**Rishi Lingampelli ** ****

**rishirao43@gmail.com**

**(816) 863-6962**

Analytical, talented and multi skilled working as ETL Lead at McDonald’s Corporation, IL,. with an overall experience above 7 years in Data Warehousing Design, ETL and Big Data Technologies currently working on Talend Integration Suite, Talend Big Data Integration.

**Summary**

* Experience in designing and developing complex mappings to extract data from diverse sources including flat files, RDBMS tables and legacy systems, experience in full life cycle development, gathering requirements, documentation, developing, designing, testing, implementing.
* Expertise in reading and writing data from multiple source systems like oracle, Microsoft SQL Server, MYSQL, AWS Redshift, Teradata and source files like delimited files, Excel, Positional, JSON, XML and CSV files.
* Experience using Talend Integration Suite 6.3.1/6.5 and using various File Input, **Logging & Error**, **Orchestration**, **Processing & System** components, Talend Data Integration & Talend Platform Setup on Windows.
* Components used comprises of tFileInputDelmited, tFileInputExcel, tFileInputJson, tFileInputXML, tFilterRow, tMap, tJoin, tReplicate,tFlowToIterate, tParallelize, tSetGlobalVar, tJava, tJavaFlex, tAggregateRow, tDie, tWarn, tLogCatcher, tStatsCatcher, tDie, tLogRow, tSystem.
* Strong programming experience in **Core java**.
* Experience working in J2EE applications and frameworks like Hibernate, Struts & Spring.
* Well versed with Big data components like **tHDFSInput**, tHDFSList,tHDFSGet, tHiveLoad, tHiveInput.
* Experience in using **Amazon cloud** components and connectors like tRedshiftInput, tRedshiftRow, tS3List, tS3Get, tS3Put for accessing data from cloud storage like Amazon S3 & **Redshift**.
* Hands on experience on **Hadoop** and its eco system components (HDFS, Map Reduce, Hive, Pig, Flume, Oozie, Sqoop)
* Used components like tRest, tRestClient to call **REST** Web Service API and process data.
* Developed reusable components like **joblets** for custom logging, context loading And reusable jobs using tRunJob for loading data into Redhsift DB.
* Experience in creating **Talend Big data** batch jobs using Spark framework.
* Hands on experience on **Spark Core**, Spark SQL and creating the Data Frames handle in SPARK with Scala.
* Proficient in supporting Data warehouse ETL activities using queries and functionalities of SQL. Solid experience in implementing complex business rules by creating re-usable transformations and robust mappings using various transformations like Unconnected and Connected lookups, Source Qualifier, Router, Filter, Expression, Aggregator, Joiner, Update Strategy etc.
* Created **Context** Groups and variables to run jobs in Dev, Stage and Production environments.
* Extensively worked on tuning Talend jobs for better performance.
* Good understanding and experience in Enterprise Data Warehousing concepts such as Change Data Capture (CDC), Slowly Changing Dimensions (SCD Type I and Type II) and Data modeling techniques like Star Schema and Snowflake modeling.
* Excellent working experience in Waterfall, Agile methodologies.
* Experience working on **Matillion ETL** for Redshift.
* Worked on Hadoop distributions like Hortonworks & **Amazon EMR**.
* Experience working in AWS Services like **Lamba, SQS, SNS**.
* Hand on **Shell scripting** experience.
* Good knowledge and experience on **Amazon EC2** & **Amazon S3**.
* **AWS** Certified Solutions Architect Associate
* Experience in setup, deployment and configuration of jobs in **Talend Administration Centre** (TAC) and Nexus.
* Well versed with TAC features & activities such as Job Scheduling, Job Conductor, Execution Plan, Publisher, Server Monitoring.
* Experience working with version control system like Git and repository management system like Stash.
* Highly skilled in time, task and data management abilities along with good communication skills.

**Technical Skills**

|  |  |
| --- | --- |
| **ETL Tools** | Talend Big Data Platform 6.3.1/6.5 (Talend Studio, Talend Adiministration Center), Datastage. |
| **Databases** | AWS Redshift, Teradata, SQL Server 2008, MySQL |
| **Dimensional modeling**  | Dimensional Data modeling, Physical and logical Modeling, Star & Snow-flake Modeling  |
| **Programming Languages/Scripting** | Java, Scala, J2EE, Shell Scripting, JavaScript, Ajax |
| **BigData Technologies**  | HDFS, MapReduce, Spark, Kafka, Hive, Pig, Flume, Oozie, Sqoop, Yarn  |
| **Hadoop Distributions**  | Hortonworks, AWS EMR  |
| **Frame Works/ J2EE Technologies** | Struts, Spring, Hibernate, Servlets, JSP, Design Patterns, ExtJS, JUnit, MRUnit, JAXB |
| **Version control** | CVS, SVN, Git. |
| **Logging Tools** | Log4j |
| **Build Tools** | Ant, Maven, SBT |
| **IDE tools** | Eclipse, Intellij IDE |
| **AWS**  | S3, EC2, Redshift |
| **Other Tools** | Stash, Confluence, JIRA tool |

**EDUCATION**

* Masters in Computer Science from University of Central Missouri, Warrensburg, MO.
* Bachelor’s Degree in Information Technology from Jawaharlal Nehru Technological University Hyderabad, India.

**CERTIFICATIONS**

* Talend Big Data Integration Developer 6.3
* AWS Certified Developer Associate
* AWS Certified Solutions Architect Associate

**Professional Experience**

**Talend ETL Lead**

**McDonalds Corporation, Chicago, IL**

**January 2018 – Present**

**Customer360**

Customer 360 provides a holistic view of the customer that allows McDonald's to realize growth opportunities by building a more personal connection with the customers. C360 dataware house hosts data from various business sources like AppAnnie, New Relic, First Data and customer data like the Customer surveys, Customer profiles, eCommerce platform data and Offers related data.

**Roles and Responsibilities**

* Designed & developed **complex ETL jobs** from various sources such as oracle, sql server , Amazon S3 and loaded into target Redshift Database.
* Interacted with business users to gather requirements on any updates they need.
* Used different **Big data components** in Talend like **thdfs**, tpigmap, thiveinput, thiveoutput, tfiledelimited, tfileoutputdelimited, tunique, tFlowToIterate, tlogcatcher, tfilelist, taggregate, tsort, tFilterRow and **webservice components** like **tRestClient, tRest**.
* Created **Context Groups** and variables to run jobs in Dev, Test and Production environments and created an utility called Context Loader using joblet feature for secured contexts access.
* Implemented custom Error handling in Talend jobs and worked on different methods of logging.
* Responsible for improving the performance of Talend jobs by tuning the sql queries, using parallelize components in case of huge source data and various **performance tuning** options provided for the individual components.
* Used the temp table features & parallel processing power of **Redshift** to perform **ELT** on huge data sets.
* Defined organization level Naming conventions for Talend job names and various components based on their functionality.
* Created **Unix Scripts** and run them using **tSystem** components for decrypting & encrypting the files received from vendors.
* Worked on **Stash** to manage the **Git version control.**
* Worked on **TAC** to publish & schedule jobs.
* Designed **joblets** (reusable components) & **Java routines** in Talend.
* Involved in the meetings on infrastructure,design architecture.

**Environment**: Talend Big Data studio 6.3.1, Talend Administration Centre(TAC), IBM DataStage, Amazon Web Services, Amazon S3, Amazon EC2, EMR 5.0, Redshift,Teradata, SQL Server 2008, Java, Stash, Confluence, Aginity Workbench

**Sr. ETL Engineer**

**McDonald’s Corporation, Chicago, IL**

**January 2017 – December 2017**

**GDAP Talend Migration Development**

Global Data and Analytics platform (GDAP) is cloud based analytics platform built in AWS. It’s a one stop solution for all the analytics required for the corporation based on Sales, Operations and Digital data in result to leverage the business. GDAP manages McDonald’s Sales, Finance, Operations, Digital data, and Master data like Restaurants, Menu Items, and Hierarchies etc. GDAP Talend Migration project is development project where **GDW DataStage jobs** and Teradata objects are re-built into Talend and AWS Services.

**Roles and Responsibilities**

* Understand ETL Mappings in DataStage and Create High level design documents and develop the respective Talend jobs.
* Created ETL jobs in Talend to load data into AWS Redshift from variety of sources, relational sources like Teradata, Oracle, SQL Server and non-relational sources like HDFS, flat files, **spreadsheets**, **xml**, **json** files using multiple transformation rules.
* Worked with various Talend Big data components like **tHDFS**, **tS3Connection**, **tRedshift**, tHive and other Talend components like **tMap**, **tFileInputdelimited**, tFileOutputdelimited, tUnique, tFlowToIterate, tlogcatcher, **tfilelist**, tFilterRow.
* Created **Context Groups** and **variables** to run **jobs in Dev, Test and Production environments**.
* Performed Unit Testing of the Developed codes and created Testing Document.
* Developed **complex SQL Queries**, UNIX shell scripts**, Redshift copy commands**, **AWS S3.**
* Review of Talend ETL codes to make sure Business logic used in **DataStage** is exactly replicated in Talend.
* Gathering of Business requirements from Customer, Requirement analysis and provide proper data integration solution.
* Closely worked with client on daily basis for requirement gathering, Task Deliverables and Task status reporting
* Daily coordinate with offshore team for Daily task updates, providing solutions, technical assistance and transitioning new functionality.

**Environment**: Talend Big Data studio 6.3.1, Talend Administration Centre(TAC), IBM DataStage, Amazon Web Services, Amazon S3, Amazon EC2, EMR 5.0, Redshift,Teradata, SQL Server 2008, Java, Stash, Confluence, Aginity Workbench

**McDonald’s Corporation, Chicago, IL**

**November 2015 – December 2016**

**Sr.Hadoop/Spark Engineer**

The Flume agents subscribe to the Rabbit MQ queues, consume messages in real time, batch them and ingest the batched files to HDFS in Amazon EMR.The batch files are read and processed by Apache Spark application in short intervals, for example every 5 minutes.The entity type of the JSON can be identified by the header fields inside the JSON after parsing.The operation(Create, Update, Delete) to the table will be identified by the header fields in the JSON message.The conversion of JSON to RDD, parsing and processing will be done in Scala using the Spark framework.Spark will perform the operations and transformations on the CRDB database based on the individual JSON’s attributes to corresponding database tables.

**Roles and Responsibilities**

* Ingestion of Daily data into Hadoop from different sources like RDBMS and SFTP.
* Implemented CDC logic (Change Data Capture) logic using **Spark Scala** and Data frames concepts.
* Used Spark for implementing the transformations need to join the daily ingested data to historic data.
* Enhanced the performance of queries and daily running spark jobs using the efficient design of partitioned hive tables and Spark logic.
* Implemented Sqoop for Ingesting the data into Hadoop from RDBMS.
* Built Workflows and sub workflows for calling sqoop, spark.
* Implemented the automated workflows for all the jobs using the Oozie and **shell script**.
* Processed streaming data using **Spark Streaming**.
* Involved in requirement and design phase to implement Streaming Lambda Architecture to use real time streaming using **Spark and Kafka**
* Developed Spark code and **Spark-SQL** for faster testing and processing of data.
* Developed a data pipeline using Kafka, AWS S3, Spark and Hive to ingest, transform and analyzing customer behavioral data.
* Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and Spark.
* Developed Spark jobs and Hive Jobs to summarize and transform data.
* Used Spark for interactive queries, processing of streaming data and integration with popular NoSQL database for huge volume of data.
* Involved in running Spark streaming jobs to process terabytes of text data.
* Implemented a streaming solution using Kafka, HBase for processing for streaming data.
* Migrated the **Map Reduce** jobs into Spark Jobs, used Spark SQL and Dataframes API to load structured and semi structured data into Spark Clusters

**Environment**: Spark, Hive, SQL, Pig, Kafka, Hbase, Kafka, Sqoop, Scala, Java, Shell Scripting, Oozie, Amazon S3, Amazon EC2, EMR 5.0, IntelliJ IDEA

**Hadoop Engineer**

The McDonald's Corporation is the world's largest chain of fast food restaurants, serving around 68 million customers daily in 119 countries across 35,000 outlets. The project objective is to extract all the sales and the products sold across all the McDonalds Restaurant’s and load it into the data warehouse called GDW using the Big Data components. It also involves loading of data from traditional databases to HDFS using various Big Data technologies.

**Roles and Responsibilities**

* Job duties involve the design, development of various modules in Hadoop Big Data Platform and processing data using MapReduce, Hive, Pig, Scoop and Oozie.
* Design, develop and test Map Reduce programs on Mobile Offers Redemptions and send it to the downstream applications like HAVI. Scheduled this MapReduce job through Oozie workflow.
* Ingest huge amount of XML files into Hadoop by Utilizing DOM Parsers with in Map Reduce. Extracted Daily Sales, Hourly Sales and Product Mix of the items sold in McDonalds Restaurant’s and loaded them into Global Data Warehouse.
* Write and test the MapReduce code to do aggregations on identified and validated data. Processed Mobile Offers data for McDonald’s stores in USA across various location and ingested the data into Hadoop HIVE tables.
* Schedule Multiple Map Reduce jobs in Oozie. Involved in extracting the promotions data for McDonalds stores within USA by writing the map reduce jobs and automating it with **UNIX shell script**.
* Schedule, monitor and debug various **MapReduce**, Pig, Hive jobs using **Oozie Workflow**.
* Work with various Hadoop Ecosystem tools like Sqoop, Hive, Pig, **Flume**, Oozie, Kafka.
* Monitor Nightly jobs to export data out of HDFS to be stored offsite as part of HDFS backup.
* Define a standard layout and standard set of attributes that are a part of all application logs.

**Environment**: Hortonworks Distribution Platfom – HDP 2.2, Hadoop, MapReduce, Sqoop, Hive, Flume, Oozie, Pig, SQL Server 2008, Java, Scala, SQL, UNIX Shell script, Eclipse IDE

**Java/Hadoop Developer**

**Fission Computer Labs, Hyderabad, India**

**January 2014 – July 2014**

Fission Labs is a Data Analytics & Cloud based company which develops scalable cloud applications to meet business challenges of the clients, combines the power of Big Data to provide solutions which helps the client’s business transcend financially. Offers an array of services across Business Intelligence and Data Analytics spectrum ranging from Data Aggregation, Data Cleansing, ETL, Data Mining & Predictive Analytics.

**Roles and Responsibilities**

* Developed **Map Reduce** jobs in java for data cleansing and preprocessing.
* Moving data from Oracle to HDFS and vice-versa using **Sqoop**.
* Collecting and aggregating large amounts of log data using Apache **Flume** and staging data in HDFS for further analysis
* Worked with different file formats and compression techniques to determine standards
* Developed **Hive** queries and **UDFS** to analyze/transform the data in HDFS.
* Developed Hive scripts for implementing control tables logic in HDFS.
* Designed and Implemented **Partitioning** (Static, Dynamic), Buckets in HIVE.
* Developed **Pig** scripts and UDF’s as per the Business logic.
* Analyzing/Transforming data with **Hive and Pig**.
* Developed **Oozie** workflows and they are scheduled through a scheduler on a monthly basis.
* Designed and developed read lock capability in **HDFS**.
* Effective coordination with offshore team and managed project deliverable on time.
* Worked on QA support activities, test data creation and Unit testing activities.

**Environment:** Java 6, Hortonworks Data Platfom – HDP 2.0, Hive, Pig, MapReduce, Oozie, Sqoop, Flume.

**Java/J2EE Developer**

**Highradius Technologies, Hyderabad, India**

**June 2011 – December 2013**

Receivables Radius is based on robust web aggregation engine that can transparently aggregate and collect information from multiple web sites. The solution can periodically traverse through all the web sites and simulate what users do manually to automatically retrieve documentation such as POD’s, BOL’s, Claims and Checks. In addition the solution can automatically match-up all the documents so than they can be bulk printed or emailed on as required basis. The solution is available as a hosted service that can be accessed as via any standard web browser .For example, users can request POD’s in bulk by simply uploading a spreadsheet with all the tracking numbers or automate the request process via a file extract from your ERP System . Further, all the retrieved documents can be stored as PDF documents on an image server or attached to a business transaction in your ERP system such as Invoice or Dispute case.

**Roles and Responsibilities**

* Involved in design, implementation and testing.
* Understand & map client’s requirements / enhancements to the product and provide technical solutions. Solution Design Phase for the clients’ requirements.
* Implemented DB inserts/updates based on constraints.
* Performed data cleaning and digesting the XML raw data using Digester & JAXB API .
* Implemented MVC design using **Struts**1.3 frameworks, JSP custom tag Libraries and various in-house custom tag libraries for the presentation layer.
* Created tile definitions, **Struts-config files**, validation files and resource bundles for all modules using Struts framework
* Used Log4J for writing into different log files Application Log and Error Log.
* Added features to send e-mail and downloading the attachments from e-mail using Java Mail API.
* Refactored code, implemented code changes in the product for performance.
* Database was handled using the **JDBC**.
* Developed various product requirements (PR), change requests (CR) using **J2EE** specifications like **Servlets, JSP, Spring and Hibernate** frameworks.
* Designed and developed the aspects (**Spring AOP**) for Logging.
* Used **Hibernate** as an ORM tool for connecting to the database and accessing information
* Used **JavaScript** and HTML to make changes on the UI based on the requirement.
* Developed the unit test cases using **JUnit**

**Environment:** Java 1.6, J2EE, JSP, Struts 1.3, Spring 3, Hibernate, MySQL server, HTML, XML, XSLT, Java Script, AJAX, JDBC, JUnit, Log4j, Eclipse