**CORE COMPETENCIES**

* Hadoop Development & Troubleshooting
* Data Analysis
* Data Visualization & Reporting in Tableau
* Real-time Streaming using Spark
* Spark Data Engineering
* Performance Tuning of Hive & Impala

**PROFESSIONAL SUMMARY**

**Hadoop Experience:**

* Certified Spark Developer with 7+ years of experience in IT industry, which includes 3+ years of experience in Hadoop/Big Data technologies and 4 years of extensive experience in JAVA, Python Database development and Data analytics.
* Hands on experience with Big Data Ecosystem like Hadoop (2.0 and YARN) framework technologies such as HDFS, MapReduce, Pig, Sqoop, Hive, Oozie, Impala, Zookeeper, NiFi, Knox.
* Experience in using Cloudera and Hortonworks distributions.
* Hands on experience with Oozie for job workflow scheduling.
* Experience in analyzing the data using Spark SQL, HiveQL, Pig Latin, Hbase, Hive UDF.
* Experience in using Pig as an ETL tool for transformations and pre- aggregations.
* Experience in importing & exporting data using Sqoop from HDFS to RDBMS and vice-versa.
* Experience with different data source files like Avro, Parquet, RC, and ORC formats and compressions like snappy, bzip.
* Hands on experience with NoSQL databases like HBase and relational databases like Oracle and MySQL.

**Spark Experience:**

* Hands on experience with Spark using Scala and Python.
* Hands on experience in Spark architecture and its integrations like Spark SQL, Data Frames and Datasets API.
* Experience in working with Tuning Spark, Level of Parallelism and Spark UDFs.
* Made POC on Spark Real Time Streaming using Kafka to subscribe the topic from multiple web servers and consume the data from broker and stored the data into HDFS.
* Hands on experience with Amazon Web Services(AWS) cloud services like EC2, S3, EBS, RDS and VPC.
* Hands on experience spinning up different AWS instances including EC2-classic and EC2-VPC using cloud formation templates.

**UNIX/JAVA/SDLC Experience:**

* Hands on experience in application development using Java, RDMS and UNIX shell scripting.
* Experience in Object Oriented Analysis Design (OOAD) and development.
* Experience in Java, JSP, Servlets, Web Logic, Web Sphere, Hibernate, Spring, JBoss, JDBC, RMI, Java Script, Ajax, JQuery, XML, HTML and have the strong understanding of Data Structures & Algorithms.
* Strong Experience in Unit Testing and System testing in Big Data.
* Hands on experience with version control software tools like SVN, Bit Bucket and Gitlab.
* Expertise in using Linux OS including flavors like CentOS, Ubuntu and Linux Mint.
* Experience in SDLC models like Agile SCRUM, Waterfall model under the guidelines of CMMI.

**CERTIFICATIONS**

* Developer Certification for Apache Spark – 2017(1.x-0834).
* Big Data University-Cognitive Class Certificate of Apache Spark.
* Big Data University-Cognitive Class Certificate of Hadoop.

**ACADEMICS**

* Masters in Computer and Information Sciences from Harrisburg University of Science and Technology, Harrisburg, PA.
* Bachelor of Engineering in Computer Science from Jawaharlal Nehru Technological University (JNTU), India.

**TECHNICAL SKILLS**

**Big Data Technologies** HDFS, MapReduce, YARN, Hive, Pig, Sqoop, Impala, Oozie,

Flume, Zookeeper, Kafka, Nifi, Zeppelin and HBase.

**Spark components** Spark, Spark SQL, Spark Streaming, Python.

**AWS Cloud Services** S3, EBS, EC2, VPC, Redshift, EMR

**Programming Languages** Java, Python, Scala.

**Databases** Oracle, MySQL, SQL Server, MySQL work bench, phpMyAdmin

**Scripting and Query Languages** Unix Shell scripting, SQL and PL/SQL

**Web Technologies** JSP, Servlets, JavaBeans, JDBC, XML, CSS, HTML, JavaScript, AJAX

**Operating Systems** Windows, UNIX, Linux distributions, Mac OS

**Other Tools** Maven, Eclipse, Tableau, GitHub, Jenkins

**PROFESSIONAL EXPERIENCE**

**Citi, Tampa, FL**

**Hadoop Engineer (May2016 - Till Date)**

**Project Description:**

Pulling the data from various data sources into Hadoop platform and standardized all data through a series of master data management processes to reach client goal. As a Hadoop developer, involved in maintaining the huge data and designing developing predictive data models for business users according to the requirement.

**Responsibilities:**

* Used Sqoop for importing and exporting data from Netezza, Teradata into HDFS and Hive.
* Imported Hive tables into Spark SQL context and converted into RDDs.
* Used Data Frames and Datasets APIs for performing analysis on Hive tables.
* Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce pattern.
* Developed Spark SQL scripts using Scala to perform transformations and actions on RDDs in spark for faster data Processing.
* Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.
* Responsible for analyzing the performance Hive queries using Impala.
* Developed Flume ETL job for handling data from HTTP source and sink as HDFS.
* Automated the Hadoop pipeline using oozie and scheduled using coordinator for time frequency and data availability.
* Monitoring of Hadoop Cluster using Cloudera Manager.
* Load and transform large sets of semi structured and unstructured data that includes sequence files and xml files and worked on Avro and Parquet file formats using compression techniques like Snappy, Gzip and Zlib.
* Worked on building hadoop cluster in AWS Cloud on multiple EC2 instances.
* Used Amazon Simple Storage Service(S3) for storing and accessing data to hadoop cluster.
* Used JIRA for Bit Bucket to check-in, Bug tracking and checkout code changes.

**Environment:** HDFS, Yarn, MapReduce, Oracle, Teradata, Sqoop, Oozie, Hive, Impala, HBase, Flume, Spark Streaming, Spark SQL, Scala, Python, Eclipse, Cloudera, AWS, S3, EC2.

**Verizon, Tampa, FL (Feb2015 – Apr2016)**

**Hadoop Developer**

**Project Description:**

The purpose of this project includes extraction, transforming and loading data into HDFS from an existing database applications. Hadoop is tested for scalability, reliability and applied transformations to generate

reports and to find pattern across sales, products and customers to perform some very key analytics to grow their sales across their organization.

**Responsibilities:**

* Worked on Large-scale Hadoop YARN cluster for distributed data processing and analysis using Sqoop, Pig, Hive and NoSQL databases.
* Participated in Gathering requirements, analyze requirements and design technical documents for business requirements.
* Responsible for importing data to HDFS using Sqoop from different RDBMS servers and exporting data using Sqoop to the RDBMS servers after aggregations for other ETL operations.
* Developed and used existing UDFs for custom implementation on table data.
* Created Partitioning, Bucketing, Map side Join, Parallel execution for optimizing the hive queries.
* Responsible for monitoring Cluster using Cloudera Manager.
* Developed Pig scripts for track data capture between arrived data and current data.
* Orchestrated hundreds of Sqoop queries, Pig scripts, Hive queries using Oozie workflows and sub-workflows.
* Responsible for handling different data formats like Avro, Parquet and ORC formats.

**Environment:** Hadoop, HDFS, Sqoop, Oozie, Pig, Hive, Oozie, Cassandra, Linux, YARN, Cloudera Manager

**BCBS, Detroit, MI (Dec2014-Jan2015)**

**Hadoop Developer**

**Project Description:**

The aim of this project is to analyze the data which is coming from multiple sources into the Hadoop data center unit. Created programs to process large volumes of data through a lot of prepay concepts, which analyze, produce suspect claims and it helps to generate datasets for visualization. The suspect claims are then verified again to save millions of dollars.

**Responsibilities:**

* Responsible for analyzing large data sets and derive customer usage patterns by developing new MapReduce programs.
* Extensively worked on Hive to analyze the partitioned and bucketed data and compute various metrics for reporting.
* Used Pig as ETL tool to do transformations, joins and some pre-aggregations before storing the data into HDFS.
* Imported all the customer specific personal data to Hadoop using Sqoop component from various relational databases like Netezza and Oracle.
* Experienced in running query using Impala and used BI tools to run ad-hoc queries directly on Hadoop.
* Worked with BI teams in generating the reports and designing ETL workflows on Tableau.
* Developed testing scripts in Python and prepare test procedures, analyze test results data and suggest improvements of the system and software.
* Experience in Job management using Autosys scheduler and developed job processing scripts using oozie workflow.

**Environment:** Hadoop, MapReduce, HDFS, Pig, HiveQL, Python HBase, Zookeeper, Oozie, Flume, Impala, Hortonworks, Storm, MySQL, UNIX Shell Scripting, Tableau.

**Cyient, Bangalore, India (June2009- Aug2013)**

**Java/J2EE Developer**

**Project Description:**

The product was a client -server application developed using the model view controller architecture. This business application software is for management of business information, which can maintain accounts, inventory and invoicing details, with user definable graphs for structured analysis.

**Responsibilities:**

* Involved in developing various data flow diagrams, use case diagrams and sequence diagrams.
* Responsible for developing and modifying the existing service layer based on the business requirements.
* Involved in database design.
* Created tables, stored procedures in SQL for data manipulation and retrieval, Database Modification using SQL, PL/SQL, Stored procedures, triggers, Views in Oracle 9i.
* Written Stored Procedures functions and views to retrieve the data.
* Involved in creating JSP pages and HTML Pages.
* Used HTTP Filtering in order to perform the filtering task on request and response.
* Worked extensively in JSP, HTML, JavaScript, and CSS to create the UI pages for the project.
* Used Maven builds to wrap around Ant build scripts.
* Created JUnit test cases for unit testing and developed generic JS functions for validations.
* Responsible to mentor/work with team members to make sure the standards and guidelines are followed and delivery of tasks in time.

**Environment:** JQuery, ETL, JSP, Servlets, Spring 2.0, JDBC, HTML, JUnit, JavaScript, XML, SQL, Maven, Web

Services,UML