**Shiva K**

Shivak1811@gmail.com

202-599-5197

[www.linkedin.com/in/shiva-k17](http://www.linkedin.com/in/shiva-k17)

**Summary**

* Passionate Data Engineer/Scientist having strong background in statistics with 7+ years of experience, seeking to solve the boatnecks of the business.
* Expertise in Statistical analysis, Predictive modeling, Text mining, Supervised learning, Unsupervised Learning.
* Highly Experienced on Data Acquisition processes, handling complex data sets with multiple file formats (CSV, JSON).
* Worked on building ETL pipelines using AWS Glue and Python to create Dashboards in Tableau.
* Experience in Big Data Analytics and developing data models using Hive, PIG, and Map reduce, SQL with strong data architecting skills designing data-centric solutions.
* Highly Skilled at python coding using SQL, NumPy, Pandas and Spark for Data Analysis and Model building, Cognitive Design, deploying and operating highly available, scalable and fault-tolerant systems using Amazon Web Services (AWS).
* Implemented & maintained the branching and build/release strategies using Version Control tools GIT, Subversion, Bitbucket and experienced in migrating GIT repositories to AWS.
* Strong experience in Data Migration, Data Cleansing, Transformation, Integration, Data Import, and Data Export.
* Implemented Apache Airflow for authoring, scheduling and monitoring Data Pipelines.
* Experience with using most common Operators in Airflow - Python Operator, Bash Operator, Google Cloud Storage Download Operator, Google Cloud Storage Object Sensor, GoogleCloudStorageToS3Operator etc.
* Expertise in Statistical analysis, Predictive modeling, Text mining, Supervised learning, Unsupervised Learning, and Reinforcement learning.
* Extensively worked on statistical analysis tools and adept at writing code in Advanced Excel, R, MATLAB, Python.
* Integrated Python with Big Data and Analytics based on Hadoop, Spark and No-SQL databases like HBase and MongoDB.
* Excellent Understanding of Spark and its benefits in Big Data Analytics
* Skilled in developing applications using Amazon Web Services like EC2, Virtual private clouds (VPCs), Storage models (EBS, S3, and instance storage), Elastic Load balancer ELBs.
* Good Understanding of other AWS services like S3, EC2, IAM, RDS Experience with Orchestration and Data Pipeline like AWS Step functions/Data Pipeline/Glue.
* Experienced in using AWS Lambda and Glue for creating highly functional Data pipelines
* Implemented AWS solutions using EC2, S3, RDS, EBS, Elastic Load Balancer, Auto scaling groups.
* Designed and engineered on-premise to off-premise CI/CD docker pipelines (integration and deployment) with ECS, Glue, Lambda, ELK, Spark DataBricks, firehose and kinesis stream.
* Developed real-time data feeds and microservices leveraging AWS Kinesis, Lambda, Spark Streaming, etc. to enhance useful analytic opportunities and influence customer content and experiences

**TECHNICAL SKILLS:**

* Cloud Infrastructure: AWS (Glue, EC2, Lambda, S3, EMR), Azure
* Containers and Orchestration: Docker, Kubernetes, AWS ECS and AWS EKS
* Tools: Toad, Erwin, Apache Airflow, Tableau
* Spark, Hive, Big Query
* Machine Learning Algorithms: Linear Regression, Logistic Regression, Decision Trees, Random Forest, NLP
* Python: Spark ML-lib, Scikit-learn, Pandas, NumPy
* SQL, TSQL, NoSQL, Postgres SQL, DynamoDB

**Professional Experience**

**Data Engineer, JPMorgan Chase – Tampa, Florida Oct 2019 – Present**

* Experienced in building and architecting multiple Data pipelines, end to end ETL and ELT process for Data ingestion and transformation in AWS and PySpark.
* Transformed the raw data into actionable insights by incorporating various statistical techniques and using data mining tools such as Python (Scikit-Learn, Pandas, NumPy, Matplotlib) and SQL.
* Performed Data Analysis, Data Migration, Data Cleansing, Transformation, Integration, Data Import, and Data Export through Python.
* Experience in Creating, Scheduling and Debugging Spark jobs using Python.
* Designed and developed architectures for data services ecosystem spanning Relational, NoSQL, and BigData technologies.
* Worked with Hadoop eco system covering HDFS, HBase, YARN, Hive, PIG, Sqoop, Flume and MapReduce.
* Performed the Data Mapping, Data design (Data Modeling) to integrate the data across the multiple databases into EDW.
* Designed and Developed ETL Processes in AWS Glue to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into AWS Redshift.
* Wrote Redshift UDFs and Lambda functions using Python for custom data transformation and ETL.
* Build servers using AWS, importing volumes, CloudFront, launching EC2, RDS, creating security groups, auto-scaling, load balancers (ELBs) in the defined virtual private connection and open stack to provision new machines for clients.
* Experienced in using AWS Boto3 SDK to build serverless functions in AWS cloud environment.
* Data Extraction, aggregations and consolidation of Adobe data within AWS Glue using PySpark.
* Developed and designed system to collect data from multiple portal using Kafka and then process it using spark.
* Experienced in full SDLC management for designing, analyzing, developing, testing, Implementation and application support.
* Built models using Python and PySpark to predict the probability of attendance for various campaigns and events.
* Performed data visualization and Designed dashboards with Tableau, and generated complex reports, including charts, summaries, and graphs to interpret the findings to the team and stakeholders.

 **Environment:** Python, SQL, Tableau, MLlib, regression, Cluster analysis, SVM, Spark, MongoDB, Unit Testing, logistic regression, Hadoop, PySpark, SNS, Elastic Load Balancing, Random Forest, regression test, Apache Airflow, AWS Glue.

**Data Engineer, Blue Cross Blue Shield - Boston, MA Feb 2019 – Sep 2019**

* Used informatica Power center for (ETL) extraction, transformation and loading data from heterogeneous source systems into the target database.
* Extensive experience in ETL design, development, and maintenance using Oracle SQL, PL/ SQL, SQL Loader.
* Created mappings using Designer and extracted data from various sources, transformed data per the requirement.
* Developed informatica Mappings and Reusable Transformations to facilitate timely Loading of Data of a Star
* Developed informatica Mappings using Aggregator, SQL Overrides in Lookups, source filer usage in source qualifiers, and data flow management into multiple targets using Router.
* Development of Unix shell scripting, PL/SQL processes (Triggers, functions, procedures, packages), AB Initio ETL jobs, Business Objects reports, and universes.
* Generated Data profile reports using Data explore in Informatica to understand the data sources and target data which is processed through ETL processes.
* Prepared Logical Data Models using ERWIN that contains set of diagrams and supporting documents containing the essential business elements, detailed definitions, and descriptions of the relationships between the data elements to analyze and document business data requirements.
* Leverage tools for capturing business & technical metadata, data linage & transformation rules for assigned data domains/data sets with special attention to data profiling, data testing, and metric reporting. Emphasis on building data scorecards.
* Expertise in principles of structuring information and in documentation, software development, and business processes.
* Worked on Data blending & extract data through connecting various sources like TM1, MS-Excel.
* Used filters, quick filters, sets, parameters and calculated fields on Tableau and PowerBI reports.
* Communicated the results with operations team for taking best decisions.

**Environment:** Python, PowerBI, AWS Glue, SQL, Tableau, MLlib, regression, Snowflake, Cluster analysis, NLP, Spark, Kafka, MongoDB, logistic regression, Hadoop, PySpark, random forest.

**Data Engineer, CEAT – Mumbai, India Nov 2016 – Dec 2018**

* Designed and implemented customized Linear regression model to predict the sales utilizing diverse data sources to predict demand, risk and price elasticity.
* Responsible for Data Cleaning, data quality, features scaling, features engineering by using NumPy and Pandas in Python.
* Developed the tools using Python Django and used MongoDB for databases. Parsers written in Python for extracting useful data from the design data base. Used Parse kit (Enigma.io) framework for writing Parsers for ETL extraction.
* Performed the statistical analysis and compared the models using R2. Created Visualization graphs and competed models based on R2 Values and Correlation.
* Developed a fully automated continuous integration system using Git, Gerrit, Jenkins, MySQL and custom tools developed in Python and Bash.
* Experimented with multiple classification algorithms, such as Logistic Regression, Support Vector Machine (SVM), Random Forest, AdaBoost and Gradient boosting using Python Scikit-Learn and evaluated the performance on customer discount optimization on millions of customers.
* Designed and developed architecture for data services ecosystem spanning Relational, NoSQL, and BigData technologies.
* Worked with Hadoop eco system covering HDFS, HBase, YARN and MapReduce.
* Performed the Data Mapping, Data design (Data Modeling) to integrate the data across the multiple databases into EDW.
* Used Informatica & SAS to extract transform & load source data from transaction systems.
* Built models using Python and PySpark to predict the probability of attendance for various campaigns and events.
* Developed Machine Learning models to build media planning and optimization specifications and implemented SQL queries to mine CRM data.
* Build back-end application with Python / Django, work with Docker, RabbitMQ, Celery, Jenkins.
* Performed data visualization and Designed dashboards with Tableau, and generated complex reports, including charts, summaries, and graphs to interpret the findings to the team and stakeholders.

**Environment:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Tableau, SQL, Software Development, Linux, Agile, Microsoft Excel, PySpark-ML, Random Forests, SVM, TensorFlow, Keras, Mulesoft, Power BI.

**Data Analyst, Zomato – Hyderabad, India June 2014 – Oct 2016**

* Created Complex ETL Packages using SSIS to extract data from staging tables to partitioned tables with incremental load.
* Collaborated with database engineers to implement ETL process, wrote and optimized SQL queries to perform data extraction and merging from SQL server database.
* Gathered, analyzed, and translated business requirements, communicated with other departments to collected client business requirements and access available data.
* Migrating data from Legacy system to SQL Server using SQL Server Integration Services 2012.
* Used C# scripts to map records.
* Involved in writing complex SQL Queries, Stored Procedures, Triggers, Views, Cursors, Joins, Constraints, DDL, DML and User Defined Functions to implement the business logic and created clustered and non-clustered indexes.
* Created and modified Stored Procedures, Functions, and Indexes.
* Developed SQL Scripts to Insert/Update and Delete data in MS SQL database tables.
* Created various ad-hoc SQL queries for customer reports, executive management reports and types of report types like tables, matrix, sub reports etc.
* Generated Sub-Reports, drill down reports, Drill through reports and Parameterized reports using 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.
* Tableau administration and development roles and helping the client with the issues of data connections, extraction and maintenance.
* Performed data mapping, logical data modelling, created class diagrams and ER diagrams and used SQL queries to filter data.
* Extensively used Inner Join, Outer join while creating SQL queries from multiple tables.
* Worked on data verifications and validations to evaluate the data generated according to the requirements is appropriate and consistent.
* Extracted data from existing data source and performed ad-hoc queries by using SQL and UNIX.
* Developed OLAP functions like sum, count, sum etc.
* Utilization of Power BI functions and Pivot Tables to further analyze in given complex data.
* Created and Monitored Batches and Sessions using Tableau Power Center Server.
* Used SSIS and T-SQL stored procedures to transfer data from OLTP databases to staging area and finally transfer into data warehouse.
* Performed data conversions from SQL server to flat file using SSIS.
* Developed reports using global variables, expressions and functions using SSRS.
* Created Complex ETL Packages using SSIS to extract data from staging tables to partitioned tables with incremental load.
* Designed high level ETL architecture for overall data transfer from the source server to the Enterprise Services Warehouse which encompasses server name, database name, accounts, tables and direction of data flow, data dictionary and Metadata using SSIS.
* Used various SSIS data flow transformations, such as Data Conversion, Conditional Split, Derived Column, and Lookup which were used for Data Scrubbing and Data Validation Checks during staging, before loading the data into Data Warehouse.

 **Environment:** SQL Server, ETL, SSIS, SSRS, Tableau, R, SAS, DDL, DML, SDLC, QlikView.

**SQL Developer, SBI – Hyderabad, India May 2013 – May 2014**

* Collaborated with database engineers to implement ETL process, wrote and optimized SQL queries to perform data extraction and merging from SQL server database.
* Created Complex ETL Packages using SSIS to extract data from staging tables to partitioned tables with incremental load.
* Gathered, analyzed, and translated business requirements, communicated with other departments to collected client business requirements and access available data.
* Migrating data from Legacy system to SQL Server using SQL Server Integration Services 2012.
* Used C# scripts to map records.
* Involved in writing complex SQL Queries, Stored Procedures, Triggers, Views, Cursors, Joins, Constraints, DDL, DML and User Defined Functions to implement the business logic and created clustered and non-clustered indexes.
* Created and modified Stored Procedures, Functions, and Indexes.
* Developed SQL Scripts to Insert/Update and Delete data in MS SQL database tables.
* Created various ad-hoc SQL queries for customer reports, executive management reports and types of report types like tables, matrix, sub reports etc.
* Generated Sub-Reports, Drill down reports, Drill through reports and Parameterized reports using 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.
* POWER BI administration and development roles and helping the client with the issues of data connections, extraction and maintenance.
* Excellent verbal communication skills, Problem-solving skills and Build Partnerships. Proven ability to learn and rapidly adapt to new challenges and followed best practices and company standards.

 **Environment:** SQL Server, ETL, SSIS, SSRS, Tableau, Linux, DDL, DML, SDLC, QlikView.