**Ram Gadde**

Contact No.: 316-305-4861

E-Mail: krishnagadde34@gmail.com

9+ years of IT experience in Data Engineering, Data Warehousing, Data Integration, and hands on experience in Cloud Data Analytics and development along with quick adaptability to new technologies.

**Professional Summary**

* 9+ years of IT experience with special emphasis on Analysis, Design, and Development of **ETL** methodologies in all the phases of the Data Warehousing life cycle.
* 5+ years of experience in Data engineering with large data sets of data, Data Acquisition, Data Validation, Data modeling, Data Visualization.
* Experienced in Extraction, Cleansing, Integration, and Loading of data from/to various data sources using **Python**, Apache **Spark** and **ETL** tools.
* Worked on Data Engineering libraries in **Python** such as Pandas, NumPy, PyMongo, PyArrow, PySpark, Matplotlib, Seaborn, Beautiful Soup, Orange, Rpy2, neurolab.
* Experience in **AWS** services like EC2, S3, EMR, Lambda, Glue, Redshift, Athena and Cloud Watch.
* Hands on experience in various big data eco stack – **HDFS,** MapReduce**, Apache Hive**, Flume, Sqoop, Zookeeper, Oozie and real-time processing framework **PySpark**.
* Extensive & diversified experience in loading and maintaining Data Warehouses and Data Marts using IBM Datastage 11.7/9.1 ETL processes**.**
* Deep understanding of Big Data using Hadoop, MapReduce, NoSQL and distributed computing tools.
* Experienced in integration of various data sources (**DB2-UDB, Hive and Oracle**).
* Extensively worked on various stages like ODBC, Transformer, Sequential file, Join, Lookup, Sort, Aggregator, Copy, Dataset, Merge, Oracle, Remove duplicates & DB2 stage.
* Experienced in Data modeling (Dimensional & Relational) concepts like Star-Schema Modeling, Snowflake Schema Modeling, and Fact and Dimension tables.
* Good understanding of NoSQL databases and hands on work experience in writing python code on No SQL data bases like **Mongo DB**.
* Experienced in **UNIX** scripting as a part of file manipulation, scheduling and text processing.
* Involved in functional requirements gathering, analysis, design, development, test and implementation.

**Technical Skills:**

|  |  |
| --- | --- |
| Languages | SQL, Python, Unix, HTML.  |
| Cloud | AWS (S3, EC2, Redshift, Glue, EMR, Lambda, AWS Data pipeline), Snowflake |
| Databases | Oracle, Hive, Snowflake, SQL Server, Mongo DB, DB2, MySQL |
| Big Data Tools | Hadoop, Hive, Sqoop |
| Libraries | Pandas, NumPy, PyArrow, PyMongo, PySpark, Bokeh. |
| ETL Tools | IBM Datastage 11.5/9.1, AWS Glue, Alteryx 11.5 |
| Database Tools | TOAD, SQL Plus, Aqua Data Studio, Db2Visualizer and Rapid SQL  |
| IDE | VS Code, PyCharm, Sublime text, Spyder |
| Scheduling | Autosys, Control-M, Airflow, ESP |

**Professional Experience:**

**USAA, San Antonio, TX May 2020 to Till Date**

**Role: Sr. Data Engineer**

**Project:** **CFO Operations Analytics**

The project objective is to automate and leverage the Enterprise data related to operating applications, financial analytical data, strategic budget allocations into Snowflake cloud Data Lake and AWS analytical environment. This is used to forecast and analyze revenue variance across the insurance line of business, issued and serviced every month pertaining to day-to-day operations according to the company’s policies.

**Responsibilities:**

* Developed ETL processes in AWS Glue to migrate regulatory affairs data from external sources like S3 to Parquet files.
* Developed the PySpark code for AWS Glue ETL jobs and for EMR.
* Created python data pipelines to extract csv and JSON files using Boto3 API to connect to S3 and then to load data into AWS Redshift.
* Designed config-driven ELT pipeline to process different file patterns from a raw AWS S3 location to Refined location using AWS server-less services, which removed redshift usage and thus, reducing cost and improving data availability.
* Created tables in Redshift Spectrum for S3 data and improved query performance of spectrum tables by S3 Seq scan and S3 Hash aggregate steps.
* Performed data cleansing and did transformations using pandas and NumPy packages in python.
* Extracted flat file files, JSON files from S3 buckets and parse the data and load to Snowflake data warehouse.
* Wrote SQL queries against Snowflake and created Snow pipe for continuous data load and used COPY to bulk load the data.
* Worked with streaming team on Kafka workflow to pick up the data from rest API server, from data lake as well as from SFTP server and send that to Kafka.
* Deployed AWS resources by incorporating Terraform, which improved infrastructure resource management.
* Worked with OpenShift platform in managing docker containers and Kubernetes clusters.
* Used Kubernetes to orchestrate the deployment, scaling and management of docker clusters.
* Worked on scheduling Control-M jobs and airflow DAGs.

**Environment:** Python 3.8.x, Snowflake, AWS - S3, Redshift Spectrum, Glue, Lambda, Athena, Kafka, PySpark, SnowSQL, Hive, HDFS, Oracle, Control-M, Docker, Kubernetes.

**Wells Fargo, Winston-Salem, NC July 2018 to May 2020**

**Role: Data Engineer**

**Project:** **Deposits and Lending Analytics**

The Enterprise blueprint data lake project is analytical data lake project which maintains the Enterprise data for all the departments of the bank. One of the subsets is deposits and lending project which consolidates the customer data under Data Management Insights (DMI) to load Data Lake to create reporting dashboards for deposits and lending analytical environment.

**Responsibilities:**

* Collaborated with cross-functional team in support of identifying modeling method (s) and determining the appropriate statistical and analytical methodologies to solve business problems within specific areas of expertise.
* Utilized Apache Spark framework to do the processing of ETL jobs to load data into Hive tables.
* Extracted, transformed and loaded as per requirement and generated CSV data files into specified AWS - S3 location.
* Created a Lambda deployment function and configured it to receive events from your S3 bucket.
* Built ETL pipelines to populate a unified data lake that serves the aftermarket.
* Automated tasks for lending department and optimized the time with Python for the big data using pandas library.
* Worked with Avro and parquet files formats and used compression techniques to leverage the storage in HDFS.
* Imported data from AWS S3 and into PySpark data frames and performed transformations and actions on data frames.
* Worked on AWS Data Pipeline to configure data loads from S3 to into Redshift.
* Developed NiFi workflow to pick up the multiple retail data files from ftp location and move those to HDFS on daily basis.
* Created external tables on top of Glue Data jobs with Athena create table statements and then load partitions with MCSK REPAIR commands.
* Worked on data extraction, aggregations and consolidation of lending data within AWS Glue using PySpark.
* Importing and exporting data into HDFS and Hive using Sqoop.
* Created Hive Fact tables on top of raw data from different retailer’s which indeed partitioned by Time dimension key, Retailer name, Data supplier name which further processed pulled by analytics service engine.
* Worked on scheduling Autosys jobs and created Data Quality Scripts using SQL and Hive to validate successful data load and quality of the data.

**Environment:** Python 3.6.x, PySpark, AWS - EC2, S3, Redshift, Glue, EMR, Athena, Lambda, Apache NiFi, Hadoop, HDFS, Hive, HBase, Flume, Sqoop, Kafka, PySpark, Autosys, SQL.

**VF Corporation, Greensboro, NC July 2017 to July 2018**

**Role: Data Engineer**

VF Corporation is one of the world’s largest global leaders in branded lifestyle apparel, footwear and accessories which outfit consumers around the world with its diverse portfolio of iconic lifestyle brands. **Order prebilling and CRM Analytics**

This project consolidates the customer data for different brands to load CRM DataMart to access information daily from cubes into one analytical environment built for General Merchandise (outdoor and action sports clothing and licensed product). This will allow stakeholders of all portfolios for respective brands to have the information for any given time frame and eliminate the activities of reconciling numbers between departments.

**Responsibilities:**

* Involved in the design, build and management of large-scale data structures and pipelines and efficient Extract/Load/Transform (ETL) workflows/feeds.
* Participated in design and developing application leveraging MongoDB which stores processed data into MongoDB.
* Work closely with data scientists to assist on feature engineering, model training frameworks, and model deployments at scale.
* Design, development and implementation of ETL pipelines using python API (PySpark) of Apache Spark on AWS Redshift.
* Developed Apache Spark jobs using python in test environment for faster data processing and used Spark SQL for querying.
* Experience in ETL operations to use Spark SQL on Hive Tables.
* Built scalable distributed data solutions using EMR cluster environment with Amazon EMR 5.6.1
* Developed Python scripts to automate data sampling process. Ensured the data integrity by checking for completeness, duplication, accuracy, and consistency.
* Wrote Python routines to log into the websites and fetch data for selected options.
* Managed Amazon redshift clusters such as launching the cluster by specifying the nodes and performing the data analysis queries.
* Created and modified batch scripts to ftp files from windows server to Linux server.
* Used Spark for Parallel data processing and better performances.

**Environment:** Python 3.6.x, PySpark, Spark SQL, MongoDB, AWS - S3, Redshift, RDS, EMR, Alteryx 11.5, IBM Data Studio 3.1.1, Oracle 11g, UNIX Shell Scripting, Jenkins.

**Erie Insurance, Erie, PA Jan 2014 to July 2017**

**Role: ETL Developer**

Erie Insurance is a multi-line insurance company which provides Personal and Commercial **P&C Insurance** offering auto, home, commercial and life insurance in 12 states.

The AdHoc project will support the Claims Management System (CMS), Policy Management System and other third-party vendors. The data is being used by reporting systems across IT and business users of functional organization.

Claims Refresh: The objective of the project is to convert and maintain the enterprise legacy customer claims from CMS (Claims Management Services) to ECC (Erie Claims Center). The data includes claims of all lines of business which are being converted and loaded to a consolidated data store which is a data repository that makes the users accessible to historical and current claims data.

**Responsibilities:**

* Understand business requirements by participating in meetings, also analyzing the data model and data elements.
* Collaborated with business team in, Low Level design document for mapping the files from source to target and implementing business logic.
* Responsible for developing, testing, supporting and maintaining for the ETL (Extract, Transform and Load) processes using Data stage Designer.
* Extensively used ETL to transfer and extract data from source files (Flat files and Oracle) and load the data into the target database.
* Involved in various projects related to Data Modeling, System/Data Analysis, Design and Development for both OLTP and Data warehousing environments.
* Design the datastage jobs extract data from ODS systems and load in to staging and later write into dimension tables.
* Converting the server jobs into parallel jobs and validate the data. After the validations, developed data stage jobs for preparing and loading into Ad-hoc application.
* Developed **multi-instance** jobs.
* Tuned transformations and jobs for performance enhancement.
* Implemented multi-node declaration using configuration files (APT\_Config\_file) for performance enhancement.
* Designed and developed job sequences to run and automate multiple jobs.
* Maintained Data Warehouse by loading dimensions and facts as part of project. Also, worked for different enhancements in FACT tables.
* Involved in ongoing production support and process improvement bydebugging of jobs and addressing production issues like data issues, ENV issues, performance tuning and enhancements.

**Environment:** IBM WebSphere Data Stage 11.5 and 8.5, DB2 UDB 9.8, IBM DB2 9.1, UNIX, Oracle, SQL Server, Embarcadero Rapid SQL.

**University of California, Oakland, CA May 2013 to Jan 2014**

**Role: ETL Datastage Developer**

University of California Office of President - Information Technology Services (UCOP - ITS) is the information technology office which carries the operations of all locations of the university. The main purpose of DSS (Decision Support Systems) data warehouse is designed to deal with many data warehouse applications such as UES (Undergraduate Experience Survey), Earnings, student accounts & payroll. The system receives input feed, process the data according to the business logic, creates load files and loads them into corresponding tables in DB2 database. This data will be used by the reporting system to generate report files.

**Responsibilities:**

* Involved as ETL developer during analysis, planning, design, development, and implementation stages of multiple projects like UCUES (UC Undergraduate Experience Survey) and UC Path using IBM WebSphere Datastage.
* Studying thebusiness requirement and worked extensively with Project Manager, BA’s and users to effectively capture the user specifications.
* Developed parallel jobs which loads dimension tables and maintained surrogate key generation in multiple runs.
* Designed ETL jobs incorporating complex transform methodologies using Data Stage tool resulting in development of efficient interfaces between source and target systems.
* Used DB2 stage to load data into mart tables and DB2 bulk load stage to load data into staging tables.
* Extensively dealt with change capture techniques for implementing slowly changing dimensions process.
* Extracted the data from source File systems and loading it into staging, base and BI area.
* Familiar with import/export of Data Stage Components (Jobs, DS Routines, DS Transforms, and Table Definitions) between Data Stage Projects and multiple jobs compile utility.
* Involved in the migration of Data Stage jobs from development to QA and then to production environment.

**Environment:** IBM WebSphere Data Stage 8.7, DB2 UDB 9.8, Flat files, UNIX-AIX, CA Erwin, Autosys, SQL, PL/SQL, UNIX, DB2 Visualizer.

**Education:**

* Master of Science in IT at Southern New Hampshire University, Manchester, NH | 2012.