

Hadoop & Spark Developer Hadoop & Spark Developer Hadoop & Spark Developer - American Express Arizona City, AZ Over 8 years of professional IT experience which includes 4+ years of experience with Hadoop Map Reduce, HDFS and Hadoop Ecosystems like Bigdata, HDFS, MapReduce, Oozie, Cassandra, Hive, Sqoop, Pig, Flume, Hbase and Zookeeper and 5 years in Java and Oracle PLSQL development. 7+ years of experience in development of applications using Object Oriented Programming. In-depth knowledge of Hadoop architecture and its components like HDFS, Name Node, Data Node, Job Tracker, Application Master, Resource Manager, Task Tracker and Map Reduce programming paradigm. Experience in cluster planning, designing, deploying, performance tuning, administering and monitoring Hadoop ecosystem. Commendable knowledge / experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems (RDBMS) and vice-versa. Experience in developing Map/Reduce jobs to process large data sets utilizing the Map/Reduce programming paradigm. Good understanding of cloud configuration in Amazon web services (AWS). Experience in database design. Used PL/SQL to write Stored Procedures, Functions, Triggers and strong experience in writing complex queries for Oracle. Proficient in writing SQL, PL/SQL stored procedures, functions, constraints, packages and triggers. Good experience in Hive tables design, loading the data into hive tables. Good understanding of HDFS Designs, Daemons, federation and HDFS high availability (HA). Good knowledge on Hadoop Cluster architecture and monitoring the cluster. Hadoop Shell commands, Writing Map reduce Programs, Verifying the Hadoop Log Files. Exposure on Query Programming Model of Hadoop. Expert on UML for Object Oriented Analysis & Design (OOAD) using MS Visio, IBM Rational. Expert on Core Java, multi-threading, debugging JVM and optimizing and profiling Java Applications. Experience on System Study, Analysis, of Business requirement, preparation of Technical design, UTP and UTC, Coding, Unit testing, Integration testing, System testing and Implementation. Experience in Object Oriented Analysis and Design (OOAD) and development of software using UML methodology. Hands on experience with Core Java with Multithreading, Concurrency, Exception Handling, File handling, IO, Generics and Java Collections. Implemented rich web applications such as HTML, XHTML, XML, XSLT,

CSS, JavaScript, AJAX(DWR), jQuery, ExtJS, JSON, and DOJO. Excellent working knowledge of MVC architecture and Struts, Spring MVC and JSF Frameworks. Developed applications using Core Java, Servlets, JSP, JDBC, Struts, Spring, Hibernate. Good understanding of the SOA technologies like SOAP, WSDL Web Services. Knowledge of Software Development Methodologies like Agile (SCRUM), Waterfall. Proficient in using application servers like JBoss and Tomcat Servers. Configured and deployed applications on IBM Web sphere, BEA Web logic, Tomcat. Excellent working knowledge of Service Oriented Architecture(SOA) , Messaging and Web Services. Experienced on developing, building and deploying applications on UNIX, Linux, Solaris and Windows platforms. Experienced in database design and development and JDBC connectivity for Oracle 11g/10g/9i/8i (SQL, PL/SQL, Stored procedures), MS SQL Server 2008/2005/2000, DB2 9.x/8.x and MySQL. Working knowledge of Java external applications like JUnit, Log4J, Apache Ant, Maven. Experienced in building and deploying applications on servers using Ant, Maven, and Perl. Worked with query tools like Toad, SQL Plus, SQL Developer. Expert level skills in Designing and Implementing web servers solutions and deploying Java Application Servers like Websphere, Web Logic, configuring Apache Web Server and configuring various Servlet engines. Comprehensive knowledge of physical and logical data modeling, performance tuning. Resourceful and skilled in analyzing and solving problems. Extensive experience in writing and executing JUnit Test cases & debugging Java/J2ee applications. Hands on working experience with different version management software such as VSS, Win CVS, Subversion, Star Team and SVN. Excellent written, verbal communication and customer service skills. Strong organizational, and interpersonal skills. And possess a high level of drive, initiative and self-motivation. A collaborative personality who enjoy working in a team-oriented environment.

Excellent debugging skills. Able to debug complex technical issues including multiple system components. Highly creative and articulate. Can adapt quickly to rapidly changing conditions. Work Experience Hadoop & Spark Developer American Express - Phoenix, AZ October 2015 to Present The American Express is an American multinational financial services corporation. The company is best known for its credit card, charge card, and traveler's cheque businesses. Amex

cards account for approximately 24% of the total dollar volume of credit card transactions in the USA. This project is designed to extract the raw weblogs into Hadoop ecosystem and provide a complete picture of customer behavior across these various work streams. About five years ago, American Express recognized that traditional databases would not be enough to effectively handle the level of data and analytics needed for their projects, and decided that a big data infrastructure would be the solution. Data volume is not only increasing, but data sources are also changing. As part of American Express's ongoing journey, they must keep up with these changes in style of interactions as well as with the increasing volume. Part of that involves making a huge number of decisions, millions every day. If American Express can become just a little bit smarter in these decisions, it can have a huge advantage to customers and to the company. This project is to expand on how they can use machine learning at large scale. With access to big data, machine learning models can produce superior discrimination and thus better understand customer behavior. The primary focus is Fraud detection, new customer acquisition and recommendation for better customer experience.

**Roles & Responsibilities:** Developed Big Data Solutions that enabled the business and technology teams to make data-driven decisions on the best ways to acquire customers and provide them business solutions. Exported the business-required information to RDBMS using Sqoop to make the data available for BI team to generate reports based on data. Migrated the existing data to Hadoop from RDBMS (SQL Server and Oracle) using Sqoop for processing the data. Developed Spark Programs for Batch and Real time Processing. Developed Spark Streaming applications for Real Time Processing. Implemented Hive custom UDF's to achieve comprehensive data analysis. Involved in installing Hadoop and Spark Cluster in Amazon Web Server. Involved in installing, configuring and managing Hadoop Ecosystem components like Spark, Hive, Pig, Sqoop, Kafka and Flume. Responsible for loading unstructured and semi-structured data into Hadoop cluster coming from different sources using Flume and managing. Responsible for Data Ingestion like Flume and Kafka. Used Hive data warehouse tool to analyze the data in HDFS and developed Hive queries. Created internal and external tables with properly defined static and dynamic partitions for efficiency. Developed MapReduce programs to cleanse

and parse data in HDFS obtained from various data sources and to perform joins on the Map side using distributed cache. Implemented daily workflow for extraction, processing and analysis of data with Oozie. Used the RegEx, JSON and Avro SerDe's for serialization and de-serialization packaged with Hive to parse the contents of streamed log data. Used Pig to develop ad-hoc queries. Responsible for troubleshooting MapReduce jobs by reviewing the log files.

Environment: Hadoop, Spark, Spark Streaming, Spark Mlib, Scala, Hive, Pig, Hcatalog, MapReduce, Oozie, Sqoop, Flume and Kafka. Hadoop Developer/Admin Merck - Kenilworth, NJ June 2014 to October 2015 Merck & Co., Inc. is an American pharmaceutical company and one of the largest pharmaceutical companies in the world. The company was established in 1891. It is the world's seventh largest pharmaceutical company by market capitalization and revenue. Its headquarters is located in Kenilworth, New Jersey. Adverse Drug Event (ADE) An adverse drug event (ADE) is an unwanted or unintended reaction that results from the normal use of one or more medications. The consequences of ADEs range from mild allergic reactions to death, with one study estimating that 9.7% of adverse drug events lead to permanent disability. However the sheer amount of data involved in this analysis is a challenge. Even restricting ourselves to analyzing pairs of drugs, there are more than 3 trillion potential drug-drug-reaction triples in the Adverse Event Reporting System (AERS) dataset, and tens of millions of triples in the data. Apache Spark is replacing Mapreduce jobs to clean, aggregate, and join data, parallelize the counting problem across multiple machines to achieve a linear speedup in the overall runtime. The faster runtime for each individual analysis allows rapid iteration on smaller models and tackle larger problems involving more drug interactions than anyone has ever looked at before. Responsibilities All the datasets are loaded from two different sources such as Oracle, MySQL to HDFS and Hive respectively on daily basis. Process 8 flat files all are delimited by Comma. Responsible in creating Hive Tables to load the data which comes from MySQL and loading data from Oracle to HDFS using Sqoop. Good hands on experience in writing core java level programming in order to perform cleaning, pre-processing and data validation. Involved in verifying cleaned data using Talend tool with other department. Experienced in creating Hive schema, external tables and

managing views. Involved in developing Hive UDFs and reused in some other requirements. Worked on performing Join operations. Involved in creating partitioning on external tables. Good hands on experience in writing HQL statements as per the user requirements. Fetching the HQL results into CSV files and handover to reporting team. Work with hive complex datatypes and involved in Bucketing. Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Scala and have a good experience in using Spark-Shell and Spark Streaming. Develop Spark code using Scala and Spark-SQL for faster testing and data processing. Import millions of structured data from relational databases using Sqoop import to process using Spark and stored the data into HDFS in CSV format. Use Spark SQL to process the huge amount of structured data. Implement Spark RDD transformations, actions to migrate Map reduce algorithms.

Expertise in running Hadoop streaming jobs to process terabytes data. Experience in importing the real time data to Hadoop using Kafka and implemented the Oozie job. Responsible in analysis, design, testing phases and responsible for documenting technical specifications. Along with the Infrastructure team, involved in design and developed Kafka and Storm based data pipeline. Develop storm-monitoring bolt for validating pump tag values against high-low and Worked on Talend Administrator Console (TAC) for scheduling jobs and adding users. Develop Kafka producer and consumers, Hbase clients, Spark and Hadoop MapReduce jobs along with components on HDFS, Hive. Good knowledge in partitions, bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance. Environment: Hadoop, Hive, MapReduce, Pig, MongoDB, Oozie, Sqoop, Kafka, Cloudera, Spark, HBase, HDFS, Solr, Zookeeper, Cassandra, DynamoDB

Hadoop Developer Home Depot - Atlanta, GA October 2012 to June 2014 The Home Depot, Inc. is a home improvement supplies retailing company that sells tools, construction products, and services. It operates many big-box format stores across the United States (including all 50 states, the District of Columbia, Puerto Rico, the Virgin Islands, and Guam), all 10 provinces of Canada, and the country of Mexico. The company is headquartered at the Atlanta Store Support Center in Cobb County, Georgia, in Greater Atlanta. Home Depot is the largest home improvement retailer in the United States, ahead of rival Lowe's. Project to predict

buying habits of customers and display ads on an online store. This was a proof of Concept developed to determine if buying habits can be extracted without adding any additional monitoring scripts and with just based on ad interaction and social interaction of the consumer. The data was all gathered from social networks Twitter, Facebook to observe how vocal consumers are about their purchases.

**Responsibilities** Worked with technology and business groups for Hadoop migration strategy. Researched and recommended suitable technology stack for Hadoop migration considering current enterprise architecture. Designed docs and specs for the near real time data analytics using Hadoop and HBase. Installed Cloudera Manager 3.7 on the clusters. Used a 60 node cluster with Cloudera Hadoop distribution on Amazon EC2. Developed ad-clicks based data analytics, for keyword analysis and insights. Crawled public posts from Facebook and tweets. Wrote MapReduce jobs with the Data Science team to analyze this data. Validated and Recommended on Hadoop Infrastructure and data centre planning considering data growth. Transferred data to and from cluster, using Sqoop and various storage media such as Informix table's and flat files. Developed MapReduce programs and Hive queries to analyse sales pattern and customer satisfaction index over the data present in various relational database tables. Worked extensively in performance optimization by adopting/deriving at appropriate design patterns of the MapReduce jobs by analysing the I/O latency, map time, combiner time, reduce time etc. Developed Pig scripts in the areas where extensive coding needs to be reduced. Developed UDF's for Pig as needed. Followed agile methodology for the entire project Defined problems to look for right data and analyze results to make room for new project. Environment: Hadoop 0.20, HBase, HDFS, MapReduce, Java, Cloudera Manager 2, Amazon EC2 classic. Java/J2ee Developer

Penn Mutual - Horsham, PA July 2011 to October 2012 The Penn Mutual Life Insurance Company, commonly referred to as Penn Mutual, was founded in Philadelphia, Pennsylvania, in 1847. It was the seventh mutual life insurance company chartered in the United States. The project was for an insurance portal which would enable Penn Mutual's customers to get a quote, Login to their policy, Report a Claim, Pay Bills, Update their Policy information, and make updates to their policies.

**Responsibilities:** Involved in the analysis, Design, Coding, Modification and implementation of

User Requirements in the Electronic Credit File Management system. Designed the application using Front Controller, Service Controller, MVC, Session Facade Design Patterns. The application is designed using MVC Architecture. Implemented the required functionality using Hibernate for persistence & Spring Framework. Used Spring Framework for Dependency Injection. Designed and implemented the Hibernate Domain Model for the services. Developed UI using HTML, JavaScript and JSP and developed Business Logic and Interfacing components using Business Objects, XML, and JDBC. Designed user-interface and checking validations using JavaScript. Involved in design of JSP's and Servlets for navigation among the modules. Developed various EJBs for handling business logic and data manipulations from database. Managed connectivity using JDBC for querying/inserting & data management including triggers and stored procedures. Developed SQL queries and Stored Procedures using PL/SQL to retrieve and insert into multiple database schemas. Developed the XML Schema and Web services for the data maintenance and structures. Wrote test cases in JUnit for unit testing of classes. Used DOM and DOM Functions using Firefox and IE Developer Tool bar for IE. Debugged the application using Firebug to traverse the documents. Involved in developing web pages using HTML and JSP. Provided Technical support for production environments resolving the issues, analysing the defects, providing and implementing the solution defects. Built and deployed Java applications into multiple UNIX based environments and produced both unit and functional test results along with release notes. Developed the presentation layer using CSS and HTML taken from bootstrap to develop for browsers. Environment: Java, Spring, JSP, Hibernate, XML, HTML, JavaScript, JDBC, CSS, SOAP Web services. Java/J2ee Developer Axis Bank - Hyderabad, Telangana May 2009 to July 2011 Axis Bank Limited is the third largest private sector bank in India and for Information Security Risk Management. The company provides wealth management, asset management, and investment banking services for private, corporate, and institutional clients worldwide, and is generally considered to be a bulge bracket bank. It deals with data related to investment banking. It loads IT Credit Risk information into oracle database using Core Java 1.7, PLSQL and Unix Shell Scripting.

Responsibilities: Developed JavaScript behavior code for user interaction. Created database

program in SQL server to manipulate data accumulated by internet transactions. Wrote Servlets class to generate dynamic HTML pages. Developed Servlets and back-end Java classes using Web Sphere application server. Developed an API to write XML documents from a database. Performed usability testing for the application using JUnit Test. Maintenance of a Java GUI application using JFC/Swing. Created complex SQL and used JDBC connectivity to access the database. Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive. Involved in the design and coding of the data capture templates, presentation and component templates. Developed Scala scripts, UDFs using both Data frames/SQL and RDD/MapReduce in Spark 1.3 for Data Aggregation, queries and writing data back into OLTP system directly or through Sqoop. Part of the team that designed, customized and implemented metadata search and database synchronization. Experience in working with versioning tools like Git CVS & Clear Case.

Used Oracle as Database and used Toad for queries execution and also Involved in writing SQL scripts, PL SQL code for procedures and functions. Environment: Java, Web Sphere 3.5, EJB, Servlets, Spark, JavaScript, JDBC, SQL, Sqoop, Git, JUnit, Eclipse IDE. Apache Tomcat 6, UDF.

Education Bachelor's Electronics Engineering D.J.Sanghvi College of Engineering Skills CASSANDRA, HDFS, MAPREDUCE, OOZIE, SQOOP, HBASE, KAFKA, SOLR, DB2, FLUME, JBOSS, JMS, MONGODB, NOSQL, APPLICATION SERVER, C/C++, C++, HBase, Hive, HTML

Additional Information SKILLS Big Data Technology HDFS, Mapreduce, HBase, Pig, Hive, SOLR, Sqoop, Flume, MongoDB, Cassandra, Puppet, Oozie, Zookeeper, Spark, Kafka Java/J2EE Technology JSP, JSF, Servlets, EJB, JDBC, Struts, Spring, Spring MVC, Spring Portlet, Spring Web Flow, Hibernate, iBATIS, JMS, MQ, JCA, JNDI, Java Beans, JAX-RPC, JAX-WS, RMI, RMI-IIOP, EAD4J, Axis, Castor, SOAP, WSDL, UDDI, JiBX, JAXB, DOM, SAX, MyFaces(Tomahawk), Facelets, JPA, Portal, Portlet, JSR 168/286, LifeRay, WebLogic Portal, LDAP, JUnit.NET Languages Java (1.4/5/6), C/C++, Swing, SQL, HTML, CSS, i18n, l10n, DHTML, XML, XSD, XHTML, XSL, XSLT, XPath, XQuery, SQL, PL/SQL, UML, JavaScript, AJAX(DWR), jQuery, Dojo, ExtJS, Shell Scripts, Perl Development Framework/IDE RAD 8.x/7.x/6.0, IBM WebSphere Integration Developer 6.1, WSAD 5.x, Eclipse Galileo/Europa/3.x/2.x,



MyEclipse 3.x/2.x, NetBeans 7.x/6.x, IntelliJ 7.x, Workshop 8.1/6.1, Adobe Photoshop, Adobe Dreamweaver, Adobe Flash, Ant, Maven, Rational Rose, RSA, MS Visio, OpenMake Meister Web/Application Servers WebSphere Application Server 8.x/7.0/6.1/5.1/5.0, WebSphere Portal Server 7.0/6.1, WebSphere Process Server 6.1, WebLogic Application Server 8.1/6.1, JBoss 5.x/3.x, Apache 2.x, Tomcat 7.x/6.x/5.x/4.x, MS IIS, IBM HTTP Server Databases NoSQL, Oracle 11g/10g/9i/8i, DB2 9.x/8.x, MS SQL Server 2008/2005/2000, MySQL NoSQL HBase, Cassandra, MongoDB, Accumulo Operating Systems Windows XP, 2K, MS-DOS, Linux (Red Hat), Unix (Solaris), HP UX, IBM AIX Version Control CVS, SourceSafe, ClearCase, Subversion, AllFusion Harvest Change Manager 7.1 Monitoring Tools Embarcadero J Optimizer 2009, TPTP, IBM Heap Analyzer, Wily Introscope, JMeter Other JBoss Drools 4.x, REST, IBM Lotus WCM, MS ISA,CA SiteMinder, BMC WAM, Mingle

Name: Christopher Johnson

Email: lweaver@example.com

Phone: 456.545.0614x481