****

**Krishna kumari Gunampalli**

**512-774-1053**

**krishna.gunampalli@gmail.com**

**https://www.linkedin.com/in/krishna-gunampalli/**

**Professional Summary:**

* 12+ years of Information Technology experience in the field of Design, Development, Support of Data Warehouse applications involving **AWS Glue, S3, Redshift, Greenplum, Snowflake, Pyspark, Teradata, Oracle, Talend, ODI, Informatica, Databricks, OBIEE, Thought spot & Tableau.**
* **Certified on AWS Solution Architect, AWS Cloud Practitioner, Talend Big Data, Data Bricks Spark Scala Developer, SAFe 6 Agilist.**
* Highly skilled and motivated data engineer with a strong background in SQL, Unix, and Linux scripting.
* Have worked on Healthcare, Oil & Gas and Finance verticals on several technologies.
* Experienced DW-BI professional with excellent business and the technology skills leading large global teams to deliver complex Business Intelligence, DW and Analytics projects.
* Experience in analyzing the impact of the various applications that would be affected by a change that is required at an application.
* Perform various data mining and data cleansing techniques on raw data by coordinating with Platform, Business & Architecture team.
* Strong experience in building data pipelines utilizing AWS cloud including S3, Glue, Redshift, Lambda, SQS, SNS. Step Functions, Event Bridge
* Created ETL jobs with Informatica, Talend, ODI, AWS Glue.
* Dedicated and experienced data engineer with a strong background in Python. Skilled in designing and implementing data pipelines and ETL processes to support data-driven decision-making.
* Extensively worked on SQL, PostgreSQL, Teradata, Redshift, Snowflake, PySpark, Python, and AWS services.
* Excellent knowledge and experience on process implementation, Automations and Best Practices.
* Implement performance tuning techniques for relevant technologies.
* Proficient in SQL, Unix, and Linux scripting.
* Hands on experience with analysis, design, development, support, and troubleshooting. Strongly worked in association with team members and provided guidance & support to peers.
* Driving and performing demonstrations of custom DWH Solutions to ensure design and implementation are on the right track and meet the business requirements.
* Strong core product development experience within Agile environment.

**Technical Profile:**

|  |  |
| --- | --- |
| **Operating System** | Windows, Linux,Unix |
| **Environment** | DWBI Systems |
| **Database** | Greenplum, Redshift, Teradata,Snowflake, Oracle, Aurora |
| **ETL Tools** | Talend, ODI, Informatica, Snow SQL, AWS Glue,Abinitio |
| **Reporting Tools** | Tableau, OBIEE, ThoughtSpot |
| **Languages** | Python, SQL and PL/SQL. |
| **Cloud** | Amazon Web services (AWS). |
| **Other** | GITHUB, Jenkins, MS Excel, Cronacle 7 & 9, Support Central, HP Quality center, SNS,Lambda,Step functions,Event Bridge |

|  |  |
| --- | --- |
| **Management and Leadership Skills** | * Proven ability to manage through others, Strong decision making and problem-solving skills.
* Excellent documentation and time management skills
* Handling new client enquiries and acting as the face of the business.
* An ability to build rapport and trust quickly with work colleagues.
* Able to prioritize tasks and workloads in order of importance.
* Track record of delivering results with deadlines.
 |
| **Training & Development**  | * Highly skilled in organizing training workshops on Greenplum, Talend, Teradata, Oracle, Informatica, OBIEE, Thought Spot.
* Highly skilled in organizing training workshops on DW & BI Architectures.
 |

**Professional Experience:**

 **Mar 2023 – Present**

**OXY- Houston, Texas**

**Data Engineer/Data Analyst/Lead developer**

**Tetris** is showing the well, bore and Rod lift services information, which is coming from different sources like Stibo, SAP, Nexus and ODS. This project requires expertise in PySpark, AWS glue, S3, Redshift, Airflow, Dynamo DB, Athena. The source data is coming in different file format like xml,csv. Need to convert this into parquet and storing this into S3 data lake after applying transformation and sending it to redshift for BI purpose.

**Responsibilities:**

* Played Data Engineer role.
* Created DBT ETL jobs and Glue jobs with PySpark scripts and created glue catalog tables using crawlers on Raw zone and curated zones.
* Converting the different type of source files like xml, csv into parquet applying the transformation and CDC on top of Raw zone before loading it into Curated zone of S3
* Created logs in Dynamo DB by creating tables for the glue jobs.
* Developed complex SQL queries for data extraction, transformation, and aggregation, improving query performance by 40%.
* Analyze on the new requirements and discuss with the Business users to identify the Gaps and suggested changes to done on the URS and suggested design recommendations as well in source side that were accepted and implemented. Developed Route caching strategies, web workers and service workers.
* Estimation for the newer enhancement requests and Project schedule preparation.
* Impact analysis on the various applications for any code changes.
* Closely working with users on the UAT approvals and testing different test scenarios.

**Environmen**t**:** AWS, RedShift, Glue, S3, Glue Crawler, Dynamo, Python, Athena, SNS, Lambda, Step functions, Event Bridge GitHub, Snowflake, DBT, Airflow, Databricks, Support Central and Oracle.

**Baker Hughes, Houston, TX May 2022– Feb 2023 (Houston, TX)**

**Data Analyst/Engineer /Lead Developer Feb 2018 – Apr 2022(Bangalore, India)**

DevOps Balanced Model provides robust support governance across Baker Hughes P&Ls along with handling enhancement and development activities for BI applications. This project requires expertise in cutting edge technologies related to Data modelling, Architecture, Design, Database technologies, especially Greenplum, Snowflake, Databricks, AWS platform, AWS S3 and Redshift, along with Talend ETL, ODI ETL, Abinitio, OBIEE, Tableau and ThoughtSpot. Skilled professional who possesses vast knowledge of Data Warehousing, Requirement Engineering, Data Analytics tools, End-to-End Architecture and Design Solutions, Client Management, Integration and Data visualization tools experience

**Responsibilities:**

* Acting as a Technical lead for Supply Chain & Finance Domain applications which uses Redshift, Snowflake, Greenplum, Databricks, Talend, ThoughtSpot and Tableau
* Created ETL jobs with Talend, ODI, Abinitio and Redshift &Green plum scripts.
* Find the root cause for runtime issues and deploy the code into production as client requirement through standard process.
* Analyze on the new requirements and discuss with the Business users to identify the Gaps and suggested changes to done on the URS and suggested design recommendations as well in source side that were accepted and implemented.
* Designed a Python script to automate data validation checks on large datasets, identifying anomalies and inconsistencies.
* Proficient in SQL, Unix, and Linux scripting.
* Estimation for the newer enhancement requests and Project schedule preparation.
* Impact analysis on the various applications for any code changes.
* Providing 100% SLA on supporting user requests.
* Closely working with users on the UAT approvals and testing different test scenarios.
* Project and Resource planning based on Release.

**Environmen**t**:** Greenplum, Snowflake, Databricks, RedShift, Glue, S3, Python, Athena, SNS, Lambda, Step functions, Event Bridge GitHub, Snowflake, DBT, Airflow, Talend, Tableau, GitHub, Support Central and Oracle.

**Capgemini June 2016 – Jan 2018**

**GE Oil & Gas, India**

**Data Engineer/Tech Lead/Sr Developer**

QWIK deals in showing Inventory Segmentation show-casing Inventory Costs at various stages WIP, Receiving and on hand. Showing Supply Demand Trend Information for different Order types like work order, planned order, purchase order. Greenplum, Talend, Informatica 9.1, Tableau, SVN, Support Central and Oracle.

Asset Install base is mainly using ERP (Pressure control, Electric Submersible Pump Rod-lift Systems, QUINN Pump & Surface Pumping System) and CRM (Service Max) Data sources. User should be able to see history of an asset deployment (Identified by part number & Serial number) and history of service records on equipment.

**Responsibilities:**

* Extensively worked on Greenplum, Talend and Tableau.
* Created Talend jobs for the Green plum scripts /excel sources/API.
* Analyze on the new requirements and discuss with the client community to identify the Gaps and suggested changes to done on the URS and suggested design recommendations as well in source side that were accepted and implemented.
* Estimation for the newer enhancement requests and Project schedule preparation.
* Impact analysis on the various applications for any code changes.
* Facilitate communication between Development team and Business stakeholders.

**Environmen**t**:** Greenplum, Talend, Informatica 9.1, Tableau, SVN, Support Central and Oracle.

**IGATE Global Solutions Ltd** (a part of **Capgemini** group)  **Oct 2010 – May 2016**

GE Healthcare

**ETL Developer / Teradata**

Americas Service Invoicing Report Generation" is a module on top of BI Siebel America and GLPROD that generates a CSV file for consumption by a third party called "Coakley". Coakley is responsible for printing and mailing or emailing the invoices to the customers per their chosen formats. Coakley uses this CSV file to generate PDFs at their end, completely outside of GE to print and/or email the invoices to our end customers.

Cost To Serve (CTS) application would provide BI Solution on Service-related product’s information and cost. User can analyze Globally and identify the Costs which were spent on various service events.

Cost to serve (CTS) is mainly used for making strategic decision on the product delivered and corresponding defects, this analysis further helps in delivering better product for tomorrow with defects fixed.

The DART Gen 2 application is a reporting tool which generates and displays Infantile Failure Rate (IFR) & Mission Failure Rate (MFR) metrics, Service costs, part failures, installed base counts, and remote service utilization information from data that reside in the GEHC Enterprise Warehouse (EDW) Teradata repository. The metrics and trends generated by this application provide Healthcare Systems (HCS), Surgery, and Healthcare IT (HCIT) P&Ls a unified approach to view and manage product quality at the product and business level.

**Responsibilities:**

* Analyzed the requirements to identify the necessary tables that need to be populated into the staging area.
* Involved in analyzing the source data coming from different Data sources such as Oracle, SQL Server, Flat files and identifying data anomalies in the source data.
* Developed and maintained ETL workflows using Apache Airflow to extract, transform, and load data from various sources into the data warehouse.
* Designed and implemented data models using SQL for efficient data storage and retrieval, facilitating data analysis and reporting.
* Optimized SQL queries and database performance, resulting in a 30% reduction in query execution time.
* Developed Python scripts for data cleansing, data transformation, and data enrichment, resulting in improved data accuracy and quality.
* Created mappings using transformations like Source Qualifier, Aggregator, Expression, Look up, Router, Filter, Update Strategy, Joiner, Union, Stored procedure, and Sequence Generator.
* Created Teradata Bteqs scripts to load the data into Data warhouse after applying transformations.
* Created Python-based ETL scripts to extract, transform, and load data from external APIs and web services into the data warehouse.
* Used Error handling strategy for trapping errors in a mapping and sending errors to an error table.
* Used FTP services to retrieve Flat Files from the external sources.
* Used Parameter files to define values for parameters and variables used in a mapping, sessions, worklet or workflow.
* Used workflow manager to create workflows, sessions and also used various tasks like command, event.
* Involved in Performance tuning at various levels including Target, Source, Mapping and Session

**Environment:** Teradata 13.0, Informatica 9.1, OBIEE 11G, SVN, Cronacle, upport Central and HP Quality center, Teradata SQL and Oracle and Shell Scripting

**Education:**

• Bachelors in Computer Science, from Sri Krishna Devaraya University India in 2003.

• Masters in Computer Applications, Sri Venkateshwara University, India in 2006.