

Big Data Tech Lead / Architect Big Data Tech Lead / Architect Big Data Tech Lead / Architect - Hilton Worldwide     Solid experience in Big data Hadoop/Spark and Java/J2EE technologies development including requirements Analysis and Design, Development, implementation, support, maintenance and enhancements in Finance & Insurance domains.     5+ years of experience as Hadoop/Spark Developer with good knowledge of Java Map Reduce, Hive, Pig Latin, Scala and Spark.     Organizing data into tables, performing transformations, and simplifying complex queries with Hive.     Performing real-time interactive analysis on massive data sets stored in HDFS.     Strong knowledge and experience with Hadoop architecture and various components such as HDFS, YARN, Pig, Hive, Sqoop, Oozie, Flume, Spark, Kafka and Map Reduce programming paradigm.     Developed many Map/Reduce programs.     Experience in analyzing data using Spark SQL, HIVEQL, PIG Latin and experience in developing custom UDF s using Pig and Hive.     Experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems and vice-versa.     Good knowledge in using job scheduling tools like Oozie.     Experienced in using IDE Tool like Eclipse 3.x, IBM RAD 7.0     Experience in requirement gathering, analysis, planning, designing, coding and unit testing.     Strong work ethic with desire to succeed and make significant contributions to the organization.     Strong problem solving skills, good communication, interpersonal skills and a good team player.     Have the motivation to take independent responsibility as well as ability to contribute and be a productive team member.

Work Experience Big Data Tech Lead / Architect Hilton Worldwide - McLean, VA September 2018 to Present Project Description: Hilton is forward thinking global leader of hospitality. Hilton is highly data driven company which uses Hadoop Big data architecture to make key business decisions. The ingested data from different upstream applications will be transformed and loaded to Redshift which will be used by MicroStrategy reporting team to generate dashboards that aid in key business decisions.

Responsibilities     Used NIFI as dataflow automation tool to ingest data into HDFS from different source systems.     Developed common process to bulk load raw HDFS files into dataframe     Developed common process to persist dataframe into S3, Redshift, HDFS, Hive     Prune the ingested data to remove duplicates by applying window functions and perform complex

transformations to derive various metrics. Used oozie scheduler to trigger spark jobs Performed POC on airflow scheduler and Involved in migration of oozie scheduler to airflow Created UDF's in spark to be used in spark sql. Performed POC to consume requests from microstrategy dashboard which will in turn trigger the spark job The spark job dynamically generates spark sql based on the level of granularity business user requests and load data into Redshift to be used on dashboard. Involved in migration of HDP to AWS and various proof of concepts to achieve it. Guide junior developers in their day to day activities and ensure delivery of project Used Spark streaming on Kafka to achieve real time data analytics. Created Dstreams & dataframes from streaming data and performed transformations. Performed performance tuning of spark jobs using broadcast joins ,correct level of Parallelism and memory tuning. Analyze and define client's business strategy and determine system architecture requirements to achieve business goals Lead the Team to complete critical and measurable mile stones of the Project. Environment: Spark 2.2, Hadoop, Hive 2.1, HDFS , Java 1.8, Scala 2.11, HDP, Elasticsearch, AWS, Redshift, Oozie, IntelliJ, ORC, Shell Scripting, bitbucket,airflow,Python, Pyspark Lead Hadoop/Spark Consultant Citibank - Tampa, FL November 2017 to September 2018 Responsibilities: Used Spark API over Cloudera Hadoop YARN to perform analytics using Scala/Pyspark programming. Created dataframes and performed various transformations to generate recommendation strategy Created hive tables and views for business to access the recommendation strategy Used Spark streaming on Kafka to achieve real time data analytics. Created Dstreams & dataframes from streaming data and performed transformations. Extract Real time feed using Kafka and Spark Streaming and convert it to RDD and process data in the form of Data Frame. Persist the streaming data on Hbase No-sql database Wrote shell scripts to automate the jobs in UNIX. Experienced in handling large datasets using Partitions, Spark in Memory capabilities, Broadcasts in Spark, Efficient joins & Transformations. Used Spark SQL to save the data into multiple Hive tables. Developed Hive queries to process the data for visualizing. Environment: Spark 1.6, Hadoop, Hive 2.0, HDFS, Kafka, Sqoop 1.4.6, Java 1.8, Scala 2.11, CDH5.8.2, Oozie, Eclipse, Elasticsearch, Parquet, Shell Scripting, bitbucket Bank of America - Charlotte, NC September 2015 to November 2017 Project

Description: Bank of America is American multinational banking and financial services corporation and is the second largest bank of United States of America. The objective of the project is to pull data from upstream and apply smoothing; transformation & value add process to generate final basel views for federal reporting. The project is being migrated from Mainframe/Teradata to Hadoop eco system.

Responsibilities: Used Spark API to perform analytics on data in Hive using Scala programming. Optimization of existing algorithms in Hadoop using Spark Context, Data Frames, Hive context. Spark RDDs are created in Scala for all the data files which then undergo transformations. The filtered RDDs are aggregated and transformed based on the business rules and converted into data frames and saved as temporary hive tables for intermediate processing. The RDDs and data frames undergo various transformations and actions and are stored in HDFS as parquet Files to create Impala views. Used Oozie scheduler to create workflows and scheduled jobs in Hadoop Cluster. Written Hive UDFs to extract data from staging tables. Involved in creating Hive tables & views to load, transform the data. Involved in writing Pig scripts. Supported and Monitored Map Reduce Programs running on the cluster. Worked on data quality framework to generate reports to business on the quality of data processed in Hadoop. Worked on developing Web UI using J2EE for reports generated on the final Basel views.

Environment: Java 1.8, Scala 2.11, Spark 1.6, Hadoop, Pig 0.12, Hive 1.1, Map Reduce, HDFS, MySQL, Sqoop 1.4.6, CDH5.8.2, Oozie, Eclipse, Avro, Parquet, Toad, Shell Scripting, Teradata, Impala, J2EE, SAS EG

Hadoop Developer Prudential Financial, NJ May 2013 to September 2015

Project Description: Prudential is one of the largest insurance and financial services institutions in the United States of America. This project is designed to extract raw data from different sources into Hadoop Eco system to create and populate the necessary Hive tables. The main aim of the project is to centralize the source of data for report generation using historical database which otherwise are generated from multiple sources.

Responsibilities: Worked on analyzing Hadoop cluster and different big data analytic tools including Pig & Hbase NoSQL database. Importing and exporting data in HDFS and Hive using Sqoop. Designed and developed Map Reduce jobs to process data from upstream. Experience with NoSQL databases. Worked on data ingestion to bring data to HDFS and Hive.

Written Hive UDFs to extract data from staging tables. Involved in creating Hive tables, loading with data. Hands on writing Map Reduce code to make unstructured data as structured data and for inserting data into HBase from HDFS. Experience in creating integration between Hive and HBase. Used Oozie scheduler to submit workflows. Review QA test cases with the QA team.

Environment: Java 6, Eclipse, Hadoop, Pig0.12, Hive 0.13, Centos 6.4, Map Reduce, HDFS, MySQL, Sqoop 1.4.4, CDH4, Hue, Oozie, Toad, HBASE Sr. Java Developer Prudential Financial - Bengaluru, Karnataka August 2010 to May 2013 Project Description: Prudential is one of the largest insurance and financial service institutions in the United States of America. PASSPORT (Plan Admin Setup System Portal) Application is built for plan Setup. This helps to setup a plan and maintain tool that handles Vision and peripheral system set-up for defined contribution plans.

Responsibilities: Involved in software development on web-based front-end applications. Involved in development of the CSV files using the Data load. Performed unit testing of the developed modules. Involved in bug fixing, writing SQL queries & unit test cases. Used Rational Application Developer (RAD). Used Oracle as the Backend Database. Involved in configuration and deployment of front-end application on RAD. Involved in developing JSP's for graphical user interface. Implemented code for validating the input fields and displaying the error messages.

Environment: Java, JSP, Servlets, Apache Struts framework, WebSphere, RAD, Oracle, PVCS, TOAD Java Developer Danske Bank - Bengaluru, Karnataka September 2008 to August 2010 Project Description: Worked on a key development project for Danske bank to develop a new IT system to automate its Trade services like Guarantee, Letter of credit etc. Responsibilities:

Participated in the implementation of efforts like coding, unit testing. Implemented a Web based application using SERVLETS, JSP. Client side validation has been done using Java Script. Involved in Unit integration and bug fixing. Involved in acceptance testing with test cases and code reviews. Developed code for handling the exceptions using exceptional handing. Involved in writing and executing queries in MySQL. Developed the application on Eclipse. Involved in deploying application on WebSphere server. Prepared test case document and performed unit testing and system testing. Environment: Java, JSP, Servlets, Java script, WebSphere, MySQL,

Eclipse, TOAD Education Executive MBA IIM Kozhikode 2013 Bachelor of Engineering in Engineering National Institute of Technology Karnataka 2004 Skills Eclipse, Java, Jboss, jquery, Jsp, Servlets, Dynamodb, Hdfs, Impala, Oozie, Sqoop, Hbase, Kafka, Cdh, Db2, Hadoop, Avro, C++, Design patterns, Hadoop Additional Information Technical Skills: Hadoop Technologies: Hadoop, HDFS, Hadoop Map-Reduce, Hive, HBase, SQOOP, Oozie, AVRO, Pig-Latin, Hue, CDH, Parquet, Impala, Scala, Spark, Python, Kafka,AWS,S3,Trifacta, DynamoDB, EMR , Apache Nifi No Sql: HBase IDE/Tools: RAD, Eclipse Web and Application Servers: Web sphere, JBOSS, Tomcat Core Competency Technologies: Java, OOPS, design patterns, JSP, servlets, JDBC, java 5 / java 6 / java 7, C, C++, shell scripting, Spark, SAS EG, Scala, Spark Streaming, Kafka Web presentation frameworks: Java Script, HTML, AJAX, jQuery, CSS, JSON Testing & Issue Log tools: JUnit 4, Bugzilla, HP Quality Centre SCM/Version control tools: PVCS, CVS, Sub Version Modeling tools: Visio 2007 Build and continuous Integration: Maven, ANT Data base: Oracle 8i/9i/10g, DB2 & MySQL 4.x/5.x OS: UNIX, LINUX, Windows, Aix

Name: Sharon Wilson

Email: henryterrell@example.com

Phone: (701)560-4137