**Srivalli G**

**Sr. Data Engineer**

**E-mail:** [**sgundapuneni11@gmail.com**](mailto:sgundapuneni11@gmail.com)

**Phone: 540-947-4069**

**Professional Summary:**

* **8 years of IT experience** in Analysis, Design, Development, Testing, Implementation and Support of Data Integration Data Migration & Data Quality Projects **using Informatica Power Center, Data Quality** software tools.
* Experienced in working with various relational data bases like **Oracle, Teradata, Db2, My Sql and SQL server**.
* Extensive experience in ETL (Extraction, Transformation and Loading) of data using Informatica Power Center from heterogeneous sources like **Flat files (Fixed width, delimited), XML files, Relational databases**.
* Extensive knowledge in SDLC (Software Development Life Cycle) and **executed projects in Waterfall and Agile methodologies**.
* Strong understanding of **Dimensional Modeling, Star, Snowflake Schema, OLAP and DW concepts**.
* Extensive knowledge in RDBMS concepts and **Relational and Dimensional Data modeling**.
* Worked closely with the data stewards to get data quality project requirements.
* Experienced in understanding the **Project scope, Dependencies, Risks and provide estimations**.
* Experienced in coordinating and leading **onsite-offshore development model**.
* Created complex mappings using transformations like **Source qualifier, Expression, Aggregator, Joiner, Look up, Normalizer, SQL, Union, Update strategy, Stored Procedure, Filter, Router, XML Generator and Parser, Java Transformations**.
* Co- ordinated, guided and leads the teams effectively in technology **migration from Informatica 8.6 to 9.1**.
* Experienced with **Teradata utilities like BTEQ, M-LOAD, F-LOAD, F-EXPORT, T-PUMP & TPT** and populate Teradata database tables.
* Very good experience and knowledge in provisioning virtual clusters under AWS cloud which includes services like EC2, S3, and EMR.
* Expertise in analyzing the source data systems for better understanding the business requirements, designing optimal ETL design.
* In depth knowledge and hands on experience of Big Data / Hadoop ecosystem (MapReduce, HDFS, Hive, Pig and Sqoop).
* Developed technical specifications of the **ETL process flow, High level and Low-Level Design and Source to Target mapping documents**.
* Designed and Developed **Audit, Control and balancing** process.
* Strong experience in handling **Slowly Changing Dimensions (SCD) and Change Data Capture (CDC)**.
* Experience with Informatica Advanced Techniques – **pushdown optimization, session partioning**.
* Extensively involved in creating reusable objects to standardize Informatica code and to bring down the development and testing efforts.
* Expertise in developing the flexible ETL code for maintenance and future enhancements by adopting **parameterization**.
* Extensively involved in **Optimization and Tuning of mappings and sessions in Informatica** by identifying and eliminating bottlenecks at Source, Transformations, Target layers.
* Experienced in setting up Shared **Informatica infrastructure and UNIX platform in co-ordination** with Infrastructure partner team.
* Extensively involved in writing Unix Shell Scripting programs like Source file validations, Files upload/Download to/from remote server, Source/Target Files Archival/purge requirements and to run ETL batch jobs from **UNIX using PMCMD command**.
* Experienced in scheduling of **ETL jobs using diff scheduling tools Maestro, Autosys, Control-M**.
* Independently perform **complex troubleshooting, root-cause analysis and provide solution to fix defects**.
* Experienced in writing sql queries using OLAP functions to improve the performance of the queries.
* Experienced in creating and using **Stored Procedures, Functions, Views and Materialized views**.
* Extensively involved right from Project engagement to till **Production Warranty support** of the project.
* Prepared Project support handover document and transitioned to **Support team to ensure smooth execution of the project in Production**.
* Generated warning and Error reports and shared with Business users to further analysis.
* Supported and guided Testing and Business users while **SIT, UAT in progress and obtained sign off from SIT, UAT Teams.**
* Extensively involved in generating automated alerts to Business users and Email notifications as per Business requirement.
* Ability to meet deadlines and handle multiple tasks, decisive with strong leadership qualities, flexible in work schedules and possess good communication skills.
* Team player, Motivated, able to grasp things quickly with analytical and problem-solving skills.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Languages** | Java 8, Python, R |
| **Python and R** | Numpy, SciPy, Pandas, Scikit-learn, Matplotlib, Seaborn, ggplot2, caret, dplyr, purrr, readxl, tidyr, Rweka, gmodels, RCurl, C50, twitter, NLP, Reshape2, rjson, plyr, Beautiful Soup, Rpy2 |
| **Algorithms** | Kernel Density Estimation and Non-parametric Bayes Classifier, K-Means, Linear Regression, Neighbors (Nearest, Farthest, Range, k, Classification), Non-Negative Matrix Factorization, Dimensionality Reduction, Decision Tree, Gaussian Processes, Logistic Regression,  Naïve Bayes, Random Forest, Ridge Regression, Matrix Factorization/SVD |
| **NLP/Machine Learning/Deep Learning** | LDA (Latent Dirichlet Allocation), NLTK, Apache OpenNLP, Stanford NLP, Sentiment Analysis, SVMs, ANN, RNN, CNN, TensorFlow, MXNet, Caffe, H2O, Keras, PyTorch, Theano, Azure ML |
| **Cloud** | Google Cloud Platform, AWS, Azure, Bluemix |
| **Web Technologies** | JDBC, HTML5, DHTML and XML, CSS3, Web Services, WSDL |
| **Data Modelling Tools** | Erwin r 9.6, 9.5, 9.1, 8.x, Rational Rose, ER/Studio, MS Visio, SAP Power designer |
| **Big Data Technologies** | Hadoop, Hive, HDFS, MapReduce, Pig, Kafka |
| **Databases** | SQL, Hive, Impala, Pig, Spark SQL, Databases SQL-Server, My SQL, MS Access, HDFS, HBase, Teradata, Netezza, MongoDB, Cassandra. |
| **Reporting Tools** | MS Office (Word/Excel/Power Point/ Visio), Tableau, Crystal reports XI, Business Intelligence, SSRS, Business Objects 5.x/ 6.x, Cognos7.0/6.0. |
| **ETL Tools** | Informatica Power Centre, SSIS. |
| **Version Control Tools** | SVM, GitHub |
| **BI Tools** | Tableau, Tableau Server, Tableau Reader, SAP Business Objects, OBIEE, QlikView, SAP Business Intelligence, Amazon Redshift, or Azure Data Warehouse |
| **Operating System** | Windows, Linux, Unix, Macintosh HD, Red Hat |

**Professional Experience:**

**Client: United Health Group, Eden Prairie, MN Nov 2016 to till date  
Role: Sr. Data Engineer**

**Responsibilities:**

* + Worked closely with the business teams to understand the requirements along with Business Analyst & ETL Team.
  + Involved in the discussions with data modelers to review the design of the data model and to create database tables.
  + Involved the in creating the technical design documents along with Business Analyst and design walk through with data architects.
  + Developed the informatica mappings using Source Qualifier, Expression, Joiner, Look up, Sql, Router, Union, Normalizer, Stored Procedure transformations.
  + Provided informatica training to team and guided them as and when necessary.
  + Extracted data from various sources Oracle, My Sql, Flat Files and loaded in to Teradata.
  + Used Teradata BTEQs to load data from staging to base database.
  + Used Teradata Utilities – M-Load, F-Load, F-Export, T-Pump & TPT both at scripting level and along with Informatica connection.
  + Perform data analysis by using Hive to retrieve the data from Hadoop cluster, Sql to retrieve data from Oracle database and used ETL for data transformation.
  + Use MLlib, Spark's Machine learning library to build and evaluate different models.
  + Perform Data Cleaning, features scaling, features engineering using pandas and numpy packages in python.
  + Develop MapReduce pipeline for feature extraction using Hive and Pig.
  + Create Data Quality Scripts using SQL and Hive to validate successful data load and quality of the data. Create various types of data visualizations using Python and Tableau.
  + Extracted the data from hive tables by writing efficient Hive queries.
  + Involved in doing performance tuning (Using Explain Plan) of long running Teradata BTEQs in prod environment as part of the support team.
* Generated error and warning reports to help business teams for further analysis of data.
  + Tuned Informatica mappings to increase the through put using various performance tuning techniques.
  + Created reusable objects /mapplets to standardize Informatica code and to bring down the development and testing efforts.
  + Created store procedures, Functions, Views as part of the business logic implementation.
  + Reviewing the Code during Peer review which is developed by the team members and provided feedback.
  + Involved in test data preparation for SIT, UAT phases.
  + Involved in Deploying the Informatica code to test and production environments.
  + Provided estimations for the ETL Tasks.
  + Participated in daily stand up calls and kick off meeting sessions at the begin of each iteration and Retrospective meetings.
  + Worked with the Scheduling/Operations teams for scheduling the ETL jobs in Maestro.
  + Supported ETL Prod environment during warranty phase using trouble tickets.

**Environment:** Informatica Power Center 10.2/9.6.1, Oracle 11g, My Sql, Flat Files, Teradata 15/14, Linux, Perl, TOAD, Teradata SQL Assistant, Win Scp, Putty, Rally.

**Client: Sheridan Healthcare, Sunrise, FL Apr 2016 to Oct 2016  
Role: Sr. Informatica Developer & Data Analyst**

**Responsibilities:**

* Developed internal and external Interfaces to send the data in regular intervals to Data warehouse systems.
* Extensively used Power Center to design multiple mappings with embedded business logic.
* Involved in discussion of user and business requirements with business team.
* Performed data migration in different sites on regular basis.
* Involved in upgrade of Informatica from 9.5.1 to 9.6.1.
* Created complex mappings using Unconnected Lookup, Sorter, and Aggregator and Router transformations for populating target tables in efficient manner.
* Attended the meetings with business integrators to discuss in-depth analysis of design level issues.
* Involved in data design and modeling by specifying the physical infrastructure, system study, design, and development.
* Supported MapReduce Programs running on the cluster.
* Extensively involved in performance tuning of the Informatica ETL mappings by using the caches and overriding the SQL queries and also by using Parameter files.
* Developed complex SQL queries to develop the Interfaces to extract the data in regular intervals to meet the business requirements and extensively used Teradata Utilities like M – load, F- Load, TPT, BTEQ and Fast Export.
* Analyzed session log files in session failures to resolve errors in mapping or session configuration.
* Written various UNIX shell Scripts for scheduling various data cleansing scripts, loading process and automating the execution of maps.
* Created transformations like Expression, Lookup, Joiner, Rank, Update Strategy and Source Qualifier Transformation using the Informatica designer.
* Performed Data Enrichment jobs to deal missing value, to normalize data, and to select features by using HiveQL.
* Developed multiple MapReduce jobs in java for data cleaning and pre-processing.
* Created mapplet and used them in different mappings.
* Worked on Flat Files and XML, DB2, Oracle as sources.
* Written PL/SQL Procedures and functions and involved in change data capture (CDC) ETL process.
* Implemented Slowly Changing Dimension Type II for different Dimensions.
* Involved in Informatica, Teradata and oracle upgrade process and testing the environment while up gradation.
* Worked with Informatica version Control excessively.
* Experience in using SVN as version control for migration.
* Written Unit test scripts to test the developed interfaces.
* Managed enhancements and coordinated with every release with in Informatica objects.
* Provided support for the production department in handling the data warehouse.
* Worked under Agile methodology and used Rally tool one to track the tasks.
* Written thorough design docs, unit test documentation, Installation and configuration guide documents.
* Performed bulk data imports and created stored procedures, functions, views and queries.

**Environment:** Informatica Power Center 9.6.1/9.5.1, Oracle 10g, DB2, Flat Files, XML, Teradata 14/12, Unix/Linux, TOAD, Teradata SQL Assistant, Win Scp, Putty, SVN, Rally.

**Client: Bank of America, Charlotte, NC Jan 2015 to Apr 2016  
Role: Sr. ETL Informatica Developer**

**Responsibilities:**

* Co-ordinated and led the interactions across Portfolio Architects, Enterprise Data Modeling Architects, DBA and Business teams in finalizing Business requirements, ETL architecture and data modeling and translate the requirements into Technical specifications.
* Worked on Technical documentation like preparing High Level Design Documents, Low Level Design Documents and Source to Target mapping sheets.
* Involved in the Logical and physical data modeling and prepared the Source data volumetric information with DBA to plan Indexes, table spaces, hashing and partition keys.
* Preparation and Review of Unit Test cases and executing Unit test cases in QC and documenting the results.
* Managing and guiding the team in design and development the ETL artifacts.
* Reviewing the Code developed by the team members and provided feedback.
* Involved in test data preparation for SIT, UAT phases.
* Participating in Reviews with the customers to get the sign off on design and mapping documents.
* Involved in technical supporting, troubleshooting data warehouse application to meet Service Level Agreement (SLA's)
* Handling of CR's (Change Requests) and enhancements for existing application and followed change management process.
* Created Informatica mappings using Source qualifier, connected/Unconnected Lookup, joiner, Sorter, Aggregator, Union and Router transformations to extract, transform and load the data to Staging, ODS, Data ware house and finally data mart area to enable Reporting capabilities.
* Created persistent look up caches to avoid building the same cache multiple times where ever applicable there by improving the performance of the ETL jobs.
* Created re-usable transformations/mapplet and used across various mappings.
* Created re-usable user defined functions to perform business validations and used across various Projects with in the portfolio.
* Worked on Slowly Changing Dimensions as per business requirement.
* Designed and developed Audit, Balancing and Control process.
* Worked on Performance Tuning of sources, targets, mappings, transformations and sessions, by implementing various techniques like partitioning techniques and pushdown optimization, and identifying performance bottlenecks from logs.
* Implemented Informatica concurrent workflow execution to improve the performance.
* Wrote UNIX Korn shell scripts for file transfer/archiving/Email notifications.
* Tuned SQL queries at source qualifier and Lookup transformation to cache required volume of data by avoiding unwanted data wherever possible to minimize the data transfer over the network.
* Guiding and Supporting during Assembly/Integration Test, QA and UAT in progress.
* Has done the Data validation and error analysis and Provide resolutions / feedback to data experts for the data to be corrected.
* Provide knowledge transfer and prepared KT documents to Production Support for ongoing maintenance of the code.
* Worked with the Scheduling/Operations teams for scheduling the ETL jobs on Control-M tool.

**Environment:** Informatica 9.5.1, Oracle, Rally, HPQC, UNIX, Toad, Putty, Control M.

**Client: American Express, Phoenix, AZ Feb 2014 to Dec 2014**

**Role: Informatica ETL Developer**

**Responsibilities:**

* Involved in envisioning phase calls with business users, SME’s and Source teams to understand the requirements and identifying the sources.
* Involved in the requirements mapping from source to target, design and ETL process flow diagrams.
* Engaged various teams like data model, Database and the shared infrastructure, Testing teams to kick start the development activities.
* Initiated and Led the discussions with Enterprise Data modeling, DBA teams to design the data model and Database.
* Involved in the Logical and physical data modeling and prepared the Source data volumetric information with DBA to plan Indexes, table spaces, hashing and partition keys.
* Prepared test Scenarios and Test cases in HP Quality Center and involved in unit testing of mappings, system testing and user acceptance testing.
* Appeared before the PRB (Project Review board) for each phase of the project to check project health checkup and to get Approvals from PFA’s and sign off from Tech Director before commencing next phase.
* Performed various data validations and cleansing activities before loading data into STAGE layer and standardizations and integration with confirmed dimensions at Data ware house and transformed data as per reporting requirement before loading to DATA MART layer.
* Created mappings using Source qualifier, Expression, Lookup, Router, joiner, Aggregator, Union, Update strategy, SQL, Java transformations.
* Implemented and maintained slowly changing dimensions as per business requirement.
* Tuned SQL queries at source qualifier and Lookup transformation to cache required volume of data by avoiding unwanted data wherever possible to minimize the data transfer over the network.
* Lookup persistent cache mechanism has been used to share the cache across sessions during the batch jobs processing to reduce the overall system resources and processing time of the jobs.
* Involved in the Historical, Daily, catch up loads once the project launched in live environment.
* Generated out bound flat files (Delimited and Fixed width) and XML files to share with downstream application teams.
* Wrote complex SQL queries to address data purging requirement.
* Has done the Data validation and error analysis and Provide resolutions / feedback to data experts for the data to be corrected.
* Provided knowledge transfer and prepared KT documents to Production Support for ongoing maintenance of the code.
* Worked with the Scheduling/Operations teams for scheduling the ETL jobs on Autosys tool.
* Write Shell script to transfer source files from remote server, Source files validation, Archive/Purge requirements and kick off TL batch jobs from UNIX environment.
* Participated in Daily stand up calls and Kick off meeting sessions at the begin of each release.
* Participated in Show and Tell and Retrospective session for each release.
* Supported in test data preparation for the SIT, UAT and fixed identified defects by the teams.
* Created Deployment documentation for the code migration to Test and Production environments.
* Monitored jobs during warranty phase and handed over to Support team with support hand over document and KT sessions.

**Environment:** Informatica Power Center 8.6.1/9.1.1, Oracle, Flat Files, Rally, UNIX, TOAD, Win SCP, Putty, Auto Sys

**Client: PNB, Hyderabad, India Oct 2012 to Jan 2014**

**Role: Informatica ETL Developer**

Punjab National Bank is an Indian multinational banking and financial services company. Bank will continue to take initiatives in transforming People, Processes and Productsthrough responsive and responsible banking.

The primary objective of this project is to capture different Customers, Policies, Claims Agents, Products and financial related data from multiple Flat files. Extracted, Transformed and Loaded data in to data warehouse using Informatica Power Center and generated various reports on a daily, weekly monthly and yearly basis. These reports give details of the various products of PNB products that are sold.

**Responsibilities:**

* Involved in understanding the Requirements of the End Users/Business Analysts and developed strategies for ETL processes.
* Responsible for requirements gathering for an enhancement requested by client. Involved in analysis and implementation for an Intranet Based Information Management Information System.
* Provided maintenance and support of Online and Batch Programs using COBOL, DB2.
* Wrote several DB2 Stored Procedure scripts to implement the business logic.
* Created [data models](http://www.linkedin.com/skills/skill/Data_Models) for information systems by applying formal data modeling techniques.
* Strong expertise in physical modeling with knowledge to use Primary, Secondary, and Join Indexes.
* Designed Fact tables and Dimension tables for star schemas and snowflake schemas using ERWIN tool and used them for building reports.
* Performed reverse engineering of physical data models from databases and SQL scripts.
* Provided database implementation and database administrative support for custom application development efforts.
* Developed OLAP reports and Dashboards using the Business intelligence tool - OBIEE.
* Synchronizing all regions PCR, Unit and System while migration changes from lower region to acceptance.
* Reviewing programs for QA and Testing.
* Responsible for Coding, Unit Test Plans, Unit Test Results, Functional Testing and Regression Testing.
* Involved in comprehensive end-to-end testing- Unit Testing, System Integration Testing, User Acceptance Testing and Regression.
* Provided 24/7 On-call Production Support for various applications and provided resolution for night-time production job failures, attend conference calls with business operations, system managers for resolution of issues.

**Environment:** Informatica Power Center 8.6.1, Oracle 10g, SQL\* Plus, TOAD, SQL\* Loader, SQL Developer, Shell Scripts, UNIX, Windows.

**HSBC Global Technology, Hyderabad Jan 2011 – Sep 2012**

**Client: HSBC Technology Services (HTSA), USA, Buffalo**

**ETL Developer**

HSBC is a global banking and financial services company headquartered in Canary Wharf, London, United Kingdom. It is the world's second-largest banking and financial services group and second-largest public.

This Project involves the creation of a Data Warehouse for the Sales Reporting System. Data from Operational Source System is extracted and loaded into data warehouse Using Teradata.

**Responsibilities:**

* Initiated sit down sessions with the Sales team for requirement gathering.
* Used Informatica to extract data from Oracle load it into Teradata staging.
* Created scripts using Fast-Load, Multi-Load to load data into the Teradata data warehouse.
* Involved in Sql scripts Macros, stored procedures in Teradata to implement business rules.
* Updated numerous Bteq/Sql scripts, making appropriate DDL changes and completed unit and system test.
* Involved in peer to peer reviews.
* Identifying Bugs and reporting defects.
* Knowledge transition to new team member.
* Daily Communication regarding the status of the work to Onsite coordinator

**Environment:** Informatica Power center 8.6.1, Oracle, Teradata 12, View Point, BTEQ, Fast Load, Multi Load, Teradata SQL Assistant, Flat Files, UNIX, Windows.

**Education:**

* Bachelor’s degree in Computer Science and Engineering From Jawaharlal Nehru Technological University (JNTU), Hyderabad, India