

Hadoop Developer Hadoop Developer Hadoop Developer - Johnson & Johnson Irvine, CA ?
Cloudera certified Spark and Hadoop Developer with around 10+ years of experience in software development, deployment and maintenance of various web based applications using Java and Big Data Ecosystems on Windows and Linux environments. ? Expertise in designing Hadoop applications and recommending the right solutions and technologies for the applications. ? Expertise in major components of Hadoop ecosystems like HDFS, MapReduce, YARN, Hive, Pig, HBase, Zookeeper, Sqoop, Spark, Kafka, Cassandra and Impala. ? Good knowledge and hands-on experience on installing, configuring and maintaining multi-node clusters on various environments and distributions of Hadoop. ? Experience working in different Hadoop distributions like Cloudera 5.5 (CDH4, CDH5) and Hortonworks distributions (HDP). ? Thorough knowledge in ETL, Data Integration and Migration, extensively used ETL methodology for supporting Data Extraction, transformations and loading using Informatica. ? Good knowledge on Statistical and quantitative analysis using tools like R Studio. ? Hands-on Experience in using version control tools like CVS, GIT, and SVN. Build tools like SBT, Ant and Maven. ? Working experience on NoSQL databases like HBase, MongoDB and Cassandra with functionality and implementation. ? Involved in developing Impala scripts for extraction, transformation, loading of data into data warehouse. ? Working experience on Spark ecosystems using spark components like Spark Core, Spark SQL, Spark Streaming, MLlib, and GraphX. ? Experienced in collecting metrics for Hadoop clusters using Ambari and Cloudera Manager. ? Extensive experience working with real time streaming applications and batch style large scale distributed computing applications, worked on integrating Kafka with NiFi and Spark. ? Developed re-usable and configurable components as part of project requirements in Java, Scala and Python. ? Good knowledge of Scala's functional style programming techniques like Anonymous Functions (Closures), Currying, Higher Order Functions and Pattern Matching. ? Hands-on experience in training, evaluating and predicting the data as a part of Machine Learning using Spark MLlib, TensorFlow, and a regular contributor to Machine Learning projects on GitHub. ? Good experience in working with cloud environment like Amazon Web Services EC2 and S3. ? Hands on experience on working with Amazon EMR framework

transferring data to EC2server. ? Strong Problem Solving and Analytical skills and abilities to make Balanced & Independent Decisions. ? Good Team Player, Strong Interpersonal, Organizational and Communication skills combined with Self-Motivation, Initiative and Project Management Attributes. ? Holds strong ability to handle multiple priorities and work load and has ability to understand and adapt to new technologies and environments faster.

Work Experience Hadoop Developer Johnson & Johnson - Irvine, CA July 2018 to Present Responsibilities: ? Worked on analyzing Hadoop cluster using different big data analytic tools including Pig, Hive and MapReduce. ? Managing fully distributed Hadoop cluster is an additional responsibility assigned to me. I was trained to overtake the responsibilities of ? A Hadoop Administrator, which includes managing the cluster, Upgrades and installation of tools that uses Hadoop ecosystem. ? Worked on Installation and configuring of Zoo Keeper to co-ordinate and monitor the cluster resources. ? Implemented test scripts to support test-driven development and continuous integration. ? Worked on POC's with Apache Spark using Scala to implement spark in project. ? Consumed the data from Kafka using Apache spark. ? Load and transform large sets of structured, semi structured and unstructured data. ? Involved in loading data from LINUX filesystem to HDFS. ? Importing and exporting data into HDFS and Hive using Sqoop. ? Implemented Partitioning, Dynamic Partitions, Buckets in Hive. ? Worked in creating HBase tables to load large sets of semi-structured data coming from various sources. ? Extending HIVE and PIG core functionality by using custom User Defined Function's (UDF), User Defined Table-Generating Functions (UDTF) and User DefinedAggregating Functions (UDAF) for Hive and Pig using python. ? Experienced in running Hadoop streaming jobs to process terabytes of xml format data. ? Involved in scheduling Oozie workflow engine to run multiple Hive and pig jobs. ? Experienced with performing CURD operations in HBase. ? Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig. ? Responsible for loading data les from various external sources like ORACLE, MySQL into staging area in MySQL databases. ? Executed Hive queries on Parquet tables stored in Hive to perform data analysis to meet the business requirements. ? Actively involved in code review and bug fixing for improving the performance. ? Good experience in handling data manipulation using python Scripts. ? Involved in

development, building, testing, and deploy to Hadoop cluster in distributed mode. ? Created Linux shell Scripts to automate the daily ingestion of IVR data ? Executed Hive queries on Parquet tables stored in Hive to perform data analysis to meet the business requirements. ? Helped the Analytics team with Aster queries using HCatalog. ? Automated the History and Purge Process. ? Created HBase tables to store various data formats of incoming data from different portfolios. ? Created Pig Latin scripts to sort, group, join and filter the enterprise wise data. ? Developed the verification and control process for daily load. Experience in Daily production support to monitor and trouble shoots Hadoop/Hive jobs Environment: Hive, SQL, Pig, Flume, Kafka, Map reduce, SQOOP, Spark, Python, Java, Shell Scripting, Teradata, Oracle, Oozie, Cassandra Hadoop/Spark Developer Citi Bank - Tampa, FL January 2016 to June 2018 This project is to capture the large currency transactions made on the various channels of the current system. All cash transactions made by non-exempt customers, above the threshold amount are reported to The Financial Crimes Enforcement Network (FinCEN) by submitting the Cash Transaction Report (CTR). ECV gets all the transactions details from various upstream data source systems into Hadoop as a Service (HAAS) environment and filters only the cash transactions. These transactions are further curated with ATM and Teller OSI data and enriched with user profile data. This data will be stored in HBase for auto generating CTRs. Responsibilities: ? Evaluated business requirements and prepared Detailed Design documents that follows Project guidelines and SLAs required procuring data from all the upstream data sources and developing written programs. ? Data files are retrieved by various data transmission protocols like Sqoop, NDM, SFTP, DMS etc., these data files are then validated by various Spark Control jobs written in Scala. ? Spark RDDs are created for all the data files and then transformed to cash only transaction RDDs. ? The filtered cash only RDDs are aggregated and curated based on the business rules and CTR requirements, 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 and in HBase for auto generating CTRs. ? Developed Spark scripts by using Scala and Python shell commands as per the requirement. ? Maintained and administrated HDFS through Hadoop - Java API, shell scripting,

Python. ? Used Python for writing script to move the data across clusters. ? Expertise in designing Python scripts to interact with middleware/back end services. ? Worked on python scripts to analyze the data of the customer. ? Involved in converting Cassandra/Hive/SQL queries into Spark transformations using Spark RDD's, and Scala Python. ? Developed monitoring and notification tools using Python. ? Wrote Python routines to log into the websites and fetch data for selected options. ? Used Collections in Python for manipulating and looping through different user defined objects. ? Wrote and tested Python scripts to create new data files for Linux sever configuration using a Python templet tool. ? Wrote shell scripts to automate the jobs in UNIX. ? Used log4j API to write log files. ? Understood the existing Oozie workflows and modified them as per new requirements. Environment: Cloudera Distribution 5.5, Hadoop Map Reduce, Spark 1.6, HDFS, Python, Hive, HBase, HiveQL, SQOOP, Java, Scala 2.10.4, Unix, IntelliJ, Maven. Hadoop Developer Sears Holding Corporation, Hoffman Estate, IL February 2014 to December 2015 With over 100 million customers and annual revenue of \$30 billion, Sears generates a huge amount of data on the transactions made by the customers. The scope of the project is to use the customer transaction data in store and online over a period to recommend items to customer, that they will find engaging. Another concept is a recommendation engine that introduces new products to a customer, which they might have not, came across before. To achieve these goals, we perform data exploration to learn about user behavior. We perform feature engineering to create new features from existing features that truly reflects the signals in the data and avoid noise. Responsibilities: ? Worked with the source team to understand the format & delimiters of the data files. ? Responsible for generating actionable insights from complex data to drive significant business results for various application teams. ? Developed and implemented API services using Python in spark. ? Troubleshoot and resolve data quality issues and maintain important level of data accuracy in the data being reported. ? Extensively implemented POC's on migrating to Spark-Streaming to process the live data. ? Ingested data from RDBMS and performed data transformations, and then export the transformed data to Cassandra as per the business requirement. ? Re-writing existing map-reduce jobs to use new features and improvements for achieving faster results. ? Analyzes

large amount of data sets to determine optimal way to aggregate and report on it. ? Performance tuned slow running resource intensive jobs. ? Worked on Data serialization formats for converting complex objects into sequence bits by using Avro, Parquet, JSON, CSV formats. ? Hands on experience working on in-memory based Apache Spark application for ETL transformations. ? Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Python. ? Developed multiple POCs using PySpark and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata. ? Developed Flume configuration to extract log data from different resources and transfer data with different file formats (JSON, XML, and Parquet) to hive tables using different SerDe's. ? Setup Oozie workflow /sub workflow jobs for Hive/SQOOP/HDFS actions. ? Experience in accessing Kafka cluster to consume data into Hadoop. ? Involved in importing the real-time data to Hadoop using Kafka and implemented the Oozie job for daily imports. ? Worked with business and functional requirement gathering team, updated user comments in JIRA and documented in confluence ? Handled tasks like maintaining accurate roadmap for project or certain product. ? Monitoring the sprints, burndown charts and completing the monthly reports. Environment: Hive, SQL, Pig, Flume, Kafka, Map reduce, SQOOP, Spark, Python, Java, Shell Scripting, Teradata, Oracle, Oozie, Cassandra Hadoop Developer Asurion Inc - Nashville, TN September 2012 to December 2013 Asurion is a privately held company that provides insurance for smartphones, tablets, consumer electronics, appliances, satellite receivers. My team was responsible for generating a Hadoop based implementation of integrating data from the Warehouse and Database as input to the algorithm used by the Business Intelligence team. Responsibilities ? Setting up the cluster, configuration and maintenance, install components of the Hadoop ecosystem. ? Exported the analyzed data to the relational databases using Sqoop and process the data for visualization and to generate reports for the BI team. ? Stored data from HDFS to respective Hive tables for business analysts to conduct further analysis in identifying data trends. ? Developed Hive ad-hoc queries and filtered data in order to increase the effectiveness of the process execution by using functions like Joins, Group By, and Having. ? Increased the time efficiency of the Hive QL using partitioning of data and reduced the time

difference of executing the sets of data by applying the compression techniques like SNAPPY for Map-Reduce Jobs. ? Created Hive Partitions for storing data for different trends under different partitions. ? Connected the Hive tables to data analysis tools like Tableau for graphical representation of the trends. ? Assisted project manager in problem shooting relevant to Hadoop technologies for data integration between different platforms like Sqoop-Sqoop, Hive-Sqoop, and Sqoop-Hive. Environment: Hortonworks, Java 7, HBase, HDFS, MapReduce, Hadoop 2.0, Hive, Pig, Eclipse, Linux, Sqoop, MySQL, Agile, Kafka. Java Developer Automatic Data Processing July 2009 to August 2012 Automatic Data Processing is a provider of business outsourcing solutions. They provide Services such as Human Capital Management, Tax & Compliance, Electronic Payment Solutions, Dealer Services and Medical Practice Services. This project is focused on Human Capital Management and particularly on Payroll Services, Human Resource Management, Talent Management, Benefits Administration, Time & Attendance, and International Solutions. This front-end application has interface with various Java applications, which are used to send and receive information to third party vendors. Responsibilities: ? Involved in Full Life Cycle Development in Distributed Environment Using Java and J2EE framework. ? Responsible for developing and modifying the existing service layer based on the business requirements. ? Involved in designing & developing web-services using SOAP and WSDL. ? Involved in database design. ? Created tables, views, triggers, stored procedures in SQL for data manipulation and retrieval ? Developed Web Services for Payment Transaction and Payment Release. ? Involved in Requirement Analysis, Development and Documentation. ? Developed front-end using JSP, HTML, CSS and JavaScript. ? Coding for DAO Objects using JDBC (using DAO pattern). ? XML and XSDs are used to define data formats. ? Implemented J2EE design patterns such as singleton, DAO for the presentation tier, business tier and Integration Tier layers of the project. ? Involved in Bug fixing and functionality enhancements. ? Followed coding and documentation standards and best practices. ? Participated in project planning discussions and worked with team members to analyze the requirements and translate them into working software modules. Environment: Java, J2EE, JSP, SOAP, WSDL, SQL, PL/SQL, XML, JDBC, Eclipse, Windows XP, Oracle Skills CASSANDRA,

AMBARI, HDFS, MAPREDUCE, OOZIE, SQOOP, HBASE, FLUME, MONGODB, TERADATA, C/C++, C++, Git, HBase, Hive, HTML, JAVASCRIPT, MapReduce, PHP, Pig Additional Information TECHNICAL SKILLS BigData Ecosystem: HDFS, MapReduce, YARN, Hive, Pig, Flume, Sqoop, HBase, Oozie, Zookeeper, Ambari, Spark (also Spark Core, Spark SQL, Spark Streaming, Spark MLlib, GraphX. Databases: Microsoft SQL Server, Oracle 12c, PL/SQL, MySQL, MongoDB, Cassandra, Teradata Programming Languages: C/C++, Java, Scala, Python, Shell Scripting, R Programming, Python, Web: HTML, CSS, PHP, JavaScript, AngularJS, NodeJS Version Control: Git, Ant, Maven Operating System: UNIX, RedHat Linux, CentOS, Ubuntu, Microsoft Windows Amazon AWS: Amazon S3, Amazon EC2, Amazon RDS Tools & IDE: Eclipse, IntelliJ, NetBeans, Maven, Jenkin, SBT

Name: Alexander White

Email: jonathan96@example.net

Phone: (405)623-9240