# ARUN CHILUVERU

 469-213-0230 | Dallas, TX | mr.arunc510@gmail.com |[LinkedIn](http://www.linkedin.com/in/arun-chiluveru)

**Summary**

* Overall 11+ years of technical IT experience in all phases of the Software Development Life Cycle (SDLC) with skills in data analysis, design, development, testing, and deployment of software systems.
* 9+ years of industrial experience in Big Data analytics, Data manipulation, using Hadoop Ecosystem tools Map- Reduce, HDFS, Yarn/MRv2, Pig, Hive, HBase, Spark, Kafka, Flume, Oozie, Avro, Sqoop, AWS, Spring Boot, Spark integration with Cassandra, Avro, Solr, and Zookeeper.
* An excellent understanding of traditional statistical modeling and Machine Learning techniques and algorithms like Regression, clustering, ensembling (random forest, gradient boosting), deep learning (neural networks), etc.
* Created AWS CloudFormation CFT and Terraform templates to build, change & version infrastructure as code.
* Working knowledge of ML & Statistical libraries (e.g. Scikit-learn, Pandas).
* Implemented an end-to-end solution for hosting the web application on AWS cloud with integration to S3 buckets.
* Developed highly scalable classifiers and tools by leveraging machine learning, Apache spark & deep learning
* Experience in developing data pipelines using AWS services, including EC2, S3, Redshift, Glue, Lambda, Step, CloudWatch, SNS, Dynamo, and SQS.
* Provisioned and configured Infrastructure using AWS CloudFormation and Terraform (JSON/YAML).
* Developed highly scalable classifiers and tools leveraging machine learning, Apache Spark & deep learning.
* Experience with Snowflake cloud data warehouse and AWS S3 bucket for integrating data from multiple source systems, which includes loading nested JSON formatted data into Snowflake table.
* Extensive experience in Tableau data visualization tools to analyze and obtain insights into large datasets and create visually compelling and actionable interactive reports and dashboards. Integrated Python dataset with Tableau to build visual analytics solutions.
* Connected Tableau with Python to provide fast visual analytics solutions with data mapping. Established Best Practices for Enterprise Tableau Environment and application intake and development processes.
* Created AWS CloudFormation CFT and Terraform templates to build, change & version infrastructure as code.
* Very Strong in Python, statistical analysis, tools, and modeling.
* Configured AWS route 53 to push traffic to Lambdas closer to customers using CloudFront Edge location.
* Invoked Lambda functions for Automate Event/Schedule Triggers, Backups, Publish & Subscribe notifications.
* My current focus is to find the best way to work with large datasets using Spark and Python, to use the computing power of the available machine memory optimally.
* Deployed and managed serverless app websites and scheduled cron jobs using Lambdas Serverless.
* Developed predictive models on large-scale datasets to address various business problems by leveraging advanced statistical modeling, machine learning, and deep learning.
* Data Engineer knowledge seeker, working on improving my machine learning and statistical skills to deal with different types and sizes of data. My career aims to optimize already found solutions and find effective mathematical solutions to business problems using machine learning algorithms.
* Developed server-side software modules and client-side user interface components and deployed them entirely in Compute Cloud of Amazon Web Services (AWS).
* Modified Redshift UDFs and Lambda functions using Python for custom data transformation and ETL processes.
* Over the years, I gained diverse software, digital marketing, and finance skills, which helped me change my role from a software engineer to a senior data engineer.

# Technical Skills:

Python, PySpark, SQL, R, Machine Learning, Tableau, Snowflake, Big Data, AWS S3, EC2, GCP, RDBMS, Oracle VirtualBox, Hadoop, Map Reduce, HDFS, Hive, Cassandra, Data Lake, Spring Boot, Swamp, Sqoop, Oozie, Kafka, Spark, Scala, Java, GitHub, Talend Big Data Integration, Solr, Impala, EMR, Salesforce.

**Professional Experience**

***CVS, Richardson, TX* Nov 2021 - Present**

 **Role: Lead Data Engineer**

 **Responsibilities:**

* Design, develop and test ETL processes in AWS Glue to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into AWS Redshift.
* Used multiple machine learning algorithms, including random forest and boosted tree, SVM, SGD, neural network, and deep learning using TensorFlow
* Very good hands-on in Spark Core, Spark SQL, Spark Streaming and Spark machine learning using Scala and Python programming languages
* Worked on AWS EMR to run spark and hive applications.
* Configured serverless deployments using AWS Fargate and Container services using ECR/ECS.
* Hands-on experience in Unified Data Analytics with Databricks, Databricks Workspace User Interface, Managing Databricks Notebooks, Delta Lake with Python, Delta Lake with Spark SQL
* Good understanding of Spark Architecture with Databricks, Structured Streaming. Setting Up AWS with Databricks, Databricks Workspace for Business Analytics, Managing Clusters In Databricks, Managing the Machine Learning Lifecycle
* In-depth knowledge of Snowflake Database, Schema, and Table structures.
* Developed the streaming applications using Spark structured streaming, Kafka, and s3 integration project to analyze real-time data.
* Define virtual warehouse sizing for Snowflake for different types of workloads.
* Developed spark applications in Python (PySpark) on the distributed environment to load many CSV files with different schema into Hive ORC tables.
* Developed multiple ETL Hive scripts for data cleansing and transformations for data.
* Exported data from Hive to AWS s3 bucket for further near real-time analytics.
* Use Sqoop to import and export data from RDBMS to HDFS and vice-versa.
* Exporting data to Teradata using SQOOP.
* Containerized Applications using Docker, EKS, ECR & ECS services for managing the application life cycle.
* Analyzed the SQL scripts and designed them using PySpark SQL for faster performance.
* Unit tested the data between Redshift and Snowflake.
* Implemented AWS IAM for managing the user permissions of applications that run on EC2 instances.
* Created AWS Lambda, EC2 instances provisioning on AWS environment implemented security groups, and administered Amazon VPC's.
* Configured ECS cluster & network for automated serverless deployments on the cloud using AWS Faregate
* Worked on performance tuning of Snowflake ETL jobs and implemented row-level security solutions in Snowflake.
* Deployed applications onto AWS lambda with HTTP triggers and integrated them with API Gateway.
* Developed and implemented predictive models of user behavior data on websites, URL categorical, social network analysis, social mining, and search content based on large-scale Machine Learning
* Involved with developing Ansible playbooks with Python and SSH as wrappers for managing AWS node configurations and testing playbooks on AWS instances.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDD, Scala, and Python.
* Experience in developing and designing POCs using Scala and deploying on the Yarn cluster compared the performance of Spark with Hive and SQL/Teradata.

***Mindtree, Hyderabad, India* Apr 2017 – Aug 2021**

# Role: Senior Data Engineer

**Responsibilities:**

* Designed and built Modern data solutions using Azure PaaS service to support data visualization.
* 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.
* Configured Jenkins to create build pipelines, new jobs, managing plugins like Maven, Docker, etc., for build configurations using Groovy and deploying docker images to AWS ECS/ECR & Google Cloud GCR.
* Implemented Lambda to configure Dynamo DB Autoscaling feature and implemented Data Access Layer to access AWS DynamoDB data.
* Developed Spark jobs using PySpark and Spark-SQL for data extraction, transformation & aggregation from multiple file formats.
* Designing and implementing data ingestion pipelines from multiple sources using Azure Databricks.
* Integrate Azure storage with Azure Databricks, Build ELT pipeline using Databricks Jobs and Workflows.
* Orchestrate Azure Databricks applications using ADF pipelines,
* Built data pipelines using ADF pipelines and Databricks.
* Implemented a CI/CD pipeline with Docker, VSTS (TFS Plugin installed), GitHub, and Azure Container Service
* Worked on AWS Services like AWS SNS to send automated emails and messages using BOTO3 after the nightly run.
* Configured Jenkins to create build pipelines, new jobs, managing plugins like Maven, Docker, etc., for build configurations using Groovy and deploying docker images to AWS ECS/ECR & Google Cloud GCR.
* Designed and implemented ADHOC data pulls using Azure Data Factory from on-premises onto SQL DW.
* Worked on moving data from Blob storage containers to Azure Synapse using ADF.
* Worked on developing tools that automate AWS server provisioning, automated application deployments, and implementation of primary failover among regions through AWS SDK’s.
* Designed and built ETL pipelines to automate structured and unstructured data ingestion.
* Optimization of Hive queries using best practices and suitable parameters and using technologies like Hadoop, Python, and PySpark.
* Evaluate Snowflake Design considerations for any change in the application
* Knowledge in performance troubleshooting and tuning Hadoop clusters.
* Advanced knowledge in performance troubleshooting and tuning of services in Azure like HD insight clusters, ADF, and Databricks Network.
* Applied unsupervised and supervised learning methods in analyzing high-dimensional data. Proficient use of Python Scikit-learn, pandas, and NumPy packages.
* Worked on Python and shell scripting to automate and schedule the workflows to run on Azure.
* Develop ETL Process using SPARK, SCALA, HIVE, and HBASE.
* Developed REST APIs using Scala, Play framework, and Akka.
* Developed Python AWS serverless lambda with concurrent and multi-threading to make the process faster and asynchronously execute the callable.
* Connected Tableau with Python to provide fast visual analytics solutions with data mapping. Established Best Practices for Enterprise Tableau Environment and application intake and development processes.
* Excellent knowledge in Data Analysis, Validation, Cleansing, Verification, and identifying data mismatch. Meet with clients to determine business, functional, and technical requirements and participate in application design, configuration, testing, and deployment.
* Experience in data analysis, data visualization using Tableau/Microsoft Power BI, model building with machine learning algorithms for prediction and forecasting using data (historical or time series with regression techniques), using statistical/econometric tools like SAS, and data mining using Python, SQL, Hadoop, Spark, Hive.

***Accenture, Bangalore, India* Apr 2014 – Mar 2017**

# Role: Software Engineer

**Responsibilities:**

* Involved in building data pipelines to ingest and transform data using spark and loading the output in multiple sources like Hive and HBase.
* Containerized Applications using Docker, ECR/ECS & PCF services.
* Developed component using Spark Structured Streaming API to form data frame from real-time streaming data on Kafka and performed aggregations on the same.
* Define virtual warehouse sizing for Snowflake for different types of workloads.
* Perform data validation and reconciliation on raw data in the Data Lake using Spark.
* Responsible for validating transactional and profile data from RDBMS, which are transformed and loaded to Data Lake using Hadoop Bigdata technologies.
* Involved in performance tuning of Spark Applications for setting the right Parallelism and memory tuning level.
* Designed, developed, and did maintenance of data integration programs in a Hadoop and RDBMS environment with both traditional and non-traditional source systems as well as RDBMS and NoSQL data stores for data access and analysis.
* Responsible for developing data pipelines with AWS cloud to extract the data from weblogs and store it in HDFS.
* Developed Spark and Python for regular expression (regex) project in the Hadoop/Hive environment with Linux/Windows for big data resources.
* Monitored containers in AWS EC2 machines using Datadog API and ingested and enriched data into the internal cache system.
* Developing Hive queries to process the data and generate the data cubes for visualization.
* Working extensively with Sqoop for importing metadata from SQL server.
* Worked on a Cluster of size 130 nodes.
* Analyzing the SQL scripts and designing the solution to implement using PySpark.
* Developed Scala scripts, UDFFs using both Data frames/SQL/Data sets and RDD/MapReduce in Spark 1.6 for Data Aggregation, queries, and writing data back into the OLTP system through Sqoop.
* Developed Spark scripts by using Scala shell commands as per the requirement.
* Developed Spark code using Scala and Spark-SQL for faster processing and testing.
* Used Reporting tools like Tableau to connect with Hive for generating daily data reports.
* Designed, developed, and maintained Tableau functional reports based on user requirements.
* Experience in the design and development of Tableau visualization solutions
* Preparing Dashboards using calculations, parameters in Tableau and creating SQL datasets for Tableau and Ad-hoc reports.
* Experienced in creating the groups, users, workbooks, and permission sets for the tableau server logins and checks.
* Extensively developed different visualizations in Tableau like Dual Axis, bar charts, Waterfall charts, Bullet graphs, box, and whisker plots.
* Collaborated with the infrastructure, network, database, application, and BI teams to ensure data quality and availability.

***Waracle, Dundee, Scotland* Dec 2011 – Oct 2013**

# Role: Software Engineer

**Responsibilities:**

* Developed an iOS doctor-assistance application that enables patients to search for nearby physicians who can provide medical advice.
* Developed client-specific e-commerce websites using HTML, CSS, JavaScript, and WordPress.
* Involved in developing iOS and Web-based applications.
* Defined worksite objectives using Java platform-based network management systems by analyzing user requirements and imagining system features and functionality.

# Education:

* **Bachelor of Technology** in **Computer Science Engineering** from **Jawaharlal Nehru Technological University**.

# Master of Science in Computer Games Technology from the University of Abertay, Dundee.