Sudhakar Pisipati   **Mobile : +1**(206) 333-7804
Data Engineer   **Email id :** lucasy@softworldusa.com

Professional Summary:

- Around 8+ years of ETL experience, proficient in Data Warehousing concepts, Data Analysis, and Web-Based application design.
- Experienced in open-source Business Intelligence Tools, specifically Pentaho, and Database management systems.
- In-depth involvement in the complete Project Life Cycle of Data Warehousing and Client/Server technologies.
- Data Warehouse development experience, including Data Migration, Data Conversion, and Extraction/Transformation/Loading using Pentaho Data Integration (Pentaho Kettle) with Oracle, SQL Server.
- Design, modeling, performance tuning, and analysis proficiency using Pentaho Data Integration/Kettle
- Strong SQL skills on Oracle, DB2, and SQL Server. Hands-on experience in Data Warehouse Star Schema Modeling, Snowflake Modeling, FACT & Dimension Tables, Physical and Logical Data Modeling.
- Adept at translating business needs into end user interactive reports thereby aiding business decisions with demonstrated experience in Analysis, Design, Development, Implementation, Modeling, Testing, and support for Data warehousing applications such as Snowflake.
- Successfully migrated over 100+ dataflows from oracle to Snowflake data warehouse while single-handedly managing the pre-validation, developing plan, post validation, migration, quality assurance, data validation, production deployment and production support.
- Ability to develop ETL pipelines in and out of data warehouse and usage of Snowflake's Snow SQL and write SQL queries against Snowflake.
- Having experience of Migrating DB objects into Snowflake environment.
- Familiar with using highly scalable parallel processing Infrastructure using IBM Infosphere Data Stage.
- Detailed and problem-solving oriented in Data Stage jobs and addressing production issues such as performance tuning and enhancement.
- Experience with using various techniques by understanding its complexion during before moving to production. This included tuning SQL in database levels, creating table indexes, breaking big jobs into smaller units, code reusability by etc.
- Proficient in creating database objects like Tables, Views, Stored Procedures, Indexes, Triggers, and Functions.
- Having 2+ years of hands-on experience in python programming in the Data Engineering domain.
- Can perform Data Cleaning, features scaling, features engineering using pandas and numpy packages in python.
- Applied Python's OOP principles to design and implement modular, reusable code, leveraging concepts such as encapsulation, inheritance, and polymorphism for efficient software development.
- Experience in analyzing data using Python, SQL, Microsoft Excel, Hive, PySpark, Spark SQL for DataMining, Data Cleansing, Data Munging.
- Building data pipelines in Python/Pyspark and constructing Python DAGs in Apache Airflow.
- Experienced in Creating, Debugging, Scheduling and Monitoring jobs using Airflow.
- Having work exposure in developing Pandas, Numpy in Python for developing data pipelines.
- Strong database expertise, covering MySQL, Oracle, Cassandra, and MongoDB, ensuring efficient data storage and retrieval.
- Implementation of monitoring solutions, including ELK, Splunk, Nagios, Prometheus, and Dynatrace, ensuring data pipeline reliability and performance optimization.
- Proficient in transforming data from various sources (flat files, XML, Oracle) to Data Warehouse using ETL tools.

- 2+ years of experience with control-m tool for scheduling and monitoring batch jobs.
- Hands-on experience with GIT, IBM UCD, and Jenkins to deploy artifacts and Database objects.
- Real-time experience in Hadoop/Big Data-related technologies.
- Strong knowledge and experience on AWS, specifically in Lambda, IAM, API Gateway, Dynamo DB, S3, Cloud Front, VPC, EC2, and Redshift.
- Experience in DevOps tools such as Subversion (SVN), Chef, Puppet, Ansible, Jenkins, Docker, Kubernetes, Terraform, Git, Gitlab, Subversion, Tomcat, Nginx, WebSphere, WebLogic, and JBoss.
- Configuration management tool expertise, including Puppet.
- Outstanding communication and interpersonal skills, ability to learn quickly, good analytical reasoning, and high compliance with modern technologies and tools.

Technical Skills:

Database	SQL Server2005/2008/2012/2016, Oracle 9i/10g, DB2, MySQL, PostgreSQL, MS Access, Snowflake
Pentaho Suite	Pentaho Data Integration (Kettle), Pentaho BI Server, Pentaho Analysis Tool, Pentaho Mondrian, Pentaho Metadata Editor, Pentaho Design Studio, Mondrian OLAP Server, Pentaho CDE Dashboard
Tools	Kubernetes, Docker, SQL Server Management Studio, , IBM IIS DataStage (Designer, Director)11.5/9.1, Pgadmin, Git, Oracle SQL Developer, SQL Plus, Oracle Enterprise Manager, SQL Profiler, ESRI ArcGIS Desktop, ArcMap, AWS S3,Lambda, EC2, Cloudwatch, SNS, SQS, Redshift, AWS Glue, Control-M, Airflow
Reporting Tools	Pentaho Report Designer, Pentaho Dashboard Designer, SSRS, Power BI
Languages	Python , T-SQL, PL/SQL, JavaScript
Operating systems	Windows 95/98/XP/Vista/7/10, Windows NT/2000, Server 2003, Centos 7,8, Redhat, Ubuntu

PROFESSIONAL EXPERIENCE:

Southwest Airlines, Dallas, Texas   **Oct 2022 - Present**

Data Engineer

Responsibilities:

- Refactored Python scripts to match training data with our database stored in AWS Cloud Search, so that to assign each document a response label for further classification.
- Implemented pandas data frame concepts to validate data across the stages of data flow.
- Created Jenkins build and deployment pipeline jobs to deploy the docker images into AWS ECR repositories and integrated with GITLAB.
- Automated AWS volumes snapshot backups for enterprise using Lambda and Python scripting. Created functions and assigned roles in AWS Lambda to run python scripts. Built S3 buckets and managed policies for S3 buckets and used S3 bucket and Glacier for storage and backup on AWS.
- Analyzing complex SQL queries to resolve issues raised during the batch load and resolving them implementing the optimization techniques.

- Monitoring performance and optimizing queries for efficiency.
- Implementing data security and disaster recovery solutions.
- Designing star schema and dimensional models for efficient data warehousing.
- Implementing data partitioning and clustering techniques for faster query execution.
- Worked in POC project on ETL Migration services by developing and deploying AWS Lambda functions for generating a serverless data pipeline which can be written to Glue Catalog and can be queried from Athena.
- Interacted with the Business Analysts to understand the process flow and the business.
- Using AWS services like AWS Glue and Lambda to extract, transform, and load data into Redshift.
- Developing and executing ETL pipelines for data movement and preparation.
- Writing SQL queries to perform data analysis and reporting within Redshift.
- Connecting Redshift with BI tools like Power BI and Tableau for interactive data visualization and reporting.
- Developing and maintaining data models and dashboards for business users.
- Implementing cost optimization strategies to manage Redshift expenses.
- Worked on AWS Data Pipeline to configure data loads from S3 to Redshift.
- Used the Data Stage Director and its run-time engine to run the solution, testing and debugging its components, and monitoring the resulting executable versions (on an ad hoc or scheduled basis).
- Tuned Data Stage transformations and jobs to enhance their performance.
- Implemented Work load Management (WML) in Redshift to prioritize basic dashboard queries over more complex longer-running adhoc queries. This allowed for a more reliable and faster reporting interface, giving sub-second query response for basic queries.
- Responsible for Designing Logical and Physical data modelling for various data sources on Redshift.
- Connecting to various data sources including Redshift, and Glue. Designing interactive reports and dashboards with visualizations like charts, graphs, and maps.
- Creating calculated measures and KPIs for deeper data analysis. Publishing Power BI reports to the Power BI service for sharing with other users.
- Setting access permissions and user roles for secure data access. Creating dashboards and reports for collaboration and decision-making across teams.
- Implementing data storytelling techniques to effectively communicate insights from data.
- Utilizing AI and machine learning capabilities within Power BI for advanced analytics.
- Building custom visuals and extensions for specialized data visualization needs.
- Installed and configured apache airflow for workflow management and created workflows in python.
- Implemented data pipelines in Apache Airflow.
- Monitoring control-m batch jobs and debug failures.

Environment: Python, AWS Lambda, S3, Redshift, Glue, IBM IIS DataStage (Designer, Director)11.5/9.1, Power BI, PL/SQL, Airflow, Windows.

Hastings Mutual Insurance Company, Hastings, Michigan   **May 2022 – Sep 2022**

Data/ETL Engineer

Responsibilities:

- Collaborated with stakeholders to meticulously map out the migration, aligning with business needs and objectives.
- Designed Data warehouse including star schema design, DW capacity planning, performance, and tuning.
- Developed and refactored python and shell scripts as per the business requirements.
- Analyzed the source on-premises environment (version, components, reports, integrations) and target Snowflake architecture (existing schema, security, access controls).
- Migrate on premise databases (MySQL, SQL Server, Oracle) to Staging schemas of Snowflake.

- Defined comprehensive data mapping strategies to translate business models and reports into the new Snowflake landscape.
- Designed efficient ETL pipelines in DataStage to extract, transform, and load data with accuracy and performance in mind.
- Developed ETL procedures to ensure compliance with standards and lack of redundancy, translated business rules and functionality requirements into ETL procedures using Data Stage.
- Used Data Stage Manager for importing metadata from repository, new job categories and creating new data elements.
- Built the Snowflake data warehouse, meticulously crafting schema objects (tables, views, materialized views) based on the mapped data models.
- Impeccably implemented DataStage jobs, utilizing appropriate stages and transformations for data extraction, cleansing, and transformation tasks.
- Optimized DataStage jobs for parallel processing and scalability, ensuring efficient data movement based on volume and processing needs.
- Created a data validation framework using Python pandas data frame concept.
- Successfully adapted Power BI reports to leverage Snowflake data sources and schema changes, preserving report formatting and functionality.
- Tuned DataStage jobs for performance by working with DBA and SQL coding, design changes within the job.
- Executed the meticulously planned migration, seamlessly transitioning users and reports to the new Snowflake environment.
- Rigorously monitored data pipelines and reports for performance and potential issues, ensuring data accuracy and integrity throughout the process.
- Trained and supported end users on the new Snowflake-based reporting system, fostering a smooth transition.
- Analyzed migration performance and identified opportunities for further optimization, continually refining DataStage jobs and the Snowflake schema.
- Monitored and maintained the migrated environment for optimal security and stability, proactively addressing any arising challenges.
- Used the Data Stage Director and its run-time engine to schedule running the solution, testing and debugging its components, and monitoring the resulting executable versions (on an ad hoc or scheduled basis).
- Developed processes for cleansing, transforming, extracting, integrating, and loading data using DataStage and Quality Stage.
- Developed ETL procedures to ensure conformity, compliance with standards and lack of redundancy, translated business rules and functionality requirements into ETL procedures using Data Stage.

Environment: Snowflake MPP, IBM IIS DataStage (Designer, Director)11.5/9.1, Control-M, Oracle 11g, MySQL, Oracle, Windows.

DXC Technologies, Hyderabad, India   **April 2019 – Jul 2021**
Client: Hastings Mutual Insurance Company, Hastings, Michigan
ETL Developer

Responsibilities:

- Designed Data warehouse including star schema design, DW capacity planning, performance, and tuning.
- Designed and developed complete BI infrastructure using Pentaho BI Suite.
- Created Enterprise Repository in Pentaho BI Server to store Pentaho ETL jobs and Reports. Responsible for managing data coming from various sources.
- Extensively worked with the business and data analysts in requirements gathering and to translate business requirements into technical specifications.

- Used Pentaho Data Integration Designer to create all ETL transformations and jobs.
- Identify and analyze data discrepancies and data quality issues and works to ensure data consistency and integrity.
- Used Pentaho Design Studio for creating custom parameters as well as generating reports.
- Implemented Slowly Changing Dimension Type 1 and Type2 in ETL jobs for certain Dimensions.
- Used Pentaho Report designer to create various reports having drill down functionality by creating Groups in the reports and drill through functionality by creating sub-reports within the main reports.
- Created several types of chart reports in Pentaho Business Analytics having Pie Charts, 3D Pie Charts, Line Charts, Bar Charts, Stacked Bar Charts and Percentage Bar charts.
- Created reports in Pentaho report designer using complex SQL queries and stored procedures against DW tables and added multiple levels of cascading prompt for both static and dynamic parameters.
- Dealt with slowly changing dimensions type 2 and multi-level hierarchical dimensions.
- Published cubes and reports onto Pentaho repository and refreshed Pentaho BI repository after uploading each object to be available for central use.
- Deployed reports on Pentaho BI Server to give central web access to the users.
- Created several dashboards in Pentaho using Pentaho Business Analytics Platform.
- Used Pentaho Schema Workbench to create Cubes, Dimensions, and fact tables in Pentaho.
- Created various ad-hoc reports using Pentaho Analyzer to allow users to generate reports on the fly by dragging and dropping dimensions onto rows and/or columns and dropping one or more measures to get the results.

Environment: Pentaho BI Server, Pentaho Report Designer, Pentaho Analyzer, Pentaho Data Integration/Kettle-Spoon, Pentaho Design Studio, Pentaho Dashboard Designer, MySQL, UNIX shell scripts, Oracle11g, SQL, Toad, Erwin, Putty, MS-Excel, Windows, Oracle SQL Developer.

DXC Technologies, Hyderabad, India   **April 2017 – Mar 2019**
Client: Uniper SE, Düsseldorf, Germany
ETL Developer

Responsibilities:

- Involved in designing Physical and Logical of Data warehouse.
- Created Data Flow Mappings to extract data from source system and Loading to Target.
- Configured and installed Pentaho Data Integration Server 4.1 hierarchically from development to production servers.
- Experienced in integrating Repository to store the Pentaho ETL jobs and Reports.
- Experienced in using Pentaho Data Integration Designer 4.1 to create all ETL transformations and jobs.
- Used Pentaho Data Integration/Kettle to design all ETL processes to extract data from various sources including live system and external files, cleanse and then load the data into target data warehouse.
- Used dimension lookup/update step to populate data into SCDs.
- Designed Data warehouse including star schema design, DW capacity planning, MySQL performance and tuning. Implemented Orders and Points DW using star schema, Orders and Points Business domain using Pentaho Metadata.
- Created and used various re-usable tasks, workflows, worklets, mapplets, and reusable transformations.
- Created cubes using schema workbench on top of DW star schema.
- Developed UNIX scripts for scheduling the delta loads and master loads using Autosys Scheduler.
- Designed various SQL queries to extract data as per the project's requirements.
- Configured error handling for transformation steps by creating Debug mode in Pentaho.
- Created Validation using Pentaho to generate reports for percentage of mismatch data and to achieve data quality.

Environment: Pentaho BI Server, Pentaho Data Integration (PDI/Kettle), Pentaho Design Studio, Pentaho Report Designer, Pentaho Dashboard Designer, Pentaho Business Analytics, Python, Oracle 11g, Oracle SQL Developer, SQL Profiler, Linux, Eclipse, RPM packaging, Unix.

DXC Technologies, Hyderabad, India   **Aug 2015 – March 2017**
Client: Coke One North America, Atlanta, USA
Software Engineer

Responsibilities:

- Interacted actively with Business Analysts and Data Modelers on Mapping documents and Design process for various Sources and Targets.
- Involved in migration of the code from IDQ to power center.
- Involved in Dimensional modeling (Star Schema) of the Data warehouse and used Erwin to design the business process, dimensions, and measured facts.
- Used various transformations like Address validator, parser, joiner, filter, matching to develop the mappings.
- Created complex mappings in Power Center Designer using Aggregate, Expression, Filter, and Sequence Generator, Update Strategy, Union, Lookup, Joiner, XML Source Qualifier and Stored procedure transformations.
- Worked on Power Center Tools like designer, workflow manager, workflow monitor and repository manager.
- Implemented SCD1, SCD2 type mappings to capture new changes and to maintain the historic data.
- Worked on extract and load type of mappings. Worked on SQL, Oracle and Teradata databases.
- Designed Workflows that uses multiple sessions and command line objects (which are used to run the UNIX scripts).
- Involved in creating stored procedures and using them in Informatica.
- Used Debugger to troubleshoot the mappings.
- Worked extensively on Unit Testing the ETL mappings and preparing efficient unit test documentation for the developed code to make sure the test results match with the client requirement.
- Responsible for solving the testing issues. Created Triggers, Cursors using T-SQL.
- Created documents that have the detailed explanation of the mappings, test cases and the expected results and the actual results.
- Involved in loading the partner flat files into the OLTP database. Provided Production Support.
- Developed the T-SQL stored procedures/functions to process these data and populate them in the appropriate destination tables.

Environment: Informatica power center 9.5.1, Informatica Data Quality 9.5.1, SQL Server 2012, DB2, UNIX, PL/SQL, Windows 7, ERWIN, SQL Developer 2005.