#  MANOJ GOWDA JAYASWAMY

**manoj.jayaswamy@gmail.com** **| +1 (469) 739-4508|Seattle, WA**

**PROFESSIONAL SUMMARY**

* Over 8+ years of experience in Data Engineering, Data Pipeline Design, Development and Implementation as a Sr. Data Engineer/Data Developer and Data Modeler.
* Strong experience in Software Development Life Cycle (SDLC) including Requirements Analysis, Design Specification and Testing as per Cycle in both Waterfall and Agile methodologies.
* Strong experience in writing scripts using Python API, PySpark API and Spark API for analyzing the data. And extensively used Python Libraries PySpark, Pytest, Pymongo, cxOracle, PyExcel, Boto3, Psycopg, embedPy, NumPy and BeautifulSoup.
* Ability to collaborate with peers in both, business, and technical areas, to deliver optimal business process solutions, in line with corporate priorities.
* Very good high exposure in building data lakes infra and set up ETL pipelines in AWS/Azure cloud infra.
* Good working knowledge of classes for centralizing all the frequently used methods, which are invoked in Lambda handlers to create and maintain custom libraries in Python
* Adept in analyzing large datasets using Apache Spark, Spark ML and Amazon Web Services (AWS).
* Proficient in working with SQL/non-SQL database, AWS EC2, S3, EMR, Lambdas, Sage Maker, Docker.
* Experience in Data Governance consulting and implementation of Data Quality frameworks, Data Catalog Metadata management, and Data Lineage.
* Big Data - Hadoop (MapReduce & Hive), Spark (SQL, Streaming), Azure Cosmos DB, SQL Datawarehouse, Azure DMS, Azure Data Factory, AWS Redshift, Athena, Lambda, Step Function and SQL.
* Strong knowledge in Spark ecosystems such as Spark core, Spark SQL, Spark Streaming libraries.
* Experience in designing Azure Cloud Architecture and Implementation plans for hosting complex application workloads on MS Azure.
* Experience working in reading Continuous json data from different source system using Ka ka into Databricks Delta and processing the iles using Apache Structured streaming, PySpark and Parallel Processing and creating the iles in parquet format.
* Created manual TestCases to check that each of the deliverables meet user's requirement.
* Good knowledge in Apache Hadoop ecosystem components Spark, Cassandra, HDFS, Hive, SQOOP, Air low.
* Experience in working with Flume and NiFi for loading log files into Hadoop.
* Experience in working with NoSQL databases like HBase and Cassandra.
* Good Experience in implementing and orchestrating data pipelines using Oozie and Air low.
* Proficiency in SQL across several dialects (we commonly write MySQL, PostgreSQL, Redshift, SQL Server, and Oracle)
* Experience in designing star schema, Snow lake schema for Data Warehouse, ODS architecture.
* Skilled in System Analysis, E-R/Dimensional Data Modeling, Database Design and implementing RDBMS speci ic features.
* Good knowledge of Data Marts, OLAP, Dimensional Data Modeling with Ralph Kimball Methodology (Star Schema Modeling, Snow-Flake Modeling for FACT and Dimensions Tables) using Analysis Services.
* Constructed data staging layers and fast real-time systems to feed BI applications and machine learning algorithms

**TECHNICAL SKILLS**

* **Hadoop Eco System: Spark, Hive, Sqoop, Oozie, Pig**
* **Azure Cloud Platform:** ADF, BLOB Storage, ADLS, Azure SQL DB, SQL server, Azure Synapse, Azure Analytic Services, Data bricks, Mapping Data low (MDF), Azure Data Lake (Gen1/Gen2), AzureCosmos DB, Azure Stream Analytics, Azure Event Hub, App Services, Logic Apps, Event Grid, Service Bus, Azure DevOps, GIT Repository Management, ARM Templates
* **AWS Cloud Platform:** AWS EC2, AWS RDS, AWS S3, AWS Lambda, AWS SNS, AWS VPC, AWS Kinesis, AWS IAM, AWS Dynamo DB, AWS SQS, AWS Athena, AWS QuickSight
* **Programming Languages:** Python, Scala, R, C, C++, Java, Shell Scripting
* **Databases and Query Languages:** Azure SQL Warehouse, Azure SQL DB, Azure Cosmos NO SQL DB, Teradata, Vertica, RDBMS, MySQL, Oracle, PostgreSQL, Microsoft SQL Server
* **Streaming Frameworks:** Kinesis, Kafka, Flume
* **Tools:** R Studio, PyCharm, Jupyter Notebook, Intellij, Eclipse, NetBeans
* **Platforms:** Linux, Windows and OS X

**Education**

**Bachelors of Technology in Electronics & Communication Engineering, Jawaharlal Nehru Technological University, 2014**

**PROFESSIONAL EXPERIENCE**

**Client: Star Seven Six(Fillmore Capital), OH Feb 2023 – Till date Sr.Data Engineer**

**Responsibilities:**

* Implemented Azure Data Factory (ADF) extensively for ingesting data from different source systems like relational and unstructured data to meet business functional requirements
* Design and developed Batch processing and real-time processing solutions using ADF, Databricks clusters and stream Analytics
* Created numerous pipelines in Azure using Data Factory v2 to get the data from disparate source systems by using different Activities like Move &Transform, Copy, filter, for each, Databricks etc. Maintain and provide support for optimal pipelines, data lows and complex data transformations and manipulations using ADF and PySpark with Databricks.
* Automated jobs using different triggers like Events, Schedules and Tumbling in ADF.
* Created, provisioned different Databricks clusters, notebooks, jobs and autoscaling.
* Performed data low transformation using the data low activity.
* Implemented self-hosted integration runtime in ADF.
* Improved performance by optimizing computing time to process the streaming data by optimizing the chaster run time.
* Perform ongoing monitoring, automation, and refinement of data engineering solutions.
* scheduled, automated business processes and work lows using Logic Apps.
* Designed and developed a new solution to process the RT data by using stream analytics, Event Hubs and Service Bus Queue.
* Created Linked services to connect the external resources to ADF.
* Worked with complex SQL views, Stored Procedures, Triggers, and packages in large databases from various servers.
* Used Azure Devops & Jenkins pipelines to build and deploy different resources(Code and Infrastructure) in Azure.
* Ensure the developed solutions are formally documented and signed off by business.
* Worked with team members to resolve any technical issue, Troubleshooting, Project Risk & Issue identification, and management.
* Developed dbt models with DataStage as a source and Snowflake as target, enabling SQL transformations for analytics.
* Implemented lifecycle management and CI/CD for DataStage jobs and dbt models using Azure DevOps and Git workflows.
* Worked on the cost estimation, billing, and implementation of services on the cloud.
* Experience managing Azure Data Lakes (ADLS) and Data Lake Analytics and an understanding of how to integrate with other Azure Services.
* Migration of on premise data (Oracle/ Teradata) to Azure Data Lake Store(ADLS) using Azure Data Factory(ADF V1/2).
* Work closely across teams (Support, Solution Architecture) and peers to establish and follow best practices while solving customer problems
* Created infrastructure for optimal extraction, transformation, and loading of data from a wide variety of data sources.
* Designed and created optimal pipeline architecture on Azure platform.
* Created pipelines using ADF to get the data from different source systems and transform the data by using many activities.
* Created Linked service to land the data from different sources to Data Factory.
* Created different types of triggers to automate the pipeline in ADF.
* Created, provisioned different Databricks clusters needed for batch and continuous streaming data processing and installed the required libraries for the clusters.

Environment: Azure Data Factory (ADF V2), Azure SQL Database, functions Apps, Azure Data Lake, BLOB Storage, SQL server, Windows remote desktop, UNIX Shell Scripting, PowerShell, Data bricks, Python, ADLS Gen 2, Cosmos DB, Event Hub,.

**Client: AT&T-Dallas, TX Feb 2022 – Jan 2023 Sr.Data Engineer**

**Responsibilities:**

* Migrate the existing data from Teradata/SQL Server to Hadoop and perform ETL operations on it.
* Responsible for loading structured, unstructured, and semi-structured data into Hadoop by creating static and

dynamic partitions.

* Worked on different data formats such as JSON and performed machine learning algorithms in Python.
* Performing statistical data analysis and data visualization using Python and R
* Implemented data ingestion and handling clusters in real time processing using Kafka.
* Imported real time weblogs using Kafka as a messaging system and ingested the data to Spark Streaming.
* Developed Hadoop streaming Map/Reduce works using Python.
* Created Hive partitioned tables using Parquet & Avro format to improve query performance and efficient space

utilization with Cybersecurity environment.

* Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
* Architect & implement medium to large scale BI solutions on Azure Devops using Azure Data Platform services (Azure

Data Lake, Data Factory, Data Lake Analytics, ADLS, Azure SQL DW, HDInsight/Databricks, NOSQL DB).

* Design & implement migration strategies for traditional systems on Azure (Lift and shift/Azure Migrate, other thirdparty tools.
* Used the Ralph Kimball Methodology for Data Warehouse and Data Mart designs.
* Used various sources to pull data into Power BI such as SQL Server, Excel, Oracle, SQL Azure etc.
* Propose architectures considering cost/spend in Azure and develop recommendations to right-size data

infrastructure.

* Responsibilities include Database Design and Creation of User Database.
* Worked with Databricks, Structured Streaming. Setting Up Microsoft Azure with Databricks, Databricks Workspace

for Business Analytics, Manage Clusters In Databricks, Managing the Machine Learning Lifecycle

* Moving ETL pipelines from SQL server to Hadoop Environment with Azure Devops
* Developed Spark applications using Parallel Processing and PySpark and Spark-SQL for data extraction,

transformation, and aggregation from multiple ϔile formats for analyzing & transforming the data to uncover

insights into the customer usage patterns.

* Experience in developing Spark applications using Spark-SQL in Databricks for data extraction, transformation,

and aggregation from multiple file formats for Analyzing & transforming the data to uncover insights into the

customer usage patterns.

* Engage with business users to Cybersecurity with gather requirements, design visualizations and provide training

to use self-service BI tools.

* Designed the data marts using the Ralph Kimball's Dimensional Data Mart modeling methodology using ER Studio.
* Implemented data quality checks using Spark Streaming and arranged bad and passable ϔlags on the data.
* Used advanced SQL methods to code, test, debug, and document complex database queries.
* Design relational database models for small and large applications.
* Extracting data from raw input Call Detail record data and loading into database using Perl DBI and DBD with

oracle.

* Loading of the CDR data into Oracle data base and generate alerts on the CDR data as per defective DR count by

absolute/percent values. Code written in Perl, SQL (oracle) on a Linux environment.

* Design and Develop code in Perl, SQL (Oracle) and Linux platform for each release.
* Perform design, development, testing and implementation of the applications using Perl, Java, Shell Script and

Oracle Database on UNIX and Windows Platforms.

* Designed and developed Scala workflows for data pull from cloud-based systems and applying transformations on

it.

* The ability to develop reliable, maintainable, efficient code in most of SQL, Linux shell, and Python.
* Implemented Apache-spark code to read multiple tables from the real-time records and ϔilter the data based on the

requirement.

* Designed and Configured Azure Cloud relational servers and databases analyzing current and future business

requirements.

* Worked on migration of data from On-prem SQL server to Cloud databases (Azure Synapse Analytics (DW) & Azure

SQLDB).

* Worked on creating tabular models on Azure analysis services for meeting business reporting requirements.
* Working with Azure BLOB, ADLS and Data lake storage and loading data into Azure SQL Synapse analytics(DW).
* Stored final computation result to Cassandra tables and used Spark-SQL, spark-dataset to perform data

computation.

* Troubleshoot and resolve complex production issues while providing data analysis and data validation with

Cybersecurity.

* Performed data wrangling to clean, transform and reshape the data utilizing pandas library. Analyzed data using

SQL, Scala, Python, Apache Spark and presented analytical reports to management and technical teams.

* Creating the High Level and Low-Level design document as per the business requirement and working with offshore

team to guide them on design and development.

* Continuously monitoring for the processes which are taking longer than expected time to execute and tune the
* rocess.
* 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.

**Environment**: Azure Cloud, Azure Data Factory (ADF V2), Azure functions Apps, Azure Data Lake, BLOB Storage, SQL

server, Teradata Utilities, Windows remote desktop, ADLS, UNIX Shell Scripting, AZURE PowerShell, Data bricks,

Python, Azure Cosmos DB, Azure Stream Analytics, Azure Event Hub.

**Client: United Airlines, Chicago, IL Feb 2021 – Jan 2022**

**Data Engineer**

**Responsibilities:**

* Maintained and developed complex SQL queries, views, functions, and reports that qualify customer requirements on Snowflake.
* Responsible for developing, maintain, and monitor automated ETL pipelines using scripting languages such as Python, Spark, SQL and AWS services such as S3, Glue, Lambda, SNS, SQS, KMS, API Gateway.
* Responsible for developing reusable ETL pipelines with Terraform (IaC), automating Glue ETL jobs.
* Worked on building & supporting AWS architecture and using services like EC2, CloudWatch, ECS, Sage maker, ECS, Steps Functions, Lambda, Postgress.
* Expertise in designing complex Mappings and have expertise in performance tuning and slowly changing Dimension Tables and Fact tables.
* Evaluate Snowflake Design considerations for any change in the application
* Build the Logical and Physical data model for snowflake as per the changes required
* Define roles, privileges required to access different database objects.
* Define virtual warehouse sizing for Snowflake for different type of workloads
* Processed large data set using Python and Spark ETL jobs with Glues and EMR cluster.
* Responsible for data security, implemented DevSecOps principles for data lake implementations including data encryption and database user access controls and logging.
* Provision infra with terraform as a code: AWS cloud services: S3, EC2, EMR, RDS, Redshift, Lambda, API-Gateway, AWS glue ETL, etc.
* Worked as COE for promote Data Science methods and processes across different business functions.
* Experience on configuring auto scalable and highly available microservices set with monitoring and logging using AWS,

Docker, Jenkins and Splunk.

* Build the Silent Circle Management System (SCMC) in Elasticsearch, Python, and Node.JS while integrating with infrastructure services.
* Involved in developing java microservices which are interconnected in the AWS cloud also involved in consuming and building web services both and SOAP and RESTful.
* Developed Restful Microservices using Flask and Django and deployed on AWS servers using EBS and EC2.
* Collaborate with application architects and Product Owners to modernize infrastructure as a service (IaaS) application to Platform as a Service (PaaS)
* Worked on setting up a POC’s with cutting edge tools stack in Kubernetes, containerization, GPUs, and ML pipeline tools like Kube low, Pachyderm, Air low, Databricks, and TensorFlow

**Environment:** Snowflake, Presto, Spark, EMR, Restful and API, Teradata, AWS (S3, Lambda, RDS, Cloud Formation IAM, shell scripting, Python, boot/boto3, MySQL, RStudio, R, Linux, TensorFlow, CI/CD, Kubernetes, Jenkins, CirlceCI, Docker, Air low, Sage maker, DataSync, Glue, API Gateway.)

**Client: Microsoft, Seattle, WA Jun 2019 – Jan 2021**

**Data Engineer**

**Responsibilities:**

* Developed pipeline using Hive (HQL) to retrieve the data from Hadoop cluster, SQL to retrieve data from Oracle database and used ETL for data transformation.
* Analyzed and gathered business requirements from clients, conceptualized solutions with technical architects, and veri ied approach with appropriate stakeholders, developed E2E scenarios for building the application.
* 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.
* 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 lot of SQL statements and PL/SQL blocks by analyzing the execute plans of SQL statement and created and modified triggers, SQL queries, stored procedures for performance improvement.
* Developed Python Modules for Ansible Customizations.
* Used Ansible Playbooks to setup Continuous Delivery Pipeline. Deployed micro services, including provisioning AWS environments using Ansible Playbooks.
* Used Ansible to document all infrastructures into version control.
* Implemented Predictive analytics and machine learning algorithms in Databricks to forecast key metrics in the form of designed dashboards on to AWS (S3/EC2) and Django platform for the company's core business.
* Created and maintained various DevOps related tools for the team such as provisioning scripts, deployment tools, an development and staging environments on AWS, Rackspace and Cloud.
* Used PowerShell for Devops in Windows-based systems.
* Designed and developed a new solution to process the NRT data by using Azure stream analytics, Azure Event Hub and Service Bus Queue.
* Created Linked service to land the data from SFTP location to Azure Data Lake.
* Created numerous pipelines in Azure using Azure Data Factory v2 to get the data from disparate source systems by using different Azure Activities like Move &Transform, Copy, ilter, for each, Databricks etc.
* Responsible for estimating the cluster size, monitoring, and troubleshooting of the Spark databricks cluster
* Participated in features engineering such as feature generating, PCA, Feature normalization and label encoding with Scikit-learn preprocessing. Data Imputation using variant methods in Scikit-learn package in Python.
* Used Sqoop to move data from oracle database into hive by creating a delimiter separated files and using these files in an external location to be used as an external table in hive and further moving the data into refined tables as parquet format using hive queries
* Used Teradata utilities such as Fast Export, MLOAD for handling various tasks data migration/ETL from OLTP Source Systems to OLAP Target Systems.
* Developed spark programs using Scala API's to compare the performance of spark with HIVE and SQL.
* Designed and created Hive external tables using shared meta-store instead of derby with partitioning, dynamic partitioning, and buckets.
* Evaluated the performance of Databricks environment by converting complex Redshift scripts to spark SQL as part of 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.

**Environment:** R, Python, ETL, Agile, Data Quality, R Studio, Tableau, Azure Cloud , Azure functions Apps , Azure stream analytics , Azure Event Hub ,Data Governance, Supervised & Unsupervised Learning, Java, NumPy, SciPy, Hadoop, Sqoop, HDFS, Spark SQL, Pandas, PostgreSQL, AWS (EC2, RDS, S3), Matplotlib, Scikit-learn, Shiny

**Client: Next Sphere Technologies – Hyderabad, IN Nov 2017 – Jan 2019 Data Engineer**

**Responsibilities:**

* Created and presented models for potential holdings to fund managers. Achieved 20% better returns v/s historical performance and Predicted stock price 25% better than traditional figures.
* Facilitated User Acceptance Testing and identified KPIs in the Reference data to predict failure of client's payments.
* Fashioned with machine learning algorithms like linear regression, KNN and decision tree for trading problems, estimated machine learning algorithm's performance for time series data (stock price data).
* Calculated and presented data visualization report for daily returns, cumulative returns, simple moving averages, Sharpe ratio, portfolio value for stock performance optimization.
* Analyzed large data sets using pandas and used regression models using SciPy to predict future data and visualized them and used SQL to manipulate data
* Pro icient in importing/exporting large amounts of data from files to Teradata and Created Teradata speci ic physical data models which included primary indexes, secondary indexes and joined indexes.
* Implemented Dimensional Data Modeling to deliver Multi-Dimensional STAR schemas and Developed Snow lake Schemas by normalizing the dimension tables as appropriate.
* Worked with data compliance teams, Data governance team to maintain data models, Metadata, Data Dictionaries; define source fields and its definitions.
* Used HIVE to do transformations, event joins and some pre-aggregations before storing the data onto HDFS.
* Used Redshift, S3 within AWS along with Informatica Cloud Services to load data into s3 bucket.
* Worked with HDFS file formats like Avro, Sequence File and various compression formats like Snappy.
* Collaborated with Data engineers and operation team to implement ETL process, Snowflake models, wrote and optimized SQL queries to perform data extraction to it the analytical requirements.
* Written SQL Scripts and PL/SQL Scripts to extract data from Database and for Testing Purposes.
* Created data pipeline package to move data from amazon S3 bucket to MYSQL database and executed MySQL stored procedure using events to load data into tables.
* Used Vagrant to implement environment for microservices deployments and testing in Docker images.
* Designed, created, executed, and kept up answers for utilizing Docker, Jenkins, Git, and Puppet for microservices and

ceaseless arrangement.

* Contribute to the client's development standard practices of Microservice architecture by developing reusable product

extensions and writing knowledge articles based on experience.

* Designed and developed Tableau graphical and visualization solutions based on business requirement and expert with the backend data retrieval team, data mart team to guide the proper structuring of data for Tableau reporting.
* Worked on metadata in Tableau desktop to alter data types, assign roles, rename fields, joins, refresh data and change extract data options to prepare data sources to be utilized by users.

**Environment:** Hadoop, ETL operations, Data Warehousing, Data Modelling, Cassandra, AWS Cloud computing architecture, EC2, S3, Advanced SQL methods, NiFi, Python, Linux, Apache Spark, Scala, Spark-SQL, HBase

**Smart IMS, Hyderabad, India Jun 2014 – Oct 2017**

**Data Analyst**

**Responsibilities:**

* Explored data of SAP users to find any hidden trends in access requirements
* Worked according to the software development life cycle.
* Gathered requirements from remotely based business users and de ined and elaborated the requirements by holding meetings with the users (who are also Fifth-third employees).
* Involved in defining the source to target data mappings, business rules, business, and data definitions
* Extensively designed Data mapping and filtering, consolidation, cleansing, Integration, ETL, and customization of data mart.
* Document various Data Quality mapping document.
* Perform small enhancements (data cleansing/data quality).
* Worked on daily basis with the main frame team and lead data warehouse developers to Created Report-Models for ad-hoc reporting and analysis.
* evaluate impact on current implementation.
* Communicated with the third-party vendor to do the programming.
* Upgraded the present application by adding new functionalities and adding new reports.
* Wrote regression test cases, did smoke testing with users.
* Coordinated (with business users, data base administrator, mainframe team and testing team) in mirror to production testing.
* Conducted User Acceptance testing (UAT) and worked with users and vendor who build the system.

**Environment:** MS Of ice 2007, MS Excel, SharePoint, PowerPoint, MS Project, SQL Server, Erwin, Business Objects, MS Outlook