DILEEP JORIGE

dileep.ods@gmail.com   |   210-995-3356

**Summary**

* 15 years of Data Warehousing experience in the areas of ETL design, development, testing and production support in domains like Finance, Healthcare, Insurance, Retail and Banking.
* 5 years of experience in Azure Databricks and Snowflake.
* 12 years of experience in IBM Infosphere DataStage.
* Extensively involved in development activities (Integration, data migration) using Azure Synapse Analytics, Azure Databricks, PySpark, DBT and Snowflake.
* Experience in both on-prem and cloud solutions. Extensive experience with Microsoft Azure ecosystem such as ADF2, Azure SQL Server, Azure SQL IAAS/SAAS/PAAS, Azure SQL DW, Azure Blob.
* Implemented Azure data services including Azure Synapse Analytics, ADF, Azure Data Lake.
* Data migration from on-premises to Cloud, Snowflake using Azure DataBricks notebooks (Pyspark & SQL), Azure Data Factory pipelines.
* Design and develop scalable data ingestion frameworks to transform a wide variety of datasets.
* Hands-on experience across all stages of Software Development Life Cycle (SDLC) including business requirement analysis, data mapping, build, unit testing, systems integration testing.
* Proficient in writing Business Requirement Documents (BRD), Functional Requirement Documents (FRD), Technical Design Document (TDD), Functional Specification Document (FSD) and Requirement Traceability Matrix (RTM).
* Experience on development of projects in Agile and Waterfall environments.
* Experience on Data Warehousing applications, directly responsible for the Extraction, Staging, Transformation, Preloading and Loading of data from multiple sources into Data Warehouse.
* Experience on IBM Infosphere DataStage Administration tasks. Experience on IBM Infosphere DataStage Version Upgrades. Migrated 1 Tier architecture IBM Infosphere DataStage 9.1 to 11.5 2 Tier architecture.
* Expert in designing Parallel jobs using various stages like Hierarchical Stage, Complex Flat File, Join, Merge, Lookup, remove duplicates, Filter, Change Capture, Sequential File, Modify, Aggregator, XML Output, XML Input, Pivot, Peek and Snowflake stages.
* Extensive involvement in different phases of the project like Requirement Analysis, Systems Analysis, Data extraction, Cleansing, Transformation and Loading into Data marts.
* Implemented Slowly Changing Dimensions Type 1 and Type 2.
* Excellent experience in developing and understanding of HIPPA EDI transaction sets like 837P, 837I, 837D, 270/271,276/277 and ANSI X12 EDI Standards.
* Experience in EDI formats like X1, EDIFACT, XML. Strong understanding of the principles of Data Warehousing using fact tables, dimension tables and star/snowflake schema modeling.
* Experience in Data validation and scheduling the DataStage jobs.

**Skills**

|  |  |
| --- | --- |
| * ETL Tools: IBM Infosphere DataStage 11.5/11.7 * Cloud: Snowflake, Azure * EAI Tools : IBM WebSphere Transformation Extender 8.0 * Databases : Oracle 9i/10g/11g/12c, Netezza 7.2, Teradata, DB2, SQL Server 2005/2017 * Languages : C, SQL, PL SQL, Python | * Operating Systems : Windows NT/2000/XP/ Solaris 5.8, 5.10, RedHat Linux 5.1/6.0 * Standards and Protocols : XML, ANSI X12, EDI * Scheduling Tools : Control-M, Autosys * Technologies & Frameworks: PySpark, Azure DataBricks, Hive, SQL |

**Experience**

Sr. Data Engineer|Hallmark - Kansas, MO|04/2023 – Till Date

Kafka, Azure DataBricks, Jenkins, PySpark, PowerBI as the primary technologies. It primarily involves the migration of data to Delta Lake (Cloud). Product, Sales, Customer data is handled.

**Responsibilities:**

* Migrate various source data using Kafka streaming service, Databricks to Azure Delta Lake.
* Azure cloud based ETL process for data integration, that allows creating meta data-driven workflows, orchestrating data movement and transforming data.
* Developed Azure Databricks Notebooks using PySpark and SQL.
* Created reusable components and templates in Azure data factory and Databricks for Integration and Analytics product teams.
* Created Azure SQL database, performed monitoring and restoring of Azure SQL database.
* Creating, validating, and reviewing data solutions leveraging Azure data factory, Azure Synapse, Azure Data Lake, Databricks, Logic Apps, Azure Log Analytics and Power BI.
* Developed notebooks for data validation and data reconciliation.
* Developed workflows to schedule the Databricks notebooks.
* Processed real time streaming data using EventHub, Stream Analytics and publishing it to Power BI.
* Collaborate with business stakeholders to identify and meet data requirements based on data model.
* Modified/Extensively used generic meta data framework which can be used for any data extractions.
* Develop and migrate to Azure, leveraging Azure Data Lake and Azure SQL pools, Apache Spark.
* Ensuring compliance with data governance policies, data quality standards, and industry regulations when working with data in Snowflake.
* Creating/updating user stories in Jira board, track, update till the closure.
* Collaborating with data engineers, data scientists, business analysts, and other stakeholders to understand data requirements and documenting data models, ETL processes, and data pipelines

**Environment:** Kafka, ADF, DataBricks, PySpark, Event Hub, Stream Analytics, PowerBI, Github and Jira.

Data Engineer|USAA - San Antonio, TX|08/2019 – 03/2023

This project is to build the Data warehouse for individual applications to consolidate the data from different source systems to enable reporting/analysis purpose

**Responsibilities:**

* Assign, Coordinate and Review work and activities of offshore programing personals.
* Using ETL/ELT tool, Workflow Charts and Diagrams to design, code, test and implement the solution for Data Storage, Data Retrieval and Data Migration between different applications utilizing USAA standards and templates.
* Create complex ETL jobs by using USAA ETL framework for Staging Area, Dimensions, Facts and Data Marts load, work extensively on different processing stages in DataStage like Filter, Sort, Aggregator, Copy, Transformer, Funnel, Join, Look Up, Merge, Pivot, Modify, Switch, Change Capture, Hierarchal Stage and Remove Duplicates etc.
* Reading Steaming data using Kafka StreamSets according the requirements of the business owner.
* Worked with Kafka to integrate data from multiple topics to database. Manage RESTful API integrate with StreamSets to move data.
* Develop job schedule using tools like – BMC Control-M and create scripts setting up dependencies and schedule ETL- DataStage jobs.
* Bulk loading from the external state (AWS S3), internal stage to snowflake using the COPY command.
* Loading PARQUET files into snowflake tables from the AWS S3 bucket.
* Loading data into snowflake tables from the internal stage using Snowsql.
* Used COPY, LIST, PUT and GET commands for validating the internal stage files.
* Used SNOW PIPE configuration and notifications for continues data ingestion from the S3 bucket.
* Created tables, views, secure views, user defined functions in Snowflake Cloud Data warehouse.
* Experience with AWS cloud services: S3, EMR and Glue.
* Stage the JSON and CSV files into Snowflake DB by FLATTEN table function for different functional services.
* Configured the models, materializations, Seeds, Snapshots, Test and Macros for data transformation using DBT.
* Used DBT to build the documentation for models.
* Responsible for Unit Testing, Functional and Performance Testing needed for data validation and completion of ETL stories.
* Provide 24/7 on call production support by identifying defects, documenting, resolving the issues and working with change management, DBA and other teams to come up with fixes.
* Troubleshoot production issues if any failure occurs for monthly and Daily processes.
* Assigned the incidents to the corresponding teams for job failures in production who owns the applications.
* Troubleshoot production deployment issues and coordination with Operations & Configuration Management teams for deploying services in production environment.
* Maintain source code/scripts/Unix configuration files in GitLab.

**Environment:** IBM Infosphere DataStage 11.5/11.7, Snowflake, Snow SQL, DBT, Jinja, StreamSets, AWS, Oracle SQL/PLSQL, Netezza, DB2, BMC Remedy Control-M 9.0, RedHat Linux 6, JIRA, GitLab, Tableau, Visual Studio Code, QTest and SQuirreL.

Sr. ETL DataStage Developer|New York Life - New York City, NY|10/2018 - 07/2019

This project is to build HR Data Warehouse (HRODS) for the SAP Success Factors modules Employee Central, Recruiting, Overtime, Performance and Position Management to consolidating the data from Success Factors to enable HR reporting/analysis

**Responsibilities:**

* Extensively used DataStage Designer for developing Parallel jobs and performed complex mappings based on user specifications
* Experience in using DataStage Director to Validate, Run, and Schedule and Monitor DataStage jobs
* Involved in data modeling for database design
* Worked on Job Sequences to control the execution of the job flow using various Activities & Triggers (Conditional and Unconditional) like Job Activity, Sequencer and Routine activity
* Worked on various stages like Transformer, Join, Pivot, Lookup, Database Connector, Sort, Filter, Funnel, Change Capture etc
* Worked on developing Materialized views to consolidate the Fact and Dimension tables for reporting purpose
* Building DataStage ETL interfaces to aggregate, cleanse and migrate data across enterprise-wide Data Warehousing systems using staged data processing techniques, patterns and best practices
* Worked on writing complex SQL queries as per the business logic
* Involved in QA testing Using HP ALM QC
* Worked on Data Exception Handling and data validation
* Migrated from 9.1 to 11.5 on 2 Tier Architecture
* Worked on configuration file changes to 4 Nodes

**Environment:** IBM Infosphere DataStage 11.5, Oracle 11g, TOAD, CSV Files, Spotfire, SAP Success Factors, RedHat Linux 6, Dell Boomi, Control-M 8.0.

Sr. ETL Datastage Developer|New York Life - New York City, NY|06/2018 - 09/2018

This project is to pull the Policy Information from different Source systems and create the Xml files for the HP-Extream to generate the letters based on Document ID and to comply with amendments to FINRA rule 4512 which is to make reasonable efforts to obtain the name and contact information of the Trusted Contact Person (TCP) for a customer's registered policy or account. This would provide NYL the opportunity to contact the TCP to address possible financial exploitation; confirm health status; or identify a legal guardian, executor, trustee, or holder of a power of attorney.

**Responsibilities:**

* Extensively used DataStage Designer for developing Parallel jobs and performed complex mappings based on user specifications.
* Experience in using DataStage Director to Validate, Run, and Schedule and Monitor DataStage jobs.
* Worked on Job Sequences to control the execution of the job flow using various Activities & Triggers (Conditional and Unconditional) like Job Activity, Sequencer and Routine activity.
* Worked on various stages like Hierarchical Stage, Complex Flat File, Xml Output, Transformer, Join, Pivot, Lookup, Database Connector, Sort, Filter, Funnel etc.
* Building DataStage ETL interfaces to aggregate, cleanse and migrate data across enterprise-wide Data Warehousing systems using staged data processing techniques, patterns and best practices.
* Build the ETL jobs through Jenkins for automation.

**Environment:** IBM Infosphere DataStage 11.5, Oracle 11g, TOAD, XML Files, Mainframe, Data Hub, HP-Exstream, Jenkins

2.121, JSON, Python 2.7

Sr. ETL Datastage Developer|C.R. BARD - New Providence, NJ|11/2017 - 05/2018

This project involves integrating the sales and finance data of different countries from various sources like Oracle, Salesforce, Flat Files and load into BD Operational Data Store (BDODS) before transforming and loading into the target BDODS Data warehouse and Data marts.

**Responsibilities:**

* Extensively used DataStage Designer for developing Parallel jobs and performed complex mappings based on user specifications.
* Experience in using DataStage Director to Validate, Run, and Schedule and Monitor DataStage jobs.
* Worked on Job Sequences to control the execution of the job flow using various Activities & Triggers (Conditional and Unconditional) like Job Activity, Sequencer and Routine activity.
* Worked on various stages like Transformer, Join, Pivot, Lookup, Database Connector, Sort, Filter, Change Capture.
* Extracted data from source transformed and loaded into target database.
* Building DataStage ETL interfaces to aggregate, cleanse and migrate data across enterprise-wide MDM ODS and Data Warehousing systems using staged data processing techniques, patterns and best practices.
* Developed sequence jobs for the batch process.
* Involved in writing SQL queries as per the business logic.
* Moving sales and finance Data from Staging area to target new BD Data Warehouse system.
* Involving in Unit Testing, SIT and UAT.
* Worked on Data Exception Handling and data validation.

**Environment:** IBM Infosphere DataStage 9.1 (Designer, Director, Administrator, Parallel Extender), Oracle 11g, TOAD, Flat Files, MDM, Micro strategy, Red Hat Linux, Control-M.

Sr. Datastage Developer|Munich Re-Insurance - Princeton, NJ|07/2017 - 10/2017

Munich Re stands for exceptional solution-based expertise, consistent risk management, financial stability and client proximity. The project is to extract all the reinsurance policies and claims data to load in to Insurance Reporting Data Warehouse (IRDW).

**Responsibilities:**

* Building Data Stage Jobs as per the requirement.
* Involving in day-to-day production support activities.
* Worked on various stages like Transformer, Join, Lookup, Sort, Filter, Change Capture, XML and Hash file.
* Worked on Renewal Rate Change Enhancements for Composite Rating and Multiple Location match.
* Extracted data from source transformed and loaded into target database.
* Experience in using DataStage Director to Validate, Run, and Schedule and Monitor DataStage jobs.
* Involved in production support to run and monitor the sequence jobs.
* Loading XML Insurance Claims and Policy data into Data warehouse.
* Developed jobs for the CAU, Glatfelter, Poulton and Binding Authority interfaces as per the business requirement.
* Working within the team to make appropriate, effective decisions related to project responsibilities and to initiate and follow-up on assigned tasks.
* Involving in Unit Testing, SIT and UAT. Worked with the users in data validations.

**Environment:** IBM Infosphere DataStage 8.7 (Designer, Director, Administrator, Parallel Extender), Oracle 11g, XML, CSV, TOAD, Visual Studio 2015, Red Hat Linux, Control-M.

Sr. Datastage Consultant|Delphi - Troy, MI|04/2017 - 06/2017

The goal of this implementation is connected vehicle components (PMS) system to integrate major financial processes at Delphi to deliver fast and easy access to the financial information, via WEB. It is a long-term solution to improve and integrate financial information, analysis, and accounting systems.

**Responsibilities:**

* Building Data Stage Jobs as per the requirement.
* Involved in the data ingestion process through DataStage to load data into HDFS from Teradata, DB2.
* Worked with Json format files by using XML, Hierarchical Data stage stages.
* Worked on various stages like Transformer, Join, Lookup, Sort, Filter, Change Capture and Hash file.
* Using Big data file stage (BDFS) to pull the data from Big data environment.
* Involving in Unit Testing, SIT and UAT. Worked with the users in data validations.
* Optimizing/Tuned DS jobs for better performance and efficiency.
* Created Logical and Physical data models using Erwin, Involved in architecture definition document, HLD and LLD design and mapping documents and Proof of concepts.
* Loaded bulk data using Teradata utilities FLOAD and MLOAD.
* Extracted data from source transformed and loaded into target database.
* Building DataStage ETL interfaces to aggregate, cleanse and migrate data across enterprise-wide MDM ODS and Data Warehousing systems using staged data processing techniques, patterns and best practices.
* Working with various application development teams to provide data integration architecture services and data modeling services.
* Conducting code review sessions with developers.

**Environment:** IBM Infosphere DataStage 11.3 (Designer, Director, Administrator, Parallel Extender), Hadoop, Hive, Teradata, DB2, Linux, Autosys.

Sr. Datastage Consultant|SEI Investments, PA - Hyderabad, India|02/2014 - 02/2017

The goal of project is to reorganize the existing system which is running on Oracle DB to a new AIX Environment.

**Responsibilities:**

* Worked closely with Business analysts and Business users to understand the requirements and to build the technical specifications.
* Extracted data from source transformed and loaded into Oracle target database
* Created jobs using the stages such as Join, lookup, Transformer, Remove duplicates, Funnel, Aggregator, filter, Sort, Modify etc.,
* Code review on DataStage jobs developed by team members.
* Involved in production support activities during weekends.
* Performed rerunning/Killing the jobs in production and monitoring the space in production.
* Removed unwanted datasets using session management activity
* Added new DataStage users and assigning roles, Grant privileges and roles to the existing DataStage users.
* Monitored the scratch disk memory during execution of DataStage jobs, clearing the space in folders on ETL server.
* Extracted data from source transformed and loaded into Oracle target database.
* Performed session management activities in DataStage.
* Involved in Unit Testing and prepared test cases.

**Environment:** IBM Infosphere DataStage 9.1, 8.5 (Designer, Director, Administrator, Parallel Extender), IBM WTX 8.3, Oracle11g, SQL/PL SQL, Netezza, UNIX.

Sr. DataStage Consultant|Arthrex, CA - Hyderabad, India|08/2012 - 01/2014

This project involves integrating the sales and marketing data of different countries from varied sources (oracle, db2, flat files) and load into operational data store before transforming and loading into the target Data warehouse and Data marts.

**Responsibilities:**

* Used Enterprise Edition/Parallel stages like Datasets, Change Data Capture, Row Generator and many other stages in accomplishing the ETL Coding.
* Designed the Source to Target mappings between sources to Operational staging areas and then to target Data warehouse.
* Developed parallel jobs to get maximum efficiency, which are implemented using Parallel Extender.
* Quickly navigate between multiple projects and products using Rally.
* Developed DataStage Parallel Jobs, Job Sequencing, and Creating Parameter Sets, creating routines, Data Cleansing and Writing transformation expressions.
* Used the DataStage Designer to develop processes for extracting, cleansing, transforming, integrating and loading data into data warehouse.
* Importing metadata from repository, importing/exporting jobs and creating new data elements.
* Used XML stage to read XML files and write to target.
* Used Kimball bottom-up approach to implement the Datastage jobs.
* Created re-usable components using shared containers for local usage/shared usage.
* Used SQL Loader and import utility in TOAD to populate tables in the data warehouse
* Imported and exported repositories across projects. Used DataStage Version Control to promote jobs from Development to Test and then to Production environment.
* Involved in Unit, Integration, System and User Acceptance Testing (UAT).

**Environment:** IBM Infosphere DataStage 8.5, Oracle 11g, DB2, Teradata, COBOL, CSV Files, Flat Files, TOAD, Rally, Autosys, Cognos, Shell, UNIX (AIX 6.1/Solaris 10) and Linux.

Sr. DataStage Consultant|TSYS, GA - Hyderabad, India|02/2011 - 07/2012

CUSTOMER DATA MANAGEMENT SERVICES [CDMS]: Projects and Jobs are developed to load the data into different applications like Customer Finance Data Integration [CFDI], Credit Data Mart [CDM], Sales Force Automation [SFA] etc. of DMS group using IIS suite products and their functionalities.

**Responsibilities:**

* Prepared Source to Target mapping documents.
* Developed parallel Jobs in DataStage Designer to Extract data from the Sources Oracle and Complex Flat Files, cleanse it, transform by applying business rules, staging it in Data marts and Load (Initial/Incremental) into Target DWH Teradata.
* Migrated all the jobs from 7.5.2 to 8.1.1.
* Used the Slowly Changing Dimension Stage itself to implement SCD Type-1 & 2 instead of implementing them using CDC Stage in 7.5.2.
* Used Job Compare option to compare the jobs developed in 8.1.1 to 7.5.2 for efficient job design.
* Tuned DataStage transformations and jobs to enhance their performance.
* For Parallel jobs, configured the multiple nodes and used parallel engine capacity efficiently as well as designed a master sequence to run multiple jobs in parallel.
* Unit tested the jobs in Test environment by running them in DataStage Director and verifying the job logs for warnings & errors.
* Exported the project from Development to Test environment using DataStage Manager.
* Exported the project from Test to Production environment using DataStage Manager.
* Used the DataStage stages Oracle Enterprise, CFF, Copy, Filter, Lookup, Transformer, Sort, Funnel, Shared Containers, Join, Dataset, Aggregator, Sequential file, Remove Duplicates, Peek.
* Involved in Unit, Integration, System and User Acceptance Testing (UAT).

**Environment:** IBM Infosphere DataStage 8.1.1/7.5.2 EE (Parallel Extender), DB2, Oracle 9i, CSV Files, SQL, PL/SQL, Control-M, RedHat Linux 5.1.

Sr. Analyst Programmer|Oregon Dental Service, OR - Chennai, India|01/2006 - 12/2008

ODS is using Trizetto's FACTS system for their business which runs on mainframe and it is in the process of transitioning to Trizetto's FACETS which is a client server system.

**Responsibilities:**

* Data Integration team is responsible for transforming heterogeneous data between ODS and its trading partners
* Responsibilities:
* Developed Type Trees for defining the source and target systems
* Developed maps using map designer to transform the data
* Worked on 837 Daisy Premier Claim, 270/271, 276/277 maps with implementation of HIPAA National Provider Identifier (NPI) regulation changes which will be mandated on May23, 2007
* Developed maps to take a COBOL copy book as input and load the subscriber detail data into Sybase database upon validation
* Developed map, which uses XML file as input and generate EDI 271 and 277 as output format
* Developed HIPPA EDI transaction sets like 837P, 837I, 837D, 270/271, 276/277 and ANSI X12 EDI Standards
* Developed maps for HIPPA transaction sets like 835, 834
* Developed Common Message Logging map to send Email and error log messages in database tables
* Performed tuning the mappings to enhance the performance
* Performed trouble shooting of maps using Audit and Trace files
* Worked with EDI Standards, XML, ftp and Java Methods
* Developed Technical Specifications as per Business Requirements
* Deployed the maps using Integration Flow Designer
* Tested the mappings to meet the desired results
* Worked with BCP Utility to load the bulk data into Sybase
* Involved in Unit testing and prepared test cases for all jobs

**Environment:** IBM WebSphere TX 8.0 (Mercator), Sybase, UNIX, Crystal Reports, FACTS, and FACETS Systems.

**Education and Training**

The University of Salford |   United Kingdom, United Kingdom | 10/2011

Master of Sciences

Nagarjuna University |   India, India | 10/2005

Bachelor of Technology