Summary

Principal Big Data Software Engineer with more than 13 years of Experience in various Oracle & Big Data Technologies.  
  
5+ years of Experience in implementing Complete Big Data solutions, including data acquisition, storage, transformation, analytics using Big Data technology that includes Hadoop, Spark, Hive, Solr, Oozie, Sqoop, Scala, Cassandra, Flume.  
  
Extensive experience of transforming the business requirements into data-driven solutions and help customers to make the decisions based on data(Big Data-Driven Decision Making)  
  
Working experience of various industries like Health Care, Retail, Banking and Manufacturing.  
  
Part of core team to successfully build the Big Data Lake distributed platform in Humana for Advanced Analytics & Machine Learning and automated the data ingestion into Hadoop from various data sources using Hadoop, Spark, Sqoop, Ozzie, Scala.  
  
Working experience of 15 PB Walmart data on 900 Nodes Hadoop cluster.  
  
Expertise in Cloudera, Big Insights and Pivotal HD distribution and Strong knowledge of NoSQL like Cassandra.  
  
Building complete data ingestion (ETL) pipeline from traditional databases and file systems into Hadoop using Sqoop, Flume, Spark, Hive, Oozie and SFTP/SCP.  
  
Extensive experience in Oracle Apps, Custom ERP solutions, RMS and Pricing Applications.  
  
Strong knowledge of Oracle Database, SQL, PLSQL, RDBMS, Oracle forms & Reports, SQL performance tuning and Advanced PLSQL concepts.  
  
Expertise in Clean, Transform, Impute and Aggregate the data in the way Data Scientist wants to feed into various Machine Learning models.  
  
~Technical Competencies  
• Big Data, Hadoop, Spark  
• Spark, Spark SQL, Hive, Oozie, Sqoop, Flume, Solr, Cassandra  
• Oracle Database, Oracle BI  
• Oracle SQL and PL/SQL  
• Oracle Forms & Reports  
• Shell Scripting, Scala, Tableau

• Transform/load data from different channels (Walmart/walmart.com, Sams/sams.com, voodoo.com) into Hadoop.  
• Analyzed customer buying patterns and store/product sales patterns based on various attributes  
• Analyzed several TB’s of data on a 15 PB data and 900-node Hadoop clusters.  
• Experience in implementing Hadoop archival system to import/export the different data sources like RDBMS data, web/server logs, live sales events and external sources from/to Hadoop systems.  
• Involved in creating Hive tables, loading and analysing data using hive queries.  
• Processing of Nielsen data into HDFS and optimized the data using hive for reporting and visualization.  
• Prepare the end to end solution and perform the required POC’s.  
• Processed over 40 TB’s of Walmart/Sams Customer scan/visits data for executive reporting.  
• Worked on performance tuning of Hive using Partitioning/Bucketing, Map & Skew joins.  
• Experience in importing/exporting of data to/from Teradata/Greenplum/Oracle/DB2/SAP HANA using Sqoop and native connectors and analysed the data using Hive/Spark/pig.  
• Involved in loading and transforming large sets of Structured, Semi-Structured and Unstructured data and analysed them by running Hive queries.  
• Used distcp to move the data across Hadoop cluster for data processing and testing.

• Part of the core team to successfully build the Big Data Lake distributed platform for Machine Learning & analytics and automated the data ingestion into Hadoop from various data sources using ETL pipelines.  
• Working with Data Engineer and Scientist to provide a scalable and efficient enterprise analytics platform and build the Data Lake for consumer analytics team, which enables the analytical and machine learning.  
• Analysed, cleaned and Stitched more than 60 interactions (Calls, Claims, Visits, Rx, Diagnosis, G&A, Web, Chats, EMR, Auth, Comm.,) between Member, provider, and Humana for Consumer 360 and Longitudinal Patient View(LPV) analytics project using Hadoop, Spark, Solr.  
• Working Experience of Processing over 1.5 PB data on a 56 Nodes Hadoop cluster  
• Waterfall analysis for Humana members to identify the Retail Segment call volume to reduce the transfers  
• Worked along with data scientist to mine the frequent pattern and FP-growth analysis using Spark to predict the future events/patterns/sequence based on member time series event data.  
• Spark implementation of Text Mining and Topic Modeling to analyze the Top trending topics for each Member from Call notes.  
• ETL ingestion of data from Oracle, Netezza, DB2, SQL Server, MySQL, SAS, network drive and FTP location into Hadoop Data Lake.  
• Performance tuning of Hive using Partition, Bucketing, Skew joins and optimized the queries using various MapReduce parameters.  
• Transforming large sets of Structured, Semi/Un-Structured, XML/JSON/Flat data sets and analyzed using Spark/Hive.  
• Experience in various Hadoop file formats like Parquet, Avro, Sequence, ORC and compression's like GZip, Snappy.  
• Optimization of Call Center and Routing Rules to identify the call pattern and call volume to reduce the operational cost.  
• Segmenting the Humana Members using various Health conditions, clinical utilization and behaviours.  
• Migrating the existing SAS project into Spark and building the SAS to Hadoop connectors.