**PAWAN NARASIMHA AVVARI**

**Phone: +1 704-281-1153**

**Email:** **pawan.avvari17@gmail.com**

**PROFESSIONAL SUMMARY**

* Over9+ years of experience in the software industry, including 4+years of experience in Awscloud services, and 4 years of experience in a Data warehouse.
* Experience in Aws Cloud, Snowflake, Data bricks,MySQL,SQL Server, S3 storage, Aws Redshift, Big Data Technologies (Hadoop and Apache Spark), and Aws Sagemaker.
* Experience in developing, supporting, and maintenance for the ETL (Extract, Transform, and Load) processes using Talend Integration Suite.
* Experience in developing very complex mappings, reusable transformations, sessions, and workflows using the Informatica ETL tool to extract data from various sources and load it into targets.
* Proficiency in multiple databases like MongoDB, Cassandra, MySQL, ORACLE, and MS SQL Server.
* Experience in Developing Spark applications using Spark - SQL in Databricks for data extraction, transformation, and aggregation from multiple file formats for analyzing and transforming the data to uncover insights into the customer usage patterns.
* Used various file formats like Avro, Parquet, Sequence, JSON, ORC, CSV and text for loading data, parsing, gathering, and performing transformations.
* Designed and created Hive external tables using a shared meta-store with Static & Dynamic partitioning, bucketing, and indexing.
* Exploring with Spark to improve the performance and optimization of the existing algorithms in Hadoop using Spark context, Spark-SQL, Data Frame, and pair RDDs.
* Extensive hands-on experience tuning spark Jobs.
* Experienced in working with structured data using HiveQL and optimizing Hive queries.
* Familiarity with libraries like PySpark, Numbly, Pandas, and Matplotlib in python.
* Writing complex SQL queries using joins, group by, nested queries.
* Experience in HBase to load data using connectors and write queries using NoSQL.
* Experience with solid capabilities in exploratory data analysis, statistical analysis, and visualization using R, Python, SQL, and Tableau.
* Running and scheduling workflows using Air flow, identifying failures and integrating, coordinating, and scheduling jobs.
* Hands-on experience on Kafka and Flume to load the log data from multiple sources directly into HDFS.
* Widely used different features of Teradata such as BTEQ, Fast load, Multifood, SQL Assistant, DDL, and DML commands and a very good understanding of Teradata UPI and NUPI, secondary indexes, and join indexes.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Big Data Technologies/Aws cloud services** | Hadoop, Map Reduce, HDFS, Hive, HBase, Flume, Kafka, Yarn, Apache Spark. Aws Ec2 instance, Aws Sagemaker, Awsdata bricks, Aws secret manager, Aws cloud watch, Aws lambda,Aws SQL Database, S3, DynamoDB. |
| **Databases** | Oracle, MySQL, SQL Server, MongoDB, Dynamo DB, Cassandra, Snowflake,  Aws Red Shift. |
| **Programming Languages** | Python, Pyspark, Shell script, Perl script, SQL, Java. |
| **Tools** | PyCharm, Eclipse, Visual Studio, SQL\*Plus, SQL Developer, SQL Navigator, SQL Server Management Studio, Eclipse, Postman. |
| **Version Control** | SVN, Git, GitHub, Maven |
| **Operating Systems** | Windows 10/7/XP/2000/NT/98/95, UNIX, LINUX, OS |
| **Visualization/ Reporting** | Tableau, ggplot2, matplotlib |

**PROFESSIONAL EXPERIENCE**

**Client:**Elevance Health**. Nov 2021 -Present**

**Role: Sr. Aws Data Engineer**

**Responsibilities**

* Architect and implement ETL and data movement solutions using Aws Service.
* Understand Business requirements, analysis and translate them into Application and operational requirements.
* Designed a one-time load strategy for moving large databases toAws RedShift Data Warehouse.
* Using Rest API to collect of resources and methods that are integrated with backend HTTP endpoints, Lambda functions, or other AWS services.
* Extract Transform and Load data from Sources Systems to Data Storage services using HDInsight.
* Created a framework to do data profiling, cleansing, automatic restart ability of batch pipeline, and handling rollback strategy.
* Experienced in database like Aws Aurora Relational Database
* Design and implement database solutions in Snowflake, Aws Redshift.
* Lead a team of six developers to migrate the application.
* Implemented masking and encryption techniques to protect sensitive data.
* Implemented SSIS IR to run SSIS packages from Aws.
* Developed mapping document to map columns from source to target.
* Created ETL pipelines using Aws s3 or On-premises Storages.
* Performed ETL using Aws Data Bricks. Migrated on-premises Oracle ETL process to Aws Synapse Analytics.
* Worked on python scripting to automate the generation of scripts. Data curation is done using Awsdata bricks.
* Worked on Aws data bricks, PySpark, HDInsight, Aws Redshift, and hive used to load and transform data.
* Used stored procedure, lookup, execute the pipeline, data flow, copy data, lambda function features in Aws services.
* Extract Transform and Load data from Sources Systems to Snowflake using a combination of S3 commands, copy commands, SQL Server, Spark SQL.
* Data Ingestion to one or more Aws Services - (Aws SQL, Aws RedShift DW) and processing the data in In Aws Databricks.
* Responsible for estimating the cluster size, monitoring, and troubleshooting of the Spark data bricks cluster.
* Creating Databricks notebooks using SQL, Python, and automated notebooks using jobs.
* Create and maintain optimal data pipeline architecture in cloud Aws Databricks

**Environment**: Hadoop, Hive, Impala, Oracle, Spark, Pig, Sqoop, Oozie, S3, Databricks and Hive, HBase, Sqoop, Flume, Map Reduce, Teradata, SQL, Kafka, Zookeeper, Pyspark, Snowflake.

**Client: Nationstar Mortgage, Dallas, TX Sep 2018 – Oct 2021**

**Role: Aws Data Engineer**

**Responsibilities:**

* Developed a data pipeline using Spark, Hive, and HBase to ingest customer behavioral data and financial histories into the Hadoop cluster for analysis.
* Working Experience on Aws Databricks cloud to organize the data into notebooks and make it easy to visualize data using dashboards.
* Performed ETL on data from different source systems to Aws Data Storage services using a combination of T-SQL, Spark SQL,S3, CLI commands. Data Ingestion to one or more Aws Services - (Aws Redshift, Aws SQL, Aws DW) and processing the data in In Aws Databricks.
* Hands-on experience implementing Spark and Hive jobs performance tuning.
* Performed Data Aggregation, Validation, and HDInsight using spark scripts written in Python.
* Generated PL/SQL scripts for data manipulation, validation, and materialized views for remote instances.
* Created partitioned tables in Hive, also designed a data warehouse using Hive external tables, and also created hive queries for analysis.
* Created and modified several database objects such as Tables, Views, Indexes, Constraints, Stored procedures, Packages, Functions, and Triggers using SQL and PL/SQL.
* Created large datasets by combining individual datasets using various inner and outer joins in SAS/SQL and dataset sorting and merging techniques using SAS/Base.
* Extensively worked on Shell scripts for running SAS programs in batch mode on UNIX.
* Wrote Python scripts to parse XML documents and load the data in the database.
* Used Hive, Impala, and Sqoop utilities and Oozie workflows for data extraction and data loading.
* Created HBase tables to store various data formats of data coming from different sources.
* Responsible for importing log files from various sources into HDFS using Flume.
* Responsible for translating business and data requirements into logical data models in support of Enterprise data models, ODS, OLAP, OLTP, and Operational data structures.
* Created SSIS packages to migrate data from heterogeneous sources such as MS Excel, Flat files, and CVS files.
* Provided thought leadership for architecture and the design of Big Data Analytics solutions for customers, actively drive Proof of Concept (POC) and Proof of Technology (POT) evaluations, and implemented a Big Data solution

**Environment**: Snowflake, Databricks and Spark, Hive, HBase, Sqoop, Flume, S3, cosmos DB, MapReduce, HDFS, Cloudera, SQL, Apache Kafka, shell script, Air Flow, Python, Power BI, Unix, SQL Server.

**Client: State Farm, Bloomington, IL Feb 2018 – Sep 2018 Role: Big data Developer**

**Responsibilities:**

* Involved in Requirement gathering, Business Analysis, and translating business requirements into technical design in Hadoop and Big Data.
* Involved in SQOOP implementation which helps in loading data from various RDBMS sources to Hadoop systems and vice versa.
* Developed Python scripts to extract the data from the web server output files to load into HDFS.
* written a python script that automates to launch of the EMR cluster and configures the Hadoop applications.
* Extensively worked with Avro and Parquet files and converted the data from either format Parsed Semi-Structured JSON data or converted it to Parquet using Data Frames in PySpark.
* Involved in Analyzing system failures, identifying root causes, and recommended course of actions, Documented the systems processes and procedures for future references.
* Involved in Configuring the Hadoop cluster and load balancing across the nodes.
* Involved in Hadoop installation, Commissioning, Decommissioning, Balancing, Troubleshooting, Monitoring, and debugging Configuration of multiple nodes using the Hortonworks platform.
* Involved in working with Spark on top of Yarn/MRv2 for interactive and Batch Analysis.
* Involved in managing and monitoring Hadoop cluster using Cloudera Manager.
* Used Python and Shell scripting to build pipelines.
* Developed data pipeline using Sqoop, HQL, Spark, and Kafka to ingest Enterprise message delivery data into HDFS.
* Developed workflow in Oozie also in Airflow to automate the tasks of loading data into HDFS and pre-processing with Pig and Hive.
* Assisted in creating and maintaining technical documentation for launching HADOOP Clusters and even for executing Hive queries and Pig Scripts.
* Integrated Hadoop into traditional ETL, accelerating the extraction, transformation, and loading of massive semi-structured and unstructured data. Loaded unstructured data into Hadoop distributed File System (HDFS).
* Created HIVE Tables with dynamic and static partitioning including buckets for efficiency. Also created external tables in HIVE for staging purposes.
* Loaded HIVE tables with data, wrote hive queries that run on MapReduce and Created a customized BI tool for management teams that perform query analytics using HiveQL.
* Aggregated RDDs based on the business requirements and converted RDDs into Data frames saved as temporary hive tables for intermediate processing and stored in HBase/Cassandra and RDBMS.

**Environment:**Hadoop 3.0, Hive 2.1, J2EE, JDBC, Pig 0.16, HBase 1.1, Sqoop, NoSQL, Impala, Java, Spring, MVC, XML, Spark 1.9, PL/SQL, HDFS, JSON, Hibernate, Bootstrap, jQuery.

**Client: United Airlines, Chicago, IL Feb 2017 – Jan 2018  
Role: Hadoop Developer**

**Responsibilities:**

* Involved in Requirement gathering, Business Analysis and translated business requirements into technical design in Hadoop and Big Data.
* Involved in SQOOP implementation which helps in loading data from various RDBMS sources to Hadoop systems and vice versa.
* Developed Python scripts to extract the data from the web server output files to load into HDFS.
* Extensively worked with Avro and Parquet files and converted the data from either format Parsed Semi-Structured JSON data and converted to Parquet using Data Frames in PySpark.
* Involved in Analyzing system failures, identifying root causes, and recommended course of actions, Documented the systems processes and procedures for future references.
* Involved in Configuring the Hadoop cluster and load balancing across the nodes.
* Involved in Hadoop installation, Commissioning, Decommissioning, Balancing, Troubleshooting, Monitoring and, debugging Configuration of multiple nodes using Hortonworks platform.
* Involved in working with Spark on top of Yarn/MRv2 for interactive and Batch Analysis.
* Involved in managing and monitoring Hadoop cluster using Cloudera Manager.
* Used Python and Shell scripting to build pipelines.
* Developed data pipeline using Sqoop, HQL, Spark, and Kafka to ingest Enterprise message delivery data into HDFS.
* Developed workflow in Oozie also in Airflow to automate the tasks of loading data into HDFS and pre-processing with Pig and Hive.
* Assisted in creating and maintaining technical documentation for launching HADOOP Clusters and even for executing Hive queries and Pig Scripts.
* Integrated Hadoop into traditional ETL, accelerating the extraction, transformation, and loading of massive semi-structured and unstructured data. Loaded unstructured data into Hadoop distributed File System (HDFS).
* Created HIVE Tables with dynamic and static partitioning including buckets for efficiency. Also created external tables in HIVE for staging purposes.

**Environment:**Hadoop,Hive, J2EE, JDBC, Pig , HBase, Sqoop, NoSQL, Impala, Java, Spring, MVC, XML, Spark, PL/SQL, HDFS, JSON, Hibernate, Bootstrap, jQuery.

**Client:Saven Technologies, Hyderabad, India May 2013 - June 2015**

**Role: Data Warehouse Developer**

**Responsibilities:**

* Creation, manipulation, and supporting the SQL Server databases.
* Involved in the Data modeling, Physical and Logical Design of Database
* Helped in the integration of the front end with the SQL Server backend.
* Created Stored Procedures, Triggers, Indexes, User-defined Functions, Constraints,etc on various database objects to obtain the required results.
* Import & Export data from one server to other servers using tools like Data Transformation Services (DTS)
* Wrote T-SQL statements for retrieval of data and involved in performance tuning of TSQL queries.
* Transferred data from various data sources/business systems including MS Excel, MS Access, Flat Files,etc to SQL Server using SSIS/DTS using various features like data conversion, etc. Also Created derived columns from the present columns for the given requirements.
* Supported team in resolving SQL Reporting services and T-SQL-related issues and Proficiency in creating different types of reports such as Crosstab, Conditional, Drill-down, Top N, Summary, Form, OLAP and Sub reports, and formatting them.
* Created logging for ETL load at package level and task level to log several records processed by each package and each task in a package using SSIS.
* Developed, monitored, and deployed SSIS packages.

**Environment**: IBM WebSphere DataStage EE/7.0/6.0 (Manager, Designer, Director, Administrator), Accentual Profile Stage 6.0, AscentialQuality Stage 6.0, Erwin, TOAD, Autosys, Oracle 9i, PL/SQL, SQL, UNIX Shell Scripts, Sun Solaris, Windows 2000.

**EDUCATION**

* Bachelor’s in information technology from LPU Punjab, India 2013.
* Master’s in Computer science and Information systems from Pace University New york, Dec 2016