**Madhur Mehta**

**SUMMARY:**

* Around **10** years of experience as a Data **Engineer** with **Data Analyst** including designing, developing and implementation of **data models** for applications and systems.
* Experience working with **Agile** and **Waterfall data modeling** methodologies.
* Good understanding and hands on experience with **MS Azure, Amazon Web Services S3** and **EC2**.
* Very Good experience working in **Azure Cloud, Azure DevOps**, **Azure Data Factory,** **Azure Data Lake Storage**, **Azure** **Synapse Analytics, Azure HD Insight**, **Big Data Technologies (Hadoop** and **Apache Spark)** and **Data bricks.**
* Experience working with **data modeling** tools like **Erwin, Power Designer** and **ER/Studio.**
* Experience in **importing** and **exporting** data using **Sqoop** from **HDFS** to **Relational Database Management Systems (RDBMS)** and from **RDBMS** to **HDFS**.
* Experience in working with **Business Intelligence** and **Enterprise Data Warehouse** **(EDW)** including **SSAS**, **Amazon Redshift** and **Azure** **Data** **Warehouse**.
* Experience in Data Migration from **RDBMS** to **Snowflake** **cloud** **data** **warehouse**.
* Expert experience and knowledge in **Master Data/Metadata Management (MDM), Data Quality.**
* Experience in developing **Map Reduce** Programs using **Apache Hadoop** for analyzing the **big** **data** as per the requirement.
* Expert in implementing the projects from end to end & in providing the **Architectural** with emphasis on requirements **analysis**, **design**, **coding**, **testing** and documentation.
* Experience in Data transformation and **Data mapping** from source to target database schemas and also **data cleansing.**
* Expertise in writing **Hadoop** Jobs to analyze data using **MapReduce,** **Hive, Pig.**
* Experience with **SQL** **Server** and **T-SQL** in constructing **Temporary Tables, Table variables, Triggers, user functions, views, Stored Procedures.**
* Good understanding and hands on experience in setting up and maintaining **NoSQL** Databases like **MongoDB,** and **HBase**
* Experience in setting up connections to different **RDBMS** Databases like **Oracle, SQL Server, DB2, Teradata** according to users’ requirements.
* Strong experience in **Data Analysis**, Data Migration, Data Cleansing, Transformation, Integration, DataImport**,** andData Export
* Extensive experience on **Excel, Pivot tables** to run and analyze the result data set and perform **UNIX** **scripting**.
* Worked on Power BI reports & Dashboards with different sources like Excel, **SQL** Server, and Tabular Model.
* Load the data using different connectivity modes like Import, Direct Query and Live connection.
* Worked on M-Language for Data Mashup operations like Split column, delete rows, delete columns, custom column, invoke custom function, pivot, un pivot, transpose, group by, append, and merge in Power BI Query Editor.
* Creation of relationship between tables in Power BI Desktop and cross filter direction.
* Expert in creating and developing Power BI Dashboards and Reports into rich look.
* Expertise in creation of reports/Visualizations like Pie, Bar, Line charts, Sub reports, Drill through, Card, Funnel, Gauge, Drill down, Tables, Matrix, Slicers and Customized Visualizations like Whisker Bar Chart, Box Plot, Power KPI, Advanced KPI, Bullet, Dial Gauge plot, etc.

**TECHNICAL SKILLS:**

* **Big Data tools:** Hadoop3.3, Hive 3.1.2, MapReduce, HDFS, Kafka 2.8, Spark3.0, Pig 0.17, Sqoop 1.4
* **Cloud Services:** Azure Data Factory, AWS Glue and Amazon Redshift
* **Data Modeling Tools:** Erwin 9.8/9.7, ER/Studio V17, Power Designer
* **Database Tools:** **SQL** Server, DB2, Oracle 12c/11g, Teradata15/14, MDM and My**SQL**.
* **ETL Tools:** SSIS, Informaticav10, Informatica Power Center, DataStage.
* **Programming Languages:** **SQL**, T-**SQL**, UNIX shells scripting, PL/**SQL**.
* **Project Execution Methodologies:** JAD, Agile, SDLC, Waterfall, and RAD
* **Reporting tools:** **SQL** Server Reporting Services (SSRS), **Tableau**, Crystal Reports
* **Operating Systems:** Microsoft Windows 10/8/7, UNIX
* **Reporting Tools:** Microsoft Power BI, Power BI Service, SSRS
* **ETL:** Power BI Query Editor, SSIS
* **Data base technologies:** Oracle, **SQL** Server, My**SQL**, No**SQL**, Snowflake, Postgre**SQL**, MongoDB, Cassandra, DynamoDB, Couchbase
* **Cloud:** Amazon Web Services: EC2, S3, ELB, EMR, Auto scaling, Elastic Beanstalk, Cloud Front, Cloud Formation, Elastic Filesystem, RDS, DMS, VPC, Direct Connect, Route 53, Cloud watch, Cloud trail, IAM, SNS, Google Cloud, OpenStack, Azure.
* **SDLC/Testing Methodologies:** Agile, Waterfall, Scrum, TDD

**PROFESSIONAL EXPERIENCE:**

**T Mobile, Seattle WA Jan 2022-Present**

**Sr. Data Engineer/ Data Analyst**

**Responsibilities:**

* As a **Data Engineer**, assisted in leading the plan, building, and running states within the Enterprise Analytics Team.
* Analyzed the **business requirements** of the project by studying the Business Requirement Specification document.
* Involved in migrating Services from **On-premises** to **Azure Cloud Environments**.
* Architected and implemented **ETL** and **data movement** **solutions** using **Azure Data Factory.**
* Worked on **Agile,** testing methodologies, resource management and scheduling of tasks.
* Worked with delivery of Data & Analytics applications involving **structured** and **un-structured** data on **Hadoop** based platforms.
* Created **ETL Pipeline** using **Spark** and **Hive** for ingest data from multiple sources.
* Created **CI/CD Pipelines** in **Azure DevOps** **environments** by providing their dependencies and tasks.
* Designed and implemented **RestAPI** to access **Snowflake DB** platform.
* Involved in implementing and managing **ETL solutions** and automating operational processes.
* Implemented solutions for ingesting data from various sources and processing the Data-at-Rest utilizing **Big Data** through **Hadoop, MapReduce, Pig** and **Hive.**
* Ingested huge volume and variety of data from disparate source systems into **Azure Data Lake Gen2** using **Azure Data Factory V2.**
* Used **Azure Data Lake** as Source and pulled data using **Azure Polybase.**
* Utilized **Oozie workflow** to run **Pig** and **Hive Jobs** Extracted files from **Mongo DB** through **Sqoop** and placed in **HDFS** and processed.
* Created **Power BI** Visualizations and Dashboards as per the requirements.
* Performed **ETL operations** in **Azure Data bricks** by connecting to different relational database source systems using jdbc connectors.
* Built the **data pipeline** using **Azure Service** like **Data Factory** to load the data from Legacy **SQL server** to Azure Data Base using **Data Factories**, **API Gateway Services**, and **Python codes**.
* Implemented Data security (Row level security (RLS)) on Dataset and Reports in Power BI.
* Responsible for Scheduling jobs for weekly, daily data refresh on different Power BI reports.
* Developed and maintained multiple Power BI dashboards in Power BI mobile and Power BI web.
* Analyzed the data flow from different sources to target to provide the corresponding design **Architecture** in **Azure** **environment**.
* Analyzed the existing process and developed **ETL DataStage** parallel jobs.
* Implemented **Data** **Validation** using **MapReduce** programs to remove unnecessary records before move data into **Hive tables.**
* Involved in data transfer using **Azure Synapse** and **Polybase**.
* Carried out data transformation and cleansing using **SQL queries**, **Python** and **PySpark**.
* Used **Polybase** for **ETL process** with **Azure Data Warehouse** to keep data in **Blob Storage** with almost no limitation on data volume.
* Extensively used **Pig** for **data** **cleansing** using **Pig** **scripts** and Embedded Pig scripts.
* Developed **Sqoop** **scripts** to extract the data from **MYSQL** and load it into **HDFS**.
* Effectively loaded real time data from various data sources into **HDFS** using **Kafka**.
* Involved in creating Stored Procedure and Scheduled them in **Azure** **Environment**.
* Provided continued maintenance and development of bug fixes for the existing and new **Power BI Reports.**
* Worked on loading data into **Snowflake DB** in the cloud from various sources.
* Created azure data factory **(ADF pipelines)** using **Azure Polybase** and **Azure blob.**
* Used stored procedure, lookup, execute pipeline, data flow, copy data, azure function features in **ADF**.
* Leveraged **Python** **development** environment for data analysis and report building.
* Worked on **Azure Cloud** technologies such as **Azure Data Lake (ADL),** **Azure Synapse**, **BLOB Storage**, **HD Insight.**
* Created several **Databricks Spark** jobs with **PySpark** to perform several tables to table operations.
* Developed **Databricks Python** notebooks to Join, filter, pre-aggregate, and process the files stored in **Azure** **data lake storage**.
* Developed Data Migration and Cleansing rules for the **Integration Architecture** **(OLTP, ODS, DW).**
* Used **Azure DevOps** to build and release different versions of code in different environments.
* Designed relational and non-relational data stores on **Azure.**
* Analyzed data where it lives by Mounting **Azure Data Lake** and **Blob** to **Databricks**.
* Designed and implemented database solutions in **Azure SQL Data Lakes**, **Azure SQL**
* Updated and manipulated content and files by using **Python scripts.**
* Designed and optimized **indexes, views, stored procedures,** and **functions** using **T-SQL.**
* Designed and implemented **ad-hoc tools** so the users could view the results and compare them against other classrooms, districts, and statewide entities.

**Capital One, Richmond VA Jan 2020-Jan 2022**

**Sr. Data Engineer/ Data Analyst**

**Responsibilities:**

* Massively involved in **Data Engineer** role to review business requirement and compose source to target **data mapping** documents.
* Used the **Agile methodology** to build the different phases of **Software development life cycle (SDLC).**
* Used different types of Filters in Power BI like Visual level Filter, Page level Filter, Report level Filter, Drill-through filter and Slicer.
* Extensively worked on Power BI transformations like Split column, delete rows, custom column, invoke custom function, pivot, un pivot, transpose, group by, append, and merge.
* Experience on power bi Service to creating Gateways, and Schedule refresh.
* Creating Content packs and enables permissions on reports. and procedures including quality and delivery methodologies.
* Implemented Data security (Row level security (RLS)) on Dataset and Reports in Power BI.
* Investigated data sources to identify new data elements needed for **data** **integration**.
* Involved in using **Apache airflow**, **data pipelines** are built to gather data from store-checkout devices into **BigQuery**.
* Worked with **google data catalog** and other **google cloud APIs** for monitoring, query and billing related analysis for **BigQuery usage**.
* Designed and developed a **Data Lake** using **Hadoop** for processing raw and processed claims via **Hive** and **Informatica**.
* Designed and implemented a fully operational production grade large scale data solution on **Snowflake Data** **Warehouse**.
* Developed **data pipeline** using **Sqoop** to ingest cargo data and customer histories into **HDFS** for **analysis**.
* Developed and deployed the outcome using **spark** and **Scala code** in **Hadoop cluster** running on **GCP**.
* Created **HBase** tables to load large sets of **structured, semi-structured** and unstructured data coming from **UNIX, NoSQL** and a variety of portfolios.
* Designed and implemented a Data Lake to consolidate data from multiple sources, using **Hadoop** **stacks** **technologies** like **Sqoop, hive.**
* Defined the key columns for the **Dimension** and **Fact** **tables** of both the **Warehouse** and **Data** **Mart**.
* Developed **MapReduce** programs to parse the raw data, populate staging tables and store the refined data in partitioned tables in the **EDW**.
* Involved in using **g-cloud function** with Python to load Data in to **BigQuery** for on arrival csv files in **GCS bucket.**
* Used **Oozie** to automate data loading into the **HDFS** and **PIG** to pre-process the data.
* Identified the entities and relationship between the entities to develop **Conceptual Model** using **Erwin**.
* Implemented **MDM** **hub** to provide clean, consistent data for a **SOA** implementation.
* Used **cloud shell SDK** in **GCP** to configure the services **Data Proc, Storage, BigQuery**
* Involved in the process of **data acquisition**, data pre-processing and data exploration of telecommunication project in **Scala**.
* Published many reports and visualizations created to the **Power BI service** and thereby developed dashboards.
* Implemented data ingestion and handling clusters in real time processing using **Kafka**.
* Worked with Analysts on various requirements gathered in **JIRA**.
* Created **BigQuery** authorized views for row level security or exposing the data to other teams.
* Worked with **data compliance teams, Data governance team** to maintain data models, Metadata, DataDictionaries define source fields and their definitions.
* Generated **DDL scripts** for database modification, **Teradata**, Macros, Views and set tables.
* Designed and documented Use Cases, Activity Diagrams, Sequence Diagrams, OOD (Object Oriented Design) using **Visio**.
* Used T-**SQL** stored procedures to transfer data from **OLTP** **databases** to staging area and finally transfer into data marts.
* Called **Data Rules** and **Transform Rules** functions using **Informatica** **Stored** **Procedure** Transformation.
* Designed new **SAS** programs by analyzing requirements, constructing workflow charts and diagrams, studying system capabilities, and writing specifications.
* Used **Excel's VLOOKUPs** to determine the customer data and created pivots to easily access and validate data.

**OPTUM HEALTH CARE, Minneapolis MN Jan 2018- Dec 2019**

**Data Engineer/ Data Analyst**

**Responsibilities:**

* Acted as a **Data Modeler/Engineer** to generate Data Models using **Erwin** and developed relational database system**.**
* Conducted and participated **JAD** **sessions** with **Business Analysis** Team, analyze and document the **Business** and **reporting** **requirements**.
* Worked on **AWS Redshift** and **RDS** for implementing models and data on **RDS** and **Redshift**.
* Implemented the model and communicated it to end-users and **stakeholders**.
* Extensively used **DynamoDB** **Stream** to **lambda** using **python**.
* Worked with both **DDL** and **DML** including Joins, Functions, Indexes, Views, Constraints, Primary Keys and Foreign Keys.
* Implemented **Data pipelines** for **big data processing** using **Spark** transformations and **Python** **API** and clusters in **AWS**.
* Exhaustively collected business and technical metadata and responsible for defining the naming standards for **data warehouse**
* Used **Python scripts** to update the content in database and manipulate files.
* Managed storage in **AWS** using **Elastic Block Storage, S3**, created Volumes and configured Snapshots.
* Used **GIT** for the version control and deployed project into **AWS**.
* Designed and built the **OLAP cubes** for **star schema** and **snowflake schema** using native **OLAP** service manager.
* Added support for **Amazon AWS S3** and **RDS** to host static/media files and the database into Amazon Cloud.
* Worked on **SQL** Server concepts **SSIS** (**SQL** Server Integration Services), **SSAS** (Analysis Services) **and SSRS** (Reporting Services).
* Prepared Dashboards using calculations, parameters, calculated fields, groups, sets and hierarchies in **Tableau**.
* Involved in importing and exporting data from different **Relational Database Systems** like **DB2** into **HDFS** and **Hive** and **vice-versa,** using **Sqoop.**
* Configured Inbound/Outbound in **AWS** Security groups according to the requirements.
* Created **stored** **procedures** using **PL/SQL** and tuned the databases and **backend** process.
* Used **AWS Glue** for the data transformation, validate and data cleansing.
* Monitored the **Data** **quality** and integrity of data was maintained to ensure effective functioning of department.
* Specified nodes and performed the data analysis queries on **Amazon redshift clusters** on **AWS.**
* Designed and generated various dashboards, reports using various **Tableau** **Visualizations**.
* Designed and Developed Oracle**,** and **Shell Scripts, Data Import/Export, Data Conversions** and **Data Cleansing.**
* Developed and deployed quality T-**SQL** codes, stored procedures, views, functions, triggers, and jobs.
* Participated in UAT sessions to educate the business users about the reports, dashboards, and the BI Systems.
* Developed calculated columns and measures using DAX (Aggregate, Date & Time, Filter, Time Intelligence, Logical, Information functions) in Power BI.
* Experienced in setup the environment and configuring Power BI gateway for both on-premises and cloud data sources.
* Designed, developed, and implemented Power BI Reports and Dashboards.
* Expertise in Power BI, Power BI Service, Power BI Pro, Power BI Premium, Power BI Mobile.
* Worked on using Power BI features such as bookmarks, selection pane, drill down drill-throughs and cross-filtering.

**BCBS Jacksonville FL May 2016- Jan 2018**

**Data Engineer**

**Responsibilities:**

* Performed **data** **analysis** and **data modeling** using **SQL** **queries** on source systems to identify data discrepancies and determine data quality.
* Developed numerous reports to capture the transactional data for the **business analysis.**
* Extensively involved in developing **logical** and **physical data models** that fit in exactly for the current state and the future state data elements and data flows using **ER/Studio**.
* Analyzed and designed the business rules for **data** **cleansing** that are required by the staging and **OLAP** & **OLTP** database.
* Performed extensive **Data Validation, Data Verification** against **Data Warehouse** and performed debugging of the **SQL**-Statements and stored procedures for business scenarios.
* Conducted design walk through sessions with **Business Intelligence** **team** to ensure that reporting requirements are met for the business.
* Wrote **UNIX shell scripts** to invoke all the stored procedures, parse the data and load into flat files.
* Designed and Developed **PL/SQL procedures, functions,** and **packages** to create Summary tables.
* Identified and documented data sources and transformation rules required to populate and maintain **data** **warehouse** **content**.
* Created **SSIS packages** to Export data from text file to **SQL Server Database.**
* Deployed **SSRS reports** to Report Manager and created linked reports, snapshots, and subscriptions for the reports and worked on scheduling of the reports.
* Load new or modified data into back-end Oracle database.
* Worked with **data investigation, discovery,** and **mapping** **tools** to scan every single data record from many sources.
* Gathered and documented the Audit trail and traceability of extracted information for **data quality.**
* Wrote **SQL scripts** for **creating tables, Sequences, Triggers, views, and materialized views.**
* Designed and developed **Tableau** **dashboards** using stack bars, bar graphs, scattered plots, and Gantt charts.
* Prepared **complex T-SQL queries,** **views,** and **stored** **procedures** to load data into staging area.
* Analyzed massive and highly complex data sets, performing **ad-hoc analysis** and **data** **manipulation.**
* Created reports analyzing large-scale database utilizing **Microsoft** **Excel** Analytics within legacy system.

**Bank of America, Raleigh NC Feb 2015- May 2016**

**Data Engineer/ Data Analyst**

**Responsibilities:**

* Involved in creation/review of **functional requirement specifications** and **supporting documents** for **business systems, experience in database design process** and data modeling process.
* Understanding the useful and technical specification.
* Deep **involvement** with the design and **advancement of Tableau visualization solutions**.
* Experience with creation of **users, groups, projects, workbooks, and the appropriate permission sets for Tableau server logons.**
* Connected to different appliances like **Oracle, MySQL**, **Tera Data**, etc.
* Used **database filters** and **context filters**.
* Expert in creating **Complex Crystal Reports**Ad Hoc Reports, Frequency Reports, Summary Reports, Drill-down, Dynamic grouping, graphical, aging reports.
* Expert in creating various **PL/SQL stored procedures, views, functions,**and temporary tables for data input to the Crystal Reports
* Deployed reports on the web using **Crystal publishing wizard** and scheduled reports using **e-Portfolio** and **Crystal management console.**
* Extensive knowledge in pre-defined functions like **string, date, numeric, conversion and null** functions of Crystal like **isnull, inrepeatedgroupheader, monthtodate, lastfullmonth, lastfullweek** etc.
* Preparing **Dashboards** using **calculations, parameters in Tableau** and Developing **Tableau** reports that provide clear visualizations of **various industry specific KPIs**.
* Understanding of **Tableau features like calculated fields, parameters, table calculations, row-level security, R integration, joins, data blending,** and **dashboard actions.**
* Develop, organize, manage, and maintain graph, table, slide and document templates that will allow for efficient creation of reports.
* Involved in supporting other **Tableau** applications which includes creation and cosmetic changes of different.
* **Migrate existing reports** from **Share-point\excel** to **Tableau Server** by recreating them in batches.
* Suggest alternate representation of **existing reports** to **highlight the KPI in depth**.
* **Interacted with business analyst** to gather **ad hoc business requirements** and designed functional and technical requirements documentation.
* Partner with the Technology teams to deliver a stable and highly available reporting platform.
* Plan and implement the overall analytics and business intelligence strategy.
* Provided **support to consultants**, Technical Support engineers in creating and **troubleshooting Micro strategy SDK customizations.**
* Provided the supporting role in designing the **Security Architecture in the Business Objects 4.1**
* Used Administrator tools like **Project Merge** and **object Manager** to move the objects from one environment to another environment.
* Worked on CMC to create the User-groups and worked on the security rights.
* Worked as a life-cycle manager for business objects Xi R3.1. Promotion Management for 4.1

**Accenture, India Jan 2014 -Jan 2015**

**Jr. Data Analyst/Engineer**

**Responsibilities:**

* Developed Scala scripts using both Data frames/**SQL**/Data sets and RDD/MapReduce in Spark for Data Aggregation, queries and writing data back into OLTP system through Sqoop.
* Extensive expertise using the core Spark APIs and processing data on an EMR cluster.
* Worked on **ETL** Migration services by developing and deploying AWS Lambda functions for generating a serverless data pipeline which can be written to Glue Catalog and can be queried from Athena.
* Developed Hive queries to pre-process the data required for running the business process.
* Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, No**SQL** and a variety of portfolios.
* Implementation of generalized solution model using AWS Sage Maker.
* Involved in designing and deploying multi-tier applications using all the AWS services like (EC2, Route53, S3, RDS, Dynamo DB, SNS, SQS, IAM) focusing on high-availability, fault tolerance, and auto-scaling in AWS Cloud Formation
* Programmed in Hive, Spark **SQL**, Java, C# and Python to streamline the incoming data and build the data pipelines to get the useful insights, and orchestrated pipelines.
* Used Data Frame API in Scala for converting the distributed collection of data organized into named columns, developing predictive analytic using Apache Spark Scala APIs
* Extensive expertise using the core Spark APIs and processing data on an EMR cluster.
* Worked on **ETL** pipeline to source these tables and to deliver this calculated ratio data from AWS to Datamart (**SQL** Server) & Credit Edge server.
* Supporting Continuous storage in AWS using Elastic Block Storage, S3, and Glacier. Created Volumes and configured Snapshots for EC2 instances.
* Experience in using and tuning relational databases (e.g., Microsoft **SQL** Server, Oracle, My**SQL**) and columnar databases (e.g., Amazon Redshift, Microsoft **SQL** Data Warehouse)
* Publishing Reports, workbooks & data source to server and exporting the reports in different Formats.

**EDUCATION**

* B. Tech