**Ravi Ch**

**2315980220 |** **tejach5252@gmail.com**

**Data Engineer**

**SUMMARY:**

* Over 6 years of experience in the IT field with extensive Data Warehousing implementations and data integration experience in developing ETL mappings and scripts using python, Scala, ETL tools like Ab Initio and Informatica.
* Experienced in analysis, design, development, testing of data warehouses and data marts. Extensive knowledge of dimensional data modeling of star schema/snowflake and relational database concepts.
* Strong knowledge of entity-relationship concept, facts and dimensions tables, slowly changing dimensions and dimensional modeling.
* Extensively worked on the ETL mappings, analysis and documentation of OLAP reports requirements. Solid understanding of OLAP concepts and challenges, especially with large data sets.
* Ab-initio knowledge in ETL, data mapping, transformation and loading from source to target databases in a complex, high-volume environment.
* Well versed with various Ab Initio components such as Join, Rollup, Partition De-partition, Dedup sorted.
* Expertise in improving performance and troubleshooting of the Ab-Initio graphs and monitoring AB-Initio run time statistics.
* Experience in integration of various data sources like Oracle, DB2, SQL server, SQL\*plus, Oracle SQL developer, XML files and flat files into staging area. Also experience in optimizing database querying, data manipulation using SQL and PL/SQL in oracle, flat files and SQL server databases.
* Developed and built frameworks/prototypes that integrate Big Data and advanced analytics to make business decisions
* Implemented Spark using Scala and Spark SQL for faster testing and processing of data.
* Extensively used different libraries like Pandas, Numpy, OS, Sub process, boto3 in Python to develop based on the mapping documents.
* Involved in performing a key management and thought leadership role in the areas of advanced data techniques, including data modeling, data access, data integration, data visualization, text mining, data discovery, statistical methods, big data design and implementation
* Extensive knowledge on AWS and good experience in migrating the ETL applications into Cloud.
* Strong working experience in Informatica power center data integration tools - Repository manager, Designer, Workflow manager, Workflow monitor.
* Good Knowledge on creating S3 buckets through AWS console and through cloud format templates.
* Experience in implementing Spark using Scala and Spark SQL for faster testing and processing of data responsible to manage data from different sources.
* Experience in L2 and L3 support environment, worked on Priority base tickets (P1, P2and P3)
* Experience on different components of Abinitio like Architecture, GDE, Co Operating System, EME, Technical Repository Management tool, Parallelism, DML, transforms, M Utilities, air commands, Variables, Partitioning and Performance Improvement techniques.
* Developing, installing and configuring Cloudera/Hadoop based Big Data solutions – Scala, Apache Spark, SBT, Kafka, Flume, HBase, HDFS, Avro, Parquet, Kite SDK, YARN, Hive, Impala and Sqoop
* Highly proficient in writing, testing and implementation of triggers, stored procedures, functions, packages, cursors using PL/SQL.
* Functional Programming experience in Scala, including Higher-order Functions, Currying, Partial Functions, Partial Application, Nested Functions
* Good Knowledge on creating and working on GCP buckets through console and Unix environment.
* Business Intelligence solution design, development and process documentation to produce required reporting and analytics deliverables using Tableau
* Experience in loading data, troubleshooting, debugging mappings, tuning mappings etc.
* Good working experience on performance tuning, identifying and resolving performance bottlenecks in various levels like sources, targets, mappings and sessions.
* Good Experience creating RDD’s and Data Frames by using Spark.
* Good Knowledge on reading Hive Tables/Data through Spark Sql component.
* Good Knowledge on Using Case Class to convert RDD’s to Data Frames.
* Worked on Informatica mappings involving Type-I, Type-II slowly changing dimensions.
* Coordinating with business users, functional design team and testing team during different phases of project development and resolving the issues.
* Self-motivated, able to set effective priorities to achieve immediate and long-term goals and meet operational deadlines. Self-starter with good interpersonal and communication skills.

**TECHNICAL SKILLS:**

**Reporting**: Tableau, Business Objects – Crystal reports, Webi, Xcelcius.

**ETL:** Ab initio, Informatica, SAP BODS and Spark/Scalla.

**Scheduling tools:** Arow, Control-M and Auto Sys

**Databases**: Snowflake Database, Oracle9i/10g/11g, MS SQL, MY SQL and MS Access.

**Data Modeling**: Dimensional Data Modeling, Data Modeling, Star Join Schema Modeling, Snowflake Modeling, FACT and Dimensions Tables, Physical and Logical Data Modeling, ERWIN 4.0/3.5, Oracle Designer, synopsis and Visio.

**Operating Systems:** Windows NT, Windows 98/XP/2000/2003, Linux & VM ware.

**Others**: MS Project, Oracle ADI 11.2, MS Access, Tableau, Big data, Hadoop, Hive, Scoop,

 MS Visio and MS Excel.

**Methodology:** Agile, Waterfall.

**Technologies:** AWS, GCP, Spark and Python.

**WORK EXPERIENCE:**

**Client Walmart, Arkansas Dec 2019- Till date**

**Role: Data Engineer**

**Responsibilities:**

* Strong understanding in product life cycle, data quality and data security for products across the enterprise from many disparate systems.
* Created source-to-target mappings for loading data warehouse/data marts in Erwin.
* Involved in the development of code in Hive, Python and Spark/SQL. Involved in developing optimized, high quality code deliverable, perform code reviews & unit test plan reviews, conduct independent testing and signoff the code release for testing stage to build the application conforming to project standard and requirements.
* Created new Automic jobs to automate the migrated scripts in Prod environment, also migrated the existing Automic Jobs to point to the new servers.
* Used Spark and Spark-SQL to read the parquet data and create the tables in hive using the Scala API.
* Integrated Maven build and designed workflows to automate the build and deploy process.
* Involved in migrating the On-prem data to GCP buckets using UDP.
* Optimized Hive QL/ pig scripts by using execution engine like Tez, Spark.
* Implemented Spark using Scala and utilizing Data frames and Spark SQL API for faster processing of data.
* Involved in designing the plan to convert and move the code in prod-17 environment to GCP.
* Involved in the development for converting the existing code from hive to Spark using Scala and Python.
* Used various spark Transformations and Actions for cleansing the input data.
* Wrote Ad-hoc queries for moving data from HDFS to Hive and analyzed the data using HIVE QL.
* Involved in creating tables in Snowflake and used SQL extensively to investigate data anomalies and support the functionalities of the ETL process. Involved in creating dashboards to see the progress on Basel monthly jobs and access the log files from the dashboard.
* Involved in building the ETL design and source to target mapping to load data into data warehouse.
* Designing and executing test scripts and test scenarios, reconciling data between multiple data sources and systems.
* Effectively understood session error logs and used debugger to test mapping and fixed bugs.
* Worked as communicator with business analysts and implementation teams regarding integration activities.

**Environment**: Hadoop, GCP, Spark, Hive, SQL, Scala, UNIX, Python, ERWIN, Scoop, ServiceNow, Automic, Data discovery, Teradata and Snowflake.

**Client Capital One, Virginia Aug 2018- Dec 2019**

**Role: Data Engineer**

**Responsibilities:**

* Responsible for integrating jobs and the underlying cost by performing Unit, System and Integration testing before deploying in Prod Environment. Involved in migrating the code from legacy environment to cloud.
* Involved in the development of code in Ab-initio, Python and Spark/SQL. Involved in developing optimized, high quality code deliverable, perform code reviews & unit test plan reviews, conduct independent testing and signoff the code release for testing stage to build the application conforming to project standard and requirements.
* Developed multiple Kafka Producers and Consumers from as per the software requirement specifications.
* Created new Arow jobs to automate the migrated scripts in Prod environment, also migrated the existing Arow Jobs to point to the new servers.
* Used Kafka for log accumulation like gathering physical log documents off servers and places them in a focal spot like HDFS for handling.
* Used Pyspark to debug any data anomalies and took appropriate measures to resolve those issues. Involved in spinning up Basel EMR servers and maintaining the EMR servers to follow the standard compliance.
* Configured Spark streaming to get ongoing information from the Kafka and store the stream information to HDFS.
* Involved in performing high- and low-level design, provide pseudo code, implement the prototype, participate in design reviews with architecture team and tech teams to deliver design documents as per customer requirements.
* Wrote Ad-hoc queries for moving data from HDFS to Hive and analyzed the data using HIVE QL.
* Rehydrating the EMR servers and maintaining them in both east and west regions. Involved the registering the S3 datasets, Snowflake tables (metadata) in Nebula.
* Involved in creating tables in Snowflake and used SQL extensively to investigate data anomalies and support the functionalities of the ETL process. Involved in creating dashboards to see the progress on Basel monthly jobs and access the log files from the dashboard.
* Created high level and detailed designs using Ab initio GDE.
* Designing and creating Hive external tables using shared meta-store instead of the derby with partitioning, dynamic partitioning and buckets.
* Involved in testing the ECE servers and migrated the code to ECE (cloud).
* Extracted the data from the flat files and other RDBMS databases into staging area and populated onto data warehouse.
* Changed the source and target databases from Teradata to Snowflake database.
* Involved in migrating the mainframe data to S3 and changed the components/logic in Ab initio to load data into S3.
* Involved in creating the nebula datasets for files in S3.
* Co-ordination with different testing groups to accommodate their testing data requirements and translate them to data selection criteria in Ab inito format.
* Created Arow documents and registered the init, clean and LR jobs.
* Worked with different Partition components like partition by range, Partition by Round Robin, partition by Expression for efficiently using Multifile system which comes under Data Parallelism.
* Performing transformations to the source data with transform components like Filter-by-Expression, Replicate, normalize, Reformat, Rollup etc.
* Used Lookups in Reformat component to fetch matching records based on for the down stream process.
* Used Sort component to sort the tables, and used Dedup Sort to remove duplicate values.
* Provided L3 supports for the Ab inito jobs and publishing the data to upstream by monitoring the Arow jobs.
* Uploading files to S3 and download the files from S3 (AWS) bucket.
* Developed and modified scripts in UNIX shell scripts and used SQL extensively to investigate data problems and to support the functionalities of the ETL process.
* Responsible for creating wrapper scripts in UNIX to run various init, clean and Load ready jobs.
* Working with large amounts of data independently executing data analysis, utilizing appropriate tools and techniques.
* Modernized the existing Ab initio code using Spark/SQL and Python/Spark.
* Effectively understood session error logs and used debugger to test mapping and fixed bugs.

**Environment:** AWS, Spark, Python, Spark-SQL, Scala, Hive, Teradata, Snowflake, Nebula, ServiceNow, Arow, data lakes, Unix and Ab initio.

**Client Charter Spectrum, NC April 2017- July 2018**

**Role: ETL developer/ Data Analyst**

**Responsibilities:**

* Responsible for creating and developing highly complex applications using oracle as back-end with expertise in design and development of oracle PL/SQL packages and procedures.
* Analyzed the functional specifications provided by the data architect and created technical specification documents for all the enhancements.
* Involved in dimensional modeling (Star Schema) of the data warehouse and used Erwin to design the business process, dimensions and measured facts.
* Involved in building the ETL design and source to target mapping to load data into data warehouse.
* Used various transformations like filter, expression, sequence generator, update strategy, joiner, stored procedure, and union to develop robust mappings in the Informatica designer
* Used type 1 SCD and type 2 SCD mappings to update slowly changing dimension tables
* Design, develop and maintain oracle database schemas, tables, standard views, materialized views, synonyms, unique indexes, non-unique indexes, constraints, triggers, sequences, implicit cursors, explicit cursors, reference cursors, and other database objects.
* Working with large amounts of data independently executing data analysis, utilizing appropriate tools and techniques.
* Developed and modified scripts in UNIX shell scripts and used SQL extensively to investigate data problems and to support the functionalities of the ETL process.
* Writing SQL queries to create end-user reports /developing SQL Queries and stored procedures in support of ongoing work and application support.
* Designing and executing test scripts and test scenarios, reconciling data between multiple data sources and systems.
* Effectively understood session error logs and used debugger to test mapping and fixed bugs.
* Worked as communicator with business analysts and implementation teams regarding integration activities.

**Environment:** Unix, Hadoop, Hive, Oracle, Teradata, Informatica, Scoop, Tableau, HPSM, Control-M and ERWIN

**Client: IDoo Global, Texas May 2016 – Mar 2017**

**Role: Data Analyst/ ETL developer**

**Responsibilities:**

* Strong understanding in product life cycle, data quality and data security for products across the enterprise from many disparate systems.
* Created source-to-target mappings for loading data warehouse/data marts.
* Formulating the QA plan for black box testing of the application including functional, regression, integration, systems and user acceptance testing.
* Involved in cleansing and extraction of data and defined quality process for the warehouse.
* Developed Informatica ETL mappings, sessions and workflows based on the technical specification document.
* Responsible for studying the existing data warehouse and working on migrating existing PL/SQL packages, stored procedures, triggers and functions to Informatica Power Center.
* Created mappings using transformations like source qualifier, joiner, aggregator, expression, SQL, rank, filter, router, lookup, update strategy and sequence generator.
* Designed and developed the logic for handling Slowly Changing Dimension tables load by flagging the record using update strategy for populating the desired.
* Developed reusable mapplets and transformations for reusable business calculations.
* Used exception handling logic in all mappings to handle the null values or rejected rows.
* Applied slowly changing dimensions Type I and Type II according to the business requirements.
* Extensively worked on performance tuning and also in isolating header and footer in single file.
* Involved in requirement gathering, design, testing, project coordination and migration.
* Prepared project estimation for ETL activity and scoping.
* Raised change requests, analyzed and coordinated resolution of program flaws and fixed them.
* Perform analysis profiling on existing data and identify root causes for data inaccuracies, impact analysis and recommendation of data quality.
* Scheduled jobs for running daily, weekly and monthly loads in a sequence with command and event tasks. Used Agile test methods to provide rapid feedback to the developers significantly helping them uncover important risks.

**Environment:** Oracle, Teradata, Informatica, Scoop, Unix, ERWIN, Hive, Python, Tableau, HPSM, Control-M and Hadoop

**Client: Accenture, Hyderabad May 2014- Jul 2015**

**Role: ETL developer**

**Responsibilities:**

* Prepared design specification documents as per the inputs received from the architect and the business analyst.
* Extracted data from heterogeneous source systems like oracle, SQL server, excel and flat files with fixed width and delimited.
* Involved in cleansing and extraction of data and defined quality process for the warehouse.
* Designed and developed the logic for handling Slowly Changing Dimension tables load by flagging the record using update strategy for populating the desired.
* Developed reusable mapplets and transformations for reusable business calculations.
* Used exception handling logic in all mappings to handle the null values or rejected rows.
* Applied slowly changing dimensions Type I and Type II according to the business requirements.
* Extensively worked on performance tuning and in isolating header and footer in single file.
* Working with large amounts of data independently executing data analysis, utilizing appropriate tools and techniques (Interpreting results and presenting them to both internal and external client.
* Writing SQL queries to create end-user reports /developing SQL Queries and stored procedures in support of ongoing work and application support.
* Designing and executing test scripts and test scenarios, reconciling data between multiple data sources and systems.
* Involved in requirement gathering, design, testing, project coordination and migration.
* Prepared project estimation for ETL activity and scoping.
* Effectively understood session error logs and used debugger to test mapping and fixed bugs in DEV in following change procedures and validation.
* Raised change requests, analyzed and coordinated resolution of program flaws and fixed them in DEV and pre-production environments, during the subsequent runs and PROD.
* Perform analysis profiling on existing data and identify root causes for data inaccuracies, impact analysis and recommendation of data quality.
* Fine-tuned ETL processes by considering mapping and session performance issues.
* Responsible for creating workflows and worklets. Created session, event, command, control, decision and email tasks in workflow manager.

**Environment:** Unix, Hadoop, Hive, Oracle, Scoop, HPSM, Control-M and ERWIN