**Soham Salvi**

Rochester, MN | (507)319-5479 | sohamsalvi01@gmail.com

**Summary**

IT professional for 6 years as Data Analyst and Developer with experience in data modeling, evaluating data sources and data warehouse/data mart design, ETL, OLTP, and client/server applications.

* Excellent experience in data mining large datasets to discover transition patterns, identifying mismatches and examine healthcare data in Epic, Cerner, McKesson Star.
* Experience in analysing data sources like Epic system, Person database, flat-files, relational and non-relational databases, and CCDA / HL7 files.
* Design logical and physical databases, data migration and source to target mapping.
* Experience in creating tables, reports, graphs and listings using various procedures and handling large databases to perform complex data manipulations.
* Experience ETL process like SSIS packages, SQL queries, stored procedures and data cleaning to transfer data between SQL Server, Oracle, and flat files.
* Strong experience in writing DDL and DML queries for creating tables, stored procedures, creating reports, UDF, views and indexes.
* Certification in OCA, Machine Learning and Data Science.
* Hands on in algorithms like Neural Network, SVM, K-means, PCA, and Recommended System in Octave. Used methods like gradient boosting, stochastic GD, RBF Kernel and Lasso-Rigid to improve performance.
* Hands on in coding R, Python and Octave. Used spark libraries like pySpark and spark.sql to prepare data/features, create data pipeline and evaluate linear and logistic models.
* Understanding of error analysis in predictive model using confusion matrix, ROC curve and regression evaluation methods.

**Technical Skills**

|  |  |
| --- | --- |
| **Languages** | SQL, T-SQL, PL/SQL, C#, C, C++, Core Java, Linux, XML |
| **BI Tools** | SQL Server 2014/2012/2008, BIDS 2010, SSIS, SSRS, Oracle 9i/10g, DB2 10.5 |
| **Modeling Tools** | IDA (InfoSphere Data Architect), UDMH Framework |
| **Development Tools** | Eclipse Mars, MS Excel, Rapid SQL, R-Studio, Juypter Notebook, Octave, SSMS, SQL\* Plus |
| **Versioning Tools** | MS Excel, GIT, JIRA, TFS, Sharepoint |
| **Big Data Tools** | Hadoop, Spark, Linux, Octave, R, Python (libraries like numpy, pandas), pySpark, spark.sql |

**Professional Experience**

**CitiusTech Inc, MN**

**Feb 2016 – Till Date**

**Technical Lead**

**Project**: Epic Retrofit at Mayo Clinic

**Technologies**: DB2 10.5, RapidSQL, Datastage, Epic Clarity, Hyperspace, SQL, MS Excel, Sharepoint.

**Accomplishments**:

* Perform data validation for hospital supplies system with epic source. Testing data elements of database and DataMart for table, keys on the basis of source to target mapping.
* Write sql queries for clarity data like patient encounter, appointments, provider schedules, hospital supplies, locations and physician.
* Modify source queries to fix data issue for supplies, manufactures, supply types/charges, location, departments, sites, specialty and providers and procedures for clarity source.
* Test DataMart’s and business test cases using source queries, source to target mapping and business requirement documents.
* Analyze data for supplies and implants and creating data mapping.
* Analyze source data to identifying business keys, lookup’s and business logic.
* Modified complex source queries and validated its output in target table.
* Write sql queries to test/analyze data for patients, encounters, appointments, and provider schedule. Write sql queries and created report in excel for users.

**Project**: DaVita Medical Group

**Technologies**: IDA, UDMH framework, Oracle 11g, Informatica Power Center, MS Excel, JIRA

**Accomplishments**:

* Analyzed healthcare codes to map to domain tables.
* Created logical model of mapped tables in IDA.
* Analyzed source tables to create mapping sheet and map to UDMH Model.

**Project**: Geisinger Subscription Notification for KeyHIE

**Technologies**: SQL Server 2014, SSIS, VS 2013, T-SQL, JIRA, GIT, Eclipse Mars.

**Accomplishments**:

* Worked as lead modeler in the project to create data model for CCDA files.
* Wrote DDL and DML scripts to migrate data to new data model. Maintained script versioning in sync with application releases.
* Maintained Dev, QA, Prod databases in sync with application fixes in different environments.
* Implemented table partitioning, tuned archiving stored procedures, and tuned UI queries.
* Used SQL Performance Dashboard and execution plan to tuned application queries.
* Implemented table partitioning and indexing to support batch loading, application queries and data migration.
* Implemented user-defined-functions, indexes, views and clustered & non-clustered indexes.
* Created SSIS packages, stored procedures and functions to extract/load flat files in batches.
* Implemented data lineage across database tables to handle batch failures.
* Monitored database performance at volume testing of application and at index rebuilding.
* Used data flow tasks like derived columns, Execute SQL Task in ssis packages.
* Used control flow tasks like File System Task to move files, and Ftp Task with winscp to encrypt and secure ftps connection.

**Med-Metrix, NJ**

**Jan 2013 – Jan 2015**

**Database Developer II**

**Technologies**: MS SQL Server 2012, BIDS 2010, Oracle 10g, SSIS, SSRS, C#.

**Accomplishments**:

* Developed, implemented and supported multiple SSIS packages, handled batch failures, and monitored data quality.
* Performed gap analysis for flat files and data mapping according to data model specification.
* Developed SSIS packages to load data into staging tables and databases.
* Monitored data quality and fixed data issues for hospital codes, account balances.
* Created C# script in script task to clean data and load into staging tables. Used C# scripting to parse non-delimited files, change file names, copy/delete files on ftp.
* Used for-loop container to load multiple files and archive them.
* Use SCD transformation task to handle type-2 transformation.
* Configured emails using SQL Mail Agent to send email alerts. Used Transaction-Option property to handle package failure and rollback transactions of task or container.
* Used sql server built-in functions to clean data elements and apply transformations.
* Used OLEDB Source to fetch data from Oracle and load into SQL Server staging table.
* Modified stored procedures to load data from staging tables to databases.
* Created logging tables, functions to log time in stored procedures. Used sql execution plan to improve performance.
* Developed stored procedures to mine data elements from large datasets.
* Handled complex logic to find anomalies in charges and transactions of patients/ATB accounts.
* Scheduled packages by implementing execute SQL Server Agent jobs, shrink database, and rebuild index tasks.
* Used Try-Catch blocks to log errors in stored procedures. Debugged / tested packages using breakpoints to handle errors.
* Used dynamic sql to create parameterized queries.

**Zodiac Solutions, PA**

**May 2012 – Dec 2012**

**SQL Developer**

**Technologies**: Teradata V2R6, MS SQL Server 2008/2005, BIDS 2005, SSIS, C#.

**Accomplishments**:

* Fetched data from Teradata using Microsoft attunity connector.
* Analysed and validated data in staging tables.
* Used C# script in Script Task to parse date, time, and decimal values; change to ascii characters from report files.
* Wrote sql queries to check data quality like min, max, null values and duplicate rows.
* Scheduled .dtsx packages using SQL Job Agent.
* Debugging and handling errors in failed packages.

**Auberle, PA**

**Sept 2011 – Apr 2012**

**SQL Developer**

**Technologies:** MS SQL Server 2005, T-SQL, Excel.

**Accomplishments:**

* Load files using bulk insert & BCP utility into SQL Server database.
* Created SQL queries to extract from multiple tables for users.
* Optimized stored procedure and queries using temp & derived tables and used indexes.

**Education**

* B.S. in Information Science & Technology, Pennsylvania State University
* B.Sc. in Information Technology, University of Mumbai

**Certifications**

* Machine Learning, Stanford University
* Introduction to Data Science, University of Michigan
* R-programming, John Hopkins University
* Oracle Certified Associate (OCA), Oracle
* C Programming, NIIT