**Kranthi Reddy**[**kranthireddy0623@gmail.com**](mailto:kranthireddy0623@gmail.com) **Phone: 832-895-9123**

**Sr. Data Engineer** <https://www.linkedin.com/in/kranthireddy2306>

**PROFESSIONAL SUMMARY**

* **9+ years** of professional experience in information technology as Data Engineer with expert hand in the areas of **Database Development**, **ETL Development**, **Data modelling**, **Report Development** and **Big Data Technologies**.
* Experience on Migrating SQL database to **Azure data Lake**, **Azure data lake Analytics**, **Azure SQL Database**, **Data Bricks** and **Azure SQL Data warehouse** and controlling and granting database access and Migrating On premise databases to **Azure SQL Data lake store** using **Azure Data factory**.
* Experience working with Amazon Web Services(AWS) cloud and its services like **Snowflake,EC2,S3,RDS,EMR,Lambda,Redshift,Data Pipeline,** ETL and other AWS Services.
* Experience in Designing Business Intelligence Solutions with **Microsoft SQL Server** and using **MS SQL Server Integration**(SSIS), **MS SQL Server Reporting**(SSRS) and **SQL Server Analysis Services**(SSAS).
* Experience in building and architecting multiple **Data pipelines**, end to end ETL and ELT process for Data ingestion and transformation in **GCP**.
* Experience in **database design** to create scalable, efficient, and robust systems for collecting, storing, and accessing large amounts of data.
* Experience in moving data between **GCP** and Azure using **Azure Data Factory**.
* Extensive experience in using ER modeling tools such as Erwin and ER/Studio.
* Excellent proficiency in **Agile/Scrum** and **waterfall** methodologies.
* Proficient in data governance, Experience in Data Ingestion projects to inject data into Data Lake using multiple sources systems using Talend BigData.
* Experience in execution of Batch jobs through the data streams to SPARK Streaming.
* Good knowledge in streaming applications using **Apache Kafka**.
* Extensive familiarity with **Pyspark**, **Python Pandas, NumPy, Scikit-learn, seaborn, and Matplotlib packages**.
* Excellent knowledge in Data Analysis, Data Validation, Data Cleansing, Data Verification and identifying data mismatch, data quality, metadata management, master data management.
* Experience on Apache Hadoop ecosystem components like **Sqoop, Spark, Hive, HBase**.
* Experience in integration of various data sources with multiple Relational Databases like **SQL Server**, **Teradata**, and **Oracle**.
* Hands on experience in **Normalization** (1NF, 2NF, 3NF and BCNF) and Denormalization techniques for effective and optimum performance in OLTP and OLAP environments.
* Proficient in crafting SQL queries, **PL/SQL Packages**, **Stored Procedures**, **Functions**, **Triggers**, and optimizing query performance. Skilled in creating partitions, indexes, and aggregating tables to meet user requirements and reporting needs.
* Expert knowledge in **SDLC** (Software Development Life Cycle) and was involved in all phases in projects.
* Solid knowledge of **Data Marts**, **Operational Data Store**, OLAP, Dimensional Data Modeling with Ralph Kimball Methodology (Star Schema Modeling, Snow-Flake Modeling for **FACT and Dimensions Tables**) using Analysis Services.
* Followed best practices for Data visualization to effectively convey messages to the intended business audience.
* Specialize in the **healthcare sector**, specifically in handling pharmacy claims data for various insurance payers including **Medicare, Medicare Advantage,** and **Commercial insurance**.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Programming/Scripting Languages** | SQL, Python, Bash, UNIX Shell script |
| **ETL/ELT** | Azure ADF, SSMS, Talend Open Studio, Informatica, Data Stage, SSIS, Apache Flink |
| **Databases** | MySQL, SQL-Server, Teradata, MongoDB, Cosmo DB Oracle 10g/11g, PostgreSQL 9.3 |
| **Cloud Technologies** | Azure Functions, Data Lake Analytics, Databricks, Azure DevOps, Azure Active Directory, Cosmos DB |
| **Data warehouses/ Data lakes** | Snowflake, Big Query, Google Cloud Storage, S3, IBM Netezza, Teradata |
| **BI Tools** | Tableau, Power BI, SSRS, Python Visualization Libraries |
| **Other Tools** | Azure Data Lake, Data Bricks, Azure Data Factory, Azure Monitoring, Active Directory, Key Vault, SQL Azure, Hadoop, Map Reduce, Hive, Scoop, Spark, Horton Works, Microsoft Office, MS Visio, Box, Share point,Alteryx. |
| **Methodologies** | Agile, Scrum, Iterative Development, Waterfall Model, UML, Design Patterns |

**Professional Experience**

**Data Engineer  
UHG, NC Apr 2020 to Present**

**Responsibilities:**

* Extract Transform and Load data from Source Systems to Azure Data Storage services using combination of **Azure Data Factory, T-SQL, Spark SQL** and **U-SQL Azure Data Lake Analytics**.
* Created Pipelines in ADF using Linked Services/Datasets/Pipeline to Extract, Transform and load data from different sources like **Azure SQL, Blob storage** and **Azure SQL Data Warehouse**.
* Created triggers for Azure Data Factory (ADF) pipelines based on time and events.
* Developed Spark applications using **Pyspark** and **Spark-SQL** for data extraction, transformation and aggregation from multiple file formats for analyzing and transforming the data to uncover insights into the customer usage patterns.
* Created, provisioned different **Databricks** clusters needed for batch and continuous streaming dataprocessing and installed the required libraries for the clusters.
* Developed JSON Scripts for deploying the Pipeline in Azure Data Factory that process the data using sql Activity.
* Developed Data Integration Platform components/processes using **Informatica** Cloud Platform Azure SQL Datawarehouse, Azure Data lake Store and Azure Blob Storage technologies.
* Implemented data quality automation to automatically rectify data inconsistencies, resulting in a significant improvement in data accuracy from 90% to 95%.
* Optimized stored procedures, tables, and views, resulting in enhanced data accuracy and processing speed in **Teradata** environment.
* Created automated python scripts to convert the data from different sources and to generate the ETL pipelines.
* Executed process improvements in data workflows using **Alteryx** processing engine and **SQL**.
* Worked on creating Adhoc reports, Database Imports and Exports using **SSIS**.
* Practical knowledge of claims, allowances, payments, patient responsibilities etc. in the medical and pharmaceutical claims areas of multiple insurers.
* Wrote MapReduce jobs to generate reports for the number of activities created on a day, during a dumped from the multiple sources and the output was written back to HDFS.
* Worked on Talend ETL to load data from various sources to Oracle DB. Used tmap, treplicate, tsort and various other features in **Talend**.
* Implemented error handling in Talend to validate the data Integrity and data completeness for the data from the Flat File.
* Involved in modeling (Star Schema methodologies) in building and designing the logical data model into Dimensional Models.
* Authored and contributed to data pipelines, ETL processes, and data infrastructure code hosted on **GitHub**, showcasing proficiency in languages like Python, SQL.
* Provided application support during the build and test phases of the SDLC for their product.

**Data Engineer  
USAA, San Antonio Sep 2018 to Mar 2020**

**Responsibilities:**

* Followed Test driven development of Agile Methodology to produce high quality software.
* Conducted JAD sessions, wrote meeting minutes and documented the requirements.
* Worked with cloud providers and APIs for **Amazon (AWS) EC2**, S3, VPC with GFS storage.
* Worked with Data ingestion, querying, processing and analysis of big data.
* Developed server-based web traffic using RESTFUL API’s statistical analysis tool using Flask, Pandas.
* Performed tuned and optimized various complex SQL queries.
* Built Scala and spark based configurable framework to connect common Data sources like MYSQL, Oracle,Postgres and SQL Server
* Developed normalized Logical and Physical database models to design OLTP system.
* Extensively involved in creating **PL/SQL** objects i.e. Procedures, Functions, and Packages.
* Performed bug verification, release testing and provided support for Oracle based applications.
* Used Model Mart of Erwin for effective model management of sharing, dividing and reusing model information and design for productivity improvement.
* Extensively used **Hive** optimization techniques like partitioning, bucketing, Map Join and parallel execution.
* Worked with Real-time Streaming using **Kafka and HDFS.**
* Worked with **Alteryx** a data Analytical tool to develop workflows for the ETL jobs.
* Designed the data marts in dimensional data modeling using star and snowflake schemas.
* Wrote, tested and implemented Teradata Fastload, Multiload, DML and DDL.
* Used various OLAP operations like slice / dice, drill down and roll up as per business requirements.
* Wrote SQL queries, stored procedures, views, triggers, T-SQL and DTS/SSIS.
* Handled importing of data from various data sources, performed data control checks using Spark and loaded data into HDFS.
* Designed **SSRS** reports with sub reports, dynamic sorting, defining data source and subtotals for the report.
* Involved in Informatica upgrade process and testing the whole existing Informatica flow in new upgrade environment.
* **Informatica** as ETL tool to pull data from source systems/ files, cleanse, transform and load data into the Teradata.
* Designed and implemented importing data to HDFS using Sqoop from different RDBMS servers.
* Worked with **Sqoop** commands to import the data from different databases.
* Gathered SSRS reports requirements and created in Tableau.
* Designed and developed Map Reduce jobs to process data coming in different file formats like XML.

**Data Analyst/Data Modeler  
Convera, Washington DC Mar 2017 to Aug 2018**

**Responsibilities:**

* As a Data Analyst /Data Modeler, I was responsible for all data related aspects of a project.
* Created report on Cloud based environment using **Amazon Redshift** and published on Tableau .
* Developed Python APIs to dump the array structures in the Processor at the failure point for debugging.
* Worked extensively on ER/ Studio in several projects in both OLAP and OLTP applications.
* Worked at **conceptual/logical/physical** data model level using Erwin according to requirements.
* Created SQL tables with referential integrity and developed queries using SQL, SQL\*PLUS and **PL/SQL**.
* Performed Data Analysis and data profiling using complex SQL on various sources systems including Oracle.
* Developed the required data warehouse model using Star schema for the generalized model
* Implemented Visualized BI Reports with **Tableau**.
* Developed the required data warehouse model using **Star schema** for the generalized model.
* Worked on stored procedures for processing business logic in the database.
* Extensively worked on Viewpoint for **Teradata** to look at performance Monitoring and performance tuning.
* Performed Extract, Transform and Load (ETL) solutions to move legacy and ERP data into Oracle data warehouse.
* Developed and maintained data dictionary to create metadata reports for technical and business purpose.
* Managed database design and implemented a comprehensive Snow flake-Schema with shared dimensions.
* Worked **Normalization** and **De-normalization** concepts and design methodologies.
* Worked on the reporting requirements for the data warehouse.
* Worked on SQL Server concepts SSIS (SQL Server Integration Services), SSAS (Analysis Services) and SSRS (Reporting Services).
* Developed complex T-Sql code such as **Stored Procedures, functions, triggers, Indexes**, and views for the business application.
* Wrote a complex SQL, PL/SQL, Procedures, Functions, and Packages to validate data and testing process

**Big Data Developer  
FDA, GA Feb 2015 to Feb 2017**

**Responsibilities:**

* Involved in the Complete Software development life cycle (SDLC) to develop the application.
* Worked with different source data file formats like JSON, CSV, and ORC etc.
* Worked on Spark for improving performance and optimization of existing algorithms in Hadoop using **Spark-SQL** and **Scala**.
* Worked on creating Hive managed and external tables based on the requirement.
* Developed Spark jobs to clean data obtained from various feeds to make it suitable for ingestion into Hive tables for analysis.
* Experience in importing data from various data sources like MySQL and Netezza using Sqoop, SFTP, performed transformations using Hive, Pig and loaded data back into HDFS.
* Import and export data between the environments like MySQL, HDFS and deploying into productions.
* Worked on partitioning and used bucketing in HIVE tables and setting tuning parameters to improve the performance.
* Involved in developing Impala scripts to do Adhoc queries.
* Experience in Oozie workflow scheduler template to manage various jobs like Sqoop, MR, Pig, Hive, Shell scripts, etc.
* Involved in importing and exporting data from HBase using Spark.
* Involved in POC for migrating ETLS from Hive to Spark in Spark on Yarn Environment.
* Utilized Hive tables and HQL queries for daily and weekly reports. Worked on complex data types in Hive like Structs and Maps.
* Created Cassandra tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios.
* Collaborated with the infrastructure, network, database, application, and BI teams to ensure data quality and availability.
* Designing ETL processes using Informatica to load data from Flat Files, Oracle, and Excel files to target Oracle Data Warehouse database.
* Worked on Hortonworks distribution and responsible for Data Ingestion, Data Cleansing, Data Standardization and Data Transformation.

**Senior Analyst  
Wide business solutions, Hyderabad Jul 2013 to Jan 2014**

**Responsibilities:**

* Spearheaded analysis, design and development for building a common architecture and data repository for Apparel and Footwear sourcing data across the Geo's for data management and reporting needs.
* Work cross functionally with analysts, business insights, report developers, and stakeholders to ensure a tight Development/QA/UAT process and quality deployments.
* Evaluate, cleanse, extract/transform data for analytic purpose within the context of Big data environment and structure large data sets by applying standard data modelling methods.
* Design and develop efficient PySpark programs using cloud-based data platforms (EMR) to extract/transform/load data in between various data warehouse applications.
* Worked on Cloud Platform like **AWS** and possess good knowledge on different types of instances for optimal usage of clusters based on the requirement.
* Engineered a solution to optimize the ETL process of Alteryx to **Snowflake** ingestion process which takes around 7+ hours into a more reliable process by using **PySpark** which simplified and drastically reduced the data ingestion time to 7 minutes.
* Designed Hive tables over the parquet and csv files for loading and analyzing data.
* Involved in developing DAGS using Airflow orchestration tool and monitored the weekly processes.
* Identify gaps in data processes and drive improvements via continuous improvement loop by ensuring good data flow between databases and backend systems.