**Harika b**

**PROFESSIONAL SUMMARY:**

* 8 plus years of IT experience in Data Analytics and Visualization- 4 Years in Tableau Desktop, 2 years in Alteryx, SQL and
* Business Intelligence and 1 plus year in Big Data Technologies like Hive and Sqoop.
* Expertise in Tableau Desktop 10.5, Tableau Server, Tableau Online, Alteryx Designer, ETL, SAP CRM 7.4, SQL server,
* Tableau Prep, Tableau Reader, Microsoft Excel and Oracle Database
* Proficient in Big Data Tools and Technologies like Sqoop and Hive.
* Experience in all phases of project development life cycle– Analysis, Design, Development, Testing and Production
* Support.
* Expertise in the design and development of Tableau visualization and dashboard solutions using Tableau Desktop and
* publishing them on Tableau Server/Public/Reader and external websites.
* Experience in creating Tableau scorecards, dashboards using stack bars, bar graphs, scattered plots, geographical maps,
* Gantt charts using show me functionality.
* Worked on extracting data from multiple data sources and created data extract files (TDE) for fast query execution in
* Tableau.
* Experience in Informatica (ETL) tools which is widely used in building enterprise data warehouse.
* Involved in Trouble Shooting, Performance tuning and resolving issues within Tableau Reports.
* Expertise in writing SQL Queries and complex joins for building CTE, Views, etc.
* Good experience in creating drill down, drill through, matrix, parameterized reports.
* Worked on data Extraction, Transformation and Loading (ETL) using Alteryx which helped to easily move and manipulate
* data from heterogeneous environments.
* Experience working with Extract &amp; Live connections with high data volumes and connecting to multiple data sources like
* Oracle, MS Access, Excel, and My SQL using Data Blending and Cross Database Joins.
* Strong experience in databases such as Oracle 10g/ 9i, SAP, MY SQL and SQL Server.
* Excellent analytical and logical programming skills, good understanding at conceptual level.
* Excellent Communication, Team Player with Organizational and Interpersonal skills.
* Good communication, problem solving and interpersonal skills, versatile team player as well as independent contributor with adaptability and understanding of business processes.
* Highly motivated team player with analytical, organizational and technical skills, unique ability to adapt quickly to challenges and changing environment.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Methodologies** | SDLC Waterfall, Scrum and Agile |
| **BI Tools** | SQL Server Integration, Analysis Services and Reporting Services (SSIS/SSAS/SSRS) (2017/2015/2012/2008), Tableau (9.1/8.3/8.2), Power BI, SQL Server Management Studio (SSMS), SQL profiler, Microsoft Visual Studio. Cognos 8 |
| **RDBMS** | MS SQL Server 2016/2014/2012/2008, MS Access, OLTP, OLAP. |
| **Languages** | SQL, T-SQL, Java, C, C++, Python |
| **Browsers** | Internet Explorer, Mozilla Firefox, Google Chrome, Safari |
| **Operating Systems** | UNIX, WINDOWS 7/8/10, Linux |
| **MS Office Tool** | MS word, MS Excel, MS PowerPoint, MS Project, TFS, GitHub |

**PROFESSIONAL EXPERIENCE:**

**Humana – Louisville, KY June 2021 – Till Date**

**Data Engineer**

**Responsibilities:**

* ETL jobs were created to load server data and data from other sources into buckets and move S3 data into the data warehouse.
* Utilizing both organized and unstructured data, reports and dashboards were produced.
* Used PySpark, to join different tables and then used EMR to perform analytics on top of them.
* Designed and developed a Scala pipeline to transform and store data in the Parquet format and allow HIVE.
* Developing controllers that perform data transformation tasks in accordance with business needs.
* Used spark streaming to transform data in real-time.
* Created several Python scripts that do SQL injection, permission checks, and analysis to discover SQL query vulnerabilities.
* By creating a customized read/write Snowflake utility function in Scala, data was transferred from an AWS S3 bucket to Snowflake.
* Participated in the creation of apps in Scala and Spark. Created Python Spark Streaming tasks to download JSON files from AWS S3 buckets and read messages from Kafka.
* Designed Python functions and Spark programs that conduct operations and transformations on data sets.
* Actively gathered project planning and development needs from business customer SMEs.
* To automate EMR Spin for Data Lake jobs, Lambda and other dependencies were deployed into AWS.
* Configure spark jobs for continuous integration and deployment to EMR clusters. Spark tasks/applications scheduled in the AWS EMR cluster.
* Spark SQL was used to connect to Hive to read data, and distributed processing was used to make the system very scalable.
* In charge of monitoring and troubleshooting the Spark Databricks cluster, as well as determining the cluster size.
* Modified and maintained SQL Server stored procedures, views, ad-hoc queries, and SSIS packages used in the search engine optimization process.
* Gathered business requirements and converted them into new T-SQL stored procedures in visual studio for database project.
* Developed Big Data solutions focused on pattern matching and predictive data modeling.
* Developed UML models to shift the focus to the teams and/or individuals who are responsible for executing the plans.
* Using Tableau/OBIEE, we wrote data and loaded it into a Data Lake environment (SNOWFLAKE) that was accessed by business users and data scientists.
* Developed a plan and moved databases to the AWS cloud from on-premises locations effectively.
* Made several dashboards in Tableau for various business requirements.
* Published dashboards, reports, and spreadsheets with interactive data visualization on Looker.
* For effective data access, partitioning, dynamic partitions, and buckets were implemented in HIVE.

**Environment:** HDFS, Hive, Scoop, Pig, Oozie, Azure, Amazon Web Services (AWS), SNS, GLUE, Python, GIT, SSIS, T-SQL, Jenkins.

**Volkswagen Credit – Libertyville, IL December 2019 – May 2021**

**Data Engineer**

**Responsibilities:**

* Performed data analysis and developed analytic solutions. Data investigation to discover correlations / trends and the ability to explain them.
* Worked with Data Engineers, Data Architects, to define back-end requirements for data products (aggregations, materialized views, tables – visualization)
* Developed frameworks and processes to analyze unstructured information. Assisted in Azure Power BI architecture design
* Experienced with machine learning algorithm such as logistic regression, random forest, XGboost, KNN, SVM, neural network, linear regression, lasso regression and k - means
* Implemented Statistical model and Deep Learning Model (Logistic Regression, XGboost, Random Forest, SVM, RNN, and CNN).
* Designing and Developing Oracle PL/SQL and Shell Scripts, Data Import/Export, Data Conversions and Data Cleansing.
* Architect & implement medium to large scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics
* Performing data analysis, statistical analysis, generated reports, listings and graphs using SAS tools, SAS/Graph, SAS/SQL, SAS/Connect and SAS/Access.
* Developing Spark applications using Scala and Spark-SQL for data extraction, transformation, and aggregation from multiple file formats. Using Kafka and integrating with the Spark Streaming. Developed data analysis tools using SQL and Python code.
* Authoring Python (PySpark) Scripts for custom UDF’s for Row/ Column manipulations, merges, aggregations, stacking, data labeling and for all Cleaning and conforming tasks. Migrate data from on- premises to AWS storage buckets.
* Agile methodology including test-driven and pair-programming concept.
* Created functions and assigned roles in AWS Lambda to run python scripts, and AWS Lambda using java to perform event driven processing.
* Involved in Installation QlikView 12.0 SR5, Nprinting 16/17 in both publisher and server.
* Involved in testing dashboards of Qlikview 11.2 version to migrate it to Qlikview 12.1. Extensive experience with Extraction, Transformation, Loading (ETL) process using Ascential Data Stage EE/8.0/
* Developed a python script to transfer data, REST API’s and extract data from on-premises to AWS S3. Implemented Micro Services based Cloud Architecture using Spring Boot.
* Worked on Ingesting data by going through cleansing and transformations and leveraging AWS Lambda, AWS Glue and StepFunctions.
* Created yaml files for each data source and including glue table stack creation. Worked on a python script to extract data from Netezza databases and transfer it to AWS S3
* Built the machine learning model include: SVM, random forest, XGboost to score and identify the potential new business case with Python Scikit-learn.
* Analyzed existing Data Model and accommodated changes according to the business requirements.
* Decommission 2 data marts and expanded the data model of an existing Oracle DW used as their Data Mining and Metrics repository.
* Implemented a top-down/bottom-up analysis approach to assess the current warehouse environment and extend the to-be data models.

**Environment:** Hadoop, Map Reduce, HDFS, Hive, Ni-fi, Spring Boot, Cassandra, Swamp, Data Lake, Sqoop, Oozie, SQL, Kafka, Spark, Scala, Java, AWS, GitHub, Talend Big Data Integration, Solr, Impala.

**BookMED – India January 2019 – October 2019**

**Data Engineer**

**Responsibilities**

* 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.
* Supporting Continuous storage in AWS using Elastic Block Storage, S3. Created Volumes and configured Snapshots for EC2 instances.
* 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.
* Developed Python 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.
* Developed Hive queries to pre-process the data required for running the business process.
* 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.
* Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios.
* Implementations of generalized solution model using AWS SageMaker.
* Extensive expertise using the core Spark APIs and processing data on an EMR cluster.
* Programmed in Hive, Spark SQL, Java and Python to streamline the incoming data and build the data pipelines to get the useful insights, and orchestrated pipelines.
* Worked on ETL pipeline to source these tables and to deliver this calculated ratio data from AWS to Datamart (SQL Server) & Credit Edge server.
* Experience in using and tuning relational databases (e.g. Microsoft SQL Server, Oracle, MySQL) and columnar databases (e.g. Amazon Redshift, Microsoft SQL Data Warehouse).

**Environment:** Hortonworks, Hadoop, HDFS, AWS Glue, AWS Athena, EMR, Pig, Sqoop, Hive, NoSQL, HBase, Shell Scripting, Python, Scala, Spark, SparkSQL, AWS, SQL Server, Tableau, ETL

**Byte Alpha Solutions – India April 2017 – December 2018**

**Data Analyst**

**Responsibilities:**

* Gathered the business requirements from Project Manager and Team to get a better
* understanding of the process.
* Optimizing the New Mexico/ Pennsylvania market using Visualization software such as Tableau
* and Power BI.
* Collection of data using Microsoft Excel, SQL Server DB, One Data, Liberty reports.
* Creating and maintaining dashboards in Microsoft PowerBI to show the data in a way that
* management can use and create plans to keep the business running efficiently.
* Created and managed SSIS Packages to migrate the data from Multiple SQL sources to the
* Destination.
* Performance improvement of the SQL/reports by creating optimized queries.
* Created Variables, Constrains and expressions in SSIS packages as per the business logics to load
* data into required tables.
* Analyzed and interpreted the data to find any abnormalities using statistical techniques, DAX
* expressions.
* Created Calculated Measures, Custom columns using Power BI Desktop to generate reports for
* Business analysis.
* Worked with Excel Pivot tables.
* Document and publish reports, troubleshoot, and escalate issues.
* Created user-oriented dashboards and meaningful, actionable data visualizations to meet the
* needs of the Market for monthly and quarterly reports.
* Increased productivity by creating new stored procedures and optimizing existing queries and
* stored procedures.
* Used Power BI Gateways to keep the dashboards and reports up to date.
* Installed and Configured Enterprise Gateway and Personal Gateway in Power Bi Services.
* Present all data insights and findings in succinct and creative ways using data visualizations,
* providing trends and forecasts, explaining processes and techniques, recommending actions.

**Environment**: Power BI, Tableau, Oracle DB,

**InfoTech Enterprises Ltd – India May 2016 – March 2017**

**Data Analyst**

**Responsibilities:**

* Analyzed the business requirements and functional specifications.
* Extracted data from oracle, Sql Server database and other file sources and staged into a single place and applied business logic to load them in target Oracle database.
* Develop and modify PL/SQL procedures, functions, sql queries per technical and functional requirements to make new enhancements.
* Develop / modify scripts to create tables and views. Use SQL Loader to load data
* Responsible for creation and maintenance of interface files which requires modifying UNIX Shell Scripts
* Created documents in each phase of the design, development, implementation and support such as Functional and Technical document, Test Cases, User Manual and Support documents.
* Wrote functional design documents and enhance documents as changes are made to meet audit requirements. Document the modifications and intended functionalities of interfaces
* Followed-up on any issues or risks and ensured resolution was identified or remediation plan was in place
* Participate in production support on-call rotation
* Participate in data management projects which requires working with cross functional teams and responsibilities includes, data mapping (source to target mapping and tracking transformations), create data flow diagrams using Visio and field mapping documents
* Defined and performed systems analysis and development tasks to improve workflow and operating efficiency
* Involved in the development of PL/SQL stored procedures, functions and packages to process business data in OLTP system.
* Worked with SSIS Import/Export Wizard for performing the ETL operations.
* Helped data modelers to prepare models and reviewed it to check if the requirements are met.
* Assisted Data Analyst in mapping task and analyzed existing reports.
* Carried out changes into Architecture and Design of Oracle Schemas for both OLAP and OLTP systems.
* Worked with Services and Portal teams on various occasion for data issues in OLTP system.
* Worked with the testing team to resolve bugs related to day one ETL mappings before production.

**Environment:** PL/SQL, Oracle 12c, Toad, ETL, Unix scripting

**Value Labs – India July 2014 – April 2016**

**Data Analyst**

**Responsibilities:**

* Worked as a part of the automation team. Aimed to reduce the overall TAT of claim disbursal.
* Used technologies like hive and sqoop to import data from IRIS system and store the output in a table which was thereafter provided to the RPA team to be consumed by the bots.
* Worked on sql queries to fetch the required information from IRIS (Oracle Database) by making use of complex queries.
* Ensured data consistency between IRIS and RxClaims (United Health Care Claims Database) by monitoring the data ingestion through the ETL process.
* Scheduled automated reports on a weekly and monthly basis as per client’s requirement by making use of cronjob using UNIX shell scripting.
* Monitored all the existing automations in production and provided the break fix whenever needed.
* Worked with Big data, Oracle and SQL server as the data source.
* Interacted with Business users, Scrum Master, Product owners and Business Analysts to understand requirements and transform them into technical specifications.

**Environment:** Hive, Sqoop, Oracle, SQL server, UNIX, Shell scripting, Putty, Winscp

**Education:** Bachelor’s Degree in India.