**Vamsi Krishna**

**Vamsik0311@gmail.com | +(862)402-0997**

**Senior Data Analyst/ Business Analyst**

**PROFESSIONAL SUMMARY:**

* Around 8 years of IT Industry experience in analysis designing testing maintaining database ETL, Big data, Cloud, and Data warehouse applications.
* Experience in designing, developing, executing, and maintaining data extraction, transformation, and loading for multiple corporate Operational Data Stores, Data warehouse, and Data mart systems.
* Strong experience in Business and Data Analysis, Data Profiling, Data Migration, Data Integration, Data governance and Metadata Management, Master Data Management, and Configuration Management.
* Expertise in generating and evaluating Business Requirements Documents (BRD), Functional Requirement Documents (FRD), System Requirement Specification (SRS), and Work Breakdown Structure (WBS) and recommending changes as per the applicable internal and external requirements gathered for process improvement.
* Strong experience in interacting with stakeholders/customers, gathering requirements, workshops, and existing system documentation or procedures, defining business processes, identifying, and analyzing risks using appropriate templates and analysis tools.
* Strong in ETL, Operations Data Store concepts, Data warehouse, data marts, and OLAP technologies.
* Fashioned on different libraries related to Data Science and Machine learning like Pandas, NumPy, SciPy, Matplotlib, Seaborn, Bokeh, Scikit - -learn, OpenCV, TensorFlow, Theano, and Keras.
* Excellence in handling Big Data Ecosystems like Apache Hadoop, MapReduce, Spark, HDFS Architecture, Cassandra, HBase, Sqoop, Hive, Pig, MLlib, and ELT.
* Experience with the use of AWS services including RDS, Networking, S3, EC2, EBS, VPC, and administering AWS resources using Console and CLI.
* Worked on ETL and Report testing using different Data warehousing tools like SSIS, Databricks, AWS, Cognos, Data Mart Worked on AWS RDS, Redshift & Tableau.
* Hands-on experience in building the infrastructure required for optimal extraction, transformation, and loading of data from a wide variety of data sources using NoSQL and SQL from AWS and Big Data technologies (Dynamo, Kinesis, S3, HIVE/Spark).
* Expertise in creating Snowflake and star schemas, FACT, and Dimension tables using Erwin for physical and logical data modeling.

|  |  |
| --- | --- |
| • | Had good experience in creating Use Case Scenarios, process modeling using Use Case Diagrams, Data Flow Diagrams, Activity diagrams, and Sequence Diagrams using MS Visio and Rational Rose. |
| • | Hands-on experience with ER Studio /Erwin for creating logical as well as physical data models and DDL. |

* Skilled in using the JIRA tool for writing and managing Issues, Epics, User Stories, and Tasks in a Scrum environment.
* Participated in Scrum ceremonies such as Sprint Planning, Daily Stand-ups, Sprint Review, and Sprint Retrospective.
* Experience in designing, developing, and publishing visually rich and intuitively interactive Tableau workbooks and dashboards for executive decision-making.
* Design and implement database solutions using Azure SQL Data Warehouse and Azure SQL.
* Practical understanding of migrating on-premise ETLs to Google Cloud Platform (GCP) using cloud-native tools such as Google Cloud Storage, BIG query, Cloud Data Proc, and Composer.
* Worked on multi-cloud on GCP and Azure, AWS.
* A self-motivated exuberant learner and adequate with challenging projects and work in ambiguity to solve complex problems independently or in a collaborative team.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Databases | AWS RDS, Teradata, Oracle, MySQL, Microsoft SQL, Postgres SQL. |
| NoSQL Databases | MongoDB, Hadoop HBase, and Apache Cassandra. |
| Programming Languages | Python, SQL, Scala, Pig C, C++, MATLAB, Java, JavaScript. |
| Cloud Technologies | AWS, Azure, Ec2, S3, Docker, GCP |
| Querying Languages | SQL, No SQL, PostgreSQL, MySQL, Microsoft SQL. |
| Deployment Tools | Anaconda Enterprise v5, R-Studio,  Machine Learning Studio, Oozie, AWS Lambda |
| Scalable Data Tools | Hadoop, Hive, Apache Spark, Pig, Map Reduce, Sqoop. |
| Operating Systems | Red Hat Linux, Unix, Ubuntu, Debian, Centos, Windows, MacOS. |
| Reporting & Visualization | Tableau, Matplotlib, Seaborn, Bokeh, gplot, pilots, Shiny. |

**PROFESSIONAL EXPERIENCE:**

**Client:** **Cardinal Health - Richmond, VA Jan 2023 to Present**

**Role: Data Analytics Engineer** Responsibilities:

* Maintained and developed complex SQL queries, views, functions, and reports that qualify customer requirements on Snowflake.
* Performed analysis, auditing, forecasting, programming, research, report generation, and software integration for an expert understanding of the current end-to-end BI platform architecture to support the deployed solution.
* Hands-on experience in resolving incident tickets related to Hadoop components like HBase, Yarn, Hive, and Kafka, and identifying root cause analysis.
* Traced and catalogue data processes, transformation logic and manual adjustments to identify data governance issues.
* Utilized Delta caching for faster data processing using Databricks and PySpark.
* Developed JSON scripts for deploying the Azure Data Factory (ADF) Pipeline, which uses the SQL Activity to process the data.
* Advanced and developed test plans to ensure successful delivery of a project. Employed performance analytics predicated on high-quality data to develop reports and dashboards with actionable insights.
* Worked with the ETL team to document the transformation rules for Data migration from OLTP to Warehouse environment for reporting purposes.
* Involved in creating Physical and Logical models using Erwin and creating CSV reports and DDL.
* Experience developing data management systems, tools and architectures using relational databases, Redshift and/or other distributed computing systems.
* Developed Databricks ETL pipelines using notebooks, Spark Data frames, SPARK SQL, and Python scripting.
* Developed Data pipelines using Python for medical image pre-processing, Training, and Testing.
* Working with Channel Sales team, reviewed accounts and contacts in Salesforce. Worked to determine and clean up Sales team data base.
* Automated execution process, scheduling, deploying of packages using SQL Server Agent by creating jobs and error reports using Alerts, SQL Mail Agent and FTP.
* Design Setup maintain Administrator the Azure SQL Database, Azure Analysis Service, Azure SQL Data warehouse, Azure Data Factory, Azure SQL Data warehouse.
* Précised Development and implementation of several types of sub-reports, drill-down reports, summary reports, parameterized reports, and ad-hoc reports using Tableau.
* Acquainted with parameterized sales performance reports, doing the reports every month, and distributing them to respective departments/clients using Tableau.
* Used Spark-SQL to load JSON data and create schema RDD and loaded it into Hive Tables and handled structured data using Spark SQL.
* Developed API's using python with flask framework and done the integration of various data sources including JDBC, RDBMS, Shell Scripting, Spreadsheets, and Text files.
* Worked on Snowflake schemas, data modeling and elements, source-to-destination mappings, interface matrix, and design elements.
* Expertize in designing and developing NoSQL Database Models in cassendra, DyanamoDB and Relational Database Models using Erwin.
* Analyzed data lineage processes to identify vulnerable data points, control gaps, data quality issues,
* and overall lack of data governance.
* Performed Dependency Analysis (by identifying any breaks in lineage), root cause analysis and remediation as part of the data lineage effort.
* Administrated and monitored the company’s Salesforce CRM application.
* Built Data Pipelines using Apache NIFI to analyze structured data by pulling it from Splunk and creating Hive tables.
* Analyze data quality issues using Snow SQL by building an analytics warehouse on top of Snowflake.
* Analyzed data using SQL, Scala, Python, and Apache Spark and presented analytical reports to management and technical teams.
* Using Linked Services/Datasets/Pipelines, Pipelines were created in ADF to extract, transform, and load data from a variety of sources, including Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool, and backward.
* Developed PySpark and SparkSQL code to process the data in Apache Spark on Amazon EMR to perform the necessary transformations based on the STMs developed.
* Creating the High Level and Low-Level design documents as per the business requirement and collaborating with the offshore team to guide them on design and development.
* Participated in the Portfolio Governance and Data Governance working group sessions to create policies.
* Experience in writing python code to schedule jobs dynamically in Apache Airflow
* Used a variety of sources, including SQL Server, Excel, Oracle, and SQL Azure to import data into Power BI.
* Optimized current pivot tables reports using Tableau and proposed an expanded set of views in the form of interactive dashboards using line graphs, bar charts, heat maps, tree maps, trend analysis, Pareto charts, and bubble charts to enhance data analysis.
* Monitor system life cycle deliverables and activities to ensure that procedures and methodologies are followed, and that appropriate complete documentation is captured.

**Technical Environment**: SQL, Snowflake, Python (Scikit -Learn/ Keras/ SciPy/ NumPy/ Pandas/ Matplotlib/ NLTK/ Seaborn), Tableau, Hive, Databricks, Airflow, PostgreSQL, Azure, JIRA, GitHub, NIfi. Kafka.

**Client: Capital One - Richmond, VA AUG 2021 to JAN 2023**

**Role: SR. Data Analyst** Responsibilities:

* Played a main role in the Elicitation of functional and non-functional requirements from stakeholders and Clients.
* Perform Data Analysis on the analytical data present in AWS S3, AWS Redshift, Snowflake, Teradata using SQL, Python, Spark, Databricks.
* Analyzed and gathered business requirements from clients, conceptualized solutions with technical architects, and verified approaches with appropriate stakeholders, developed E2E scenarios for building the application.
* Assisted the PO (Product Owner) in scope creation, requirements analysis, and creating MVP, during backlog grooming meetings by decomposing the Epics into user stories and assisted the team in writing user stories.
* Derived data from relational databases to perform complex data manipulations and conducted extensive data checks to ensure data quality. Performed Data wrangling to clean, transform, and reshape the data utilizing NumPy and Pandas library.
* Experienced in BA with Salesforce implementations from Insurance and financial services.
* Worked on various POCs to adopt new technologies like Apache Airflow, Snowflake, and Terraform for infrastructure management.
* Implemented SDLC which included Requirements, Specifications, Design, Development, Analysis, and Testing.
* Worked on building the data model using Erwin as per the requirements. Designed the grain of facts depending on reporting requirements.
* Used mount functionality in spark to get access to the One Lake datasets for consumption purposes using control plane with particular, ASV, BAPCI and LOB ARN IAM role for a given databricks NPI cluster.
* We have worked with datasets of varying degrees of size and complexity including both structured and unstructured data and Participated in all phases of Data mining, Data cleaning, Data collection, variable selection, feature engineering, developing models, Validation, Visualization, and Performed Gap analysis.
* Optimized a lot of SQL statements and PL/SQL blocks by analyzing the execution plans of SQL statements and created and modified triggers, SQL queries, and stored procedures for performance improvement.
* Transformed batch data from several tables containing tens of thousands of records from SQL Server, MySQL, PostgreSQL, and csv file datasets into data frames using PySpark.
* Implemented Predictive analytics and machine learning algorithms in Databricks to forecast key metrics in the form of designed dashboards on AWS (S3/EC2) and Django platform for the company’s core business.
* Perform Data Cleaning, features scaling, and features engineering using pandas and numpy packages in python.
* Documented the complete data lineage from origination to the warehouse.
* Worked on catapulting data from teradata to snowflake to consume on Databricks.
* Analyzed data lineage processes to identify vulnerable data points, control gaps, data quality issues, and overall lack of data governance.
* Create several types of data visualizations using Python and Tableau.
* Participated in Facets Table data Modelling planning, designing, and implementation of the data warehouse and conducted testing by developing complex SQL queries.
* Developed spark programs using Scala APIs to compare the performance of spark with HIVE and SQL.
* Designed and created Hive external tables using a shared meta-store instead of derby with partitioning, dynamic partitioning, and buckets.
* Experience in building scalable distributed data solutions using an EMR cluster environment with Amazon EMR.
* Evaluated the performance of Databricks environment by converting complex Redshift scripts to spark SQL as part of a new technology adaption project.
* Lead engagement planning: developed and managed Tableau implementation plans for the stakeholders, ensuring timely completion and successful delivery according to stakeholder expectations.
* Managed workload and utilization of the team. Coordinated resources and processes to achieve Tableau implementation plans.

**Technical Environment**: R, Python, ETL, Agile, Data Quality, R Studio, Tableau, Data Governance, Supervised &amp; Unsupervised Learning, Java, NumPy, SciPy, Hadoop, Sqoop, HDFS, Spark SQL, Pandas, PostgreSQL, AWS (EC2, RDS, S3), Matplotlib, Scikit-learn, Shiny.

**Client: Caterpillar Inc – Peoria, IL NOV 2019 to JUL 2021**

**Role: Data Analyst / Business Analyst**

Responsibilities:

* Worked in the Data transformation team for an internal application called NEXUS where I was responsible for all the Data Development, Data Analysis.
* Ability to model business processes and data related to the full end-to-end business operation, including the order to cash, procure to pay process, and record to report.
* Data Quality and Data Maintenance efforts of the application which my team was highly focusing on.
* Created various Boto scripts to maintain applications in the cloud environment and automated them.
* Analyzed the Incident, Change, and Job data from Snowflake and created a dependency tree-based model on the occurrence of incidents for every application service present internally.
* Helped businesspeople minimize the manual work they were doing and created Python scripts like LDA sourcing, One Lake, and SDP. S3, Databricks, Data Bench, and Snowflake to get the cloud metrics and make their efforts easier.
* Experience in SSIS, SSRS, SPSS in solving business problems environments: Oracle 9i, SQL Server 2005, SQL Loader, XML, Erwin, Windows 2000/NT, UNIX.
* Worked on moving data from external legacy system into Salesforce application using Informatica cloud, Apex Data loader& CSV files.
* Coaching product teams towards best practices and learnings. Ensures delivery of messaging, communication, and customer outreach as necessary for B2B integrations
* Helped design data warehouse solutions and business intelligence dashboards and scorecards.
* Substantial experience with the use of relational databases using SQL for data extraction, management, and queries.
* Automate the process to send the data quality alerts to slack channel and email using Databricks, Python and HTML. This will alert users if there are any issues with data.
* Participated in the Data Governance working group sessions to create Data Governance Policies.
* Used Git, GitHub, and AWS EC2 and deployment and used extracted data for analysis and carried out various mathematical operations for calculation purpose using python library - NumPy, SciPy.
* Merged with D4 Rise Analytics team and helped them to remove all the manual work they were doing on Incident data by generating metrics pulling data from various sources like Service Now, Snowflake, and a few other API calls and creating an Incident Dashboard with a lot of intelligence built in it.
* Working on getting the data from on-prem SQL server database to a Raw Zone in Azure Data Lake by using Azure Data Factory.
* Implemented metadata standards, data governance and stewardship, master data management, ETL, ODS, data warehouse, data marts, reporting, dashboard, analytics, segmentation, and predictive modelling.
* Utilized AWS EMR to transform and transfer huge volumes of data between Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB, among other AWS data storage and databases.
* Designed and developed a security framework that provides fine-grained access to objects in AWS S3 using AWS Lambda, and DynamoDB.
* Used Reverse Engineering approach to redefine entities, relationships, and attributes in the data model as per new specifications in Erwin after analyzing the database systems currently in use.
* Perform data comparison between SDP(Streaming Data Platform) real time data with AWS S3 data and Snowflake data using Databricks, Spark SQL, and Python.
* Conducted A/B testing using predictive models to identify the most effective marketing strategies and tactics, increasing campaign ROI.
* Used Spark SQL to migrate the data from hive to Python using the Pyspark library.
* I have used Hadoop technologies like spark and hive Including using the Pyspark library to create spark data frames and converting them to normal pandas data frames for analysis.
* Created complex SQL Stored Procedures, Triggers, HIVE queries, and User user-defined functions to manage incoming data as well as to support existing applications.
* Queried and analyzed large amounts of data on Hadoop HDFS using Hive and Spark
* Lead the intern's project by getting them started on AWS, and best practices, suggesting other simplified methods, and monitoring their work weekly.
* Worked extensively on Airflow for orchestration and scheduling of the ingestion scripts.
* Built a prototype for a project containing various kinds of dashboard metrics for all the applications to have a central metric tracker.
* Built a process to get the usage of One Lake usage for all the LOBs for higher level VPs that includes Kibana data source, EFK, and Kinesis streams and created AWS Lambda functions that automate in a daily manner.
* Built interactive dashboards and stories using Tableau Desktop for accuracy in report generation applying advanced Tableau functionality: parameters, actions, and tooltip changes.

**Technical Environment**: Python, Tableau (9.x/10.x), Hadoop, HDFS, PySpark, Teradata, PostgreSQL, AWS, Jenkins, SQL, Snowflake, JIRA, GitHub, Agile/ SCRUM, Kafka.

**Client: Wissen Infotech- Bangalore, India APR 2017 TO JUL 2019**

**Role: Data Analyst Responsibilities:**

* Collaborated with stakeholders to understand business requirements and worked closely to perform data analysis and reporting.
* Built predictive models such as logistic regression, and random forests to predict the likelihood of the customer journey as well as automated it as a tool to generate the probability week over week.
* Created SQL queries to extract and manipulate data from the Teradata database.
* Designed and implemented automated reporting processes using Teradata.
* Created data governance templates and standards for the data governance organization.
* Developed and deployed predictive models using machine learning algorithms to forecast customer behavior, such as purchase propensity, churn likelihood, and lifetime value.
* Worked in supporting integrated testing of interfaces and validation of data mapping, data migration, as well as data conversion activities conducted during pre-go-live and post-go-live activities.
* Conducted customer segmentation using predictive modeling to identify customer groups with similar characteristics and behavior patterns, resulting in more targeted and effective marketing campaigns.
* Built and maintained customer response models to optimize marketing campaigns and promotions, increasing response rates and reduction in marketing costs.
* Interacted with business SMEs, enterprise architects, developers, business intelligence, and data warehouse teams regarding project requirements explanation and interpretation.
* Extensively use ETL methodology for performing Data Migration, Data Profiling, Extraction, Transformation and Loading using Talend and designed data conversions from wide variety of source systems like SQL Server, Oracle, DB2 and non - relational sources like XML, flat files, and mainframe Files.
* Conducted market basket analysis using predictive modeling to identify cross-selling and upselling opportunities, resulting in a 15% increase in revenue.
* Migrated an existing on-premises application to AWS. Used AWS services like EC2 and S3 for small data sets processing and storage, Experienced in Maintaining the Hadoop on AWS EMR.
* Conducted propensity modeling to identify high-value prospects and inform lead generation efforts, resulting in a 15% increase in lead conversion rates.
* Involved in translating the business requirements into data requirements across different systems.
* Maintained Excel workbooks, such as development of pivot tables, exporting data from external SQL databases,
* producing reports and updating spreadsheet information.
* Managed and implementation of data governance using Microsoft Project, Visio, PowerPoint, SharePoint tools, and multiple training programs.
* Built and maintained data pipelines in Snowflake to ingest and store marketing data from multiple sources, such as web analytics, customer databases, and third-party vendors.
* Monitored email performance and provided regular reporting and analysis to clients, leveraging Select Target's reporting and analytics capabilities to identify optimization opportunities.
* Collaborated with clients to develop email marketing strategies, including audience segmentation, messaging, and content, leveraging Select Target's advanced segmentation capabilities.
* Managed end-to-end email campaigns for various clients using ExactTarget's Select Target feature, achieving a good open rate.
* Demonstrated exceptional written and verbal communication skills, including the ability to present data-driven insights and recommendations to senior leaders.
* Partnered with Client, Sales, and Marketing teams to understand their process and find relevant gaps by performing opportunity analysis.

**Technical Environment**: Python, R, SQL, Jupyter Notebook, Anaconda, PyCharm, RStudio, Teradata, Snowflake, MySQL, PostgreSQL, AWS, Azure, GCP, Select Target, ExactTarget, Jira, Confluence, Tableau, Power BI.