**Srikanth**

**Sr. Data Engineer,**

**Phone: +1**

**Email:** [**srikanthg9457@gmail.com**](mailto:srikanthg9457@gmail.com)‬‬‬‬‬‬‬‬‬‬‬

**Professional Summary:**

* **Over 10+ Years of experience as Data Engineer and Data Modeler with excellent knowledge on BI, Data warehouse, ETL, Cloud and Big - Data technologies.**
* **In - depth knowledge of Hadoop architecture and its components like YARN , HDFS, Name Node, Data Node, Job Tracker, Application Master, Resource Manager , Task Tracker and Map Reduce programming paradigm.**
* **Extensive experience in Hadoop development of enterprise level solutions utilizing Hadoop components such as Apache Spark, MapReduce, HDFS, Sqoop, PIG, Hive, HBase, Oozie, Flume, NiFi, Kafka, Zookeeper, and YARN.**
* **Profound experience in performing Data Ingestion, Data Processing (Transformations, enrichment, and aggregations). Strong Knowledge on Architecture of Distributed systems and parallel processing, In-depth understanding of MapReduce programming paradigm and Spark execution framework.**
* **Big Data - Hadoop (MapReduce & Hive), Spark (SQL, Streaming), and SQL Data warehouse, AWS Redshift, Athena, Lambda, Step Function and SQL.**
* **Experience in Developing Spark applications using Spark - SQL in Data bricks for data extraction, transformation and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.**
* **Experienced with the Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark Context , Spark-SQL , Dataframe API , Spark Streaming, MLlib , Pair RDD 's and worked explicitly on PySpark and Scala.**
* **Experienced working on Big-Data technologies like Hive, Sqoop, HDFS and Spark streaming Kafka, and No-SQL Databases like MongoDB, Cassandra, Good exposure on cloud technologies like AWS S3, AWS EC2, AWS Redshift.**
* **Handled ingestion of data from different data sources into HDFS using Sqoop, Flume and perform transformations using Hive, Map Reduce and then loading data into HDFS. Managed Sqoop jobs with incremental load to populate HIVE external tables. Experience in importing streaming data into HDFS using Flume sources, and Flume sinks and transforming the data using Flume interceptors.**
* **Experience in Oozie and workflow scheduler to manage Hadoop jobs by Direct Acyclic Graph (DAG) of actions with control flows.**
* **Implemented the security requirements for Hadoop and integrating with Kerberos authentication infrastructure- KDC server setup, creating realm /domain, managing.**
* **Highly proficient in Data Modeling retaining concepts of RDBMS, Logical and Physical Data Modeling until 3NormalForm (3NF) and Multidimensional Data Modeling Schema (Star schema, Snow-Flake Modeling, Facts and dimensions).**
* **Expertise in the Data Analysis, Design, Development, Implementation and Testing using Data Conversions, Extraction, Transformation and Loading (ETL) and SQL Server, ORACLE and other relational and non-relational databases and experience in developing ETL applications on large volumes of data using different tools: MapReduce, Spark-Scala, PySpark, Spark-Sql, and Pig.**
* **Experience with commercial BI tools, such as Tableau, Cognos, as well as open source and experienced working with various data sources such as Oracle, SQL Server, DB2, and Teradata & Netezza.**
* **Demonstrable architecture experience, specifically within both Standard CRM packages & ideally within the financial services industry.**
* **Experience with various ETL engines, such as Oracle OWB, and Microsoft (SQL Server Integration Services), as well as strong experience writing the complete ETL process with PL/SQL, along with scripting languages (Bash/Python/Perl)**
* **Hands-on experience in handling database issues and connections with SQL and NoSQL databases such as MongoDB, HBase, Cassandra, SQL server, and PostgreSQL. Created Java apps to handle data in MongoDB and HBase. Used Phoenix to create SQL layer on HBase.**
* **Experience in designing and creating RDBMS Tables, Views, User Created Data Types, Indexes, Stored Procedures, Cursors, Triggers and Transactions.**
* **Expert in designing ETL data flows using creating mappings/workflows to extract data from SQL Server and Data Migration and Transformation from Oracle/Access/Excel Sheets using SQL Server SSIS.**
* **Expert in designing Parallel jobs using various stages like Join, Merge, Lookup, remove duplicates, Filter, Dataset, Lookup file set, Complex flat file, Modify, Aggregator, XML.**
* **Experienced in Consolidating and auditing Metadata from disparate tools and sources, including business intelligence (BI), extract, transform, and load (ETL), relational databases, modeling tools, and third-party metadata into a single repository.**
* **Experienced in using distributed computing architectures such as AWS products (e.g. S3, EC2, Redshift, and EMR), Hadoop and effective use of Map-Reduce, SQL and Cassandra to solve Big Data type problems.**
* **Expertise in writing SQL Queries, Dynamic-queries, sub-queries and complex joins for generating Complex Stored Procedures, Triggers, User-defined Functions, Views and Cursors and extensive experience in advanced SQL Queries and PL/SQL stored procedures.**
* **Experienced in deploying and scheduling Reports using SSRS to generate all daily, weekly, monthly and quarterly Reports including current status and experienced in designing and deploying reports with Drill Down, Drill Through and Drop down menu option and Parameterized and Linked reports.**
* **Experienced in fact dimensional modeling ( Star schema, Snowflake schema ), transactional modeling and SCD (Slowly changing dimension)**
* **Experienced with JSON based RESTful web services, and XML/QML based SOAP web services and also worked on various applications using python integrated IDEs like Sublime Text and PyCharm.**
* **Developed web-based applications using Python, DJANGO, QT, C++, XML, CSS3, HTML5, DHTML, JavaScript and JQuery.**
* **Extensive working experience in Normalization and De-Normalization techniques for both OLTP and OLAP systems in creating Database Objects like tables, Constraints (Primary key, Foreign Key, Unique, Default), Indexes.**
* **Experienced in working in SDLC, Agile and Waterfall Methodologies.**
* **Building and productionizing predictive models on large datasets by utilizing advanced statistical modeling, machine learning, or other data mining techniques.**
* **Developed intricate algorithms based on deep-dive statistical analysis and predictive data modeling that were used to deepen relationships, strengthen longevity and personalize interactions with customers.**
* **Strong skills in analytical, presentation, communication, problem solving with the ability to work independently as well as in a team and had the ability to follow the best practices and principles defined for the team.**

**Skills:**

|  |  |
| --- | --- |
| **Big Data** | Hadoop, MapReduce, HDFS, Sqoop, PIG, Hive, HBase, Oozie, Flume, NiFi, Kafka, Zookeeper, Yarn, Apache Spark, Mahout, Sparklib |
| **Databases** | Oracle, MySQL, SQL Server, MongoDB, Cassandra, DynamoDB, PostgreSQL, Teradata, Cosmos |
| **Programming** | Python, PySpark, Scala, Java, C, C++, Shell script, Perl script, SQL |
| **Cloud Technologies** | AWS |
| **Frameworks** | Django REST framework, MVC, Hortonworks |
| **Tools** | PyCharm, Eclipse, Visual Studio, SQL\*Plus, SQL Developer, TOAD, SQL Navigator, Query Analyzer, SQL Server Management Studio, SQL Assistance, Eclipse, Postman |
| **Versioning tools** | SVN, Maven, Git, GitHub |
| **Operating System** | UNIX AIX 5.3, OS/390 z/OS 1.6, Windows 95/98/NT/ME/00/XP, UNIX, MS-DOS, Sun Solaris 5.8, Linux 8x |
| **Database Modelling** | Dimension Modeling, ER Modeling, Star Schema Modeling, Snowflake Modeling |
| **Monitoring Tool** | Apache Airflow |
| **Visualization/ Reporting** | Tableau, ggplot2, matplotlib, MS Office, SSRS and Power BI |
| **Build Tools** | Jenkins, Toad, SQL Loader, PostgreSQL, Talend, Maven, ANT, RTC, RSA, Control-M,Oozie, Hue, SOAP UI |
| **Web Technologies** | HTML, JavaScript, CSS, J2EE, JQuery |
| **Methodologies** | Agile, Waterfall |
| **Web/Application server** | Apache Tomcat, WebLogic, WebSphere Tools Eclipse, NetBeans |
| **IDE** | IntelliJ, Eclipse and NetBeans |

**Professional Experience:**

**Client: US BANK, GA Sept 2022 – TILL DATE**

**Role: Sr. Data Engineer**

**Responsibilites:**

* Extensive experience in working with GCP cloud Platform and ON perm.
* Using Spark Context, Spark-SQL, Spark MLlib, Data Frame, Pair RDD and Spark YARN.
* Expertise in writing Scala code using higher order functions for iterative algorithms in Spark for Performance considerations.
* Developed Pre-processing job using Spark Data frames to flatten JSON documents to flat file.
* Used Hive QL to analyze the partitioned and bucketed data, Executed Hive queries on Parquet tables.
* Stored in Hive to perform data analysis to meet the business specification logic.
* Written multiple MapReduce programs for data extraction, transformation and aggregation from multiple file formats including XML, JSON, CSV & other compressed file formats.
* Developed ETL pipelines in and out of data warehouse using combination of Python and Snowflakes SnowSQL.
* Developed Streaming applications using PySpark and Kafka to read messages from Azure Data Lake & write the JSON data to Azure Synapse
* Created source-to-target mappings for loading data warehouse/data marts in Erwin.
* Involved in the development of code using Spark, Hive, Python and Spark/SQL in GCP. Involved in developing optimized, high quality code deliverable, perform code reviews & unit test plan reviews, conduct independent testing and signoff the code release for testing stage to build the application conforming to project standard and requirements.
* Converted PL/SQL procedures containing CCPA data using pyspark and migrated them from SQL server to GCP.
* Used Spring Kafka API calls to process the messages smoothly on Kafka Cluster setup.
* Responsible to Design & Develop the Business components using Java.
* Automated the Pyspark jobs using Chronoton scheduler and produced the necessary data to create the tableau reports.
* Worked with Onetrust data, created workflows to get the data from Onetrust portal, process the data and load it in the final tables which is used to create Tableau reports.
* Created new Automic jobs to automate the migrated scripts in Prod environment, also migrated the existing Automic Jobs to point to the new servers.
* Used Spark and Spark-SQL to read the parquet data and create the tables in hive using the Scala API.
* Integrated Maven build and designed workflows to automate the build and deploy process.
* Involved in migrating the On-prem data to GCP buckets using UDP.
* Optimized Hive QL/ pig scripts by using execution engine like Tez, Spark.
* Implemented Spark using Scala, Python and utilizing Data frames and Spark SQL API for faster processing of data.
* Creation of Java classes and interfaces to implement the system.
* Experience in automating to create infrastructure for Kafka clusters different instances as per components in the cluster using Terraform for creating multiple Azure instances.
* Involved in designing the plan to convert and move the code in prod-17 environment to GCP.
* Involved in the development for converting the existing code from hive to Spark using Scala and Python.
* Used various spark Transformations and Actions for cleansing the input data.
* Developed PySpark scripts to collect data from source systems and store it on HDFS.
* Wrote Ad-hoc queries for moving data from HDFS to Hive and analyzed the data using HIVE QL.
* Involved in creating tables in Snowflake and used SQL extensively to investigate data anomalies and support the functionalities of the ETL process. Involved in creating dashboards to see the progress on Basel monthly jobs and access the log files from the dashboard.
* Involved in building the ETL design and source to target mapping to load data into data warehouse.
* Designing and executing test scripts and test scenarios, reconciling data between multiple data sources and systems.
* Effectively understood session error logs and used debugger to test mapping and fixed bugs.
* Worked as communicator with business analysts and implementation teams regarding integration activities.

**Client: Neiman Marcus, TX June 2020 – Sept 2022**

**Role: Sr. Data Engineer**

**Description.** Neiman Marcus is the company’s signature retail business. I was assigned to work on Data Analytics Platform application which allows to access trusted data, get insights and recommendations to make informed business decisions .The core intent is to provide data consumers within Neiman Marcus a consistent, reliable and secure set of data regardless of the technology platform that is optimal of their use cases. Engineering team build processing pipelines to consolidate data from various source systems like store registers and E-commerce order databases to create centrally managed well understood data asset providing associates fast and convenient access to tools and applications that allow them to discover, visualize and Analyze data necessary for making evidence base decisions and also worked on a migration project to load data from various sources like Teradata, SQL Server, Hadoop to AWS via UDP framework to S3 and finally from S3 to Snowflake

**Responsibilities:**

* Extensive experience in working with AWS cloud Platform (EC2, S3, EMR, Redshift, Lambda and Glue).
* Working knowledge of Spark RDD, Data Frame API, Data set API, Data Source API, Spark SQL and Spark Streaming.
* Developed Spark Applications by using Python and Implemented Apache Spark data processing Project to handle data from various RDBMS and Streaming sources.
* Worked with the Spark for improving performance and optimization of the existing algorithms in Hadoop.
* Using Spark Context, Spark-SQL, Spark MLlib, Data Frame, Pair RDD and Spark YARN.
* Used Spark Streaming APIs to perform transformations and actions on the fly for building common.
* Learner data model which gets the data from Kafka in real time and persist it to Cassandra.
* Developed Kafka consumer API in python for consuming data from Kafka topics.
* Consumed Extensible Mark-up Language (XML) messages using Kafka and processed the xml file using Spark Streaming to capture User Interface (UI) updates.
* Developed Pre-processing job using Spark Data frames to flatten JSON documents to flat file.
* Load D-Stream data into Spark RDD and do in memory data Computation to generate output response.
* Experienced in writing live Real-time Processing and core jobs using Spark Streaming with Kafka as a Data pipe-line system.
* Migrated an existing on-premises application to AWS. Used AWS services like EC2 and S3 for data sets processing and storage.
* Experienced in Maintaining the Hadoop cluster on AWS EMR.
* Loaded data into S3 buckets using AWS Glue and PySpark. Involved in filtering data stored in S3 buckets using Elastic search and loaded data into Hive external tables.
* Configured Snow pipe to pull the data from S3 buckets into Snowflakes table.
* Bulk loading from the external stage (AWS S3), internal stage to snowflake cloud using the COPY command.
* Loading data into snowflake tables from the internal stage using SnowSQL.
* Used COPY, LIST, PUT and GET commands for validating the internal stage files.
* Used import and Export from the internal stage (snowflake) from the external stage (AWS 53).
* Writing complex snows scripts in snowflake cloud data warehouse to business analysis and reporting.
* Used FLATTEN table function to produce a lateral view of VARIANT, OBJECT, and ARRAY column.
* Used SNOW PIPE for continuous data ingestion from the S3 bucket.
* Performed data quality issue analysis using Snow SQL by building analytical warehouses on Snowflake
* Experience with AWS cloud services: EC2, S3, EMR, RDS, Athena.
* Stored incoming data in the Snowflakes staging area.
* Created numerous ODI interfaces and load into Snowflake DB. worked on Amazon Redshift for shifting all Data warehouses into one Data warehouse
* Good understanding of Cassandra architecture, replication strategy, gossip, snitches etc.
* Designed columnar families in Cassandra and Ingested data from RDBMS, performed data transformations, and then exported the transformed data to Cassandra as per the business requirement.
* Used the Spark Data Cassandra Connector to load data to and from Cassandra.
* Worked from Scratch in Configurations' of Kafka such as Managers and Brokers.
* Experienced in creating data-models for Clients transactional logs, analyzed the data from Cassandra.
* Tables for quick searching, sorting and grouping using the Cassandra Query Language.
* Tested the cluster Performance using Cassandra-stress tool to measure and improve the Read/Writes.
* Used Hive QL to analyze the partitioned and bucketed data, Executed Hive queries on Parquet tables.
* Stored in Hive to perform data analysis to meet the business specification logic.
* Used Apache Kafka to aggregate web log data from multiple servers and make them available in downstream systems for Data analysis and engineering type of roles.
* Worked in Implementing Kafka Security and boosting its performance.
* Experience in using Avro, Parquet, RCFile and JSON file formats, developed UDF in Hive.
* Developed Custom UDF in Python and used UDFs for sorting and preparing the data.
* Worked on Custom Loaders and Storage Classes in PIG to work on several data formats like JSON, XML, CSV and generated Bags for processing using pig etc.
* Developed Sqoop and Kafka Jobs to load data from RDBMS, External Systems into HDFS and HIVE.
* Developed Oozie coordinators to schedule Hive scripts to create Data pipelines.
* Written several Map reduce Jobs using PySpark, Numpy and used Jenkins for Continuous integration.
* Setting up and worked on Kerberos authentication principals to establish secure network communication.
* On cluster and testing of HDFS, Hive, Pig and MapReduce to access cluster for new users.
* Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.

**Environment::** Spark, Spark-Streaming, Spark SQL, AWS EMR, mapR, HDFS, Hive, Pig, Apache Kafka, Sqoop, Python, PySpark, Shell scripting, Linux, MySQL Oracle Enterprise DB, SOLR, Jenkins, Eclipse, Oracle, Git, Oozie, Tableau, MySQL, Soap, Cassandra and Agile Methodologies.

**Client: Capital One, Richmond, VA June 2019 – May 2020**

**Role: Sr Data Engineer**

**Description:** Capital One Financial Corporation is a bank holding company specializing in credit cards, auto loans, banking and savings products. The main purpose of project Floga is to encrypt sensitive data related to Auto Finance applications from various source systems such as MySQL, Teradata, and PostgreSQL and migrate to AWS s3 buckets and load them into Snowflake cloud EDW.

**Responsibilities:**

* This allowed for a more reliable and faster reporting interface, giving sub - second query response for basic queries.
* Analyzing, designing, and developing ETL strategies and processes, writing ETL specifications, Informatica development, and administration.
* Expert in implementing advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala.
* Expertized in implementing Spark using Scala and Spark SQL for faster testing and processing of data responsible to manage data from different sources.
* Implemented data ingestion and handling clusters in real time processing using Kafka. Customized Tableau Dashboards, integrating Custom SQL from Teradata and Confidential, Hadoop and performing data blending in reports.
* Implemented scripts that load Google Big Query data and run queries to export data.
* Analyzed clickstream data from Google analytics with Big Query.
* Developed Spark Programs using Scala and Java API's and performed transformations and actions on RDD’s.
* Implemented a 'server less' architecture using API Gateway, Lambda, and Dynamo DB and deployed AWS Lambda code from Confidential S3 buckets. Created a Lambda Deployment function, and configured it to receive events from your S3 bucket.
* Designed the data models to be used in data intensive AWS Lambda applications which are aimed to do complex analysis creating analytical reports for end-to-end traceability, lineage, and definition of Key Business elements from Aurora.
* Worked and learned a great deal from AWS Cloud services like EC2, S3, EBS, RDS.
* Migrated an existing on-premises application to AWS. Used AWS services like EC2 and S3 for small data sets processing and storage, Experienced in Maintaining the Hadoop cluster on AWS EMR.
* Imported data from AWS S3 into Spark RDD, Performed transformations and actions on RDD's.
* Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS
* Experienced with AWS AZURE services to smoothly manage application in the cloud and creating or modifying the instances.
* Created data pipeline for different events of ingestion, aggregation and load consumer response data in AWS S3 bucket into Hive external tables in HDFS location to serve as feed for tableau dashboards.
* Used EMR (Elastic Map Reducing) to perform big data operations in AWS.
* Creating customized Tableau Dashboards, integrating Custom SQL from Teradata and Confidential, Hadoop and performing data blending in reports.
* Designed, developed, tested, and maintained Tableau reports and dashboards based on user requirements.
* Deploying Spark cluster and running PySpark jobs using Dataproc and scheduling batch jobs through Cloud.
* Collaborated with Architects to design Spark model for the existing MapReduce model and migrated them to Spark models using Scala. Worked on writing Scala Programs using Spark-SQL in performing aggregations.
* Developed Web Services in play framework using Scala in building stream data Platform.
* Worked with Apache Spark which provides fast engine for large data processing integrated with Scala.
* Experience working with SparkSQL and creating RDD's using PySpark. Extensive experience working with ETL of large datasets using PySpark in Spark on HDFS.
* Developed ETL programs to load data from Confidential to Snowflake using Informatica snowflake.
* Developing ETL pipelines in and out of data warehouse using combination of Python and Snowflakes SnowSQL Designed physical and logical data models based on Relational (OLTP), Dimensional on snowflake schema using Erwin modeler to build an integrated enterprise data warehouse.
* Experienced in creating data pipeline integrating Kafka with spark streaming application. Created Kafka broker for structured streaming to get structured data by schema.
* Developed Elastic Search Connector using Kafka Connect API with source as Kafka and sink as elastic search.

**Environment:** Apache Spark, Spark SQL, PySpark, Scala, Kafka, Teradata, SQL, AWS, AWS Lambda, AWS S3, Tableau, Big Query, HDFS, Snowflake, Elastic Search

**Client: Verizon, Tampa, FL June 2017 – May 2019**

**Big Data Engineer**

**Description:** The Fiber Store is a single repository of VZ fiber enabled locations, with existing & prospect customer information and the services available to generate intelligent leads that is controlled, financially certified and federated, to increase penetration in the buildings and monetize the fiber assets.

**Responsibilities:**

* Gathered the business requirements from the Business Partners and Subject Matter Experts
* Analyzed large and critical datasets using Cloudera, HDFS, MapReduce, Hive, Hive UDF, Pig, Sqoop and Spark.
* Developed Spark Applications by using Scala and Implemented Apache Spark data processing project to handle data from various RDBMS and Streaming sources.
* Worked with the Spark for improving performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Spark MLlib, Data Frame, Pair RDD's, Spark YARN.
* Used Spark Streaming APIs to perform transformations and actions on the fly for building common learner data model which gets the data from Kafka in near real time and persist it to Cassandra.
* Consumed XML messages using Kafka and processed the xml file using Spark Streaming to capture UI updates. Developed Dashboard reports on Tableau.
* Developed Pre-processing job using Spark Data frames to flatten JSON documents to flat file.
* Experienced in writing live Real-time Processing and core jobs using Spark Streaming with Kafka as a data pipe-line system.
* Worked and learned a great deal from AWS Cloud services like EC2, S3, EBS, RDS.
* Migrated an existing on-premises application to AWS. Used AWS services like EC2 and S3 for small data sets processing and storage, Experienced in Maintaining the Hadoop cluster on AWS EMR.
* Imported data from AWS S3 into Spark RDD, Performed transformations and actions on RDD's.
* Ingested data from RDBMS and performed data transformations, and then export the transformed data to Cassandra as per the business requirement.
* Written multiple MapReduce programs for data extraction, transformation and aggregation from multiple file formats including XML, JSON, CSV & other compressed file formats.
* Developed automated processes for flattening the upstream data from Cassandra which in JSON format. Used Hive UDFs to flatten the JSON Data.
* Optimized MapReduce Jobs to use HDFS efficiently by using various compression mechanisms
* Developed PIG UDFs to provide Pig capabilities for manipulating the data according to Business Requirements and worked on developing custom PIG Loaders and Implemented various requirements using Pig scripts.
* Experienced on loading and transforming of large sets of structured, semi structured and unstructured data
* Created POC using Spark Sql and MLlib libraries.
* Developed a Spark Streaming module for consumption of Avro messages from Kafka.
* Experienced in Querying data using SparkSQL on top of Spark Engine, implementing Spark RDD’s in Scala.
* Expertise in writing Scala code using higher order functions for iterative algorithms in Spark for Performance considerations.
* Experienced in managing and reviewing Hadoop log files
* Worked with different File Formats like TEXTFILE, AVROFILE, ORC, and PARQUET for HIVE querying and processing.
* Create and Maintain Teradata Tables, Views, Macros, Triggers and Stored Procedures.
* Expertise in snowflake to create and Maintain Tables and views.
* Monitored workload, job performance and capacity planning using Cloudera Distribution.
* Worked on Data loading into Hive for Data Ingestion history and Data content summary.
* Created Impala tables and SFTP scripts and Shell scripts to import data into Hadoop.
* Created Hive tables and involved in data loading and writing Hive UDFs. Developed Hive UDFs for rating aggregation
* Provided ad-hoc queries and data metrics to the Business Users using Hive, Pig.
* Did various performance optimizations like using distributed cache for small datasets, partition and bucketing in hive, doing map side joins etc.
* Worked on importing and exporting data from snowflake, Oracle and DB2 into HDFS and HIVE using Sqoop for analysis, visualization and to generate reports.
* Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS
* Experienced with AWS AZURE services to smoothly manage application in the cloud and creating or modifying the instances.
* Created data pipeline for different events of ingestion, aggregation and load consumer response data in AWS S3 bucket into Hive external tables in HDFS location to serve as feed for tableau dashboards.
* Used EMR (Elastic Map Reducing) to perform big data operations in AWS.
* Worked on Apache spark writing python applications to convert txt, xls files and parse.
* Developed Python scripts, UDF's using both Data frames/SQL and RDD/MapReduce in Spark for Data Aggregation, queries and writing data back into RDBMS through Sqoop.
* Installed the application on AWS EC2 instances and configured the storage on S3 buckets.
* Loading data from different source (database & files) into Hive using Talend tool.
* Implemented Spark using Python/Scala and utilizing Spark Core, Spark Streaming and Spark SQL for faster processing of data.
* Scheduled Oozie workflow engine to run multiple Hive and Pig jobs, which independently run with time and data availability
* Worked on custom Pig Loaders and Storage classes to work with a variety of data formats such as JSON, Compressed CSV etc.
* Involved in running Hadoop Streaming jobs to process Terabytes of data
* Used JIRA for bug tracking and CVS for version control.

**Environment:** Hadoop, Map Reduce, Hive, HDFS, PIG, Sqoop, Hortonworks, Flume, HBase, Oracle, Snowflake, Teradata, Tableau Unix/Linux, Hadoop, Hive, PIG, SQOOP, Flume, HDFS, Oracle/SQL & DB2, Unix/Linux, JIRA, AWS

**Client: Insert V Consultancy Services PVT LTD, Hyderabad, IND Sep 2011 – Nov 2015**

**Role: SQL Server Developer**

**Project 1: First American Title Insurance Company**

**Responsibilities:**

* Translated business needs into data analysis, business intelligence data sources and reporting solutions for different types of information consumers.
* Involved in Planning, Defining and Designing data base using Erwin on business requirement and provided documentation.
* Developed complex T-SQL code such as Stored Procedures, functions, triggers, Indexes and views for the application.
* Used Joins, correlated and non-correlated sub-queries for complex business queries involving multiple tables & calculations from different databases.
* Created many complex Stored Procedures/Functions and used them in Reports directly to generate reports on the fly.
* Created SSIS package to load data from XML Files and SQL Server 2005 to SQL Server 2008 by using Lookup, Fuzzy Lookup, Derived Columns, Condition Split, Term Extraction, Aggregate, Pivot Transformation, and Slowly Changing Dimension.
* Converted Crystal Reports to SQL Server Reports.
* Used SSIS to create ETL packages to Validate, Extract, Transform and Load data to Data Warehouse and Data Mart Databases.
* Developed Unit Test Plans for SSIS packages for checking the functionality of each component.
* Modified the existing SSIS packages to meet the changes specified by the users.
* Scheduled Jobs for executing the stored SSIS packages which were developed to update the database on Daily basis using SQL Server Agent.

**Project 2: Charter Communications**

**Responsibilities:**

* Gathered user requirements to develop the business logic and analyzed and designed database solutions.
* Generated Performance reports on SSIS packages, Stored Procedures and Triggers.
* Created OLAP cubes using SSAS to analyze the business condition of the sales.
* Wrote complex queries for the drill down reports and used conditional formatting using SSRS.
* Used SSIS to transfer data from various sources like SQL Server, MS Access to be used for generating reports using SSRS 2008.
* Formatted reports for various Output formats as requested by the management using SSRS.
* Actively supported business users for changes in reports as and when required using SSRS.
* Analyzed reports and fixed bugs in stored procedures using SSRS.
* Created linked reports and managed snapshots using SSRS.
* Performed various calculations using complex expressions in the reports and created report models.
* Designed and optimized indexes, views, stored procedures and functions using T-SQL.
* Performed maintenance duties like performance tuning and optimization of queries, functions and stored procedures.
* Designed high level SSIS architecture for overall data transfer from the source server to the Enterprise Services Warehouse.

**Environment:** SQL, T-SQL, SQL Server, SSIS, SSAS, SSRS, Crystal Reports, Unit Test, Erwin.

**Education:**

* **Master’s in Information Security and Intelligence, US, May 2017**
* **Bachelor’s in Electrical and Electronics Engineering, JNTU, India, 2011**