Hadoop Developer Hadoop Developer Hadoop Developer - Sutter Health Sacramento, CA 7+ years of experience in all phases of Software Application requirement analysis, design, development and maintenance of Hadoop/Big Data application like SPARK, KAFKA, EMR, Hive, Sqoop and applications using java and scala to tailored with industry needs. Hands on experience with Spark Core, Spark SQL, Spark Streaming. Used Spark-SQL to perform transformations and actions on data residing in Hive. Used Kafka & Spark Streaming for real-time processing. Experience with migrating data to and from RDBMS and unstructured sources into HDFS using Sqoop. Good Knowledge in Apache Spark data processing to handle data from RDBMS and streaming sources with Spark streaming. Experience in Data Warehousing and ETL processes and Strong database, SQL, ETL and data analysis skills. Good understanding/knowledge of Hadoop architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and Map Reduce programming paradigm. Have good skills in writing SPARK Jobs in Scala for processing large sets of structured, semi-structured and store them in HDFS. Good Knowledge in Spark SQL queries to load tables into HDFS to run select queries on top. Experience in writing Hive Queries for processing and analyzing large volumes of data. Experience in importing and exporting data using Sqoop from Relational Database Systems to HDFS and vice-versa. Implemented several optimization mechanisms like Combiners, Distributed Cache, Data Compression, and Custom Practitioner to speed up the jobs. Authorized to work in the US for any employer Work Experience Hadoop Developer Sutter Health - Sacramento, CA September 2018 to Present Sutter Health is a not-for-profit health system in Northern California, headquartered in Sacramento. It includes doctors, hospitals and other health care services in more than 100 Northern California cities and towns. Major service lines of Sutter Health-affiliated hospitals include cardiac care, women's and children's services, cancer care, orthopedics and advanced patient safety technology. This project seek to harness these massive, complex buckets of data to obtain more focused knowledge and insights into the world of healthcare. Big Data attempts to make more sense of this information overload and provide better insights from the expanding volumes and sources of data. Responsibilities: Interacting with multiple teams understanding their business requirements for designing flexible and

common component.    Validating the source file for Data Integrity and Data Quality by reading header and trailer information and column validations.    Implemented Spark SQL to access hive tables into spark for faster processing of data.    Used Hive to do transformations, joins, filter and some pre-aggregations before storing the data.    Validating and visualizing the data in Tableau. Using hive extensively to create a views for the feature data.    Working with platform and Hadoop teams closely for the needs of the team.    Using Kafka for Data ingestion for different data sets. Experienced in importing and exporting data into HDFS and assisted in exporting analyzed data to RDBMS using SQOOP.    Developed sqoop jobs to import the data from RDBMS and file servers into Hadoop.    Environment: Hadoop, Cloudera, Amazon AWS, HDFS, Hive, Impala, Spark, Kafka, s3, Sqoop. Spark/Hadoop Developer Nike - Portland, OR November 2017 to August 2018 Nike, Inc. is an American multinational corporation that is engaged in the design, development, manufacturing, and worldwide marketing and sales of footwear, apparel, equipment, accessories, and services. The company is headquartered near Beaverton, Oregon, in the Portland metropolitan area. Nike Inc project provides up-sell and cross-sell recommendations for -customers enabling the increase in online purchases by recommending relevant products and promotions in real time. Products recommend based on what other similar customers have bought - providing upsell, cross-sell or "next best offer" opportunities. Responsibilities:    Interacting with multiple teams understanding their business requirements for designing flexible and common component.    Validating the source file for Data Integrity and Data Quality by reading header and trailer information and column validations.    Used Spark SQL for creating data frames and performed transformations on data frames like adding schema manually, casting, joining data frames before storing them. Implemented Spark SQL to access hive tables into spark for faster processing of data.    Worked on Spark streaming using Apache Kafka for real time data processing.    Experience in creating Kafka producer and Kafka consumer for Spark streaming.    Used Hive to do transformations, joins, filter and some pre-aggregations before storing the data into HDFS.    Worked on three layers for storing data such as raw layer, intermediate layer and publish layer.    Creating external hive tables to store and queries the data which is loaded.    Optimizations techniques include partitioning, bucketing.

Using Avro file format compressed with Snappy in intermediate tables for faster processing of data. Used parquet file format for published tables and created views on the tables. Created sentry policy files to provide access to the required databases and tables to view from impala to the business users in the dev, uat and prod environment. Automated the jobs with Oozie and scheduled them with Autosys. Experience in AWS to spin up the EMR cluster to process the huge data which is stored in S3 and push it to HDFS. Participated in evaluation and selection of new technologies to support system efficiency. Participated in development and execution of system and disaster recovery processes. Environment: Hadoop, Cloudera, Amazon AWS, HDFS, Hive, Impala, Spark, Kafka, s3, Sqoop, Java, Scala, Eclipse, Tableau and Maven, SBT Spark/Hadoop Developer Capital one - Richmond, VA September 2016 to October 2017 Retail Enterprise Credit Risk application calculates Bank's retail data such as credit cards, auto, student and home loans for risk domains including Enterprise Capital Management. Data comes from different System of records mainly from Teradata. This data will undergo several cleansing and value added processing and then finally views will be created on this data as Hadoop warehouse. This data will be consumed by downstream like ECM for analyzing and generating reports. Responsibilities: Interacting with multiple teams understanding their business requirements for designing flexible and common component. Validating the source file for Data Integrity and Data Quality by reading header and trailer information and column validations. Used Spark SQL for creating data frames and performed transformations on data frames like adding schema manually, casting, joining data frames before storing them. Implemented Spark SQL to access hive tables into spark for faster processing of data. Worked on Spark streaming using Apache Kafka for real time data processing. Experience in creating Kafka producer and Kafka consumer for Spark streaming. Used Hive to do transformations, joins, filter and some pre-aggregations before storing the data into HDFS. Used Sqoop for importing and exporting data from Netezza, Teradata into HDFS and Hive. Worked on three layers for storing data such as raw layer, intermediate layer and publish layer. Creating external hive tables to store and queries the data which is loaded. Optimizations techniques include partitioning, bucketing. Using Avro file format compressed with Snappy in

intermediate tables for faster processing of data.    Used parquet file format for published tables and created views on the tables.    Created sentry policy files to provide access to the required databases and tables to view from impala to the business users in the dev, uat and prod environment.    Automated the jobs with Oozie and scheduled them with Autosys.    Experience in AWS to spin up the EMR cluster to process the huge data which is stored in S3 and push it to HDFS.    Participated in evaluation and selection of new technologies to support system efficiency. Participated in development and execution of system and disaster recovery processes. Environment: Hadoop, Cloudera, Amazon AWS, HDFS, Hive, Impala, Spark, Kafka, s3, Sqoop, Java, Scala, Eclipse, Tableau and Maven, SBT. Java Developer Keane India LTD - Bengaluru, Karnataka April 2014 to May 2016 I worked as an associate java developer on ATS (Asset Tracking Software) which manages and tracks the assets of a company. It manages separate login functionality for asset registry and transactions including orders, invoices and returns. It also allows the vendor to closely monitor the inventory and sales.  Responsibilities:    Involved in the complete SDLC software development life cycle of the application from requirement gathering and analysis to testing and maintenance.    Developed the modules based on MVC Architecture.    Developed UI using JavaScript, JSP, HTML and CSS for interactive cross browser functionality and complex user interface.    Created business logic using servlets and session beans and deployed them on Apache Tomcat server.    Created complex SQL Queries, PL/SQL Stored procedures and functions for back end.    Prepared the functional, design and test case specifications.    Performed unit testing, system testing and integration testing.    Developed unit test cases. Used JUnit for unit testing of the application.    Provided Technical support for production environments resolving the issues, analyzing the defects, providing and implementing the solution defects. Resolved more priority defects as per the schedule.    Environment: Java, JSP, Servlets, Apache Tomcat, Oracle, SQL Java Developer Senmex - Hyderabad, Telangana February 2011 to March 2014 Developed with a flexible design which provides the platform to offer on-line integrated financial services to the bank customers. Using this web based application customers can conduct activities like on-line retail banking, Secure messaging, Credit Card Payments, Shopping, Corporate banking and even

payment to third party individual sand institutes. Responsibilities: Involved in design, development and analysis documents in sharing with Clients. Developed web pages using Struts framework, JSP, XML, JavaScript, Hibernate, springs, Html/ DHTML and CSS, configure struts application, use tag library. Developed Application using Spring and Hibernate, Spring batch, Web Services like Soap and restful Web services. Used Spring Framework at Business Tier and also spring's Bean Factory for initializing services. Used AJAX, JavaScript to create interactive user interface. Implemented client side validations using JavaScript & server side validations. Developed Single Page application using angular JS & backbone JS. Implemented Hibernate to persist the data into Database and wrote HQL based queries to implement CRUD operations on the data. Developed an API to write XML documents from a database. Utilized XML and XSL Transformation for dynamic web-content and database connectivity. Database modeling, administration and development using SQL and PL/SQL in Oracle 11g. Coded different deployment descriptors using XML. Generated Jar files are deployed on Apache Tomcat Server. Involved in the development of presentation layer and GUI framework in JSP. Client-Side validations were done using JavaScript. Involved in configuring and deploying the application using WebSphere. Involved in code reviews and mentored the team in resolving issues. Undertook the Integration and testing of the various parts of the application. Developed automated Build files using ANT. Used Subversion for version control and log4j for logging errors. Code Walkthrough, Test cases and Test Plans Environment: HTML5, JSP, Servlets, JDBC, JavaScript, Json, Spring, SQL, Oracle 11g, Tomcat, Eclipse IDE, XML, XSL, ANT, Tomcat 5. Education Bachelor of Technology in Electronics and Communication Engineering in Electronics and Communication Engineering Anna University - Chennai, Tamil Nadu May 2009 Skills Hdfs, Impala, Sqoop, Kafka, Hadoop Additional Information Technical Skills: Big Data/Hadoop HDFS, Hive, Sqoop, Impala, Kafka, Map Reduce, Cloudera, Amazon EMR. Spark Components Spark Core, Spark SQL, Spark Streaming. Programming Languages SQL, Scala and Java Databases MySQL, Hive-QL, RDBMS. Cloud Amazon EMR, EC2, S3. Operating Systems Windows, Unix, Red Hat Linux.

Name: Deanna Williamson

Email: paulkaren@example.net

Phone: 001-862-916-9200x387