|  |  |
| --- | --- |
| **KRISHNAPRASATH RAVICHANDRAN** | Mail: krishna2prasath@gmail.com |
| ***Data Engineer*** | Mobile: +1 (646)322-3072 |
|  | <https://www.linkedin.com/in/krishna2prasath/> |

|  |
| --- |
| **Career Objective** |

To use every opportunity available to prove my proficiency and to take every challenge as a learning process to update my technical and soft skills.

|  |
| --- |
| **Key Skills and Competencies** |

* IT professional with **10+ years** of Experience in **ETL Development**, Data analysis and DWBI Integration applications
* Extensive experience in ETL tools like **Data stage / Informatica on Cloud-Administrator, Designer** and **Director**
* Worked extensively in migrating data from legacy database to **Azure** **Snowflake** cloud database.
* Experienced in designing ETL solutions for applications handling real time data (thru API’s and extracts)
* Experienced in orchestrating and scheduling data pipelines using Apache **Airflow**.
* Proficient in designing and configuring complex workflows (DAGs) in **Airflow** for data extraction, transformation, and loading (ETL) processes.
* Proficient in configuring and managing connections between various data sources and destinations in **Airbyte**.
* Experienced in projects related to ETL development, Data Modeling, System/Data Analysis, Design and Development for Data warehousing environments.
* Extensive experience in **data modeling** and designing ETL solutions for data warehousing environments.
* Skilled in **Python** programming and scripting for automation processes and proof of concepts (POCs)
* Expertise in designing end-to-end dynamic flows using ETL to implement process automation and regression testing.
* Possess strong knowledge of database management in writing **complex SQL** queries, **stored procedures** (SP), database tuning, query optimization and resolving key performance issues in SQL, Oracle and Teradata databases.
* Worked extensively with **Data migration** and extraction using **ETL Processes** for DWH to handle 100+ Million data.
* Skilled in using Python for automation of various processes, including data pre - processing, data cleaning, and data validation.
* Extensive experience in data modeling tools such as Visio to design and maintain data models.
* Extensive experience in Python libraries like NumPy, Pandas, SciPy, and Matplotlib for data manipulation, analysis, and visualization
* Familiarity with Python-based data visualization tools like Seaborn and Plotly
* Expertise in designing Sequence and Parallel jobs using Datastage to populate tables in Data Warehouse and Data marts.
* Expertise in Master Data Management concepts, Methodologies and building **MDM** solutions.
* Experienced in scheduling sequence and parallel jobs, using **UNIX/Python scripts** and **Scheduling tool - Autosys.**
* Extensive experience in **Unit Testing, Functional Testing, Integration Testing** and **Regression Testing**
* Experienced in creating components using Markit EDM and created **web interface (UI)** using UI workflow to facilitate ease access of the Markit EDM reports to the end-users.

|  |
| --- |
| **Educational Qualification** |

* Pursuing Post Graduate Diploma in **Artificial Intelligence and Machine Learning** in **University of Texas, Austin**
* Completed **Bachelor of Engineering (B.E.,)** from Anna University, Chennai, India

|  |
| --- |
| **Technical Skills** |

**Domain** : Banking, Retail, Real Estate and Asset Management

**ETL Tools** : Informatica (IICS), Datastage 11.7,Data Build Tool (DBT), Airbyte

**DB** : Oracle, PostgreSQL, DB2, Teradata, SQL Server, My SQL, Snowflake, Netezza, Airbyte

**Cloud** : Google Cloud Platform (GCP),Azure

**Computer Language** : UNIX, Python, C, Reltio MDM, Mainframes, Reltio, HTML and Java

**Scheduler** : Autosys, Control - M, Airflow

**Repository** : SVN, Tortoise SVN, GitHub, Bit Bucket, Urban Code Deploy

**Development Methodology** : Agile model and Waterfall model

**Ticketing** : JIRA

|  |
| --- |
| **Project Details** |

**USAA, Plano, TX (May 2022 - May 2023)**

***Project* : SQL Development and Snowflake Data Migration**

***Role* : Data Engineer Lead**

***Tools used* :** Datastage 11.7, Azure Snowflake, DBT, Python, DB2, Control - M, Jira, GIT, Urban Code Deploy, Netezza

**Description**: Project deals with development of transformation SQL in DB2 for month end flows for reporting purposes. It also deals with conversion of transformation SQLs in Snowflake which is already running in production in Netezza. Generic data validation jobs are developed in IBM Datastage to compare the data loaded in DB2 and Snowflake databases.

**Roles and responsibilities**

* Extensively worked with Data Modeler in designing intermediate tables and it’s DDL in Snowflake
* Developed transformation SQL in Snowflake to intermediate tables.
* Involved in designing end to end architecture for ETL Skelton
* Involved in migrating transformation SQL from Netezza database to Snowflake database.
* Extensively worked with stakeholder’s right from requirement gathering till production deployment
* Designed **dynamic ETL flows** using run-time column propagation (**RCP**) to reuse components to validate the data loaded in DB2 and Snowflake
* Experienced in designing dimensional data models for data warehousing environments.
* Experienced in data modeling tools Visio to design and maintain data models.
* Communicate the project risks or issues encountered during the Analysis, Design, and Development phases and bring them to the Project Core team.
* Developed ETL jobs to **implement** data warehouse loads using Slowly Changing Dimensions **(SCD) Type 2**
* **Performed code reviews** against other developer’s Datastage jobs and to suggest best suited solution.
* Documented artifacts including requirement analysis, unit test plans and traceability matrix to ensure that the output is as per the requirements.
* Skilled in using Python for automation of various processes, including data pre - processing, data cleaning, and data validation.
* Created **job sequencers** and **Control -M** jobs to schedule various ETL Batch processes using **Control -M** scheduler.
* Provided Root Cause Analysis (**RCA**) for technical issues raised and provided solid updates to the stakeholders.
* **Agile**: Follow Agile methodology to breakdown EPIC into technical Stories, Tasks, Analyze and Finalize Requirements, Solution and Design Review approvals, Development, Testing, Autosys jobs, Pre-load data validation, DS & Non-DS Migration, Historical data load, Production data validation and Ongoing Support activities
* Prepare Implementation plan, backups and back out plan and support the release activities.

**Bank of America, Plano, TX (Jun 2021 - Apr 2022)**

***Project :* MDM Decommission & MDM Bridge LLE Automation**

***Role* : ETL Datastage Designer cum Developer**

***Client*** : **Bank of America**

***Tools used* :** Datastage 11.7, DB2, Autosys, UNIX, Python, Tortoise SVN, Reltio MDM

**Description:** Project is to implement Reltio MDM logic of product and pricing domain using IBM Datastage and to create post load validation jobs to compare the MDM table load and IBM Datastage table load. It also deals with scaling up an environment similar to Production environment, to ensure all processes of Batch cycle runs in scheduling automation to eradicate manual error and intervention and to ensure that files are delivered within deadline along mailer notification in automated way. A new process is built integrating Jar and Data stage to compare the data of the tables across upstream and our application.

**Roles & Responsibilities:**

* Initiated the study and documentation of existing functional flows of **Product** and **Pricing** domains in **Reltio MDM** amidst limited guidance, laying the groundwork for future team collaborations.
* Devised **pilot ETL Datastage** jobs for mappings of Product and Pricing domains in Reltio MDM, showcasing proficiency in handling complex transformation logic.
* Advocated for ETL Datastage utilization to reduce Reltio MDM dependency, successfully demonstrating cost-effectiveness to stakeholders.
* Led the transition of mappings from **Reltio MDM to ETL IBM Datastage**, requiring in-depth application knowledge and technical expertise.
* Developed intricate mappings spanning from low to high complexity tasks, including enhancements of older processes to bolster the overall ETL pipeline.
* Established ETL data validation flows to validate table loads from Reltio MDM and IBM Datastage, safeguarding data integrity and consistency.
* Authored Python scripts for automation of data pre-processing, cleaning, and validation tasks, improving data quality and operational speed.
* Orchestrated job sequencers and **Autosys** JILs for ETL Batch process scheduling using Autosys scheduler, enhancing automation and reducing manual errors.
* Saved the client 150K USD annually by migrating the entire process to ETL from Reltio MDM and eliminating manual intervention and support team dependency.
* Showcased strong team collaboration skills, contributed to problem-solving brainstorming sessions, and played a significant role in fostering an innovative work culture.
* Collaborated with stakeholders for requirement gathering, ensured their accurate implementation by effective team communication.
* Actively participated in comprehensive testing processes, from unit testing to system testing, to ensure the stability and reliability of ETL processes.

**UST Global, India (Jul 2018– May 2021)**

***Project*  : Merchandising Allowance Dashboard**

***Role* :ETL Datastage Lead Developer**

***Client* :** Albertsons is an American grocery company founded and headquartered in Boise, Idaho

***Domain* :**Retail

***Tools used* :**DataStage 11.3, Oracle 11g, Azure Snowflake, UNIX, Autosys, SVN

**Description:** This project eradicates the manual effort to find the total allowance offered to customers/stores using POWER BI reports. Scope of ETL flow is to extract from source tables, implement the business logic which involves complex analytical and arithmetic operations based on allowance types, to calculate allowance amount and to load the same in Teradata tables, which is the source of the Power BI system.

**Roles & Responsibilities:**

* Played vital role in **all phases of Software Development Life Cycle** (SDLC) such as Requirements analysis, Design, Development, Testing (Unit, Integration, Regression, and User Acceptance), Production Deployment for data warehousing and business intelligence projects.
* Facilitated technical sessions with stakeholders to gather relevant inputs required for design and development.
* Analyze source data, technical metadata, profile source data and create data mapping and transformation rules based on business requirements.
* Designed and developed ETL jobs by identifying various source files and applying transformations and **ETL logic** to them to load **Oracle** databases.
* Extensively used Datastage Designer to develop various jobs to cleanse the data from source tables.
* Designed complex Datastage jobs and UNIX scripts that handled required business transformations in an efficient manner.
* Created complex **SQL queries** using **CTE** and **Sub-queries** to implement business logic for tables with low data volume.
* Performed ETL **Datastage** and **SQL tunin**g to achieve better performance adhering to coding standards.
* Performed rigorous Unit and Integration testing on developed jobs for all possible positive and negative scenarios which resulted in *zero defects in system testing phase.*
* Coordinating with Business Governance team representing the vendor for all technical discussions
* Provide post-deployment support for the jobs once implemented in production.

***Project* : Albertsons – Haggen Merger**

***Role* :ETL Datastage Developer**

***Tools used* :**DataStage 11.3, Oracle 10g, Teradata, Unix, SVN, Autosys

**Description:** Project deals with creation of centralized data warehouse data by migrating data from both acquired organization and parent organization. Migrating data involves complex transformation, where product details and inventory details of both stores need to be merged as a single entity without duplication and need to be priced specific to the stores present in the Geo, which needs to be handled in ETL transformations.

**Roles & Responsibilities:**

* Working towards the methodologies and principles of Data Warehousing and ETL processes
* Collaborated with EDW team to gather the Business requirements.
* Played a key role in analysis of the current system and Design solutions to **create new centralized data warehouse** as well as planning migration solutions from the current system to the new system.
* Created the **ER diagram**, mapping and Functional Specification documents.
* Involved in creating new centralized **Data Warehouse** in **snowflake schema** and **junk dimension schema.**
* Extracted data from heterogeneous sources like **Teradata, Oracle, Flat File** and performed data validation, cleansing in staging area and then loaded into **Data Warehouse** using **Datastage.**
* Performed the Unit Testing and Integration testing with the help of Integrity queries.
* Managed software releases after each stage of the testing and defect removal. As well as resolving issues from **UAT** and System testing.
* Documented artifacts including requirement analysis, design documents (HLD, LLD), unit test plans and traceability matrix to ensure that the output is as per the requirements.
* Created **job sequencers** and **Autosys JIL’s** to schedule various ETL Batch processes using Autosys scheduler.
* **Performed code reviews** against other developer’s Datastage jobs and to suggest best suited solution.

**Infosys Limited, India (Nov 2013 - Jun 2018)**

***Project* : Consent data migration & extract**

***Role* : Data Engineer Lead**

***Domain* :** Insurance

***Tools used* :** Datastage 11.7, Snowflake, DB2, Autosys, UNIX, Python, Jira, GIT hub, Bit Bucket

**Description**: Project deals with migration of 100+ Millions of data to new DB Snowflake, as part of one-time migration in stipulated time on daily basis using API calls. As per terms, consent signed on a daily basis will be handed over to our application, which is processed and stored through API calls. This will be extracted on Daily and Monthly basis and functional logic will be applied on the extracted data and delivered to the consuming application.

**Roles and responsibilities**

* Extensively worked with stakeholder’s right from requirement gathering till production deployment
* Designed **dynamic ETL flows** using run-time column propagation (**RCP**) to reuse components for daily, monthly and one-time processes – to run the same **concurrently.**
* Collaborate with the data modeling team to ensure that the data models support the requirements and business rules for the ETL processes.
* Developed **ETL jobs** and **Unix scripts** to handle **100+ Million data** in an efficient way by utilizing minimal resources.
* Communicate the project risks or issues encountered during the Analysis, Design, and Development phases and bring them to the Project Core team.
* Developed ETL jobs to **implement** data warehouse loads using Slowly Changing Dimensions **(SCD) Type 2**
* **Performed code reviews** against other developer’s Datastage jobs and to suggest best suited solution.
* Documented artifacts including requirement analysis, design documents (HLD, LLD), unit test plans and traceability matrix to ensure that the output is as per the requirements.
* Created various scripts using Python for automation of various processes, including data pre - processing, data cleaning, and data validation.
* Created **job sequencers** and **Autosys JIL’s** to schedule various ETL Batch processes using Autosys scheduler.
* Provided Root Cause Analysis (**RCA**) for technical issues raised and provided solid updates to the stakeholders.
* Developed **python scripts** for load file creation to implement the parallel processing.
* Created pilot flows in Datastage to validate the data load between existing database and new database.
* **Agile**: Follow Agile methodology to breakdown EPIC into technical Stories, Tasks, Analyze and Finalize Requirements, Solution and Design Review approvals, Development, Testing, Autosys jobs, Pre-load data validation, DS & Non-DS Migration, Historical data load, Production data validation and Ongoing Support activities
* Prepare Implementation plan, backups and back out plan and support the release activities.
* Keep up to date with industry trends and technologies related to data integration, ETL, and data warehousing, and provide recommendations for improvements to the ETL processes and infrastructure.

***Project* : Reference Data Management (RDM)**

***Title* : ETL Datastage lead cum developer**

***Client*** : **Bank of America**

***Tools used* :** DataStage 11.3, Oracle, DB2, UNIX, Python, Tortoise SVN, Autosys, MDM, Reltio

**Description:** This project is about storing hierarchical data in a single table and creating extracts. We store country code formats of different standards such as ISO, IRS, etc. and create mapping between the standards. Internal applications of banks were using different standards instead of generic standards and faced challenges while communicating to other applications. We have created a master database in DB2 tables, which contains all standards, and designed ETL jobs to load the data, extract the data and create mapping files across standards.

**Roles & Responsibilities:**

* Identified problems, devised ideas, analyzed feasibility and piloted ETL flows which brought more business.
* Extensively worked in Data Modeling, **ER diagram creation** and table creation
* Designed and developed dynamic ETL jobs in Datastage, DB and UNIX
* Worked in **all phases of Software Development Life Cycle** (SDLC) from analysis until implementation of artifacts in production environment.
* **Developed reusable components** to utilize maximum performance using **run-time column propagation** (RCP), The **consuming application** can be **on-boarded** with a **DML entry without an ETL change.**
* Performed SQL tuning to achieve better performance in the queries.
* Developed code and documented artifacts including unit test plans and ensured that the output is as per the given specifications and SLAs.
* Coordinated with Business Governance team regarding load data and handled follow-up projects of RDM.
* Aided the team members wherever necessary.

|  |
| --- |
| **Certification** |

* Completed “**Snowflake for Data Analyst**” certification from Pluralsight.
* Completed “**Python Programming”** certification from Great Learning
* Completed “**Agile Software Development**” certification from LinkedIn Learning
* Completed “**Software Design: From Requirements to Release**” certification from LinkedIn Learning
* Completed “**Learning SQL Programming**” certification from LinkedIn Learning and evaluated by QAS Self Study