**Mohammed**

**Sr. Data Engineer**

**Email:** [**mohammed.engineer58@gmail.com**](mailto:mohammed.engineer58@gmail.com) **Phone: (315) 274 4606**

**SUMMARY:**

* 8+ years of experience in Data Modeling, business and Data Analysis, production support, Database Management, strategic analysis, requirements gathering, data mapping and data profiling.
* Experienced working with Technical Architects and Database analysts for the Design of Summary tables required for efficient Report Design.
* Experience in analyzing data using Hadoop Ecosystem including HDFS, Hive, Spark, Spark Streaming, Elastic Search, Kibana, Kafka, HBase, Zookeeper, PIG, Sqoop and Flume.
* Hands-on experience in architecting and data modeling for AWS Redshift, AWS Oracle RDS, AWS PostgreSQL and AWS Aurora and Good understanding and hands on experience with AWS S3 and EC2.
* Experienced on Python for statistical computing and also experience with MLlib (Spark), Matlab, Excel, Minitab, SPSS, and SAS
* Very good experience and understanding in development of Conceptual, Logical and Physical Models for Online Transaction Processing and Online Analytical Processing (OLTP & OLAP).
* Experience in developing Entity-Relationship diagrams and modeling Transactional Databases and Data Warehouse using tools like ERWIN, ER/Studio and Power Designer
* Hands on experience in Normalization and De-Normalization techniques upto 3NF for optimum performance in relational and dimensional database environments.
* Experienced in writing SQL queries and optimizing the queries in Oracle, SQL Server, Netezza, Teradata and Big Data.
* Experienced in automating the infrastructure in AWS using web services and experience in building policies for access control and user profiles using AWS IAM, S3 controls with bucket policies.
* Expertise in designing Star schema (identification of facts, measures, and dimensions), Snowflake schema for Data Warehouse, ODS architecture by using tools like Erwin data modeler, Power Designer, and E-R Studio.
* Skillful in Data Analysis using SQL on Oracle, MS SQL Server, DB2 & Teradata and performed Data analysis and Data profiling using complex SQL on various sources systems including Oracle and Teradata
* Extensive experience in loading and analyzing large datasets with Hadoop framework (MapReduce, HDFS, Pig, Hive, Flume, Sqoop).
* Experienced in logical/physical database design and review sessions to determine and describe data flow and data mapping from source to target databases coordinating with End Users, Business Analysts, DBAs and Application Architects.
* Proficient in using Python, SQL, Hadoop ecosystem for extracting data and building predictive models.
* Strong background in various Data Modeling tools using ERwin, ER Studio and MS Visio.
* Experienced with both DDL and DML including Joins, Functions, Indexes, Views, Constraints, Primary Keys and Foreign Keys.
* Experienced in Performance tuning on oracle databases by leveraging explain plans, and tuning SQL queries and excellent experience in writing SQL queries to validate data movement between different layers in data warehouse environment.
* Extensive ETL testing experience using Informatica 9x/8x, Talend, Pentaho and expert in the Data Analysis, Design, Development, Implementation and Testing using Data Conversions, Extraction, Transformation and Loading (ETL) and SQL Server, ORACLE and other relational and non-relational databases.
* Very good experience in designing and developing Data Marts by following Star Schema and Snowflake Schema Methodology, using industry leading Data Modeling tools like ERwin.
* Strong experience in using Excel and MS Access to dump the data and analyze based on business needs.
* Managed full SDLC processes involving requirements management, workflow analysis, source data analysis, data mapping, metadata management, data quality, testing strategy and maintenance of the model.
* Excellent knowledge on creating reports on SAP Business Objects, Webi reports for multiple data providers.
* Extensive knowledge and experience in producing tables, reports, graphs and listings using various procedures and handling large databases to perform complex data manipulations.

**Skills Table:-**

**Programming Languages:** Python, SQL, PL/SQL, UNIX shell Scripting, VBScript, PERL, AWK, SED  
**Databases:** Oracle 12c/11g/10g/9i, Teradata R12, R13, R14, R15, MS SQL Server 2005/2008, MS Access  
**Bigdata and Cloud Technologies:** Hadoop Framework(Hive, Pig, MapReduce, Sqoop, HDFS, Flume, NoSQL Databases), AWS Glue, AWS S3, AWS Redshift, AWS EMR, Azure, Azure Data Factory.

**Tools:** MS-Office suite (Word, Excel, MS Project and Outlook), VSS  
**Testing and defect tracking Tools:** HP/Mercury (Quality Center, Win Runner, Load Runner, Quick Test Professional, Performance Center, VU Scripting, Business Availability Center), Requisite, MS Visio & Visual Source Safe, Salesforce.  
**Operating System**s: Windows Vista/XP/2000/98/95, Dos, Unix  
**ETL/Data warehouse Tools:** Informatica9.5/9.1/8.6.1 (Repository Manager, Designer, Workflow Manager, and Workflow Monitor), SAP Business Objects XIR3.1/XIR2, Web Intelligence  
**Data Modeling:** Star-Schema Modeling, Snowflake-Schema Modeling, FACT and dimension tables, Pivot Tables, Erwin, SAP Power Designer.  
**Tools & Softwares:** TOAD, MS Office, BTEQ, Teradata SQL Assistant  
**Business Intelligence tools**: SSDT-BI, SSIS/SSRS/SSAS solutions

**Work Experience:**

**BMO Harris, Chicago IL April 2018 to Till Date**

**Sr. Data Engineer**

**Responsibilities:**

* Involved in Relational and Dimensional Data modeling for creating Logical and Physical Design of Database and ER Diagrams with all related entities and relationship with each entity based on the rules provided by the business manager using ERWIN r9.6.
* Analysis of functional and non-functional categorized data elements for data profiling and mapping from source to target data environment. Developed working documents to support findings and assign specific tasks
* Accomplished financial tests to ensure compliance with CCAR, BASEL, Dodd-Frank and Sarbanes-Oxley using SQL, Oracle, SAS, DB2, Teradata, and MS Access; being proficient with business intelligence tools such as SSIS, SSRS, TOAD, SAS Enterprise Guide, Teradata SQL Assistant, VBA, Tableau, and Actimize.
* Extracted Mega Data from Amazon Redshift, AWS, and Elastic Search engine using SQL Queries to create reports.
* Designed and developed architecture for data services ecosystem spanning Relational, NoSQL, and BigData technologies.
* Developed Spark/Scala, Python for regular expression (regex) project in the Hadoop/Hive environment with Linux/Windows for big data resources.
* Involved with writing scripts in Oracle, SQL Server and Netezza databases to extract data for reporting and analysis and Worked in importing and cleansing of data from various sources like DB2, Oracle, flat files onto SQL Server with high volume data
* Worked with data investigation, discovery and mapping tools to scan every single data record from many sources.
* Executed Hive queries on Parquet tables stored in Hive to perform data analysis to meet the business requirements and developed Map Reduce programs to cleanse the data in HDFS obtained from heterogeneous data sources to make it suitable for ingestion into Hive.
* Developed automated data pipelines from various external data sources (web pages, API etc) to internal data warehouse (SQL server, AWS), then export to reporting tools like Datorama by Python.
* Used Informatica power center for (ETL) extraction, transformation and loading data from heterogeneous source systems and studied and reviewed application of Kimball data warehouse methodology as well as SDLC across various industries to work successfully with data-handling scenarios, such as data
* Connected to AWS RedShift through Tableau to extract live data for real time analysis and worked on Normalization and De-normalization concepts and design methodologies like Ralph Kimball and Bill Inmon's Data Warehouse methodology.
* Designed and developed architecture for data services ecosystem spanning Relational, NoSQL, and Big Data technologies.
* Used Spark Data frames, Spark-SQL, Spark MLLib extensively and developing and designing POC's using Scala, Spark SQL and MLlib libraries
* Developed normalized Logical and Physical database models to design OLTP system for finance applications and used SQL tools to run SQLqueries and validate the data loaded in to the target tables.
* Worked on analyzing Hadoop cluster and different big data analytic tools including Pig, HBase database and Sqoop.
* Extensively used ERwin for developing data model using star schema methodologies and collaborated with other data modeling team members to ensure design consistency and integrity.
* Involved in Planning, Defining and Designing data base using Erwin on business requirement and provided documentation.
* Performed data analysis and data profiling using complex SQL on various sources systems including Oracle and Netezza and worked on Teradata Appliance Backup Utility (ABU) and ARC to backup data to and from Teradata nodes.

**Environment:** ERWIN 9.6, UNIX, DB2, SQL server, PL/SQL, Informatica Power Center 9.6.1, HBase, Teradata 15.0, Oracle 12c, DB2, SQL server2014, AWS Redshift, RDS, MS Visio, MS Outlook, Spark, Scala, Python, Hadoop, HiveQL Queries, MapReduce, AWS S3, AWS EMR, Sqoop, Oozie, SQL, MongoDB and HBase.

**STERIS Corporation, Mentor OH Jan 2017 to Mar 2018**

**Sr. Data Engineer**

**Responsibilities**

* Involved in all phases of SDLC using Agile and participated in daily scrum meetings with cross teams
* Wrote complex Hive queries to extract data from heterogeneous sources (Data Lake) and persist the data into HDFS.
* Created data integration and technical solutions for Azure Data Lake Analytics, Azure Data Lake Storage, Azure Data Factory, Azure SQL databases and Azure SQL Data Warehouse for providing analytics.
* Involved in all phases of data mining, data collection, data cleaning, developing models, validation and visualization.
* Designed and develop Big Data analytic solutions on a Hadoop-based platform and engage clients in technical discussions.
* Installed, Configured and Maintained the Hadoop cluster for application development and Hadoop ecosystem components like Hive, Pig, HBase, Zookeeper and Sqoop.
* Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig.
* Build data platforms, pipelines, and storage systems using the Apache Kafka, Apache Storm and search technologies such as Elastic search.
* Worked on creating Hive tables and written Hive queries for data analysis to meet business requirements and experienced in Sqoop to import and export the data from Oracle & MySQL.
* Utilized SAS and SQL extensively for collecting, validating and analyzing the raw data received from the client.
* Worked on Hive queries to categorize data of different wireless applications and security systems.
* Responsible in loading and transforming huge sets of structured, semi structured and unstructured data.
* Extensively involved in writing PL/SQL, stored procedures, functions and packages.
* Responsible for Big data initiatives and engagement including analysis, brainstorming, POC, and architecture and created linked services to connect to Azure Storage, on-premises SQL Server and Azure HDInsight
* Designed and implemented scalable Cloud Data and Analytical architecture solutions for various public and private cloud platforms using Azure.
* Implemented logical and physical relational database and maintained Database Objects in the data model using Erwin.
* Worked with NoSQL databases like HBase in creating tables to load large sets of semi structured data coming from source systems.
* Developed numerous MapReduce jobs in Scala for Data Cleansing and Analyzing Data in Impala.
* Managed the Metadata associated with the ETL processes used to populate the DataWarehouse.
* Created Hive queries and tables that helped line of business identify trends by applying strategies on historical data before promoting them to production.
* Configured Azure SQL database with Azure storage Explorer and with SQL server.
* Designed Data Marts by following Star Schema and Snowflake Schema Methodology, using industry leading Data modeling tools like Erwin.

**Environment:** Erwin 9.6, Python, SQL, Hadoop 3.0, SDLC, Azure, HBase 1.2, Pig 0.17, Sqoop 1.4, Zookeeper, Oozie 4.3, SQL, HDFS, Hive 2.3, PL/SQL, , Scala, Apache Nifi, ETL, Excel, Flume 1.8, SSRS and Informatica.

**Target, Minneapolis MN Sept 2014 – Dec 2016**

**Sr. Data Engineer/Data Modeler**

**Responsibilities:**

* Involved in creating Physical and Logical models using ER Studio and presented the data scenarios via, Erwin logical models and excel mockups to visualize the data better.
* Provided subject matter expertise as appropriate to ETL requirements, information analytics, modeling & design, development, and support activities.
* Designed the ER diagrams, logical model (relationship, cardinality, attributes, and, candidate keys) and physical database (capacity planning, object creation and aggregation strategies) for Oracle and Teradata as per business requirements using ER Studio
* Worked with cloud based technology like Redshift, S3, AWS, EC2 Machine, etc. and extracting the data from the Oracle financials and the Redshift database.
* Worked on a MapR Hadoop platform to implement Bigdata solutions using Hive, Map reduce, shell scripting and Pig.
* Created Data Pipeline using Processor Groups and multiple processors using Apache Nifi for Flat File, RDBMS as part of a POC using Amazon EC2.
* Designed 3rd normal form target data model and mapped to logical model and developed the logical data models and physical data models that capture current state/future state data elements and data flows using ER Studio.
* Implemented HQL Scripts in creating Hive tables, loading, analyzing, merging, binning, backfilling, cleansing using hive.
* Used Spark Data frames, Spark-SQL, Spark MLLib extensively and developing and designing POC's using Scala, Spark SQL and MLlib libraries.
* Involved in extensive DATA validation using SQL queries and back-end testing and generated DDL statements for the creation of new ER/studio objects like table, views, indexes, packages and stored procedures.
* Designed, configured and deployed Amazon Web Services (AWS) for a multitude of applications utilizing the Amazon Web Services focusing on high-availability, fault tolerance and auto-scaling.
* Designed and developed cubes using SQL Server Analysis Services(SSAS) using Microsoft Visual Studio and tested the database to check field size validation, check constraints, stored procedures and cross verifying the field size defined within the application with metadata.
* Performed transformations of data using Spark and Hive according to business requirements for generating various analytical datasets.
* Worked on Teradata SQL queries, Teradata Indexes, Utilities such as Mload, Tpump, Fast load and Fast Export.
* Worked with cloud based technology like Redshift, S3, AWS, EC2 Machine, etc. and extracting the data from the Oracle financials and the Redshift database.
* Designing and customizing data models for Data warehouse supporting data from multiple sources on real time. Requirements elicitation and Data analysis. Implementation of ETL Best Practices.
* Troubleshoot test scripts, SQL queries, ETL jobs, and data warehouse/data mart/data store models.
* Implemented solutions for ingesting data from various sources and processing the Data-at-Rest utilizing Big Data through Hadoop, Map Reduce Frameworks, HBase, and Hive.
* Responsible for different Data mapping activities from Source systems to Teradata and Reverse Engineered DB2 databases and then forward engineered them to Teradata using ER Studio.
* Translate business requirements into conceptual, logical data models and integration data models, model databases for integration applications in a highly available and performance configuration using ER/Studio.
* Worked closely with the ETL SSIS Developers to explain the complex Data Transformation using Logic and create and deploy reports using SSRS.

**Environment:** ER- Studio, Oracle 12c, Microsoft SQL Server 2016,AWS, SQL Developer, SQL Manager, SQL Developer Data Modeler, SQL, PL/SQL Developer, ETL, SSIS, Tableau, Hadoop, Hive, Pig, MapReduce, MongoDB, HBase, Spark, Scala, Python, AWS EMR, AWS S3.

**Mckesson Health - Fort Wayne, IN Jan 2013 – Aug 2014**

**Sr. Data Modeler/Analyst**

**Responsibilities:**

* Highly provide hands-on Database, Data Modeling, Data Warehousing, ETL expertise and mentoring and involved in gathering/analyzing the business/user requirements, designing ODS tables in consultation with the Data Architect, analyzing the source/target dependencies & Production Troubleshooting.
* Worked with Big Data Hadoop Ecosystem in ingestion, storage, querying, processing and analysis of big data and conventional RDBMS.
* Responsibilities also include Gathering the Business requirements, Business Analysis, Design, QA testing and final promotion to the production as well.
* Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios.
* Responsible for interaction with business stake holders, gathering requirements and managing the delivery, covering the entire Tableau development life cycle.
* Worked with project management, business teams and departments to assess and refine requirements to design/develop BI solutions using MS Azure.
* Worked on Amazon Redshift and AWS and architecting a solution to load data, create data models and extracted Mega Data from Amazon Redshift, AWS, and Elastic Search engine using SQL Queries to create reports.
* Developed Stored Procedures, Functions & Packages to implement the logic at the server end on Oracle and Performed Application/ SQL Tuning using Explain Plan, SQL Tracing & TKPROF. Also, used Materialized View for the Reporting Requirement.
* Created data masking mappings to mask the sensitive data between production and test environment.
* Handled importing of data from various data sources, performed transformations using Hive, MapReduce, loaded data into HDFS and Extracted the data from Oracle into HDFS using Sqoop.
* Involved in OLAP model based on Dimension and FACTS for efficient loads of data based on Star Schema structure on levels of reports using multi-dimensional models such as Star Schemas and Snowflake Schema
* Worked using Apache Hadoop ecosystem components like HDFS, Hive, Sqoop, Pig, and Map Reduce
* Involved in Data profiling and performed Data Analysis based on the requirements, which helped in catching many Sourcing Issues upfront and performed Data Analysis and data profiling using complex SQL on various sources systems including Oracle 10g/11g and Teradata.
* Extracted data from Oracle 11g and upload to Teradata tables using Teradata utilities FASTLOAD & Multiload.
* Worked with supporting business analysis and marketing campaign analytics with data mining, data processing, and investigation to answer complex business questions.
* Defined ETL architecture which includes load pattern for staging and ODS layer using Microsoft SSIS ETL tool, file archival process, data purging process, batch execution process.
* Performed logical data modeling, physical Data Modeling (including reverse engineering) using the ERWIN Data Modeling tool.
* Created data models for AWS Redshift and Hive from dimensional data models and worked on Data modelling, Advanced SQL with Columnar Databases using AWS.
* Involved using ETL tool Informatica to populate the database, data transformation from the old database to the new database using Oracle.
* Developed Data mapping, Data Governance, Transformation and Cleansing rules for the Master Data Management Architecture involving OLTP, ODS and OLAP and involved in the creation, maintenance of Data Warehouse and repositories containing Metadata.
* Involved in Providing SQL to tollgate and data quality check to ETL to test the data inserted by the Java rules engine into staging table as per requirement and logic to automate the process.
* Developed detailed ER diagram and data flow diagram using modeling tools following the SDLC structure and involved in Provided PL/SQL queries to developer as source queries to identify the data provided logic to assign.

**Environment:** ERWIN, SQL, PL/SQL, XML, Excel, Hadoop, Hive, Pig, MapReduce, HBase, MongoDB, Cassandra, Sqoop, Oozie, AWS, AWS S3, AWS EMR, AWS Redshift, Informatica, ETL, Tableau and SSRS, Oracle, Teradata, Netezza, SQL Server and Kafka.

**Pearson Iowa City IA Sept 2010 – Dec 2012**

**Data Analyst/Data Modeler**

**Responsibilities:**

* Gathered high level requirements and converted into business requirements and created and maintained the data dictionaries and worked with supporting business analysis and marketing campaign analytics with data mining, data processing, and investigation to answer complex business questions.
* Created Entity Relationship Diagrams (ERD), Functional diagrams, Data flow diagrams and enforced referential integrity constraints and created logical and physical models using Erwin.
* Expertise in the Forward Engineering of the logical models to generate the physical model using Erwin and generate Data Models using ERwin and subsequent deployment to Enterprise Data Warehouse.
* Designed Normalized data up to 3rd Normal form and Participated in brain storming sessions with application developers and DBAs to discuss about various de-normalization, Partitioning and Indexing Schemes for Physical Model.
* Created tables, views, sequences, indexes, constraints and generated SQL scripts for implementing physical data model.
* Star schema was developed for proposed central model and normalized star schema to snowflake schema.
* Data mapping documents between Legacy, Production, and User Interface Systems were developed and data sourcing and data staging are performed.
* Working along with ETL team for documentation of transformation rules for data migration from OLTP to warehouse for purpose of reporting.
* Created and maintained Logical Data Model (LDM) for the project. Includes documentation of all entities, attributes, data relationships, primary and foreign key structures, allowed values, codes, business rules, glossary terms, etc.
* Responsible for indexing the tables in the data warehouse and performed data modeling within information areas across the enterprise including data cleansing and data quality using data modeling methods and processes.
* Developed operational data store to design data marts and enterprise data warehouses and metadata repository was maintained and updated based on change requests.
* Involved in Data profiling and performed Data Analysis based on the requirements, which helped in catching many Sourcing Issues upfront.
* Defined ETL architecture which includes load pattern for staging and ODS layer using Microsoft SSIS ETL tool, file archival process, data purging process, batch execution process.
* Served as a resource for analytical services utilizing SQL Server and TOAD/Oracle and Created SQL queries using TOAD and SQL Navigator and also created various databases object stored procedure, tables, views.
* Involved using ETL tool Informatica to populate the database, data transformation from the old database to the new database using Oracle and involved in the creation, maintenance of Data Warehouse and repositories containing Metadata.

**Environment:** Oracle , MS Visio, PL-SQL, Microsoft SQL Server, Rational Rose, ata warehouse, OLTP, OLAP, ERWIN 9.2, Informatica 9.x, Windows, SQL, PL/SQL, SQL Server, Talend Data Quality, Talend Integration Suite 4x, Oracle 10g/9, Flat Files, Windows XP, iReport, SVN