

Big Data Developer Big Data Developer Big Data Developer - Molina HealthCare Over 8 years of experience in application development and design using Hadoop ecosystem tools and Java /J2EE Technologies. Developed and built frameworks that integrate big data and advanced analytics to make business decisions. Extensive experience in installing, configuring and using ecosystem components like Hadoop Map reduce, HDFS, Hive, Pig, Sqoop and Spark and good knowledge on Kafka Preprocessed and cleansed big data for better analysis. Certified Cloudera Spark and Hadoop Developer Experience in Cloudera distributions (CDH) and Hortonworks Data Platform (HDP) Created various use cases using massive public big data sets. Ran various performance tests for verifying the efficacy of Map Reduce, PIG and HIVE Migrated to Azure cloud and created end-to-end architecture for running in Cloud. Have experience on ADF, ADLS, Blob Storage, HD Insights, Ranger, S3, IR, IoT Hub, Stream Analytics, etc. Good knowledge of Amazon Web Services (AWS) components like EC2, EMR, S3, CloudWatch etc. Proficiency in developing applications using Java, JSP, JavaScript, JDBC, Selenium, Oracle ADF, Python Strong coding and debugging skills in Java Platform Experience in shipping enterprise products, web/mobile UI applications to a large customer base Experienced in Full Life Cycle development of software products Good at Servlets, JSPs and MVC framework Have excellent analytical and problem-solving skills and ability to learn new technologies quickly

**Work Experience**

**Big Data Developer Molina HealthCare - Long Beach, CA November 2018 to Present**

Molina Healthcare is a health insurance company. It provides quality, affordable health care to individuals and families covered by government programs, such as Medicaid and Medicare for over 30 years. In Molina, as part of Enterprise Information Management (EIM) team, we are responsible for developing the semantic layers for claims and curating, transforming the data for the business users to build dashboards and essentially, the "Single source of truth" for Molina. I was part of the Claims and Pharmacy pipelines.

**Responsibilities:**

- Worked on a live 30 node (Prod) and 6 node (UAT) big data production cluster CDH 5.13.3
- Developed and maintained the complex Claims Semantic Pipeline for weekly full load and incremental loads
- Talend ETL jobs run to clean, validate and load the data into inbound area.
- Weekly full load of claims is validated against Netezza and verified for

any discrepancies      Resolved the state issue (Universal and Medicare state) in the data set for reference for all the pipelines      Developed the aggregated datasets and lookup columns from Claims dataset and all reference tables      Integrated SIU pipeline into the existing Claims pipeline and retired the SIU pipeline      Used windowing techniques and UDFs in SparkSQL in scala. Implemented Spark using Scala and SparkSQL for faster processing of data.      Developed and incorporated the enhancements into the existing claims pipeline      Monitored and maintained weekly talend job and resolve any failures to meet the SLAs      Converted existing SQL logic to SparkSQL for Pharmacy pipeline and optimize it      Improved the performance of Provider datasets and incorporated all the provider data into claims      Worked with PARQUET file formats using SNAPPY compression to fasten network transfer of big data      Created Hive tables and views using Impala. Implemented partitioning, bucketing in Hive for better organization of data      Build Power BI dashboards to validate the data against Netezza      Currently in the process of automating the check before the start of pipeline to validate the L0 data      Collaborated with Data Management team on the business requirements and retirement of Netezza      Follow Agile Scrum methodology in JIRA during project      Gained very good business knowledge on claim processing

**Big Data Developer**  
British Petroleum - Houston, TX September 2017 to October 2018  
British Petroleum is a British multinational oil and gas company. It is vertically integrated company operating in all areas of oil and gas industry, including exploration and production, refining, distribution and marketing, petrochemicals, power generation and trading. In BP, as part of the Upstream Data Lake team, we are committed to migrating the on-prem data into Azure cloud leveraging the services such as Azure Data Factory, Azure Data Lake Store (ADLS), Blob Storage and HDInsight cluster. Analyze the real-time data using IoT Hub and Stream Analytics. Responsibilities:      Involved in the complete SDLC of Big data project that includes requirement analysis, design, coding, testing and production.

Worked on a live 24 nodes and 4 nodes (Test) big data cluster of type Hadoop 3.6 on Linux. Experience working on both Non-domain and domain joined clusters.      Worked with highly unstructured, structured and semi structured data of 30 TB in size (90 TB with replication factor of 3)

Ingested structured data from TIBCO Composite Data Virtualization tool into ADLS using Sqoop

Created Shell scripts to automate the Sqoop jobs. Developed Ambari workflows for scheduling and orchestrating the ETL process. Worked with ORC file formats using ZLIB compression to fasten network transfer of big data. Ingested structured big data from Teradata, Oracle, Netezza, Postgres, SQLServer into ADLS using Azure Data Factory (ADF). Created pipelines in ADF to create cluster, ingest, create hive tables, enable daily triggers. Involved in converting Hive queries into Spark transformations using Spark Structured API. Used PySpark (Python) for analyzing the data in Non-domain joined Spark 2.3 cluster. Scripted Python Code to transfer data from Hive tables into Data Science Sandbox using SFTP. Used Azure Data Catalog to create data catalog and data lineage. Very good experience in monitoring and managing the Hadoop cluster using Ambari. Created dashboards in Power BI based on the Incident record data to generate metrics and Hive tables using ODBC connection. Gained very good business knowledge on oil and gas industry, well pad, weather, mud pressure and exploration analysis. Collaborated with Digital Security, Data Scientists, Palantir and Catalog team to ensure data quality and availability. Follow Agile Scrum methodology in Visual Studio Team Services during the course of project.

Hadoop Developer ServiceNow, Inc - Orlando, FL January 2017 to August 2017 ServiceNow is a company that provides service management software as a service. It specializes in IT services management (ITSM), IT operations management (ITOM) and IT business management (ITBM). In ServiceNow, I was part of the Performance analytics team. There I had performed Attribute Value Mapping, to identify the tables for analysis.

Responsibilities:

- Worked on a live 80 nodes Hadoop cluster running CDH5.10
- Worked with structured and semi structured data of 150 TB in size (450 TB with replication factor of 3)
- Created and worked Sqoop jobs with incremental load to populate Hive External tables.
- Extensive experience in writing Pig scripts to transform raw data from several data sources into forming baseline data.
- Developed Hive queries and UDFs to analyze/transform the data in HDFS.
- Very good understanding of Partitions, bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance
- Established custom MapReduce programs to analyze data and used Pig Latin to clean unwanted data
- Used Pattern matching algorithms in PIG to recognize the fraudulent customer across different sources and built

risk profiles for each customer and stored the result data into HDFS. Used Oozie to orchestrate the MapReduce jobs and worked with HCatalog to open up access to Hive's Metastore. Software Development Engineer Kony India Pvt.Ltd - Hyderabad, Telangana June 2013 to May 2016 Kony is an app development software company that offers tools for all phases of mobile app development, from design to management. Kony's cross-platform, low-code solution empowers businesses to develop and manage their own apps to better engage with their customers, partners and employees by rapidly delivering apps across the broadest array of devices and systems, today and in the future.

In Kony, I worked on a pilot project for Citi where huge data sets are analyzed to help the organization get a competitive advantage by finding out the customer spending and payment trends. We collected the data from all the customer apps and logs and performed the analysis by using various traits such as demographics, credit and debit patterns, etc. I worked on Kony Visualizer and Kony Studio. Both the products provide an integrated development environment for building native, hybrid apps across phones, tablets and desktops - all from a single code base - Write Once and Run Everywhere. I shipped Kony Visualizer versions of 2.0, 2.5 and 3.0. Responsibilities: Project #: Hadoop for Citi. Worked on 10 nodes Hadoop Cluster. Worked on semi structured and structured data of 15TB in size (45TB with replication factor of 3). Loaded data from disparate data sets using Sqoop and flume. Used sqoop to import/export data between RDBMS and hive tables. Imported logs from web servers with Flume to ingest the data into HDFS. Created Sqoop jobs with incremental load to populate Hive External tables. Have a very good understanding of Partitions, bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance. Writing Pig Latin Scripts to perform transformations as per the use case requirement. Worked with different file formats and compression techniques. Environment: Cloudera Enterprise, Hadoop, MapReduce, Pig, Hive, Avro, Sqoop, HBase. Project #: Visualizer. Automate the scenarios commonly used while developing an application using Visualizer (Using Java, Selenium, JavaScript). UI validation of the app creation scenarios. Created Action editor for app development using Kony APIs. Create Services and map them to the widgets. Code reviews of every commit in the areas owned by me. Create apps specific to iOS, Android and desktop for

both mobile and tablets using Visualizer. Create applications in different areas to find any pain points for customers. Create a Proof of concept (POC) and plan for automation framework to integrate Photoshop with Visualizer automation. Guided and trained the Softserve team (Russia) about Kony Visualizer and its automation framework. Proposed the UX issues with Photoshop integration. Dealt with the customer tickets and was active on the Kony forum to resolve the problems. Environment: Java, JavaScript, Selenium, GIT, Eclipse, Maven, Photoshop, UNIX, Node JS, XML, TestNG, JSON Member Technical Staff Oracle India Pvt.Ltd - Bengaluru, Karnataka July 2008 to April 2011 Oracle Corporation is an multinational computer technology corporation that specializes primarily in developing and marketing database software and technology, cloud engineered systems and enterprise software products. In Oracle, I was part of the Enterprise Content Management team (ECM) as a developer. ECM helps in organizing and storing the documents and it offers solutions throughout the lifecycle of the content providing strategies, methods and tools. I worked on the creation of User Interface for Grid Control and Oracle Deployment Server. Grid Control offers a single-window solution that allows end-to-end monitoring, administration of the complete Oracle IT Infrastructure from a single console. Responsibilities: Created functional and design specification documents. Analyzed on how to display the data/metrics collected on Enterprise Management (EM) and develop the relevant pages. Worked on User-Interface using JSPs and Servlets for the Enterprise Manager framework. Discover all the Universal Content Management servers installed on the content server and identify their statuses. Extracted the configuration details of the server. Integrate the targets (SOA, WebLogic, WebCenter) to the EM Tree. Create Dynamic Monitoring Services (DMS) messages for the Content Management. Add the DMS instrumentation to the Content Server code to extract the metrics and validating and testing them. Identified the cached queries, active databases, documents waiting, and number of service requests in the Content Server. Analyzed the system performance and monitor system status. Used Oracle Application Development Framework (ADF) for end-to-end Java-based application development. Resolve the issues on the server based on the priority. Environment: Java, J2EE(Servlets), OOPS concepts, Oracle DB, JDBC Internship

Oracle India Pvt.Ltd August 2007 to January 2008 I worked on the Client-Side Java implementation of Diameter Base Protocol. Diameter is an AAA protocol for computer networks. It is used for online and offline charging. It was done using Socket Programming and Agile Scrum Methodology.

Responsibilities: Prepare Requirement, Functional and Design Specification documents. Worked on Oracle JDeveloper, which is a free integrated development environment. Dynamic peer discovery has to be done both statically and dynamically (Using SLP and NAPTR) Created Realm and Peer routing tables. Invoked TCP Connection to send and receive data over it. Test each method with JUnit. Used EMMA Code Coverage to help improve the coverage of the Project.

Implemented Failover and Failback procedures Environment: Java, J2EE (Socket Programming), Design Patterns, Seagull Traffic generator Education MBA in (Finance and Operations), Post Graduate Institute of Public Enterprise 2011 to 2013 Bachelor of Engineering (Honors) in Computer Science BITS-Pilani 2004 to 2008 Skills Hdfs, Mapreduce, Sqoop, Flume, Hadoop, Jms, Tibco, Visual studio, Design patterns, Git, Hadoop, Hive, Json, Mapreduce, Object oriented, Pig, Xml, Oracle, Sql, Eclipse Additional Information Technical Skills: Learning: Can rapidly adapt to new environments and designs. Apache Hadoop: HDFS, Hive, Pig, MapReduce, Flume, Sqoop and Spark Cloud: HDInsight, ADLS, ADF, S3, EMR, EC2, NACL, Security groups Programming Language & Scripts: Java, J2EE, UNIX, Java Script, SQL, UML, XML, CSS, JSON Enterprise Java: JSP, Servlets, JSF, EJB, JMS, Socket Programming, Java Beans Software Design: Design Patterns, Data Structures, Object Oriented design Tools & Framework: TIBCO Composite, JSF, Spring, Web Services, Selenium, JUnit, Maven, Ant Web Servers: Weblogic, Web Sphere, Tomcat, Oracle OC4J, Oracle Weblogic Server IDE: Eclipse, Visual Studio, XCode, GIT

Name: Daniel Hill

Email: robertherman@example.com

Phone: 4355020012