**Sandeep Vinjamoori**

**9402995108**

[**Vinjamoorisandeep@gmail.com**](mailto:Vinjamoorisandeep@gmail.com)

**www.linkedin.com/in/sandeep-vinjamoori-658060201**

**SUMMARY**

• Overall 6+ years of professional IT experience in Software Development. This also includes 4+ years of experience in Ingestion, Storage, Querying, Processing and Analysis of Big Data using Hadoop technologies and solutions.  
• Excellent understanding/knowledge of Hadoop architecture and various components of Hadoop ecosystem such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Map Reduce &YARN.  
• Hands on experience in using Hadoop ecosystem components like Map Reduce, HDFS, Hive, Pig, Sqoop, Spark, Flume, Zookeeper, Hue, Kafka, Storm & Impala.

• Experience working with Azure Data Factory (ADF), Azure Data Lake (ADL), Azure Data Lake Analytics, Azure SQL database, Azure Databricks, Azure Synapse, Azure SQL Data Warehouse, Azure BLOB Storage.  
• Experience with Agile Methodology.  
• Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse and Controlling and granting database access and Migrating On premise databases to Azure Data lake store using Azure Data factory.

• Expertise in job workflow scheduling and monitoring tools like Oozie.  
• Developed simple to complex Map/Reduce jobs using Hive and Pig to handle files in multiple formats like JSON, Text, XML, Sequence File etc.  
• Worked extensively on creating combiners, Partitioning, Distributed cache to improve the performance of Map Reduce jobs.  
• Experience in working with different data sources like Flat files, XML files, log files and Database.  
• Very Good understanding and Working Knowledge of Object Oriented Programming (OOPS).  
• Expertise in application development using Scala, RDBMS, and UNIX shell scripting.  
• Experience developing Scala applications for loading/streaming data into NoSQL databases (HBASE) and into HDFS.  
• Worked on ingesting log data into Hadoop using Flume.  
• Experience in managing and reviewing Hadoop log files.  
• Experience in importing and exporting data using Sqoop from HDFS to Relational Database Management System and vice-versa.  
• Using Apache Flume, collected and stored streaming data(log data) in HDFS..  
• Experience with scripting languages (Scala, Python and Shell) to manipulate data.  
• Worked with relational database systems (RDBMS) such as My SQL, and No SQL database systems like HBase and had basic knowledge on MongoDB and Cassandra.  
• Hands on experience in identifying and resolving performance Bottlenecks in various levels like sources, Mappings and Sessions.  
• Highly Motivated, Adaptive and Quick learner.  
• Ability to adapt to evolving Technology, Strong Sense of Responsibility and Accomplishment.

Authorized to work in United States for any employer.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Programming Languages** | Python, Scala, SQL |
| **Operating Systems** | Windows 95/98/NT/2000/2003/XPs, UNIX/Linux. |
| **Hadoop Eco-System** | HDFS, Nifi, Map Reduce, Oozie, Hive/Impala, Pig, Sqoop, Zookeeper and Hbase, Spark, Scala, Kafka, Apache Flink, AWS- EC2, S3, EMR, snowflake. |
| **Cloud Technologies** | Azure Databricks, Azure Data factory, Azure Synapse, Azure SQL Data warehouse, AWS EC2, IAM, S3, Auto Scaling. |
| **ETL Tools** | Data Stage, Ab-initio, Informatica |
| **Databases** | ESSBASE11.x/ 9 BI+, Oracle 9i/10g, DB2, SQL Server 2008/2014/2016, MS Access |
| **IDE** | Eclipse, IntelliJ, SQL Developer, Microsoft Visual Studio. |
| **Methodologies** | Data Modeling – Logical/Physical/Dimensional, Star/Snowflake Schema, OLAP, Agile/Waterfall |

**EXPERIENCE & PROJECTS**

**Capital One**

**Sr. Azure Data Engineer Sept 2021- Present**

**Description:**

Capital One Financial Corporation is an American bank holding company specializing in credit cards, auto loans, banking, and savings accounts, headquartered in McLean, Virginia with operations primarily in the United States

**Responsibilities**:

* Providing End to end business intelligence solutions by using Microsoft technologies using Azure Data Lake, Azure Databricks, Azure Data Factory, Azure SQL data warehouse, Azure Synapse.
* Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL and U-SQL Azure Data Lake Analytics. Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Databricks.
* Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform and load data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards.
* Developed Azure Synapse Notebooks using Pyspark and Spark-SQL for data extraction, transformation and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
* Responsible for estimating the cluster size, monitoring and troubleshooting of the Spark data bricks cluster.
* Developed JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data using the SQL Activity.
* Hands-on experience on developing SQL Scripts for automation purpose.
* Created Build and Release for multiple projects (modules) in production environment using Visual Studio Team Services (VSTS).

**Environment**: Azure Data Factory, Azure Synapse, Azure Databricks, Azure Lake Gen2,Azure SQL database, Hadoop, Map Reduce, Yarn, Hive, Python, Spark, Scala, SSMS, Horton Works, Data Lake, Databricks.

**Synchrony Apr 2019-June 2021**

**Sr. Data Engineer**

**Description:**

Synchrony Financial is a consumer financial services company headquartered in Stamford, Connecticut, United States. The company offers consumer financing products, including credit, promotional financing and loyalty programs, installment lending to industries, and FDIC-insured consumer savings products through Synchrony Bank, its wholly owned online bank subsidiary.

**Responsibilities**:

* Requirement gathering, analysis and designing solution to meet the goals. Worked with Business Users to gather requirements and provide possible solutions.
* Worked with Hortonworks distribution of Hadoop for setting up the cluster and monitored it using Ambari.
* Involved in the development of real time streaming applications using PySpark, Scala, Kafka, Hive on distributed Hadoop Cluster.
* Created ODBC connection through Sqoop between Hortonworks and SQL Server.
* Implemented Spark Scripts using Scala, Spark-SQL to access hive tables into spark for faster processing of data.
* Great hands-on experience with Pyspark for using Spark libraries by using python scripting for data analysis.
* Extensively worked on Hive, Pig, Map Reduce, Sqoop, Oozie in an optimized way of distributed processing.
* Worked extensively on Hive to create, alter, and drop tables and involved in writing hive queries.
* Developed Scala scripts, UDFs using both Data frames/SQL and Data sets in Spark for Data Aggregation.
* Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and AWS cloud. Injecting large volumes of data from heterogenous sources into the Data Lake.
* Successfully migrated the data from AWSS3 source to the HDFS sink using Flume.
* Created Airflow Scheduling scripts in Python.
* Processing data in AWS EMR using pyspark data and exporting the data to NoSQL Db.
* Used Jenkins pipelines to drive all microservices builds out to the Docker registry and then deployed to Kubernetes.
* Wrote Hive and Pig scripts for joining the raw data with the lookup data and for some aggregative operations as per the business requirement.
* Good Knowledge in using NiFi to automate the data movement between different Hadoop systems.
* Tested Apache Tez, an extensible framework for building high performance batch and interactive data processing applications, on Pig and Hive jobs.
* Used Sqoop to import data from RDBMS to HDFS cluster using custom scripts.

**Environment**: Hadoop, Map Reduce, Yarn, Hive, Python, NumPy, Jenkins Pig, Flume, Sqoop, AWS, EMR, Core Java, Spark, Scala, Kafka, MongoDB, Tez, Elastic search 5.x, Horton Works, Data Lake, Airflow, UNIX, Bitbucket, Java, Databricks

**Experian Feb 2018 – Mar 2019**

**Hadoop Developer**

**Description:**

Experian is an American Irish multinational data analytics and consumer credit reporting company.

**Responsibilities**:  
• Data Ingestion implemented using SQOOP, SPARK, loading data from various RDBMS, CSV, XML files.  
• Data cleansing, transformations tasks are handled using SPARK using SCALA and HIVE.  
• Data Consolidation was implemented using SPARK, HIVE to generate data in the required formats by applying various ETL tasks for data repair, massaging data to identify source for audit purpose, data filtering and store back to HDFS.  
• Responsible for design development of Spark SQL Scripts based on Functional Specifications.  
• Exploring with the Spark improving the Performance and Optimization of the existing algorithms in Hadoop.  
• ETL development to normalize this data and publish it in IMPALA.  
• Involved in converting Hive/SQL queries into Spark RDD using Scala.  
• Responsible for Job management using Fair scheduler and Developed Job Processing scripts using Oozie Workflow.  
• Responsible for Performance Tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and Memory tuning.  
• Optimizing of existing algorithms in Hadoop using Spark Context, Spark-SQL and Pair RDD's.  
• Responsible in handling large datasets using Partitions, Spark in Memory capabilities, Broadcasts in Spark, Effective & efficient Joins, Transformations and other during Ingestion process itself.  
• Importing and exporting data into HDFS and HIVE, PIG using Sqoop.  
• Involved in creating Hive Tables, loading with data and writing Hive queries which will invoke and run Map Reduce jobs in the backend.  
• Implemented the workflows using Apache Oozie framework to automate tasks.  
• Worked with No SQL databases like HBase. Creating HBase tables to load large sets of semi structured data coming from various sources.  
• Worked with different file formats such as Text, Sequence files, Avro, ORC and Parquet.  
• Responsible to manage data coming from different sources.  
• Responsible on loading and transforming of large sets of structured, semi structured and unstructured data.  
• Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.

**Environment**: Scala, Hive, HBase, Flume, Java, Impala, Pig, Snowflake, Spark, Oozie, Oracle, Yarn,, Unix, Cloudera, Flume, Sqoop, HDFS, Java, Python.

**Wells Fargo (Infosys) Mar 2016- Feb 2018**

**Hadoop developer**

**Description:**  
**Wells Fargo** is a diversified financial service holding company that offers retail and wholesale banking, and wealth management services to individuals, businesses, high-net-worth individuals, and institutions, through its subsidiaries.

**Responsibilities**:  
• Worked on Spark SQL, reading/writing data from JSON file, text file, parquet file, schema RDD.  
• Identifying data sources and create appropriate data ingestion procedures.  
• Transformed the data using Spark, Hive, Pig for BI team to perform visual analytics according to the client requirement.  
• Developed the service to run the Map-reduce jobs as per the requirement basis.  
• Importing and exporting data into HDFS and HIVE using Sqoop.  
• Populated big data customer marketing data structures.  
• Developed Spark scripts by using Python as per requirements.  
• Performed joins on tables in Hive with various optimization techniques.  
• Implemented lateral view in conjunction with UDFs in Hive according to the client requirements.  
• Created Hive tables as per internal requirements in static and dynamic partitions.  
• Implemented the workflows using Apache Oozie framework to automate the tasks.  
• Developing design documents considering all possible approaches and identifying best of them.  
• Developed scripts and automated data management from end to end sync up between the clusters.  
• Imported the data from different sources like HDFS/Hbase into Spark RDD.  
• Experienced with Spark context, Spark-SQL, Data frame, Pair RDD's, Spark Yarn.  
• Involved in converting Hive/SQL queries into Spark transformations using Spark RDD, Scala.  
• Automated the jobs for pulling data from FTP server to load data into Hive tables using Oozie workflows.  
  
**Environment**: Hive, Sqoop, Python, Shell scripting, Spark, Oozie, Snowflake, Scala, Java.

**EDUCATION**

**University of North Texas, Denton -Texas**

*Master of Science, Data Science in 2022*

**Osmania University-Hyderabad**

*Bachelor of Technology, Mechanical Engineering in 2017*