**Venkateswara**

**Phone:** **9452459232**

**SUMMARY:**

* Software professional with 13 years of experience in providing comprehensive business reports with detailed analysis for executive review to make business decisions
* Skill in Database Design and Development, Data Flow Diagrams, Normalization, Report Automation, SQL Reporting, Requirements Analysis, Data Quality Assurance, Data Modeling, Data Warehousing, Data Analysis, Agile/Scrum, OLAP, OLTP, Multidimensional Databases, KPIs
* Proficient in data transformation, processing, and extraction using **Python, SQL, ETL’s** and **Macros**
* Good experience of software development in Python and IDEs: PyCharm, sublime text, Jupyter Notebook
* Experienced in using python libraries like Pandas, NumPy, SQLalchemy, **PySpark**, Boto3 and Matplotlib
* Hands on experience in developing **SPARK** applications using Spark tools like RDD transformations, Spark core, Spark MLlib, Spark Streaming and Spark SQL.
* Good understanding of **Amazon Web Services** (AWS) cloud computing platform and migrating the application from existing systems to AWS
* Worked on **Amazon Elastic Cloud Compute** (EC2) infrastructure for computational tasks and **Simple Storage Service** **(S3)** as Storage mechanism.
* Hands-on experience with AWS services such as S3, EC2, Redshift, Glue, and Athena for building scalable and reliable data solutions.
* Migrated data from legacy systems (SQL Server, Teradata, Cassandra) to cloud based (AWS) Snowflake Data Warehouse
* Designed and Built ETL pipelines using Azure **Databricks** and ingested data from different sources like Snowflake, Teradata, Vertica, Oracle ETC to azure data lake storage and monitored ongoing **Databricks** jobs
* Experience in writing Snowflake’s SnowSQL
* Develops DBT models using SQL to transform data stored in AWS data warehouses such as Amazon Redshift, Amazon RDS, or Amazon Athena.
* Running of Apache Hadoop, CDH and Map-R distributions, **Elastic Map Reduce** (EMR) on (EC2).
* Expert in using **ELK** Stack; Elasticsearch for deep search and data analytics, Logstash for centralized logging, log enrichment and parsing and Kibana for powerful and beautiful data visualizations.
* Good knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, and Map Reduce concepts
* Configured alerting rules and setup **PagerDuty** alerting for Elasticsearch, Kafka, Logstash and different Microservices in Kibana
* Highly experienced in creating different types of Tabular Reports, Crystal Reports, Matrix Reports, Drill-Down, Cross Tab Reports and distributed reports in multiple formats using **SSRS**
* Strong knowledge on creating Extract, Transform and Load (ETL) packages in SQL Server Integration Services for data migration between various databases and Building data sources for **Tableau Desktop**
* Experience including analysis, modeling, design, and development of Tableau reports and dashboards for analytics. Expertise in using **Tableau Server**, **Tableau Desktop** and **Tableau Public**.
* Strongly followed PEP-8 coding standard and test a program by running it across test cases to ensure validity and effectiveness of code using **Pylint**

**PROFESSIONAL EXPERIENCE:**

**Lead Data Engineer,Houston,TX**

**Duration: January 2022 – Till Date**

**Company:** Citi Bank (Remote)

* Design robust, reusable and scalable data driven solutions and data pipeline frameworks to automate the ingestion, processing and delivery of both structured and unstructured batch and real-time streaming data using Python Programming.
* Design and develop ETL Data Pipelines using Pyspark, Spark SQL, Snowflake SQL
* Develop Lambda functions in python using AWS Lambda, Kinesis and Elastic Search which invokes python scripts to perform various transformations and analytics on large data sets in AMAZON EMR.
* Develop Spark applications using Spark tools like RDD transformations and Spark SQL, Hadoop HDFS, Apache Hadoop YARN.
* Spin up an EMR cluster in AWS to run PySpark applications and SSH to master instance to perform spark-submit options to run application in Client and Cluster mode.
* Designed and implemented multiple dashboards for internal metrics using Azure Synapse – PowerPivot & Power Query tools.
* Develop, test, and deploy Kafka streaming applications using Kafka Streams API or alternative frameworks like Apache Spark.
* Implement data transformations, aggregations, and joins within Kafka streams.
* Integrate Kafka with other data systems
* Designed, Built and Analyzed data solutions using Azure and Azure Databricks services. Provided data Visualization using Azure Databricks Dashboards and Power BI and provided data Insights for Business
* Iterate rapidly and work collaboratively with product owners, developers, and other members of the development team.
* Developed DAGs and Setup production environment for the Apache Airflow for scheduling and automation system that managed ETL and reporting
* Utilized Snowflake SQL and ipython Jupyter notebooks to extract, transform, clean and load data to the target to enable effective reporting and business intelligence functions.
* Defect tracking and resolving bug-fixes, hot-fixes by merging the feature branch code to the master branch in GIT source control management.
* Designed and Built ETL pipelines using Azure Databricks and ingested data from different sources like Snowflake, Teradata, Vertica, Oracle ETC to azure data lake storage and monitored ongoing Databricks jobs
* Use Amazon Elastic Cloud Compute (EC2) infrastructure for computational tasks and Simple Storage Service (S3) as Storage mechanism.
* Implemented Data quality checks using Azure Synapse workflows by using Pyspark and spark-SQL by triggering Azure Databricks notebooks.
* Work with different types of data sources such as application data, process data and formats like CSV, JSON, Parquet etc.
* Building processes around CICD pipelines involving integrations with Jenkins, testing frameworks, GitHub, etc. End to end understanding and work experience on CICD pipeline.
* Building Dev (develop), Quality Assurance Testing, User Acceptance Testing, Production environments to test the processes and data in multiple stages
* Provide financial data feeds to Anaplan tool and Cognos reporting
* Integration of S3 with other AWS services like EMR, Lambda, or Glue for data processing workflows.
* Effectively used numpy, pandas, SQLalchemy and sci-kit packages to support the migration in Python
* Troubleshooting, problem solving and performance tuning of queries accessing Snowflake data warehouse
* Supported real-time data handling by making use of Amazon S3 buckets for storing and accessing process results
* Create and maintain source codes in GitHub, kept track of source codes, explored and shared the changes of coding scripts, notes in GitHub.

**Data Engineer,Hyderabad,India**

**Duration: November 2019 – December 2021**

**Company: Dell**,

* Migrate data (Amazon Connect Call Metrics) from legacy systems (Cassandra) to cloud based (AWS) Snowflake Data Warehouse
* Prioritize the development of new tables, schemas, and data structures
* Proactively identify new opportunities to support the business with data, aggregate various data sets to inform business decisions
* Migrate data from legacy systems (Salesforce) to cloud based (AWS) Snowflake Data Warehouse and develop metrics to provide data insights as per business requirement
* Develop ETL pipelines in and out of data warehouse using combination of Python and Snowflake’s SnowSQL
* Write sql queries to perform data quality checks/validation in AWS snowflake as compared to legacy data
* Collaborate with the Business Intelligence Analyst to support data modeling efforts
* Develop presentation layer and Create Tableau scorecards, dashboards using stack bars, bar graphs, scattered plots, geographical maps, Gantt charts using show me functionality
* Design and Optimize Data Connections, Data Extracts, Schedules for background Tasks and Incremental Refresh for the weekly and monthly dashboard reports on Tableau Server
* Data validation of the results in Tableau by validating the numbers against the data in the database tables by querying on the database.
* Complied interactive dashboards in Tableau Desktop and published them to Tableau server which allowed Story telling with Quick filters for on demand needed information and data insights with just one click of a button.
* Schedule AWS snowflake reports through Python airflow, data visualization in AWS S3 cloud box. Automatically upload S3 cloud box files to databases in AWS snowflake through Python airflow.
* Create and maintain source codes in GitHub, kept track of source codes, explored and shared the changes of coding scripts, notes in GitHub.
* Develop a layer of applications modules over the Python - Pandas library, delivered various data frame visualization tools, Data wrangling and cleaning using Python Pandas
* Optimize SQL queries and DBT models for performance within AWS data services
* Schedule DBT runs using AWS CloudWatch Events or AWS Glue workflows.
* Automate DBT model runs using AWS services such as AWS Step Functions or AWS Lambda.
* Develop DAGs and Setup production environment for the Apache Airflow for scheduling and automation system that managed ETL and reporting
* Utilize Snowflake SQL and ipython Jupyter notebooks to extract, transform, clean and load data in the target tables to enable effective reporting and business intelligence functions.
* Effectively used numpy, pandas, SQLalchemy and sci-kit packages to support the migration in Python
* Troubleshooting, problem solving and performance tuning of queries accessing Snowflake data warehouse
* Supported real-time data handling by making use of Amazon S3 buckets for storing and accessing process results

**Data Engineer,Hyderabad,India**

**Duration: May 2017 – October 2019**

**Company: Dell International Services**,

* Migrating an entire oracle database to Big Query and using of power bi for reporting.
* Build data pipelines in airflow in GCP for ETL related jobs using different airflow operators.
* Experience in GCP Dataproc, GCS, Cloud functions, Big Query.
* Experience in moving data between GCP and Azure using Azure Data Factory.
* Experience in bui
* Experience in building power bi reports on Azure Analysis services for better performance.
* Used cloud shell SDK in GCP to configure the services Data Proc, Storage, Big Query
* Coordinated with team and Developed framework to generate Daily adhoc reports and Extracts from enterprise data from Big Query.
* Designed and Co-ordinated with Data Science team in implementing Advanced Analytical Models in Hadoop Cluster over large Datasets.
* Wrote scripts in Hive SQL for creating complex tables with high performance metrics like partitioning, clustering and skewing
* Work related to downloading Big Query data into pandas or Spark data frames for advanced ETL capabilities.
* Worked with google data catalog and other google cloud APr's for monitoring query and billing related analysis for Binary usage.
* Worked on creating POC for utilizing the ML models and Cloud ML for table Quality Analysis for the batch process.
* Knowledge about cloud dataflow and Apache beam.
* Good knowledge in using cloud shell for various tasks and deploying services.
* Created Big Query authorized views for row level security or exposing the data to other teams.
* Expertise in designing and deployment of Hadoop cluster and different Big Data analytic tools including Pig, Hive, SQOOP, Apache Spark, with Cloudera Distribution.

**ETL tester,Chennai,India**

**Duration: Oct 2012 – Apr 2017**

**Company:** Panasonic,

* Involved in understanding of Business logic.
* Involved in running the ETL jobs for ETL process.
* Prepared test data for testing
* Designed SQL Queries for data validations
* Involved in extensive DATA validation using SQL queries and back-end testing.
* Used SQL tools to run SQL queries and validate the data loaded into the target tables.
* Prepared status summary reports with details of executed, passed and failed test cases.
* Defect Analyzing and Reporting in QC
* Interactions with BA & Development teams to resolve the issues.
* Reporting daily testing status report
* Involved in understanding of Business logic.
* Involved in running the ETL jobs for ETL process.
* Prepared test data for testing
* Designed SQL Queries for data validations
* Involved in extensive DATA validation using SQL queries and back-end testing.

**EDUCATION**:

**MCA master of computers applications from JNTU Kakinada**

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Programming Languages | C, C++, Visual Basic, SQL, PYTHON, JavaScript, UNIX Shell, HTML and CSS, Scala |
| Operating System | Windows 2008/2012/2013, XP/Vista/7, Linux |
| Database | MS Access, MySQL, MS SQL SERVER, Oracle, Cassandra |
| Business Intelligence Tools | SSRS, Tableau, SSIS, SSMS, Query Analyzer, BIDS, Cognos |
| Cloud Services | Amazon Web Services (AWS), Kinesis, RedShift CloudWatch, RDS Amazon EC2, AmazonS3,Glue,Lambda,Snowflake, Amazon ELK, Salesforce, Big Data Technologies, Big Query, Cloud Compose |
| IDE’s | Jupiter Notebook, PyCharm, Sublime Text,DataBricks |
| Documenting and Modeling Tools | UML 2.0, MS Project, MS Office, MS Visio, MS SharePoint |
| Area of Expertise & Tools | SDLC, Business Analysis, Database Testing & SQL workbench, GitHub, Rally, JIRA, HP ALM, Confluence, Airflow, PagerDuty |