**Vijay Bhaskar K**

**Phone: +1 (732) 347 6970 | Email ID: vijayabhaskarreddy840@gmail.com**

**---------------------------------------------------------------------------------------------------------------------------  
SUMMARY:**

* Data Engineering professional with 10+ years of experience designing and implementing scalable data pipelines across GCP, AWS, and Azure, with expertise in building ETL/ELT solutions for structured, semi-structured, and unstructured datasets.
* Proficient in leveraging cloud-native tools including GCP (BigQuery, Dataflow, Pub/Sub, Dataproc, GCS), AWS (S3, Glue, Lambda, DMS), and Azure (Data Factory, Synapse, Event Hubs, Databricks) for large-scale data ingestion, processing, and analytics.
* Proven expertise in **data modeling,** including design and implementation of conceptual, logical, and physical models across Snowflake, BigQuery, and Databricks using star and snowflake schemas.
* Hands-on experience using **data modeling tools and frameworks** like dbt (Data Build Tool), Great Expectations, and Datafold for building, testing, and maintaining scalable data models.
* Designed and implemented cloud-based data warehouses using Snowflake, BigQuery, Redshift, and Synapse Analytics, optimizing query performance with partitioning, clustering, and materialized views.
* Developed complex mappings, tasks, and workflows in IICS for real-time and batch data integration, supporting data migration and synchronization between Salesforce, AWS S3, and Snowflake.
* Led the migration of GraphQL services from Hasura to The Guild's modular GraphQL stack, enabling greater customization, fine-grained access control, and improved performance monitoring.
* Strong command over **data governance and metadata management,** utilizing platforms like GCP Data Catalog, Azure Purview, and Collibra to ensure lineage tracking, DLP compliance, and secure access control.
* Designed and developed interactive dashboards and visualizations using Qlik Sense to deliver actionable insights from complex datasets across various business domains.
* Experienced Data Modeler with expertise in Erwin, delivering robust data architecture solutions for CPG enterprises by translating complex business processes into scalable, high-integrity data models.
* Experience translating complex business requirements into data models to support analytics, KPIs, and dashboarding across **insurance and financial domains.**
* Experienced in implementing infrastructure-as-code using Terraform, GitLab CI/CD, and CloudFormation for automated deployment and monitoring of data platforms, supporting blue-green rollouts.
* Hands-on experience with modern BI and analytics tools such as Tableau, Palantir, Epik, and DBT for transforming raw data into actionable insights across financial and healthcare domains.
* Migrated legacy data systems like Teradata and SQL Server to modern cloud data platforms including AWS Redshift and Snowflake, modernizing architecture and improving performance.
* Strong SQL and Python developer with deep expertise in window functions, joins, stored procedures, and custom UDFs across Snowflake, Oracle, Redshift, and BigQuery for large-scale data transformations.
* Designed and implemented scalable Big Data pipelines using Hadoop, Spark, and Hive to process and transform large volumes of structured and unstructured data for analytics and reporting.
* Designed and deployed Delta Live Tables (DLT) pipelines in Databricks for structured streaming use cases, simplifying orchestration and enhancing data reliability.
* Integrated Great Expectations into ETL workflows to validate schema consistency, null constraints, and data profiling metrics across raw and transformed datasets.
* Built modular SQL transformation layers using dbt Cloud with version control and CI/CD integration, supporting scalable and testable data modeling practices.
* Implemented Apache Hudi-based data lakes to support upserts and incremental ingestion with ACID-compliant, scalable storage for event-driven data.
* Collaborated with stakeholders to create **semantic data layers** and reusable data marts for BI tools including Tableau, Power BI, and Looker.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| ***Google Cloud Platform*** | GCS, Cloud Pub/Sub, Dataflow, Dataproc, BigQuery, Cloud Functions,Cloud Run, Dataprep, Cloud Composer, Data Catalog, KMS, IAM,Cloud Monitoring, Logging, Build. |
| ***AWS Cloud Platform*** | AWS S3, Kinesis, EMR, Athena, Redshift, Lambda, Step Functions,DMS, Glue, Sage Maker, CloudWatch, IAM, KMS, CloudFormation,CDK, EKS, CloudWatch, MSK. |
| ***Azure Cloud Platform*** | Azure Data Factory, Synapse Analytics, Azure Data Lake, Event Hubs,Databricks, Functions, Cosmos DB, Key Vault, Active Directory,Azure Monitor, Log Analytics, AKS. |
| ***Hadoop Core Services*** | HDFS, Apache Hive, HiveQL, Sqoop, Apache Kafka, Apache Flume,Map Reduce, Spark, YARN. |
| ***Hadoop Distribution*** | Apache Hadoop, Spark, Flink, Beam, Cloudera, Hortonworks. |
| ***On-Premises*** | SAS, DB2, Teradata, Netezza, Oracle. |
| ***Databases*** | HBase, Spark-Redis, Cassandra, Oracle, MySQL, PostgreSQL,Teradata. |
| ***Data Services*** | Hive, Pig, Impala, Sqoop, Flume, Kafka. |
| ***Data Warehousing*** | Big Query, Snowflake, Redshift, Azure Synapse Analytics. |
| ***Scheduling Tools*** | Apache Airflow, Zookeeper, Oozie. |
| ***Monitoring Tools*** | Prometheus, Grafana, Cloud Monitoring, Cloud Logging, Cloud Watch,Azure Monitor |
| ***Data Visualization*** | Looker, Tableau, Data Studio, Power BI |
| ***Cloud Computing Tools*** | AWS, Azure, GCP. |
| ***Programming Languages*** | Java17, Python, Scala, SQL, PL/SQL, Pig Latin, HiveQL, ShellScripting. |
| ***Operating Systems*** | UNIX, Windows, LINUX. |
| ***Build Tools*** | Terraform, Docker, Kubernetes, Git, Cloud Build, Jenkins, Maven, Ant. |
| ***ETL Tools*** | Apache NiFi, Talend, Sqoop, SSIS, IBM DataStage, Robot Scheduler. |
| ***Development Tools*** | Eclipse, NetBeans, Microsoft SQL Studio, Toad. |

**PROFESSIONAL EXPERIENCE:**

**Sr. Data Modeler Duration: July 2023 to Present**

**Client Swiss Re, Mo**

**Roles and Responsibilities:**

* Designed and maintained conceptual, logical, and physical data models across Snowflake, BigQuery, and Databricks to support actuarial, underwriting, and claims analytics.
* Built star and snowflake schema models to structure data marts for reporting and self-service analytics using Tableau, Power BI, and Looker.
* Used dbt to implement scalable transformation logic, version control, and CI/CD for model deployment.
* Designed and maintained conceptual, logical, and physical data models using Erwin to support business intelligence and analytics initiatives in the CPG domain.
* Migrated Hive-based legacy models to cloud warehouses and optimized them using partitioning, clustering, and materialized views.
* Integrated real-time Change Data Capture (CDC) from PostgreSQL and Oracle using Kafka and Pub/Sub to maintain model freshness.
* Applied Great Expectations and Datafold for schema validation, regression testing, and ensuring model consistency.
* Implemented CI/CD pipelines to support GraphQL schema stitching, validation, and deployment during platform migration, reducing release time and ensuring schema integrity.
* Maintained metadata and enforced governance using GCP Data Catalog, Snowflake RBAC, and Azure Purview.
* e-architected data access layer during the transition from Hasura to Grafbase, ensuring schema compatibility, query optimization, and minimal disruption to existing front-end applications.
* Designed and implemented end-to-end data integration workflows using IICS, enabling seamless data movement across cloud and on-premise systems, improving data processing efficiently.
* Optimized Qlik data models and implemented advanced expressions, set analysis, and scripting techniques to improve dashboard performance and enhance user experience.
* Developed end-to-end data ingestion and processing workflows leveraging Big Data technologies to enable real-time and batch data processing for enterprise-scale analytics platforms.
* Implemented data lineage tracking and DLP (Data Loss Prevention) policies via Dataplex and Collibra for regulatory compliance.
* Performed data validation using Great Expectations and implemented data profiling checks with Datafold for regression control.
* Modeled datasets from Adobe Analytics and campaign platforms for customer segmentation and marketing attribution.
* Deployed ML pipelines using Delta Live Tables and TFX for automated fraud model training and scoring on new claim events.
* Worked with ML engineers to design feature models integrated into fraud detection pipelines using Delta Live Tables and TFX.
* Migrated Hive-based ETL logic into Snowflake and GCP Dataproc, simplifying workflows using DBT for transformation logic.
* Collaborated with cross-functional teams to model data assets aligned with CPG-specific processes such as supply chain, sales forecasting, and inventory management.
* Integrated MongoDB and Kafka to process IoT device logs with Spark, enabling real-time alert generation and monitoring.
* Ingested Adobe Analytics and campaign datasets using Apache NiFi into BigQuery for marketing attribution and segmentation analysis.
* Collaborated with stakeholders to define KPIs and build semantic layers that enable consistent, business-friendly data access.
* Implemented robust error handling, parameterization, and reusable components in IICS, ensuring scalability, reusability, and easier maintenance of integration processes.
* Monitored ETL workflows using Prometheus, Stackdriver, and OpenTelemetry, enabling proactive resolution of job failures.
* Generated automated extracts for Anaplan and Cognos for weekly financial reporting and scenario forecasting.
* Collaborated with data scientists to deploy Delta Live Tables and TensorFlow Extended (TFX) pipelines for machine learning model inference and retraining workflows.
* Managed metadata cataloging and access control with Data Catalog, Purview, and Snowflake RBAC policies.

**Sr Data Engineer Duration: Dec 2019 To May 2023**

**Nationwide, OH**

**Roles and Responsibilities:**

* Migrated large-scale data from Cassandra, Oracle, and Salesforce to AWS Snowflake and GCP BigQuery using PySpark and SnowSQL to improve data accessibility and reduce latency.
* Built scalable ETL pipelines using DBT, Snowflake SQL, and Python to transform raw claims, eligibility, and provider data into structured formats for healthcare analytics.
* Developed real-time and batch pipelines using Apache Beam (Dataflow) and Spark (Dataproc) to support cross-cloud processing across GCP and AWS environments.
* Configured Azure Data Factory pipelines to ingest data from Azure Data Lake into Azure SQL Staging and Production layers, supporting hybrid cloud ETL workflows.
* Containerized Python-based ETL services using Docker, deployed via Cloud Run and Cloud Build for secure and repeatable processing.
* Implemented automated data quality checks using Dataplex rules, Great Expectations, and DBT tests to ensure clean and validated data delivery.
* Ensured data integrity and consistency by applying industry-standard modeling techniques and establishing governance practices within large-scale CPG datasets.
* Built business dashboards using Tableau, Looker, and Data Studio by sourcing from Snowflake and BigQuery, visualizing KPIs, fraud scores, and operational trends.
* Created standardized Dataprep transformation recipes to clean and normalize data from 15+ ingestion sources, improving downstream analytics accuracy.
* Orchestrated workflows using Cloud Composer (Airflow) and managed real-time events with Pub/Sub and Kafka for end-to-end data movement.
* Processed streaming files using Java 8, Spark SQL, and Spring Boot, applying transformation logic before persisting into Snowflake and BigQuery.
* Scheduled DBT model runs and Airflow DAGs using AWS CloudWatch Events, Step Functions, and Glue workflows for automated data refreshes.
* Created incremental models, federated queries, and materialized views in Snowflake and BigQuery for high-performance querying.
* Managed metadata and access controls using GCP Data Catalog, Snowflake RBAC, and Azure Purview to ensure governance and compliance.
* Designed Tableau dashboards with advanced charts (e.g., Gantt, maps, bar/stacked bar), published to Tableau Server with background task scheduling.
* Enabled data exploration for business users by integrating TypeScript UIs with BigQuery and Snowflake for interactive querying.
* Tuned Spark and SQL performance using job profiling tools and DBT model optimization techniques to reduce job runtimes.
* Conducted regression and integration testing across AWS and GCP environments, validating outputs via SQL, Jupyter, and Informatica.
* Led QA automation for DBT models and ETL workflows with Jenkins, GitHub, and Airflow for test scheduling and CI/CD integration.
* Implemented Talend jobs and API integrations to support data exchange between Snowflake, Redshift, and external systems.
* Designed and built an enterprise data warehouse using Star and Snowflake schema modeling to support scalable BI and analytics reporting.

**Data Engineer Duration: March 2017 To Oct 2019**

**Vanguard, NC**

**Roles and Responsibilities:**

* Developed streaming and batch pipelines using Apache Beam with orchestration through Cloud Composer (Airflow) to support asynchronous job control and automated retries.
* Built scalable data pipelines using BigQuery, Pub/Sub, Cloud Storage, and Dataflow to process real-time clinical, claims, and eligibility datasets across GCP.
* Integrated Epik with GCP, AWS, and Azure to manage multi-cloud ETL workflows and automate data movement across environments.
* Designed clinical data models in BigQuery following HL7 and FHIR standards to support provider performance analysis and claims adjudication.
* Delivered Tableau, Power BI, and Data Studio dashboards using GCP data sources, enabling real-time business insights and operational monitoring.
* Implemented predictive models in Python for product demand forecasting using AWS Kinesis Firehose and S3 for real-time data streaming.
* Wrote optimized SQL scripts for healthcare data validation, anomaly detection, and regulatory compliance reporting in BigQuery and Snowflake.
* Built and scheduled Informatica IICS workflows for data transformation, cleansing, and load processes supporting regulatory and reporting needs.
* Used Dataproc and Hadoop ecosystem tools (Hive, Pig, SQOOP, Spark) for batch processing large healthcare datasets.
* Designed high-performance Hive tables using partitioning, clustering, and skew handling to reduce query latency and storage cost.
* Built ingestion pipelines with Apache Kafka and implemented ELK stack for monitoring high-throughput log data.
* Automated ETL and validation scripts using PySpark, SQL, and Python, integrated with infrastructure provisioned through Terraform.
* Validated and optimized data queries across BigQuery and Azure Synapse, reducing BI dashboard load times and improving accuracy.
* Worked with data scientists to deploy advanced analytical models in GCP Hadoop clusters over large datasets.
* Set up CI/CD pipelines using GitHub Actions, Terraform, and Cloud Build to enable repeatable and secure deployment of workflows.
* Orchestrated job logic using Cloud Functions, Pub/Sub, and Airflow operators to manage asynchronous task scheduling and data readiness checks.
* Used GCP Cloud Shell and CLI tools for deployment tasks, monitoring, and service management across development environments.
* Built and maintained Splunk dashboards for monitoring pipeline health, enabling real-time alerting and operational visibility.

**Jr Data Engineer Duration: March 2012 To Nov 2015**

**Client: Reliant Hyderabad, India**

**Roles and Responsibilities:**

* Engineered ETL pipelines using SQL, Talend, and Epik to extract, transform, and load data from multiple source systems into centralized reporting and analytics platforms.
* Designed and implemented scalable data models in collaboration with architects and analysts to support operational data stores (ODS) and enterprise data marts.
* Built and configured an enterprise data lake for secure storage and processing of high-volume structured and semi-structured datasets used in analytics and BI reporting.
* Integrated Apache Kafka with Spark to develop real-time data ingestion and log processing pipelines using producer-consumer models.
* Performed exploratory data analysis and profiling using Python libraries (NumPy, Pandas, Seaborn, Matplotlib) to identify trends, outliers, and data inconsistencies.
* Applied Dataplex security policies and transformation rules to meet GDPR, HIPAA, and internal data privacy compliance requirements.
* Wrote optimized SQL queries for large-scale data validation, reconciliation, and transformation tasks across staging and target environments.
* Developed reusable Talend workflows for automating batch and incremental data loads, minimizing manual intervention in ETL execution.
* Conducted rigorous testing of ETL jobs by preparing test data, creating SQL-based test cases, and documenting execution results and defect logs.
* Utilized Quality Center (QC) for defect tracking and resolution, coordinating with QA, BA, and development teams to ensure data quality.
* Automated data validation logic and created lineage tracing documentation to ensure end-to-end data integrity across reporting systems.
* Participated in daily Agile standups, sprint reviews, and release planning to align data engineering tasks with business priorities.
* Supported production ETL workflows by debugging failures, rerunning critical jobs, and ensuring timely delivery of analytics-ready datasets.
* Built monitoring dashboards using Splunk and scheduled alerts for ETL job failures, helping reduce issue resolution time and improve pipeline reliability.
* Collaborated with QA teams to automate test case execution using shell scripts and SQL validation scripts, improving regression coverage for data workflows.
* Created parameterized and reusable SQL scripts to validate data transformations across multiple data domains, improving consistency in data quality checks.
* Implemented performance tuning techniques for SQL queries and Talend ETL jobs, reducing data load times by over 30% in high-volume tables.
* Maintained version control of ETL code and SQL scripts using Git, ensuring traceability of changes and smoother collaboration across development and QA teams.

**EDUCATION**:

• Bachelors in Computers Science and Engineering from **Veltech University** **(2011)**