**Sr. Data Analyst**

**Name**: Mir Yousuf Sultan

**Email**: yousufmir888@gmail.com

**Contact**: +1 210-783-7710

**Professional Summary**

* Over 10+ years of professional experience in Data Modeling, Business Data Analysis and design of OLTP systems.
* Proficient in data mart design, creation of cubes, identifying facts & dimensions, star & snowflake schemes and canonical model. Solid hands on experience with administration of data model repository, documentation in metadata portals in such as Erwin, ER Studio and Power Designer tools. Strong working Experience with Agile, Scrum, Kanban and Waterfall methodologies. Extensive experience using R packages like ggplot2.
* Capture, validate and publish metadata in accordance with enterprise data governance policies. Extensive experience in in - depth data analysis on different data bases and structures. Excellency in MS Excel with proficiency in Vlookups, Pivot Tables and understanding of VBA Macros.
* Responsible for data governance rules and standards to maintain the consistency of the business element names in the different data layers.
* Extensive experience in development of T-SQL, Oracle PL/SQL Scripts, Stored Procedures and Triggers for business logic implementation.
* Proficient in Normalization/De-normalization techniques in relational/dimensional database environments and have done normalizations up to 3NF. Expertise in SQL Server Analysis Services (SSAS) and SQL Server Reporting Services (SSRS) tools.
* Experience in conducting Joint Application Development (JAD) sessions with SMEs, Stakeholders and other project team members for requirements gathering and analysis. Conduct data analysis, mapping transformation, data modeling and data-warehouse concepts.
* Software Development Life Cycle (SDLC) experience including Requirements, Specifications Analysis/Design and Testing as per the Software Development Life Cycle. Experience in working with relational databases (Teradata, Oracle) with advanced SQL programming skills.
* Solid Excellent experience in creating cloud based solutions and architecture using Amazon Web services (Amazon EC2, Amazon S3, and Amazon RDS) and Microsoft Azure. Experience in designing visualizations using Tableau software and publishing and presenting dashboards.
* Experience in designing star schema, Snowflake schema for Data Warehouse, ODS architecture by using tools like Erwin Data Modeler, Power Designer, Embarcadero E-R Studio and Microsoft Visio. Hands on experience with big data tools like Hadoop, Spark, Hive, Pig, PySpark, Spark SQL Experience in Big Data platforms like Hadoop platforms (Map-R, Hortonworks& others), Aster and Graph Databases
* Adept and deep understanding of Statistical modeling, Multivariate Analysis, model testing, problem analysis, model comparison, and validation.
* Have ability to build advanced statistical and predictive models, such as generalized linear, decision tree, neural network models, ensembles models, Support Vector Machines (SVM), and Random Forest. Hands on experience in SQL queries and optimizing the queries in Oracle, SQL Server, DB2, and Netezza. Excellent knowledge on creating reports on SAP Business Objects, Web reports for multiple data providers.
* Skilled in managing entire data science project life cycle and actively involved in all the phases of project life cycle including data acquisition, data cleaning, data engineering, features scaling, features engineering, statistical modeling (decision trees, regression models, clustering), dimensionality reduction using Principal Component Analysis, testing and validation using ROC plot, K - fold cross-validation and data visualization. Experience in backend programming including schema and table design, stored procedures, Triggers, Views, and Indexes.
* Experience with a variety of NLP methods for information extraction, topic modeling, parsing, and relationship extraction with developing, deploying, and maintaining production NLP models with scalability. Creative thinking and propose innovative ways to look at problems by using data mining approaches on the set of information available. Good Knowledge on SQL queries and creating database objects like stored procedures, triggers, packages and functions using SQL and PL/SQL for implementing the business techniques.
* Experience and Technical proficiency in Designing, Data Modeling Online Applications, Solution Lead for Architecting Data Warehouse/Business Intelligence Applications. Strong experience in Software Development Life Cycle (SDLC) including Requirements Analysis, and design
* Identifies/creates the appropriate algorithm to discover patterns, validate their findings using an experimental and iterative approach.
* Extensive experience in Text Analytics, generating data visualizations using R, Python and creating dashboards using tools like Tableau.
* Experienced in transforming business requirements into analytical models, designing algorithms, building models, developing data mining and reporting solutions that scale across a massive volume of structured and unstructured data. Expertise in transforming business requirements into analytical models, building models, designing algorithms developing data mining and reporting solutions that scale across a massive volume of structured and unstructured data. Implemented, tuned and tested the model on AWS EC2 with the best algorithm and parameters.

**Technical Skills**

|  |  |
| --- | --- |
| **Cloud Management** | **AmazonWeb** Services(**AWS**), **Amazon** Redshift |
| **Data Modeling Tools:** | ER/Studio V17, **Erwin** 9.7, Power Sybase Designer 16.6. |
| **Programming Languages:** | **SQL**, PL/**SQL**, **UNIXshell** Scripting, PERL, AWK, SED |
| **Databases:** | **Oracle** 12c, **Teradata** R15, **MSSQLServer** 2017, DB2. |
| **Testing and defect tracking Tools:** | HP/Mercury, Quality Center, Win Runner, **MS** Visio 2016 & Visual Source Safe |
| **Operating System:** | Windows 10/8, **Unix**, Sun Solaris |
| **Methodologies:** | RAD, JAD, RUP, UML, System **DevelopmentLifeCycle** (**SDLC**), **Agile**, **Waterfall** Model. |
| **ETL Tools:** | **Informatica**: 9.6/8.6/7.1.4/6.0/5.0, **SSIS**, **Data** Services 3.2, Datastage 8.5 |
| **Utilities/Application:** | **MS** Office Excel, Outlook, PowerPoint, **MS** Project, **MS** Access |
|  |  |
|  |  |

**Educations Details:** Bachelors in computer science From Osmania University India -2012

**Professional Experience:**

**Role: Sr. Data Analyst Oct 2021 to Present**

**Client- BOSCH USA, Farmington Hills, Michigan**

**Responsibilities**:

* Agile methodology was used throughout the project and had daily scrum meeting and bi-weekly sprint planning and backlog meeting.
* Extensively involved in almost all the phases of Project Life Cycle (SDLC) right from Requirements Gathering to Testing and Implementation, Reporting etc. Implemented Exception Handling Mappings by using Data Quality, data validation by using Informatica Analyst.
* Learned the company’s EDW architecture which consists of sources, middleware, staging area, data warehouse, and reporting tools.
* Worked on data profiling & various data quality rules development using Informatica Data Quality. Created Proof of concepts for the different user in the tableau server. Build and developed reports based on the user stories given by the product manager and product owner.
* Held responsibility for integration points and the order in which these must be executed. Evaluated Data Mart’s for the tableau visualization.
* Created extracts from the data source and refresh extracts on the monthly data. Used Type 1 SCD and Type 2 SCD mappings to update slowly Changing Dimension Tables. Creating user filters on the reports and restrict the data and the visualization based on the user requirements.
* Used data blending, groups, sets, dual axis, and synchronized axis to enhance the data visualization. Used Debugger in Informatica Power Center Designer to check the errors in mapping. Wrote SQL, PL/SQL, stored procedures for implementing business rules and transformations.
* Implemented Change Data Capture using Informatica Power Exchange 9.1. Built compelling, interactive dashboards in Tableau that answer key business questions used dashboard actions for the drill down of various reports and URL action to connect to the tableau server.
* Designed, developed Informatica Power Center 9.5 mappings to extract, transform and load the data into Oracle 11g target tables.
* Extensively worked on Informatica tools like source analyzer, mapping designer, workflow manager, workflow monitor, Mapplets, Work lets and repository manager. Closely moved with the Micro strategy reporting team and helped them to get the data for creating report.
* Created test cases and completed unit, integration and system tests for Data warehouse. Involved in Performance tuning for sources, targets, mappings, sessions and server. Managed different server such as production server, Development Server and QA Server.
* Used JavaScript API for the deployment of the visualization in the HTML Page. Prepared an ETL technical document maintaining the naming standards. Published reports from the tableau desktop to the tableau server and also managing the sites and projects on the tableau server.
* Implemented various transformations like Joiner, Aggregator, Expression, Lookup, Filter, Update Strategy, Sequence Generator, and Router etc.
* Worked on ILM to create and assign rules for data masking and generating the workflows. Worked on different databases and flat files such as Oracle, DB2, VSAM, mainframe flat files. Used Constraint Based Loading, partitioning, performance tuning on existing mappings
* Created deployment groups in one environment for the Workflows, Work lets, Sessions, Mappings, Source Definitions, Target definitions and imported them to other environments. Worked closely with Subject Matter Experts, application owners and end users to get detailed working aspects of their respective apps and masking requirements. Used parameters and variables at different levelslike mappings and sessions
* Worked extensively on Data Masking transformation applying all the rules available to mask data in QA/DEV/SIT environments depending on the requirement from the SME. Worked with DBA's for performance tuning and to get privileges on different tables in different environments.
* Used Power Exchange to create and maintain data maps for each file and to connect Power center to mainframe
* Developed Mapplets and created mappings in a pipeline structure to implement reusable transformational logic
* Developed Worklets to run multiple sessions. Migrated repository objects, services and scripts from development to production environment.

**Role: Sr. Data Analyst Nov 2018 to Sept 2021**

**Client: Medco Health Solutions, Irving, Texas**

**Responsibilities**

* Performed complex data analysis in support of ad-hoc and standing customer requests Worked with Systems Development Life Cycle (SDLC)/ Software as a Service (SaaS) delivery models. Queried and analyzed data from Cassandra for quick searching, sorting and grouping
* Designed and implemented secure data pipelines into a Snowflake data warehouse from on-premise and cloud data sources
* Implemented Data Lake in Azure Blob Storage, SAS, Azure Data Lake, Azure Analytics, Data bricks Data load to Azure SQL Data warehouse using Polybasic, SAS, Azure Data Factory Designed and implemented effective Analytics solutions and models with Snowflake.
* Involved in Data profiling, Data analysis, data mapping and Data architecture artifacts design. Worked with Azure Data Factory (ADF) since it’s a great SaaS solution to compose and orchestrate Azure data services. Wrote python scripts to parse XML documents and load the data in database. Used Erwin Data Modeler tool for relational database and dimensional data warehouse designs.
* Designed data pipelines using Azure Data Factory, SAS, IBM DataStage / Info sphere, Azure Copy, Poly base, Multi region data replication
* Involved in complete SSIS life cycle in creating SSIS packages, building, deploying and executing the packages all environments.
* Developed MDM integration plan and hub architecture for customers, products and vendors, Designed MDM solution for three domains.
* Implemented Custom Azure Data Factory pipeline Activities and SCOPE scripts. Used DSE SQOOP for importing data from RDBMS to Cassandra
* Setup storage and data analysis tools in Amazon Web Services cloud computing instances (AWS EC2) and setup storage buckets in the AWS S3 service. Documented data dictionaries and business requirements for key workflows and process points
* Developed reports for users in different departments in the organization using SQL Server Reporting Services (SSRS). Developed and supported on Oracle, SQL, PL/SQL and T-SQL queries. Tested the ETL process for both before data validation and after data validation process.
* Translated business concepts into XML vocabularies by designing XML Schemas with UML Created Reports for marketing analytics team using Adobe analytics. Interacted with stake holders on clearing their doubts regarding the reports in PowerBI
* Worked on creating Ad-hoc reports using SQL server. Actively involved in SQL and Azure SQL DW code development using T-SQL
* Created publishing reports for stake holders using PowerBI. Analyzed escalated incidences within the Azure SQL database.
* Worked on the enhancing the data quality in the database. Designed and developed a horizontally scalable API’S using Python Flask.

**Role: Data Analyst Mar 2016 to Oct 2018**

**Client: Arcelormittal, Chicago, Illinois**

**Responsibilities**:

* Generated Corporate portal dashboards and On boarding reports using PowerBI. Extensively used SQL queries to analyze on boarding data and develop data reports. Used SQL queries extensively to perform validations, code fixes, data extraction, and transformation.
* Optimized the data view that enhanced overall performance by 23% with the creation of real-time dashboards using Power BI.
* Administered and pre-processed data hosted on Blob storage to facilitate Mix panel dashboard development to monitor user health and wellness trends. Created, Deleted and Modified Azure subscriptions, resources, and resources groups as per client request.
* Created pipelines in ADF using Linked Services/Datasets/pipeline to extract, transform and load data from different sources like Azure Blob storage, Azure SQL, Azure Cosmos dB and backwards. Supervised Python scripts for KSAA Survey.
* Ingested data into Azure platform and leveraged Azure ML for healthcare forecasting and cutting healthcare cost for corporations. Built models using techniques like Regression, Tree based ensemble methods, Time Series forecasting, KNN, Clustering and Isolation Forest methods
* Generated reports and dashboards using Tableau that provided actionable insights across departments leading to 15% increase in revenue.
* Automated data pipelines jobs using Azure Data Factory, API as source and blob store as sink and flatten complex JSON files using Data Factory dataflow into Azure SQL for landing before ETL into Azure Synapse. Created detailed BRD reports and created test cases and use case
* Engaged in extensive discussions with clients for enhancements and code fixes and deployed them into Azure cloud.
* Created a detailed business analysis report, outlining problems, opportunities and solutions for a business
* Programmed SQL queries utilizing multiple joins, nested queries to extract and analyze data based on project requirement
* Performed Data cleanup and transformation to support the business process as required
* Developed dynamic visualizations using MS Excel on Time-series and Quantitative data to evaluate the model’s performance and suggest changes Updated reports in MS Excel using pivot tables and LOOKUP, INDEX, MATCH, SUMIFS and other functions.
* Established Data profiling stored procedures using dynamic SQL as well as generated complex scripts to schedule various maintenance tasks.
* Designed and developed Entity-relationship Diagram and Data Dictionary up to date after any alterations were made to the metadata and data in each table. Developed SSRS Reports like Drill through Reports, Drilldown Reports, linked reports and parameterized reports
* Created Data-storage migration space to move physical data block from one to another disk for visualization purpose.
* Created multiple automated reports and dashboards sourced from data warehouse using Tableau/SSRS.
* Developed story telling dashboards in Tableau Desktop and published them on to Tableau Server, which allowed end users to understand the data on the fly with the usage of quick filters for on demand needed information.

**Role: Data Analyst Dec 2014 to Feb 2016**

**Client: Elan Financial Services, Dallas, Texas**

**Responsibilities:**

* Involved in Data mapping specifications to create and execute detailed system test plans. The data mapping specifies what data will be extracted from an internal data warehouse, transformed and sent to an external entity.
* Analyzed business requirements, system requirements, data mapping requirement specifications, and responsible for documenting functional requirements and supplementary requirements in Quality Center.
* Setting up of environments to be used for testing and the range of functionalities to be tested as per technical specifications.
* Tested Complex ETL Mappings and Sessions based on business user requirements and business rules to load data from source flat files and RDBMS tables to target tables.
* Responsible for different Data mapping activities from Source systems to Teradata
* Created the test environment for Staging area, loading the Staging area with data from multiple sources.
* Responsible for analyzing various data sources such as flat files, ASCII Data, EBCDIC Data, Relational Data (Oracle, DB2 UDB, MS SQL Server) from various heterogeneous data sources.
* Delivered file in various file formatting system (ex. Excel file, Tab delimited text, Coma separated text, Pipe delimited text etc.)
* Executed campaign based on customer requirements
* Followed company code standardization rule
* Performed ad hoc analyses, as needed, with the ability to comprehend analysis as needed
* Involved in testing the XML files and checked whether data is parsed and loaded to staging tables.
* Executed the SAS jobs in batch mode through UNIX shell scripts
* Created remote SAS sessions to run the jobs in parallel mode to cut off the extraction time as the datasets were generated simultaneously
* Involved in code changes for SAS programs and UNIX shell scripts
* Reviewed and modified SAS Programs, to create customized ad-hoc reports, processed data for publishing business reports.

**Role:** Data Analyst/Tableau developer**Aug 2013 to Nov 2014**

**Client: Liberty Global, Englewood, Colorado**

**Responsibilities:**

* Successfully accomplished a wide range of development activities involving requirement gathering Developed interactive Tableau Dashboards per business requirements Designed and created databases, tables, views, store procedures and triggers
* Expertly wrote complex Tableau calculations using LOD, Date functions and applying complex, compound calculations to large, complex data sets. Implemented and Scheduled Replication process for updating our parallel servers in autonomous sites
* Make recommendation on systems hardware and software implementation for management approval and took responsibility for complete and timely implementation Use Industry Best Practices when configuring metrics, dashboards, reports, and forms.
* Organized and co-ordinates daily operations and project delivery meeting to identified all possible problems and ensure they are rectified to keep project on schedule and report any project variance to the appropriate upper management personnel.
* Actively involved in data extraction, data cleaning, and setup/prepare workspace, update/modify SQL codes and perform calibration and generated outputs. Performed data migration between different environments and Third-Party Databases.
* Develop and maintain reporting using MS Excel queries and SQL. Worked on data profiling, data analysis and validating the reports sent to third party. High level SQL experience developing tables, organizing data, running queries in an enterprise level setting.
* Develop database objects like tables, views, procedures, triggers, and functions using T-SQL to provide definition, structure and to maintain data efficiently. Collected and analyzed data on current and prospective customers.