

Contact

www.linkedin.com/in/sandeep-reddy-k-702330147 (LinkedIn)

Top Skills

Hadoop
Apache Spark
hive

Certifications

Programming for Everybody (Getting Started with Python)
Python Data Structures
Cloudera Certified Developer for Apache Hadoop (CCDH)
DataBricks Certified Developer- Apache Spark 2.X for Scala
Neo4j Certified Professional

Sandeep Reddy K.

Big Data Engineer at Microsoft
Redmond, Washington

Summary

Experience in Batch and Streaming Analytics using Hadoop environment include spark, Map Reduce, HDFS, Hive, Pig, Zookeeper, and Sqoop.

In depth understanding of Hadoop Architecture and its various components such as Resource Manager, Node Manager, Applications Master, Name Node, Data Node concepts.

Experience in importing and exporting data using Sqoop from Relational Database Systems to HDFS and vice-versa. Hands on experience with SQL Server 2012 Always on High Availability Groups. Extensive experience in developing Stored Procedures, Views and Complex Queries on SQL Server 2014, 2012, 2008. Job workflow scheduling and monitoring using tools like Azure Data Factory.

Good experience in Cloudera, Hortonworks & Apache Hadoop distributions.

Worked with relational database systems (RDBMS) such as MySQL, MSSQL, Oracle.

Used Shell Scripting to move log files into HDFS. Developed Scala scripts, UDF's using both Data frames/SQL and RDD/MapReduce in Spark for Data Aggregation, queries and writing data back into RDBMS through Sqoop.

Good understanding in the processing of real-time data using Spark. Import the data from different sources like HDFS into Spark RDD. Good knowledge of Data Modelling and Data Mining to model the data as per business requirements. Expert in Data Visualization development using Tableau to create complex and innovative dashboards. Generated ETL reports using Tableau and created statistics dashboards for Analytics.

Working knowledge of Agile and waterfall development models.

Worked with different software version control, Jira, bug tracking and code review systems like Git

Experience

Microsoft
Senior Big Data Engineer
June 2019 - August 2019 (3 months)

T-Mobile
Senior Big Data Engineer
December 2018 - May 2019 (6 months)
Greater Seattle Area

Microsoft
Senior Big Data Engineer
June 2017 - November 2018 (1 year 6 months)
Greater Seattle Area

- Understand, articulate and present business requirements into technical solutions
- Analyze and profile current data to ensure no requirements are missed
- Create data integration and technical solutions for Azure Data Factory, Azure SQL databases and Azure SQL Data Warehouse for providing analytics and reports for improving marketing strategies.
- Create linked services to connect to Azure Storage, on-premises SQL Server and Azure HDInsight
- Configured Azure SQL database with Azure storage Explorer and with SQL server
- Creating ADL, Azure Storage, ADB and resource groups in the Azure portal
- Preparing and uploading the source data via Azure storage explorer.
- Leverage Azure Blob storage accounts via PowerShell or Azure storage explorer for storing the different types of data like Blob storage, Table storage, Queue storage and File storage.
- Create ADF accounts for the resource groups
- Create Input and Output data sets to manage the data type, data structures, availability and policy of input and output files
- Created U-SQL script for transform activities and developed complex queries to transform the data from multiple sources, and outputted the data into Azure Data warehouse.
- Customizing extractors, reducers, processors to extract, reduce, and process data files such as text, CSV, parquet, JSON and structured data.
- Monitoring the Pipeline and Data Slices for ADF lifecycle
- Manage ADF Custom activities.
- Work with OnPrem-Azure SQL Databases and Azure SQL Data Warehouse.
- Usage of Key Vault for storing and retrieving security-related information.

- Experience working in an agile team environment
- Understanding of code promotion, code turnover to operations, version control.
- Conduct code walkthroughs, peer reviews, and produce technical documentation
- Validated the reports in dashboard and shared those reports to team members.
- Exposure on COSMOS.

PayPal

Senior Big Data Engineer

January 2016 - June 2017 (1 year 6 months)

San Francisco Bay Area

- Worked on Hadoop technologies like Hive, Sqoop, SparkSQL and Big Data testing.
- Developed automated scripts for ingesting the data from Teradata around 200TB bi-weekly refreshment of data.
- Developed Hive scripts for end-user/analyst requirements for ad-hoc analysis.
- Used of Partitions, Bucketing concepts in Hive and designed both Managed and External tables in Hive for optimized performance.
- Extensively used Apache Sqoop for efficiently transferring bulk data between Apache Hadoop and relational databases (Oracle) for product level forecast.
- Worked in tuning Hive to improve performance. Developed UDFs using JAVA as and when necessary to use HIVE queries.
- Extracted the data from Teradata into HDFS using Sqoop.
- Created Tableau Dashboards with interactive views, trends and drill downs along with user-level security
- Used Impala to read, write and query the Hadoop data in HDFS.
- Developed programs in Spark based on the application for faster data processing than standard MapReduce programs.
- Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshooting, manage and review data backups, manage and review Hadoop log files.
- Worked with application teams to install the operating system, Hadoop updates, patches, version upgrades as required.
- Configured Hadoop system files to accommodate new sources of data and updated the existing configuration Hadoop cluster

- Involved in gathering business requirements and prepare detailed specifications that follow project guidelines required to develop written programs.
- Worked on importing and exporting data from different databases like Oracle, Teradata into HDFS and Hive using Sqoop.
- Actively participating in the code reviews, meetings and solving any technical issues.

UnitedHealth Group

Big Data Engineer

August 2014 - December 2015 (1 year 5 months)

San Francisco Bay Area

- Imported data from different relational data sources like RDBMS, Teradata to HDFS using Sqoop.
- Designed and implemented Incremental Imports into Hive tables.
- Involved in converting Map Reduce programs into Spark transformations using Spark RDD's on Scala.
- Implemented Spark using Scala and Spark SQL for faster testing and processing of data.
- Experienced in working with various kinds of data sources such as Teradata and Oracle. Successfully loaded files to HDFS from Teradata, and load loaded from HDFS to Hive.
- Develop wrapper using shell scripting for Hive, Pig, Sqoop, Scala jobs.
- Worked on developing Unix Shell scripts to automate Spark-SQL
- Performed advanced procedures like text analytics and processing, using the in-memory computing capabilities of Spark using Scala.
- Worked in Loading and transforming large sets of structured, semi-structured and unstructured data.
- Written Hive jobs to parse the logs and structure them in tabular format to facilitate effective querying on the log data.
- Involved in creating Hive tables, loading with data and writing hive queries that will run internally in MapReduce way.
- Experienced in managing and reviewing the Hadoop log files.
- Successfully ran all Hadoop MapReduce programs on Amazon Elastic MapReduce framework by using Amazon S3 for input and output.
- Involve in Data Asset Inventory to gather, analyze, and document business requirements, functional requirements and data specifications for Member Retention from sources SQL / Hadoop.
- Migrated ETL jobs to Pig scripts do Transformations, even joins and some pre-aggregations before storing the data onto HDFS.

- Exported data from HDFS environment into RDBMS using Sqoop for report generation and visualization purpose.
- Created and maintained Technical documentation for launching HADOOP Clusters.

Environment(s): Hadoop, HDFS, Pig 0.10, Hive, MapReduce, Sqoop, Java Eclipse, SQL Server, Shell Scripting

Education

Alfred University

International Technological University (ITU)

Master's degree, Engineering Management (Management Information Systems)

Mahatma Gandhi Institute of Technology

Bachelor of Engineering - BE