

Developer & Admin Developer & Admin Developer & Admin - Hadoop Cleveland, OH ? 5+ years of IT experience that includes Data Analysis and Hadoop Ecosystem ? Experience in components of Hadoop ecosystem including HDFS, MapReduce, Sqoop, Hive, Pig, HBase, Oozie, Flume, Kafka, Zookeeper, and Spark. ? Expertise in Hadoop Ecosystem, HDFS Architecture and Cluster technologies such as YARN ? Management, HDFS, HBase, MapReduce, Hive, Pig, Flume, Oozie, Sqoop, Zookeeper and Ranger ? Experience in developing software solutions to build out capabilities on a Big Data Platform ? Experience in configuring cluster and installing the services, monitoring the cluster by eliminating the compatibility errors ? Experience in different Hadoop distributions like Cloudera and HortonWorks Distributions (HDP) ? Highly capable of processing large sets of Structured, Semi-structured and Unstructured datasets and supporting Big Data applications. ? Experience with NoSQL databases like HBase, MapR and Cassandra as well as other ecosystems like Zookeeper, Oozie, Impala, Storm, Spark- Streaming/SQL, Kafka, Hypertable, Flume ? Expertise in transferring data between a Hadoop ecosystem and structured data storage in a RDBMS such as MYSQL, Oracle, Teradata and DB2 using Sqoop. ? Good experience with Hive concepts like static/dynamic partitioning, bucketing, managed, and external tables, join operations on tables. ? Proficient in building user defined functions (UDFs) in Hive and Pig, to analyze data and extended HiveQL and Pig Latin functionality. ? Experience in implementing unified data ingestion platform using Kafka producers and consumers. ? Proficient with Flume topologies for data ingestion from streaming sources into Hadoop ? Ability to adapt to evolving technology, strong sense of responsibility and accomplishment. ? Has very good development experience with Agile Methodology. ? Strong experience in distinct phases of Software Development Life cycle (SDLC) including Planning, Design, Development and Testing during the development of software applications. ? Excellent leadership, interpersonal, problem solving and time management skills. ? Excellent communication skills both written (documentation) and verbal (presentation). ? Very responsible and good team player. Can work independently with minimal supervision. Work Experience Developer & Admin Hadoop - Cleveland, OH June 2018 to Present Williams generates a large amount of unstructured data through R&D. As the part of a team, worked within processing of

huge unstructured datasets of the chemically formulated paints; analyzing the large data; transforming the data and loading data to the R&D. Responsibilities: ? Understanding the scope of the project and requirements gathering ? Using MapReduce to Index the large amount of data to easily access specific records ? Loading log data into HDFS using Flume ? Installed and configured Hadoop cluster of Hortonworks Data Platform using Ambari Server and maintained it. ? Maintaining the Operations, installation, configuration of clusters with MapR distribution. ? Creating MapReduce jobs to power data for search and aggregation ? Writing Apache PIG scripts to process the HDFS data ? Writing Mapreduce Code for filtering data ? Creating Hive tables to store the processed results in a tabular format ? Developing Sqoop scripts to make the interaction between Pig and Oracle ? Writing script files for processing data and loading to HDFS. ? Working with Sqoop for importing data from Oracle ? Utilizing Apache Hadoop ecosystem tools like HDFS, Hive and Pig for large datasets analysis ? Developing Pig and Hive UDF to analyze the complex data to find specific user behavior ? Using Pig for data cleansing and developed Pig Latin scripts to extract the data from web server output files to load into HDFS ? Developing MapReduce ETL in Java/Pig and data validation using HIVE ? Working on Hive by creating external and internal tables, loading it with data and writing Hive queries. ? Creating HBase tables to store data from various sources ? Developing workflow in Oozie to automate the tasks of loading data into HDFS and pre-processing with Pig and Hive ? Working with various Hadoop file formats, including Text, Sequence File, RCFILE and ORC File. ? Configured Zookeeper for Cluster coordination services Environment: Hadoop, MapReduce, HDFS, Pig, Hive, HBase, Zookeeper, Flume, Kafka, Spark, Elastic Search, Oozie, Java(jdk1.6), Cloudera, Oracle 11g/10g, Windows, UNIX Shell Scripting. Graduate Assistant Long Island University - New York, NY September 2017 to May 2018 Worked on a research project on Social Media Analytics. Performed data analytics through R-programing. Performed various techniques to get the accurate results using Twitter Data. Responsibilities: ? Conducted research in Social Media Analytics ? Involved in collecting, processing, analyzing and reporting social media data of specific research topic ? Worked on Tracking Community Development from Social Media using R ? Explored Social Media Analysis on Community Development Practices based on the

results from R ? Performed data mining, data cleaning & explored data visualization, techniques on a variety of data stored in spreadsheets and text files using R and plotting the same using with R packages ? Hands-on statistical coding using R and Advanced Excel Environment: R-Studio, RPods, Java (v1.8), ShinyApps, Excel Hadoop Developer Eventcy - IN October 2013 to August 2015 India Responsibilities: ? Worked on Hadoop Ecosystem using different big data analytic tools including Hive, Pig. ? Involved in loading data from LINUX file system to HDFS. ? Importing and exporting data into HDFS and Hive using Sqoop. ? Implemented Partitioning, Bucketing in Hive. ? Worked on different file formats (ORCFILE, TEXTFILE) and different Compression Codecs (GZIP, SNAPPY, LZO). ? Worked with multiple Input Formats such as Text File, Key Value, and Sequence File Input Format. ? Experienced in running Hadoop Streaming jobs to process terabytes of json format data. ? Involved in scheduling Oozie workflow engine to run multiple Hive and Pig jobs ? Executed Hive queries on Parquet tables stored in Hive to perform data analysis to meet the business requirements. ? 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. ? Worked collaboratively with different teams to smoothly slide the project to production. Environment: HDFS, Pig, Hive, Sqoop, Shell Scripting, HBase, Zoo Keeper, MySQL. Software Developer Aetins - IN September 2011 to July 2013 India Responsibilities: ? Performed analysis for the client requirements based on detailed design documents ? Developed Use Cases, Class Diagrams, Sequence Diagrams and Data Models using Microsoft Visio ? Developed STRUTS forms and actions for validation of user request data and application functionality ? Developed a WebService using SOAP, WSDL, XML and SoapUI ? Developed JSP with STRUTS custom tags and implemented JavaScript validation of data ? Involved in developing business tier using stateless session bean ? Used JavaScript for the web page validation and Struts Validator for server side validation ? Designing the database and coding of SQL, PL/SQL, Triggers and Views using IBMDB ? Design patterns of Delegates, Data Transfer Objects and Data Access Objects ? Developed Message Driven Beans for asynchronous

processing of alerts ? Used ClearCase for source code control and JUNIT for unit testing ? The networks are simulated in real-time using an ns3 network simulator modified for multithreading across multiple cores, which is implemented on generic Linux machine ? Involved in peer code reviews and performed integration testing of the modules Environment: Struts, JSP with Struts, JDBC, Struts Valuator, SQL, PL/SQL, IBMDB, JUNIT, Java, JSP, Servlets, EJB 2.0, SQL Server, Oracle 9i, JBoss, WebLogic Server 6, JavaScript Education Master in Computer Science in Computer Science Long Island University 2016 to 2018 Bachelors in Production Engineering in Production Engineering Osmania University 2007 to 2011 Skills CASSANDRA, MAPREDUCE, OOZIE, SQOOP, HBASE, KAFKA, DATA VISUALIZATION, FLUME, HADOOP, NOSQL, POWER BI, C++, Hadoop, HBase, Hive, HTML, JAVASCRIPT, MapReduce, PHP, Pig, SQL, XML, Sql Server Certifications/Licenses Driver's License Additional Information SKILLS Languages C, C++, Java (Core), J2EE, Asp.Net, Python, Scala, UNIX Shell Scripting Scripting HTML, PHP, JavaScript, CSS Hadoop Ecosystem MapReduce, HBASE, HIVE, PIG, SQOOP, Zookeeper, OOZIE, Flume, HUE, Kafka, SPARK-SQL Hadoop Distributions Cloudera, Hortonworks, MapR Database MySQL, NoSQL, Oracle DB, Cassandra Virtualization / Cloud Amazon AWS, VMware, Virtualbox Data Visualization Power BI, Tableau IDE Eclipse, Net Beans, VisualStudio Methodologies Agile, SDLC

Name: Randy Crawford

Email: lindayoung@example.org

Phone: 819-666-1536