Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - Genuent, LLC Philadelphia, PA   Having 8+ years of overall IT experience of comprehensive experience as an Apache Hadoop Developer.   Expertise in writing Hadoop Jobs for analyzing structured and unstructured data using HDFS, Hive, HBase, Pig, Spark, Kafka, Scala, Oozie and Talend ETL.   Good knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, YARN and MapReduce concepts.   Experience in working with different kind of MapReduce programs using Hadoop for working with Big Data analysis.   Experience in analyzing data using Hive QL, Pig Latin and custom MapReduce programs in Java.   Experience in importing/exporting data using Sqoop into HDFS from Relational Database Systems and vice-versa.   Extensive knowledge and experience on real time data streaming techniques like Kafka, Storm and Spark Streaming.   Working experience on designing and implementing complete end-to-end Hadoop Infrastructure including PIG, HIVE, Sqoop, Oozie, Flume and zookeeper.   Good Knowledge in providing support to data analyst in running Pig and Hive queries.   Experience in writing shell scripts to dump the shared data from MySQL servers to HDFS.   Experience in designing both time driven and data driven automated workflows using Oozie.   Knowledge in performance tuning the Hadoop cluster by gathering and analyzing the existing infrastructure. Experience in working with various Cloudera distributions (CDH4/CDH5), Hortonworks and Amazon EMR Hadoop Distributions.   Extensively worked on Hive and Sqoop for sourcing and transformations.   Experience in automating the Hadoop Installation, configuration and maintaining the cluster by using the tools like Puppet.   Hands on experience knowledge in NoSQL databases like HBase, Cassandra, Mongo db.   Experience in working with flume to load the log data from multiple sources directly into HDFS.   Strong debugging and problem solving skills with excellent understanding of system development methodologies, techniques and tools.   Wrote Flume configuration files for importing streaming log data into HBase with Flume.   Processing this data using Spark Streaming API with Scala.   Worked in complete Software Development Life Cycle (analysis, design, development, testing, implementation and support) in different application domain involving different technologies varying from object oriented technology to Internet programming on

Windows NT, Linux and UNIX/ Solaris platforms and RUP methodologies.   Familiar with RDBMS concepts and worked on Oracle 8i/9i, SQL Server 7.0., DB2 8.x/7.x   Involved in writing shell scripts, Ant scripts for Unix OS for application deployments to production region.   Used Spark API over Hortonworks Hadoop YARN to perform analytics on data in Hive.   Experience in Hadoop Distributions like Cloudera, Hortonworks, Big Insights, MapR Windows Azure, and Impala.   Hands on experience in developing the applications with Java, J2EE, J2EE - Servlets, JSP, EJB, SOAP, Web Services, JNDI, JMS, JDBC2, Hibernate, Struts, Spring, XML, HTML, XSD, XSLT, PL/SQL, Oracle10g and MS-SQL Server RDBMS.   Having very good POC and Development experience on Apache Flume, Kafka, Spark, Storm, and Scala.   Good understanding in using data ingestion tools-such as Kafka, Sqoop and Flume.   Good working knowledge on Hadoop hue ecosystems.   Good knowledge in evaluating big data analytics libraries and use of Spark-SQL for data exploratory. Exceptional ability to quickly master new concepts and capable of working in-group as well as independently with excellent communication skills. Work Experience Sr. Hadoop Developer Genuent, LLC - Houston, TX October 2017 to Present The Northwestern Mutual is a financial services mutual organization. Its products include life insurance, long-term care insurance, disability insurance, annuities, mutual funds, stocks, bonds and employee benefit services.   Responsibilities:

   Good at working on Hadoop, MapReduce, and Yarn/MRv2 developed multiple MapReduce jobs for structured, semi-structured and unstructured data in java.   Involved in Configuring Hadoop cluster and load balancing across the nodes.   Developed MapReduce programs in Java for parsing the raw data and populating staging Tables   Created Hive queries to compare the raw data with EDW reference tables and performing aggregates.   Experienced in developing custom input formats and data types to parse and process unstructured and semi structured input data and mapped them into key value pairs to implement business logic in Map-Reduce.   Experience in implementing custom sterilizer, interceptor, source and sink as per the requirement in Flume to ingest data from multiple sources.   Experience in setting up Fan-out workflow in flume to design v shaped architecture to take data from many sources and ingest into single sink.   Analyzing the requirement to setup a cluster.   Importing and exporting data into HDFS and Hive using Sqoop

Experienced in analyzing data with Hive and Pig   Experienced knowledge over designing Restful services using java based API's like JERSEY.   Developed Pig Latin scripts to extract the data from the web server output files to load into HDFS.   Integrating bulk data into Cassandra file system using MapReduce programs.   Got good experience with NoSQL databases HBase, Cassandra. Involved in HBase setup and storing data into HBase, which will be used for further analysis. Expertise in designing, data modelling for Cassandra NoSQL database.   Experienced in managing and reviewing Hadoop log files.   Experienced in defining job flows using Oozie workflow   Involved in working with Spark on top of Yarn/MRv2 for interactive and Batch Analysis   Worked closely with AWS EC2 infrastructure teams to troubleshoot complex issues   Expertise in writing the Scala code using higher order functions for the iterative algorithms in spark for performance consideration Experienced in analyzing and Optimizing RDD's by controlling partitions for the given data   Good understanding on DAG cycle for entire spark application flow on Spark application WebUI Experienced in writing live Real-time Processing using Spark Streaming with Kafka   Developed custom mappers in python script and Hive UDFs and UDAFs based on the given requirement Used HiveQL to analyze the partitioned and bucketed data and compute various metrics for reporting   Experienced in querying data using SparkSQL on top of Spark engine   Experience in managing and monitoring Hadoop cluster using Cloudera Manager   Supported in setting up QA environment and updating configurations for implementing scripts with Pig, Hive and Sqoop Environment: CDH , Java(JDK1.7), Hadoop, MapReduce, HDFS, Hive, Sqoop, Flume, HBase, Cassandra, Pig, Oozie, Kerberos, Scala, Spark, SparkSQL, Spark Streaming, Kafka, Linux, AWS, Shell Scripting, MySQL Oracle 11g, PL/SQL, SQL*PLUS Hadoop Developer Saint Luke's Health System - Kansas City, MO September 2015 to October 2017 Saint Luke's Health System includes 10 hospitals and campuses across the Kansas City region, home care and hospice, behavioural health care, dozens of physician practices, a life care senior living community, and more. Responsibilities: Installed and configured Hadoop Mapreduce, HDFS, Developed multiple MapReduce jobs in java for data cleaning and pre-processing.   Supported Map Reduce Programs those are running on the cluster.   Involved in loading data from UNIX file system to HDFS.

Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.   Develop MapReduce jobs for the users. Maintain, update and schedule the periodic jobs which range from updates on periodic MapReduce jobs to creating ad-hoc jobs for the business users.   Importing and exporting data into HDFS and Hive using Sqoop.   Experienced in defining job flows.   Experienced in managing and reviewing Hadoop log files.   Extracted files from Couch DB through Sqoop and placed in HDFS and processed.   Experienced in running Hadoop streaming jobs to process terabytes of xml format data.   Load and transform large sets of structured, semi structured and unstructured data.   Responsible to manage data coming from different sources.   Got good experience with NOSQL database.   Developed a custom File System plug in for Hadoop so it can access files on Data Platform.   This plugin allows Hadoop MapReduce programs, HBase, Pig and Hive to work unmodified and access files directly.   Designed and implemented Mapreduce-based large-scale parallel relation-learning system   Extracted feeds form social media sites such as Facebook, Twitter using Python scripts.   Setup and benchmarked Hadoop/HBase clusters for internal use.   Gained very good business knowledge on health insurance, claim processing, fraud suspect identification, appeals process etc.   Involved in review of functional and non-functional requirements.   Facilitated knowledge transfer sessions.

Environment: Java , Eclipse, Oracle , Sub Version, Hadoop, Hive, HBase, Linux, MapReduce, HDFS, Hive, Java (JDK), Hadoop Distribution of Horton Works, Cloudera, MapReduce, DataStax, IBM DataStage, Oracle, PL/SQL, SQL*PLUS, UNIX Shell Scripting. Hadoop Developer Utica National Insurance Group - Utica, NY August 2013 to September 2015 Utica National Insurance Group is a Top 100, nationally recognized insurer providing personal and commercial insurance products and services, with the second-largest errors and omissions business in the United States. Utica National sells its products through more than 2,200 independent insurance agents and employs over 1,200 people countrywide.   Responsibilities: Worked on Spark and Cassandra for the User behavior analysis and lightning speed execution   Installed and configured Hadoop Map-Reduce, HDFS and developed multiple Map-Reduce jobs in Java for data cleansing and preprocessing.   Importing and exporting data into HDFS and Hive using Sqoop.   Used UDF's to

implement business logic in Hadoop   Extracted files from Oracle and DB2through Sqoop and placed in HDFS and processed.   Load and transform large sets of structured, semi structured and unstructured data.   Responsible to manage data coming from different sources.   Supported Map-Reduce Programs those are running on the cluster.   Involved in loading data from UNIX file system to HDFS.   Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.   Developed mapping parameters and variables to support SQL override.   Used existing ETL standards to develop these mappings.   Extracted the data from the flat files and other RDBMS databases into staging area and populated onto Data warehouse.   Worked on JVM performance tuning to improve Map-Reduce jobs performance   Environment: Hadoop, MapReduce, HDFS, Hive, Oracle, Java, Struts, Servlets, HTML, XML, SQL, J2EE, JUnit, Tomcat. Java Developer AOP and Spring Security - Chicago, IL July 2011 to August 2013 Reva has a history of success in reselling, architecting, designing, developing, implementing and supporting solutions across multi-platforms. With partnerships across Tier-1 platforms in enterprise content management and mobility, Reva is known for delivering on-time and on-budget.   Responsibilities: Used Spring framework for implementing IOC/JDBC/ORM, AOP and Spring Security to implement business layer.   Developed and Consumed Web services securely using JAX-WS API and tested using SOAP UI.   Extensively used Action, Dispatch Action, Action Forms, Struts Tag libraries, Struts Configuration from Struts.   Extensively used the Hibernate Query Language for data retrieval from the database and process the data in the business methods.   Developed pages using JSP, JSTL, Spring tags, JQuery, Java Script & Used JQuery to make AJAX calls.   Used Jenkins continuous integration tool to do the deployments.   Worked on JDBC for database connections.   Worked on multithreaded middleware using socket programming to introduce whole set of new business rules implementing OOPS design and principles.   Involved in implementing Java multithreading concepts.   Developed several REST web services supporting both XML and JSON to perform task such as demand response management.   Used Servlet, Java and Spring for server side business logic.   Implemented the log functionality by using Log4j and internal logging API's.   Used Junit for server side testing.   Used Maven build tools and SVN for version control.

Developed frontend of application using Bootstrap, AngularJS and Node.JS frameworks. Implemented SOA architecture using Enterprise Service Bus (ESB). Designed front-end, data driven GUI using JSF, HTML4, JavaScript and CSS Used IBM MQ Series as the JMS provider. Responsible for writing SQL Queries and Procedures using DB2. Connection with Oracle, MySQL Database is implemented using Hibernate ORM. Configured hibernate, entities using annotations from scratch. Environment: Core Java, EJB, Hibernate , AWS, JSF, Struts, Spring, JPA, REST, JBoss, DB2, Oracle, XML, JUnit, HTML4, CSS, JavaScript, Apache Tomcat 5x, Log4j Java Developer Oracle Corporation - Santa Clara, CA June 2009 to July 2011 GTL, a Global Group Enterprise, is a leading Infrastructure Services company focused on telecom and Power. In the telecom segment the company provides Network Services to Telecom Operators, OEM's and Tower Companies. In the power sector the company offers EPC services, Distribution Franchisee and Smart Grid solutions to Utilities and distribution companies. Responsibilities: Used Struts framework to generate Forms and actions for validating the user request data. Developed Server side validation checks using Struts validators and Java Script validations. With JSP's and Struts custom tags, developed and implemented validations of data. Developed applications, which access the database with JDBC to execute queries, prepared statements, and procedures. Developed programs to manipulate the data and perform CRUD operations on request to the database. Used message driven beans for asynchronous processing alerts to the customer. Worked on developing Use Cases, Class Diagrams, Sequence diagrams, and Data Models. Developed and Deployed SOAP Based Web Services on Tomcat Server Coding of SQL, PL/SQL, and Views using IBMDB2 for the database. Working on issues while converting JAVA to AJAX. Supported in developing business tier using the stateless session bean. Extensively used JDBC to access the database objects. Using Clear case for source code control and JUNIT testing tool for unit testing. Reviewing the code and perform integrated module testing. Environment: Java , J2EE ,AJAX, Struts , Web Services, SOAP, HTML, XML, JSP, JDBC, ANT, XML, IBM, Tomcat, JUNIT, DB2, Rational Rose, Eclipse Helios, CVS. Education Bachelor's Skills ORACLE (7 years), SQL (9 years), XML (8 years), JAVA (9 years), TOMCAT (6 years) Additional Information Technical

Skills: BigData Technologies Hadoop, MapReduce, HDFS, Hive, Pig, Zookeeper, Sqoop, Oozie, Flume, IMPALA, HBASE, Kafka, Storm.   Big Data Frameworks HDFS, YARN, Spark.   Hadoop Distributions Cloudera (CDH3, CDH4, CDH5), Horton works, Amazon EMR, EC2.   Programming Languages Java, shell scripting, Scala.   Databases RDBMS, MySQL, Oracle, Microsoft SQL Server, Teradata, DB2, PL/SQL, CASSANDRA, MongoDB.   IDE and Tools Eclipse, NetBeans, Tableau. Operating System Windows, Linux/Unix.   Frameworks Spring, Hibernate, JSF, EJB, JMS.   Scripting Languages JSP & Servlets, JavaScript, XML, HTML, Python.   Application Servers Apache Tomcat, Web Sphere, Web logic, JBoss.   Methodologies Agile, SDLC, Waterfall.   Web Services Restful, SOAP.   ETL Tools Talend, Informatica.   Others Solr, elastic search.

Name: Lee Davis

Email: benjaminpeter@example.net

Phone: +1-928-674-6599x64175