ASWINI REDDY
Email: Nandini@o2inc.net
Phone: 949-502-2810
Big Data Developer

SUMMARY:

- Creative **Data Engineer** with a master's degree in computer science having around **6+ years** of experience in building and maintaining **Big Data applications**.
- Hands-on experience and strong knowledge in Echo-systems like **Spark, HDFS, Hive, Pig, Sqoop, HBase, Map Reduce, Spark SQL, Kafka.**
- Experience in integration of various **relational and non-relational** sources such as **DB2, Oracle, SQL Server, NoSQL-MongoDB, XML and Flat Files**, to data warehouse.
- Efficient in writing **UDFs** for both Hive & PIG.
- Performed **Hive** operations on large datasets with proficiency in writing **HiveQL** queries using transactional and performance efficient concepts: **UPSERTS, Partitioning, Bucketing, Windowing**, etc.
- Extensive experience with **ETL** and Query big data tools like **Pig Latin** and **Hive QL**.
- Hands on experience in big data ingestion tools like **Flume** and **Sqoop**.
- In-depth understanding of **Spark** Architecture and performed several batch and real-time data stream operations using **Spark (Core, SQL, Streaming)**.
- Experienced in handling large datasets using **Spark** in-memory capabilities, Partitions, Broadcast variables, Accumulators, Effective & Efficient Joins. Used **Scala** to develop Spark applications.
- Experienced in Amazon Web Services (**AWS**) such as **AWS EC2, S3**
- Experienced in analyzing business requirements and translating requirements into functional and technical design specifications.
- End to end experience in designing and deploying Analytical **Dashboards** and **data visualizations** using **Tableau** Desktop, **Tableau** Server, **Tableau** Reader and **Tableau** Public.
- Delivered products and services using **Agile development** methodologies.
- Excellent problem-solving skills, high analytical skills, good communication and interpersonal skills.
- Experienced in working with different file formats like xml, parquet, orc
- Expert in performing business analytical scripts using HQL
- Experience in working with tools like Autosys and AWP

TECHNICAL SKILLS:

Programming Skills	Python, Scala, R, SQL
Big Data Ecosystem	Spark, Hadoop, HDFS, Hive, Map-Reduce, Pig, YARN, HBase, Scala, Kafka, Sqoop, Pyspark
Databases	HBase, Cassandra, Mango DB, MySQL, PostgreSQL
Tools & Utilities	Git, Eclipse, Putty, SQL Developer, SQL client, Autosys, AWP
Methodologies	Agile
Operating Systems	Windows 10, UNIX and LINUX

WORK EXPERIENCE:

Citigroup,   **Irving,**   **TX**
Jan 2022 Till Now
Role: Big Data Developer

Responsibilities:

- Experience in creating and loading data into Hive tables with appropriate static and dynamic partitions, intended for efficiency
- Worked on various File formats like Parquet, ORC and Avro
- Experience in Data Warehousing applications, Extractions, Transformations and Loading (ETL) of data from multiple sources.
- Worked in optimizing Hive SQL and spark jobs.
- Advanced knowledge of Autosys Scheduling techniques in Distributed Environment.
- Created JIL and updated existing jobs, calendars, resources, variables that are supported by autosys
- Analyzed SQL scripts and designed the solution using pyspark
- Used Oozie workflow to co-ordinate hive scripts
- Created pyspark code to ingest data from .dat files into Scorecard tables
- Ran the jobs using AWP

Environment: PySpark Hive, Sqoop, Scala, Shell Scripting, Spark SQL, Autosys, AWP

Cigna, Kansas City, KS
Dec 2019 to Jan 2022
Role: Big Data Developer
Employer, Prokarma Inc.,

Responsibilities:
- Worked in development of a client pricing Applications
- Created Ingestion tool which is used to ingest data from Teradata to Hadoop
- Developed different Scala APIs which works on top of Spark and Hadoop Ecosystem.
- Created UNIX scripts for or ingesting and manipulating data between Teradata and Hadoop
- Proficiency and knowledge of best practices with Hadoop (Yarn and HDFS)
- Work in an agile environment and continuously improve the agile processes
- Used Spark and Spark-SQL to read the ORC data and create the tables in hive using the Scala API.
- Implemented Spark using Scala and Spark SQL for faster testing and processing the data
- Responsible for performing sort, join, aggregations, filter, and other transformations on the datasets using Spark.
- Involved in converting Hive/SQL queries into Spark transformations using Spark
- RDDs and Scala.
- Designed and developed Security Framework to provide fine grained access to objects in AWS S3 using AWS Lambda, DynamoDB.

Environment: Python, Hive, Sqoop, Scala, Shell Scripting, Spark SQL, Oracle

Mindtree Limited, Bangalore, India
Oct 2014 to April 2016
Role: Hadoop/Python Developer
Project-People Health Services

Responsibilities:
- Created **Hive** tables, loaded data, executed **HQL** queries and developed **MapReduce** programs to perform analytical operations on data and to generate reports.
- Created Hive internal and external tables, used **MySQL** to store table schemas. Wrote custom **UDFs** in **Python**.
- Moved data between **MySQL** and **HDFS** using **Sqoop**.
- Developed **MapReduce** jobs in **Python** for log analysis, analytics, and data cleaning.
- Wrote complex MapReduce programs to perform operations by extracting, transforming, and aggregating to process terabytes of data.
- Wrote many **SQL**, Procedures, Triggers and Views on top of **Oracle**.

- Involved in the implementation of the Software development life cycle (**SDLC**) that includes Development, Testing, Implementation, and Maintenance Support.

Environment: Python, Hive, Sqoop, MapReduce, SQL, MySQL, Oracl

Advantage One Human Resourcing Pvt Ltd, India
June-2013 to May 2014
Role: Data Analyst
Project-Audience IQ
Audience IQ is a leading Digital Marketing Intelligence product which combines the Audience data with the touchpoint data and applies attributions on it.

Responsibilities:
- Involved in the complete life cycle of design, development, testing, and deployment of web service development.
- Aggregating, manipulating, and analyzing large data sets to deliver recurring reporting.
- Customizing **SQL** codes according to desired reporting.
- Created a database warehouse for new and ongoing projects according to company specifications and simplified the **ETL** process.
- Organized and managed project work to drive deliverables.
- Helped in creating **AGILE** methodology for ongoing projects.
- Used tools like tableau, power BI and Microsoft Visio for delivering intelligent business solutions.
- Developed macros and excel workbooks to automate day to day tasks.
- Designed and developed complete workflow for Audience IQ, where end users will be able to analyze their online users (V.1).
- Redesign V.1 to support daily attribution and any cadence Touch point reports with completed processing in Hadoop with on a demand **EC2** cluster.
- Designed and developed to process huge volumes of Audience data through **NIFI**.
- Configured **SQS**, **SNS AWS** messages from **S3** buckets.
- Developed **MySQL**, **Hive** , **Spark SQL** scripts for data processing and attribution models
- Used Sqoop and Distcp tools for data transfers.

Environment: HDP 2.4, Hadoop 2.4, Spark 2.2, Hive 2.6, Sqoop. NIFI 1.2, Eclipse

EDUCATION SUMMARY:

- Master of Computer Science

   Illinois Institute of Technology – 2019