

Tech Lead Tech Lead Tech Lead - Elevate Credit Inc Work Experience Tech Lead Elevate Credit Inc - Addison, TX November 2018 to Present Environment: Hive, StreamSets, Cloudera, Kafka, SQL Server, Jython The main objective of this project was to design and develop a rule engine to execute business rules to detect and flag fraudulent application for loan or line of credit. The implementation of this project involved developing a near real time data pipeline using StreamSets and rules execution engine in SQL servers. I was responsible for data modeling, design and also the implementation of the data pipelines. This project is still in progress and finer details are still being worked on.

Project: Credit Application Underwriting Re-Design Using Cloud Technologies Senior Hadoop Developer Elevate Credit Inc - Dallas, TX September 2017 to November 2018 Environment: Hive, StreamSets, Cloudera, AWS S3, AWS EMR, Spark, JAVA, Kafka, R The main objective of this project was to design and develop next generation underwriting process which is scalable and future proof leveraging cloud technologies. From big data perspective my role was to lead the development effort for design & development of a near real time data pipeline to process a credit application, create base data attributes, run it through R data models developed by data scientists and return the credit score back to the underwriting services in a very strict SLA. The implementation of this project involved developing a Spark application in JAVA. Kafka message queues were used for integration between various services. The data format used for the files/requests were XML, JSON. Once scored the application data, relevant attributes were stored in AWS S3 buckets. This project also involved developing various housekeeping support data pipelines to download data from the AWS EMR cluster to HADOOP on-premise cluster in batch mode.

RESPONSIBILITIES Requirement analysis and documentation Collaborate with design teams in design and architectural meetings & discussions Design & development Spark application to consume credit data with highlights:

- ? Custom XML parser
- ? REST API calls to get more data and join with the credit data
- ? Parse JSON data request and create JSON reply
- ? Multiple DStreams to handle Kafka Queues for parallelization
- ? R integration with Spark
- ? AWS S3 data storage for persistence and auditing

Develop data pipeline to download data from AWS S3 to on-premise HADOOP cluster Senior Hadoop Developer Elevate Credit Inc - Addison, TX March

2017 to August 2017 Environment: Hive, StreamSets, SQOOP, Cloudera, AWS S3, AWS RedShift  
The objective for this project was to work in collaboration with the Marketing team and external vendors to build a real time & batch data pipeline to ingest marketing data captured from our product websites. The main challenge was to come up with an efficient solution to capture and download data from AWS RedShift & S3 location to on premise HADOOP clusters. RESPONSIBILITIES

Data profiling, data schema design and document      Design & develop bash Linux script using psql cli commands & HDFS/ beeline commands to pull data from AWS S3 buckets to on premise HADOOP cluster.      Schedule/ Monitor batch scripts using JAMs      Develop JAVA application to download data from Google BigTable. Senior Hadoop Developer Elevate Credit Inc - Addison, TX

July 2016 to March 2017 Environment: Hive, Nifi, SQOOP, Flume, Cloudera, Spark, PIG, StreamSets, Informatica      The objective for this project was to build a data lake to store data from third party sources i.e. Credit Bureaus & internal transactional data including marketing, collections, sales, finance data for different products. This project involved profiling and data analysis of various data sources. Develop, document, test, debug data schemas, ingestion process, data streams. It involved a lot of cross functional communication which was one of the challenges apart from various technical challenges. This data lake hosted data to various downstream process including Marketing, Reporting, Data Science modeling etc. RESPONSIBILITIES      POC's including flume,

Nifi, StreamSets, Spark, Hive, AWS S3, PIG, Kafka      Collaborate with ETL developers to develop pipelines to deliver data to various Data Marts      Work closely with Data Warehouse team to develop data delivery platform.      Work with BI team to identify the reporting data points and expose them in flat tables & views for dashboards.      Extend custom Hive SerDe to parse XML      Extend/ Debug JAVA application to develop Hive DDL from a given XML, JSON files.      Mentor & support offshore

team members, contractors      Onboarding training new resources      Training business partners in using HADOOP data tools for data exploration Senior Software Engineer STMicroelectronics, Inc - Dallas, TX January 2014 to June 2016 Environment: Hive, Pig, Linux, Sqoop, Flume, Horton Works

2.3.2      The objective was to build a digital media services data collection & analytical framework which enables the client to analyze the subscriber's usage trend and create reports and dashboards.

The solution collects various events/logs from the digital media gateway & customer premise equipment (CPE) devices and after processing them it generates load ready files. These load ready files are then sent to enterprise data warehouse for data mining (billing and reporting). Each and every activity on the Set Top Box generates an Event. Most of this data is used to capture the TV habits of the Clients. Some of this data is also used for the auditing purposes. The solution processes more than 200 Million events per day and generates load ready files which are in gigabytes. The solution needed to be robust to make sure that data generated by IPTV Platform is accurate and events are processed without any loss of data.

**RESPONSIBILITIES**

Created SQOOP (version 1.4.4) incremental job to import operational data from databases, CRM system with incremental load to Hive external tables. Developed FLUME (version 1.5.2) with Fan in architecture using AVRO source and sink with multiple collector tiers-based ingestion scheme for ingesting near Real Time STB box tuning data and system logs. The RAW data stored in PARTITION by day. The FLUME pipeline was implemented using file type channels for reliability and with failover functionality using load balancing sink group processor. Interceptors were used to do some filtering of data anomalies. Transferred data using SQOOP from Teradata to be used in HDFS and in hive. Used Teradata Connector to Hadoop (TDCH 1.3) to increase performance to sqoop large datasets from Teradata to Hadoop. Did POC on TDCH (1.4) to get the data directly in ORC format in hive tables. On boarded new feeds for collection by interfacing with respective teams, analyze and agree on a log format to be collected and distributed. Providing the support for Hadoop Jobs deployed on production. Ensure no failures and smooth run of each Job on Daily execution. Prepared, evaluated and optimized the workflow and efficiency of Hadoop Production Jobs which leads to minimizing the manual intervention when failure occurs.

Senior Software Engineer STMicroelectronics, Inc - Dallas, TX January 2013 to December 2014 Environment: Hive, Pig, Linux, Sqoop, Flume, Oozie, DATAMEER, Wireshark

Developed an end-to-end network analysis solution for efficient & stable operations of MSO networks. The main objective was to create a framework to monitor the status and performance of a network proactively to diagnose and resolve connectivity or service issues. To detect network security threats to block unauthorized

access. To achieve the objective, the network data was captured, ingested, cleansed and stored for further analysis and run various throughput tests for TCP/IP protocols.    RESPONSIBILITIES: Developed flume & Sqoop based ingestion scheme for efficiently moving large amounts of network sniffer captures recorded in libpcap format to HDFS.    Evaluation of various custom SerDe for most efficient query performance and writes.    Hive schema development for storing the TCP/IP packets for optimal query.    Development of Various PIG scripts & Hive queries for per protocol analysis. Various UDF's developed for manipulating the IP header data.    Basic visualization done using Datameer    Worked on various POC's using AWS EC2, EMR & S3. Software Engineer STMicroelectronics, Inc - Dallas, TX January 2007 to December 2012 Environment: Java, C, Verilog, Tcl/Tk, MIB Browser, RF Sniffer, Spirent packet Generator, Cisco CMTS This project was done for the Cable Division designing Cable Modem, Set-top box and gateways solutions. The main objective was to create in house test, measurement and performance metrics collection capability for next generation DOCSIS devices.    RESPONSIBILITIES:    Developed automated validation procedures conforming to standards developed by Cable Labs in TCL/Tk. Expanding CL-TEPs to cover all ATPs.    Developed API's for automating the handling and of RF Packet Sniffer, Packet Generator    Represented ST in various interoperability runs.    Developed expect based scripts to remotely login to various equipment for configuration.    Developed a simulation environment for simulating cable modem termination system in C.    Involved in implementation of various vendor specific SNMP MIB's & counters for performance monitoring and events logging.    Developed various GUI's for performance metrics visualization for customer demonstration.    Developed scripts to dump critical performance data for troubleshooting.    Developed OpenSSL based image signing tool for software upgrade in the field. Education Bachelor's Skills Java, Ecommerce Links <http://bcert.me/sbvladkl>

Name: Mia Washington

Email: jsnow@example.com

Phone: 976-320-9340