**DHEERAJ REDDY**

**Data Engineer**

**pdheerajkreddy@gmail.com**

**+1(312) 620-4638**

**PROFESSIONAL SKILLS:**

A dynamic professional with 7 years of diversified experience in the field of Information Technology with an emphasis on Big Data/Hadoop Eco System, ETL, SQL/NO-SQL databases, Spark/Scala, Python technologies and tools using industry accepted methodologies and procedures.

* **Experienced in Healthcare industry in various Domains such as Membership, Providers, Clinical, Claims etc.**
* Experienced in Healthcare Effectiveness Data and Information Set **(HEDIS), Medicare, Medicaid** and **Retail** data.
* Experienced in **Financial domains**.
* Knowledge of management consulting methodology, tools and techniques (includes interviewing for information, issue-based consulting, report writing, project planning, scope management, cost management).
* Experienced in Data vault modeling methodology.
* Experienced in Ab initio for ETL.
* Work to tight deadlines while delivering high quality work product.
* Works closely with project manager to develop work plan for Data lake projects and keep the manager aware of any issues
* Hands on Expertise on Analyzing insights of Various Domains, formats and sources of data.
* Analyzed the data is various file formats like fixed width, xml, text, json etc.
* Good understanding of CDC data, overlapping records, Manufacturing History for the data.
* Experience in Changed Data Capture (CDC) process for Live and Streaming data by using IBM Infosphere Data Replicator (IIDR) and Kafka.
* Good understanding and knowledge of NOSQL databases like HBase.
* Experience in Parsing the XML, Fixed width and Cobol files in Spark using Scala and storing in HDFS and Hive tables.
* Used Tableau for Reporting the data trends.
* **Data Modelling/Profiling**: Experience in Profiling, Modelling and Mapping Various data sets according to Business requirements.
* Created various Data Models and Mappings for various domains of data in healthcare and Financial industries.
* Expertise on SQL and HQL for Data Analysis and worked on Oracle, Db2, Teradata, Hive and Impala.
* Good Knowledge on **Data Science** tools like Juypter notebook and Zeppelin notebook.
* Experience on Supervised and Unsupervised algorithms like "Regression", "Classification" “Clustering” and “Non-clustering” problems.
* Good Knowledge on Linear Algebra, Probability and Statistics.
* **Bigdata/ETL Developer:** Extensively worked on Hadoop tools which include Sql, Pig, Hive, Sqoop, Spark, Scala, Python, HBase, Zena and Kafka.
* Good Knowledge on Data Science tools like Juypter notebook and Zeppelin notebook.
* Experience with AWS cloud services.
* Experience in creating tables on top of data on AWS S3 obtained from different data sources and providing them to analytics team.
* **Hadoop Distributions:** Worked with Apache Hadoop along enterprise version of Hortonworks and cloud platform Azure. Good Knowledge on Cloudera distribution.
* **Methodologies:** Handful experience in working with different software methodologies like Agile and Waterfall.
* Extensively worked in Agile Environment like Jira, Agile/Scrum methodology, Conducted Sprint planning, Sprint Refinement, Backlog grooming, User stories, story points etc.
* **CI/CD:** Used GitHub, Jenkins and UCD for maintaining code versioning and Deployments.
* **ServiceNow:** Used Service now for Creating Change orders, Change Tasks and Incidents For Production Deployments and worked closely with Release management team for Successful Deployment.

**TECHNICAL SKILLS:**

**Programming Languages**: C, SQL, Python, Scala, Shell Scripting.

**Big Data Technologies**: AWS, Apache Hadoop, HDFS, MapReduce, Spark, HIVE, PIG, Sqoop,

 Python, Kafka, Zena and Abi initio.

**Data Science tools** Juypter notebook and Zeppelin notebook.

**Database**: Oracle, MySQL, DB2, Teradata.

**CI/CD** GitHub, Jenkins, UCD

**No SQL Database:** HBase, Cassandra

**Hadoop Distributions:**  Hortonworks, Cloudera.

**Cloud Platform** Azure.

**Web Technologies:** ServiceNow, HTML, DHTML, XML, WSDL, SOAP

**Operating Systems:** UNIX, LINUX, Mac OS and Windows Variants.

**PROFESSIONAL EXPERIENCE:**

**SunTrust Bank (TRUIST), Atlanta Mar 2020 – Present**

**Sr. Bigdata Developer**

* Captured Business Requirements, Application requirements and Designed Architecture for Data transfers between two companies SunTrust and BB&T.
* Built the various Data pipelines by leveraging Hadoop, Spark/Scala, Hive, Kafka technologies to ingest, Integrate and transform the data.
* Used HBase for storing the Audit logs.
* Used Oozie and CA7 for job orchestration and Scheduling.
* Experienced in Ab Initio for ETL and ingested to HDFS
* Migrated the jobs to CDH6 environment.
* Worked on CI/CD tools like GitHub, Gitlab, Jenkins and Harvest.
* Involved in Unit testing and QA for the jobs and deployed to production.
* Experience in Production support activities during Cluster outages, Server down issues, Installing new Software’s or Hardware’s etc.

Environment: CDH6, Hadoop, Hive, Kafka, PDOA, DB2, MYSQL, Spark, Scala, Python, Sqoop, Ab initio, HBase and Shell Scripting.

**Health Care Service Corporation (HCSC), Chicago Jan 2018 – Feb 2020**

**Sr. Bigdata Consultant/Data Analyst**

* **Experienced in various Domains such as Membership, Providers, Clinical, Claims etc.**

**Responsibilities:**

• Capture business and application requirements and Product requirements.

* Document functional requirements and operations (calculations, technical details, data manipulation and processing etc.).
* Data Analysis, Modelling and Mapping the tables from Source to Target based on consumption or Business requirements.
* Created various audit scripts for Data Analysis to capture record counts, corrupted records, missing data etc.
* Used Tableau to report the Data analysis for Business teams on weakly basis.
* Analyzed the various data issues in production and fixed them on priority basis to reduce the impact on the team.
* Expertise in processing data from HDFS, HBase using Hive, Sqoop and Pig components.
* Transformed Text, Xml, Sequence, Fixed width, json, orc and various kinds of Data.
* Used Hive tables for loading the data in Text, json and Orc file formats for querying and creating the extract for Business and consumption requirements.
* Developed pipeline for various Data sets for ETL purpose.
* Developed Sqoop and Kafka jobs for loading the data from source databases to HDFS.
* Creation of tables in Hive Databases and working on it to make transformations.
* Developing Shell Scripts for Populating the data into Hive tables.
* Automate Scheduling using Zena Schedular.
* Extract data from DB2 with IBM IIDR using Sqoop and Kafka to map the data into target business warehouse.
* Transforming the data based on Mapping documents.
* Developed spark programs using Scala, involved in creating Spark SQL Queries and Developed Zena workflow for spark jobs.
* Got pleasant experience with NOSQL databases like HBase.
* Implemented Spark using Scala, Java and utilizing Data frames and Spark SQL API for faster processing of data.
* Developed Spark code and Spark-SQL/Streaming for faster testing and processing of data.

Environment: HDP 2.3, Hadoop, Hive, Kafka, Oracle, DB2, IIDR, Spark, Scala, Python, Pig, Sqoop, Zena, HBase and Shell Scripting.

**LearnBeyond Consulting, New Jersey Jan 2017– Dec 2017**

 **Hadoop Developer**

**Responsibilities:**

* Responsible to manage data ingested from different sources.
* Analyzed different formats of Data by parsing the files and storing into Hive tables.
* Created Mapping documents for various tables which includes joins, Transformations etc.
* Created Data profiling for banking data.
* End-to-end performance tuning of Hadoop clusters and Hadoop MapReduce routines against very large data sets.
* Worked on AWS services like S3, EC2, EMR, RDS, Lambda, Redshift etc.
* Implemented AWS solutions using IAM, Elastic Compute cloud (EC2), Lambdas, EMR, Simple Storage Service (S3).
* Developed Hive queries for the data analysis.
* Used Spark/Scala for processing high volume data sets.
* Used Kafka and spark streaming for streaming data.
* Worked on creating Mapping documents by joining the tables and other functionalities for populating the Business required data in the tables.
* Developed Scripts and Batch Job to schedule various Hadoop Programs.
* Created Hive tables and working on them using Hive QL.
* Imported data using Sqoop to load data from MySQL to HDFS on regular basis.
* Involved in creation of Test Cases for JUnit Testing and carried out Unit testing.

**Environment:** AWS, HDP 2.3, MapReduce, HDFS, Hive, Spark, Python, Sqoop, Sql, DB2.

**Alcorn State University Jan 2016 – Dec 2016**

**Graduate Assistant**

**Responsibilities:**

* Involved in maintaining databases of students
* Retrieving student’s data to ensure the course intakes for that semester and scholarship details
* Created various tables in SQL to load the data.
* Involved in administration tasks such as scheduling professor’s classes and maintaining department needs
* Creating presentations, documentation of **Big Data Technologies** for professor related subjects.

**Environment:** Hadoop,Spark, UNIX Scripting, Oracle, Sql, DBMS.

**Superior Software and Technology Solutions, Hyderabad - India Jan 2015 – Dec 2015**

**SQL Developer**

**Responsibilities:**

* Analyzed the data on SQL server by Querying and joining the tables.
* Handled moving of data from various data sources and perform transformations using Pig.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Responsible for troubleshooting issues in the execution of MapReduce jobs by inspecting and reviewing log files.
* Expertise in developing web services like SOAP and REST.
* Used Sqoop to transfer data between RDBMS and Hadoop Distributed File System.
* Used NoSQL Databases HBASE.
* Assisted in troubleshooting and optimization of MapReduce jobs and Pig Latin scripts.
* Effectively used Oozie to develop automatic workflows of Sqoop, MapReduce and Hive jobs.
* Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting.
* Optimized Map/Reduce Jobs to use HDFS efficiently by using various compression mechanisms.
* Implementing various advanced join operations using Pig Latin.
* Worked on Sequence files, ORC files, Map side joins, bucketing, partitioning for Hive performance enhancement and storage improvement.
* Involved in collecting and aggregating large amounts of streaming data into HDFS using Flume.
* Continuous monitoring and managing the Hadoop cluster through Ambari.

Environment: SQL, Shell Scripting, HADOOP, PIG, SQL, ORACLE, DB2, SQOOP, Cassandra, LINUX.

**EDUCATION:**

**Master of Computer and Information Sciences Jan 2016 - May 2017**

Alcorn State University