advanced joins, window functions, stored procedures), Snowflake, Tableau, Python, PySpark, Spark SQL, Pandas, NumPy, ETL Design, Data Modelling, Data Warehousing
Databases: MySQL, MongoDB, Oracle, SQL Server, CosmosDB
Data Transformation & ETL Tools: Informatica, SSIS, Collibra, Azure Data Factory (ADF), Databricks
Containerization & Orchestration: Docker, Kubernetes, AWS Fargate
CI/CD & DevOps: Jenkins, Git, GitHub, BitBucket, Terraform, CI/CD Pipelines
Data Visualization & Reporting: Tableau, Power BI
Monitoring & Logging: ElasticSearch, ELK Stack , Stackdriver, Prometheus, Splunk
Data Modelling: Star Schema, Snowflake Schema, Data Warehousing
API Integration: RESTful Services
Project Management & Collaboration: Agile, Scrum, Kanban, JIRA, Confluence, Slack, Microsoft Teams, MS Visio, Draw.io, ADO.NET
Productivity Tools: MS Office 365 Suite (Excel, Word, Outlook, PowerPoint), MS Office
Certifications: Microsoft Certified: Power BI Data Analyst Associate, Azure Data Fundamentals, Certified ScrumMaster, Bloomberg Certification – Bloomberg Market Concepts

WORK EXPERIENCE:

HCA Health Care – Rahway, NJ,
Data Analyst |
Responsibilities:

- Implemented scalable and secure data storage solutions with Amazon S3, supporting diverse datasets in the cloud.
- Experience in healthcare data migration and analytics, including handling EHR/EMR data, ensuring compliance with regulatory standards (HIPAA, HL7, FHIR), and managing cloud-based data integration, monitoring, and troubleshooting.
- Managed and optimized AWS RDS relational database environments, achieving a 30% reduction in query response time and enhancing system reliability for production workloads.
- Optimized MPP query performance in Amazon Redshift and Hive by applying advanced SQL tuning techniques, partitioning, and distribution strategies.
- Developed and optimized Apache Spark pipelines, enhancing data transformation and analytics processes and reducing processing latency by 35%.
- Developed and optimized AWS Glue Data Catalog for metadata management and automated schema discovery using AWS Glue Crawlers, improving ETL efficiency.
- Developed and deployed AWS Lambda functions for serverless data processing, reducing latency and operational costs.
- Enabled interactive querying across multiple data sources with Presto, improving query performance and scalability.
- Integrated ElasticSearch with log pipelines for real-time indexing and search, enabling efficient debugging and operational monitoring across distributed systems.
- Collaborated with domain analysts to deliver scalable Snowflake datasets for operational and compliance reporting.
- Partnered with product managers and portfolio teams to model advisory data including proposals, securities holdings, asset allocation strategies, and portfolio construction plans aligned to regulatory standards..
- Built performance reporting dashboards using Power BI to track returns, benchmarks, and compliance metrics across structured advisory portfolios.
- Documented data lineage and metadata mapping to support regulatory audits, stakeholder reviews, and internal governance dashboards using Confluence and Visio.
- Supported data reporting pipelines for compliance with federal partner standards including CDC, APHL, and HRSA, ensuring interoperability between state-level health data systems and national surveillance networks.
- Conducted impact analysis and root-cause investigation for clinical and claims datasets, including Medicaid member records and dental claims retrieved through EHR/EMR feeds. Validated CDT codes, service utilization trends, and benefit design logic to support internal claims reconciliation and coverage audit workflows.
- Designed real-time data streaming applications with AWS Kinesis, achieving high throughput and low-latency ingestion for time-sensitive data pipelines.
- Conducted advanced data wrangling and preprocessing with Python, improving data quality and reducing preprocessing time by 40%.
- Interpreted Medicaid claims data and encounter records to support policy evaluation and utilization-based scoring for performance metrics across care delivery programs.
- Collaborated with clinical analysts to model service coverage limits and detect anomalies in benefit expansion scenarios using historical Snowflake datasets.
- Developed queries and dashboards in ElasticSearch to analyze system logs and metrics, enhancing observability and supporting root cause analysis for data pipeline failures.
- Enhanced large-scale data query processing efficiency by 30% using Spark SQL across distributed data systems.
- Containerized and deployed data processing applications using Docker, ensuring reproducibility and scalability across environments.
- Leveraged ElasticSearch within the ELK Stack to support data quality validation, log aggregation, and performance monitoring across AWS-based ETL workflows.
- Orchestrated data workflows and managed resource allocation with Kubernetes, applying cloud infrastructure concepts to scale distributed systems and ensure workload reliability.
- Designed metadata harvesting pipelines using AWS Glue Data Catalog, extracting schema details from AWS Lake Formation, Redshift, EMR, Aurora PostgreSQL, and S3 for improved data governance.
- Monitored data pipelines and cloud infrastructure using AWS CloudWatch, proactively ensuring high availability and troubleshooting issues.

- Streamlined deployment automation using AWS CodePipeline and CodeBuild, integrating data processing pipelines for continuous integration and delivery.
- Integrated LabWare and Clarity LIMS platforms with Snowflake data marts to unify lab results with patient records, improving traceability and analytical accuracy for public health workflows.
- Built clinical and operational dashboards in Tableau and Power BI using Snowflake-derived models to monitor utilization trends, benefit outcomes, and patient progression. Designed executive views for claims optimization scenarios and service delivery gaps.
- Integrated epidemiological data sources with Merlin disease surveillance system to enable real-time tracking of infectious disease outbreaks and automated report generation for public health stakeholders.
- Optimized Hadoop-based data querying using Apache Hive, improving data analysis efficiency by 20% in large-scale systems.
- Translated business value streams into semantic data models and governed Azure data assets using cataloging practices to support analytics in a regulated environment.
- Developed Tableau dashboards to track medical inventory reconciliation and healthcare supply chain fulfilment, utilizing SQL-driven logic and real-time clinical data to improve operational transparency and ensure regulatory compliance.
- Built predictive models and applied statistical analysis techniques in Python, delivering actionable business insights.
- Utilized Pandas and NumPy for efficient data manipulation, enabling 50% faster data preparation and transformation workflows.
- Implemented NoSQL solutions with DynamoDB, reducing query latency for high-velocity applications by 30%.
- Designed, developed, and optimized large-scale ETL workflows using Databricks and PySpark, processing both batch and streaming data for healthcare analytics. Used SQL window functions and stored procedures in Snowflake for post-ingestion transformations and business rule enforcement.
- Developed cloud-native pipelines to support NGS and AMD data processing, enabling large-scale genomic and genetic sequencing analytics in compliance with state bioinformatics standards.
- Built and maintained reusable Python ETL frameworks within Databricks notebooks, enabling consistent PySpark pipeline development across multiple domains.
- Supervised cross-functional agile teams and ensured timely delivery of project deliverables aligned with client expectations and compliance requirements.
- Delivered training and technical assistance to clinical and analytics teams on data platform usage, file transfer protocols, and genomic data interpretation using cloud tools and LIMS integration.
- Implemented fine-grained access control policies in AWS Lake Formation, leveraging IAM roles, encryption, and column-level security to ensure secure data access across AWS Glue, Redshift, and S3, meeting HIPAA and enterprise compliance standards.
- Partnered with compliance and governance teams to enforce data quality rules, lineage standards, and attestation procedures across regulated datasets.
- Participated in lineage validation for patient and claims data workflows, documenting transformation points and manual interventions using Visio and Confluence.
- Coordinated task progress with department supervisor and submitted monthly performance reports to the Contract Manager, documenting sequencing data integration milestones and database updates.
- Created attestation templates and transitioned control responsibilities to operational teams post-project delivery.

Environment: Apache Spark, AWS Glue, AWS Lambda, Amazon S3, AWS RDS, Apache Hive, Presto, AWS CloudWatch, AWS CodePipeline, AWS CodeBuild, DataBricks, Amazon Redshift, HDFS, AWS Kinesis, Python, Spark SQL, Hadoop, Docker, Kubernetes, AWS Lake Formation, MapReduce, Glue Data Catalog, Redshift Spectrum, Apache Beam, Pandas, NumPy, DynamoDB, Databricks, SQL.

Data Analyst |
Responsibilities:

- Developed scalable AWS Lambda functions to automate data processing tasks, enhancing efficiency in cloud-native workflows.
- Developed end-to-end ETL pipelines using PySpark and Databricks, integrated with AWS S3 and Redshift for cloud-native data transformations.
- Applied advanced SQL optimization on MPP platforms including Amazon Redshift and Google BigQuery, improving complex query performance and storage efficiency.
- Used SQL to develop transformation logic, create stored procedures, and perform data quality validations on staging tables before loading into production layers.
- Created data models in Snowflake using dbt for claims eligibility, prior authorization trends, and denial rates, improving payer-provider alignment by 20%.

- Acted as subject matter expert for advisory and transactional data, mapping accounts, positions, instruments, and transaction flows into Snowflake marts using dbt and modular SQL logic.
- Delivered analytics on securities transactions, exposure profiles, and position-level drilldowns using Snowflake SQL and Power BI.
- Partnered with data architects to ensure semantic consistency and lineage traceability across key domains such as client holdings, investment proposals, and transactional activities.
- Modeled intercompany transactions with support for multi-entity reporting and currency conversions, enabling consistent data aggregation and reconciliation across subsidiaries.
- Designed and implemented AWS Step Functions and Google Cloud Workflows to automate multi-step cloud data processes, reducing execution time and improving orchestration efficiency.
- Implemented advanced partitioning and lifecycle policies for Amazon S3 and Google Cloud Storage, reducing query latency and storage costs by optimizing data access patterns.
- Designed and implemented secure data lakes using AWS Lake Formation, enforcing IAM-based access control, encryption, and compliance. Prior experience with Google Cloud Dataplex for multi-cloud governance.
- Cloud Security & Identity Management: Applied strict IAM policies in AWS and Google Cloud IAM, managing role-based data access controls and ensuring secure multi-cloud environments.
- Developed distributed data processing pipelines using Apache Spark on AWS EMR, applying infrastructure concepts like VPC configuration, storage classes, and distributed compute tuning for performance.
- Designed and optimized data models for Snowflake and Google BigQuery, ensuring efficient querying and fast data retrieval across large-scale datasets.
- Managed data ingestion pipelines on Azure using Event Hubs and Synapse Pipelines, delivering real-time insights aligned with agile sprint goals in wealth management engagements.
- Automated ETL workflows using AWS Glue and Google Cloud Dataflow, reducing manual data transformations and improving data pipeline efficiency.
- Managed Apache Airflow for ETL orchestration and automation in AWS, defining DAGs for scheduled transformations and integrating on-prem and cloud systems through modular pipelines.
- Utilized Amazon Athena for querying semi-structured data in S3, improving ad hoc analysis performance for large event-based datasets.
- Utilized data build tool with Snowflake to orchestrate transformation logic, enforce data modeling standards, and enable team-wide version control using Git.
- Optimized query execution using Amazon Athena and Google BigQuery, reducing data retrieval times for analytical queries.
- Data Transformation & Preprocessing: Utilized Python-based data wrangling with Pandas and NumPy, improving data preprocessing efficiency for analytics workloads.
- Designed interactive dashboards using Matplotlib and Google Data Studio, providing real-time data insights for business stakeholders.
- Data Encryption & Security Compliance: Implemented encryption mechanisms in AWS KMS and Google Cloud KMS, ensuring secure data handling and compliance with GDPR and enterprise security policies.
- Implemented cost-saving measures in AWS reducing compute, storage, and query costs through optimized data architecture.
- Containerized and deployed data processing applications using Docker and Kubernetes (EKS on AWS), ensuring scalability and reproducibility.
- Actively collaborated in Agile and Scrum ceremonies, coordinating sprints with onshore stakeholders and offshore delivery teams.
- Established data governance frameworks using AWS Glue Data Catalog and Google Cloud Dataplex, ensuring data integrity, security, and lineage tracking.
- Supported enterprise data governance initiatives using Collibra, tagging critical datasets and stewarding business glossary updates to align with compliance and reporting requirements.
- Modeled consumer and institutional investment flows using economic indicators and Monte Carlo simulations, integrating results into Snowflake for advisory performance analytics.
- Managed MongoDB collections to store semi-structured advisory and client data, enabling rapid iteration on portfolio segmentation and personalized insights delivery.
- Tracked project requirements across the SDLC lifecycle and proactively identified and documented project risks along with mitigation strategies.
- Integrated Concur, Salesforce, and ADP data with cloud-based analytics platforms for unified reporting; incorporated Avalara tax data into financial ETL workflows for compliance and reconciliation.

Environment: AWS Lambda, AWS Step Functions, Amazon S3, AWS Lake Formation, MongoDB, AWS IAM, Apache Spark, Snowflake, Kafka Streams, Kafka Connect, MySQL, Amazon Athena, Python, Jenkins, Matplotlib, Amazon EMR, Amazon RDS, Apache Hadoop.

Data Engineer Analyst |

Responsibilities:

- Developed and maintained scalable data pipelines leveraging Azure Synapse Analytics, optimizing data integration for efficient analytics and reporting.
- Implemented ETL processes using Azure SQL Database to automate data transformation and streamline workflows across business systems.
- Developed ETL workflows to extract, validate, and transform AR/AP, GL, and revenue recognition data; supported audit-readiness and ensured reporting compliance with GAAP standards, including fixed asset reconciliation.
- Developed analytics dashboards for financial products and reconciliations, integrating Snowflake with Power BI to monitor product lifecycle metrics and investment flows.
- Integrated cost accounting logic into ETL pipelines to accurately allocate COGS and support inventory valuation, enabling margin analysis and financial reporting.
- Azure Data Lake Storage (ADLS) was used to manage unstructured data, ensuring effective storage and accessibility for large-scale data processing.
- Managed compliance and auditing practices to meet industry regulations, ensuring data usage and governance adhered to best practices.
- Supported finance teams by aligning Snowflake data marts with NetSuite OneWorld structures; enhanced reconciliation between sub-ledgers and general ledger modules through automated validation rules.
- Configured NetSuite OneWorld environments to support multi-entity finance workflows, including GL, AR/AP, and revenue recognition aligned with GAAP and audit-readiness standards.
- Designed and implemented data validation frameworks to ensure high-quality, accurate data transformations and minimize errors in data processing.
- Created and deployed Power BI dashboards to provide real-time business intelligence and actionable insights for key decision-makers.
- Served as NetSuite Functional Analyst liaison between finance, IT, and supply chain, translating process improvement goals into system configuration requirements and workflow enhancements.
- Collaborated with cross-functional teams to define health chain business requirements and develop actionable Tableau dashboards for warehouse stock monitoring, demand planning, and delivery performance metrics.
- Aligned Snowflake data models with manufacturing operations by integrating work order, BOM, and MRP datasets from ERP platforms, supporting improved forecasting and material planning.
- Collaborated with Manufacturing Engineering teams to map ERP-based BOM and MRP workflows into analytical models; drove process alignment with NetSuite's Manufacturing Process Alignment framework for supply planning and COGS analysis.
- Translated business value streams into semantic data models using SQL, dbt, and Data Vault modeling techniques.
- Created and managed Epics, Features, and User Stories in JIRA, collaborated with technical and business stakeholders through Confluence documentation.
- Authored BRDs, test cases, and Visio-based process flows for finance system enhancements, ensuring traceability from requirements through QA in Agile delivery cycles.
- Developed detailed BRDs and test plans for NetSuite integrations with Salesforce, Concur, ADP, and Avalara, enabling end-to-end finance and tax workflow automation across internal systems.
- Implemented dbt models in Snowflake for healthcare KPIs and claims analytics, including slowly changing dimensions and patient cohort segmentation using modular SQL logic.
- Optimized SQL queries for high-performance data extraction from cloud databases, enhancing the speed and reliability of data pipelines.
- Automated the deployment of data engineering solutions using Azure DevOps, promoting continuous integration and continuous delivery (CI/CD) across environments.
- Coordinated version control using BitBucket, facilitating smooth collaboration and code integration between cross-functional development teams.
- Applied Azure Functions to automate serverless data workflows, reducing infrastructure management overhead and increasing operational efficiency.
- Conducted real-time data processing using Apache Spark, improving processing speed and scalability of big data applications.
- Enforced data security policies across cloud platforms, implementing role-based access controls (RBAC) to restrict unauthorized access to sensitive data.
- Utilized Azure Stream Analytics to process streaming data in real time, providing up-to-date insights for operational decision-making.
- Architected scalable data processing pipelines using PySpark on Databricks, enabling parallel transformation of large datasets and optimized execution through SQL-based transformations and partitioning.
- Designed and implemented data architecture strategies by applying infrastructure concepts such as data lake zones, cloud-native pipelines, and resource provisioning to support complex workflows and scalability.

- Collaborated with business teams to define business intelligence requirements and design solutions that deliver insights from large, diverse datasets.
- Managed data lakes for optimal storage and retrieval of diverse data sets, ensuring fast data processing and compliance with retention policies.
- Performed advanced data processing tasks using Pandas and NumPy, ensuring efficient data transformations and statistical analyses.
- Designed and maintained modular dbt models within a medallion architecture to transform and standardize Medicaid claims, EHR, and dental service utilization data. Incorporated testing and documentation practices to ensure auditability and reproducibility.
- Used PowerShell for automating infrastructure provisioning and managing cloud resources, reducing manual intervention and improving system reliability.
- Designed and implemented data quality frameworks, using automated testing to ensure clean, accurate, and consistent data throughout the pipeline.
- Led cross-functional collaboration with development, data science, and business teams, ensuring alignment between data engineering and business objectives.
- Established data governance practices for metadata management, ensuring data integrity, traceability, and compliance across cloud storage solutions.
- Integrated machine learning models into the data pipeline, using data from Azure SQL Database to generate predictive analytics and insights.
- Enabled real-time analytics by streamlining data flow from multiple sources using Azure Data Lake Storage and processing it with Spark.
- Led requirements-gathering sessions through stakeholder interviews, workshops, and document analysis to capture detailed data engineering and analytics needs.
- Designed & implemented Azure Data Factory pipelines to migrate large datasets from SQL Server to Azure Synapse Analytics, ensuring seamless data transformation & validation using Azure Databricks (PySpark).
- Led requirements-gathering sessions and documented data control procedures using Confluence and JIRA, ensuring transparency in data lineage and rule validation.

Environment: Azure Synapse Analytics, Azure SQL Database, ADLS, Azure Functions, Apache Spark, Power BI, Azure DevOps, BitBucket, Azure Stream Analytics, PySpark, PowerShell, Pandas, NumPy, SQL, Role-Based Access Control (RBAC).

Data Reconciliation, Data Governance, Agile, Hedge Accounting