**ABHIRAM NIMMAGADDA**

# PROFESSIONAL SUMMARY

* Self-driven and creative person with about 9**+ years** of experience in data analysis, data visualization, and data engineering. Highly skilled in leveraging a wide range of tools and technologies, including **Python, PySpark, Apache Kafka, Azure, AWS, Spark Streaming, Tableau, PL/SQL, and Oracle, to extract, transform (ETL), and analyze complex datasets.**
* Proficient in data processing, **Spark SQL**, Data Frame Operations, RDD Transformations, and handling a variety of file types, including Parquet, Flat, Split, delimiter, JSON, XML, etc.; has experience with both **Apache Spark** and **PySpark**.
* extensive practical experience in data management, data strategy, data governance, and data operations.
* Knowledge in various **GCP** services including **BigQuery**, **Pub/Sub**
* Practical knowledge of relational and dimension data modeling concepts, such as fact and **dimension tables**, **snowflake** schema modeling, and star schema modeling.
* Experienced with **Snowflake** as a **data warehousing** solution, proficient in optimizing data processes, performance tuning, and ensuring data integrity.
* Proficient in **PL/SQL**, with skills in creating and managing views, schemas, and cursors for data transformation. Optimized data extraction with efficient window functions, joins, and subqueries.
* created and managed **PL/SQL** packages, procedures, functions, triggers, and indexes to maintain a reliable database system. conducted data extraction and transformation using **PL/SQL** and **SQL** from ETL.
* expertise in developing **Extract, Transform**, and **Load** (ETL) solutions for Data Warehousing **Applications** and **Data Integration** and Workflow Solutions using **SQL Server** Integration Services (SSIS).
* competent in data security implementation, query optimization, schema design, and data modeling, ensuring **data accuracy, performance**, and **compliance** across the data lifecycle.
* Proven capacity to manage massive data sets, up to terabytes in size, and optimize data processing for improved performance, all of which support effective data workflows.
* knowledgeable about utilizing cloud-based services to process and store data seamlessly, allowing for scalable and affordable solutions. Examples of these services include **Amazon S3, Redshift, AWS Glue, Azure Data Factory, and Azure SQL Database.**
* Strong familiarity with data visualization and reporting tools like **Tableau, PowerBI, and AWS Glue**, which helps to provide stakeholders with actionable insights.
* competent with version control systems such as **Git** and **Azure DevOps**, guaranteeing teamwork and **code integrity**, **code automation**
* outstanding teamwork and communication abilities, as well as a track record of converting business needs into technical solutions and operating successfully in **cross-functional teams**.
* Proven track record of optimizing data processing speed, reducing operational costs, and enhancing data quality through innovative solutions.
* For proactive issue resolution, **error-handling** mechanisms were developed, as well as monitoring

and alerting systems, strong **problem-solving** and **analytical thinking**

# TECHNICAL SKILLS

* **Hadoop/Spark Ecosystem**: Hadoop, MapReduce, Hive/impala, YARN, Kafka, Flume, Sqoop, Oozie,

Zookeeper, Spark

* **Databases:** Oracle, MySQL, SQL Server, PostgreSQL, HBase, Snowflake, Cassandra, MongoDB
* **Cloud computing:** Amazon Web Services (AWS), Google Cloud Platform (GCP), Amazon Redshift, MS Azure, Azure blob storage,
* Azure Data Factory, Azure Synapse
* **BI Tools:** Business Objects XI, Tableau 9.1, Power BI
* **Query Languages:** SQL, PL/SQL, T-SQL
* **Scripting Languages:** Unix, Python
* **Operating Systems**: Linux, Windows, Ubuntu, Unix
* **SDLC Methodology**: Agile, Scrum, Waterfall, UML

# PROFESSIONAL EXPERIENCE

**Elevence Health, Virginia, VA NOV 2021- Current**

**AWS Big Data Engineer**

**Responsibilities:**

* Create and implement data pipelines for both individual and zipped encryption and decryption. files using boto3 in an S3 bucket. data that has undergone **PySpark** transformations.
* Developed numerous complex views and **adhoc** SQL queries based on business needs in Snowflake, utilizing complex functions like array agg, split by table, and map agg to cut down on query latency by 82%.
* Developed Airflow Dag’s with **Python** for Task Schedule and automated daily run.
* Migrated on-premises **SAP ASE database** to **AWS EC2** using **AWS database**

**migration service (DMS).**

* Designed and developed internal data infrastructure and analytics support from the

ground up by deploying custom **Python Batch ETL** subsystems using **Docker**, establishing a data warehouse on **AWS redshift**.

* Developed **ETL pipelines** using SSIS to acquire call volume data from multiple sources and build a dataset to be used for call volume forecast by data scientists.
* Created **Python** scripts for data access and analysis to aid in the process and system monitoring, and reporting.
* Integrated **SAP SRM** with **SAP IBO** for improved process monitoring for a major telecom company with 100% defect-free code at the time of Go-Live.
* Developed ERMS (Employee Management Response System) which helped to reduce downtime on tickets handled by the customer service agent.
* Implemented **SAP CRM** (Customer Relationship Management) Service Module.
* Designed dashboard using **Tableau** to monitor the movement of customers between deciles based on their propensity score to inform businesses to take necessary actions.
* Engineered and executed end-to-end **ETL pipelines** using **SSIS** to provide a platform for the alternative dataset.
* Building and optimizing **OLAP** cubes for analytical queries
* Worked to standardize documentation formats and practices across the team, ensuring consistency
* and ease of use
* Integration of **JSON** processing and **REST API** interactions into SSIS projects, demonstrating
* expertise in handling diverse **JSON** structures and **API** response formats
* Optimized the Hive tables using optimization techniques like **partitions** and **bucketing** to provide better performance with **HiveQL queries**.
* Implemented POC to develop a Data OPS solution for **AWS** Services using **S3, Glue, Lambda, Code Pipeline, and Terraform**.
* Developed intent classification model using **python NLP** libraries spacy, **NLTK** to identify the customer intent for smart routing of email to service reps based on tags like inquiry, feedback, and dissatisfaction.

**Environment:** Linux, Python, R, pandas, R, Hadoop, SAPCRM, Erwin, SRM, NLP, spacy, NLTK, Genism, SQL, Tableau, AWS IAM, S3, SNS, SQS, EC2, VPC,EMR, LAMBDA, Redshift, Glue, BigQuery, Crawler, Athena, QuickSight, IAM, CLI, Databricks, Data Lake, Dynamo DB, Apache Spark, Spark SQL, Hadoop, Hive, Kafka, Apache Airflow, HDFS, kudu, Scala, Cloudera, CDSW, Splunk, Data Lake, Power BI, Tableau, Kubernetes

**Infosys |Richardson, TX Feb 2019-OCT 2021**

**Azure Data Engineer**

**Responsibilities:**

* Implemented large Lambda architectures using Azure Data platform capabilities like **Azure Data** **Lake, Azure Data Factory, HDInsight, Azure SQL Server**.
* Experience in developing spark applications using **spark-SQL** in **Databricks** for data extraction, transformation and aggregation from multiple formats for analyzing and transforming the data to uncover insights into the customer usage patterns.
* Analyzing the Data from different sources using Big Data Solution Hadoop by implementing Azure **Data Factory, Azure Data Lake, Azure Data Lake Analytics, HDInsight’s, Hive, and Sqoop**.
* Used notebooks, **Spark Data frames, SPARK SQL** and python scripting to build ETL pipelines in **Databricks**.
* Developed and maintained the data pipeline on Azure Analytics platform using **Azure Databricks**, PySpark, and Python.
* Involved in building an Enterprise **Data Lake** utilizing Data Factory and Blob storage, empowering different groups to work with more perplexing situations and ML solutions.
* Engaged on Azure Functions to write the code in Python to perform different transformations task and performed analytics in HD Insight on large volume of data.
* Automated the loading of data into Blob Storage with **Data Factory, PySpark**, and extracted the required data in Blob Storage using HD Insight and loaded it into HDFS.
* Developed an automated process in Azure cloud which can ingest data daily from web service and load into **Azure SQL DB**.
* Analyzed & transformed data to uncover insights from multiple file formats with data transformation and aggregation with spark applications using **PySpark** and **SparkSQL.**
* Monitor Resources using Azure Automation and create alarm for **VM, Blob Storage, ADF**, **Databricks**, **Synapse Analytics** based on different events that are occurring.
* Designed end to end scalable architecture to solve business problems using various Azure Components like **HDInsight, Data Factory, Data Lake, Azure Monitoring, Key Vault, Function app** and **Event Hubs**.
* Good experience in tracking and logging end to end software application build using Azure DevOps.
* Used **Terraform** script for deploying applications for higher environments.
* Involved in various **SDLC** Life cycle phases like development, deployment, testing, documentation, implementation & maintenance of application software.
* Responsible for estimating the cluster size, monitoring, and troubleshooting of the **Spark data** **bricks cluster**.
* Executed on an Azure copy to load data from an on-premises SQL server to an Azure SQL Data warehouse.
* Worked on re-designing the existing architecture and implementing it on **Azure SQL**.
* Experience with Azure SQL database configuration and tuning automation, vulnerability assessment, auditing, and threat detection.
* Collaborated with business stakeholders to gather data requirements and provide data solutions aligned with Syneos Health's strategic objectives for clinical research.

**Environment:** Spark (PySpark, SparkSQL, SparkMLIib), Python (Scikit-learn, NumPy, Pandas), Tableau, GitHub, AWS EMR/EC2/S3/Redshift, Pig, Azure SQL, Spark data bricks, BigQuery, Spark SQL, Terraform, Spark data bricks, SDLC, Key Vault, PyArrow

**KTREE COMPUTER SOLUTIONS- HYDERABAD**

**Data Stage Developer Sep 2018- Jan 2019**

**Responsibilities:**

* Lead the architectural design and implementation of a scalable data warehousing solution using **Amazon Redshift**. Ensured the **seamless integration** of diverse data sources into the data warehousing environment.
* **Developed** and **optimized** end-to-end **ETL processes**, utilizing **Python** for scripting and Spark for large-scale data transformations.
* **Implemented** efficient data extraction **strategies**, transforming **raw data** into structured formats suitable for analytical purposes.
* Leveraged Snowflake Cloud as a central component for cloud-based data warehousing, ensuring the utilization of its unique features for data storage and processing.
* Designed, implemented, and managed complex workflows using Apache Airflow for orchestrating **ETL** tasks. Ensured effective **scheduling, monitoring**, and **error handling** within the workflow management system.
* Worked on multi-tier applications using **AWS services (EC2, S3, RDS, Dynamo DB, SNS, SQS, IAM)** focusing on high-availability, **fault tolerance**, and **auto-scaling** in **AWS Cloud Formation**.
* Implemented comprehensive data profiling to ensure data quality and identify anomalies or inconsistencies. Utilized **AWS** **Glue** for automated ETL job execution, simplifying the development, testing, and maintenance of data pipelines.
* Integrated **Amazon S3** for scalable and cost-effective storage, facilitating seamless data storage and retrieval. Conduct in-depth data analysis using **Informatica**, exploring and understanding complex relationships within datasets.
* Architect and implemented data warehousing solutions using **Snowflake** Cloud Data Warehouse. Utilized external cloud storage solutions like **AWS S3** or **Azure Blob** Storage for data staging and **integration**.
* Integrated **version** **control** systems to manage code changes systematically.
* Optimized **Glue** jobs for efficient data transformation, leveraging **PySpark** for complex processing tasks. Implemented **CI/CD** pipelines to automate the testing and deployment of **ETL** processes, ensuring a reliable and repeatable release process.
* Created an **AWS S3** **buckets**-based **Data** **Lake**. Object tagging and a data **catalog** for metadata were used to manage data and Isaac was implemented to automate the provisioning of **EMR** **clusters** and resources like roles and **security** **groups**.
* **Monitored** and **optimized** the performance of data **pipelines** and data warehousing platforms. Implemented **tuning** and **optimization** strategies to enhance overall system efficiency.

**Environment** Jira dashboard, Linux, AWS lambda, Python, GIT, Spark, Scala, MySQL, PostgreSQL, Snowflake data warehouse, AWS S3, AWS kinesis, Snow SQL, Microsoft Word/Excel, Flask, Snowflake, DynamoDB, Athena, Lambda, MongoDB, Pig, Sqoop, Tableau, Power BI and UNIX, Docker, Kubernetes.

**HCL Technologies-BANGALORE**

**HSBC Aug 2016 – Aug 2018**

**Informatica Developer**

**Responsibilities:**

* Designed the **ETL processes** using **Informatica** to load data from **Oracle** Flat Files **XML** Files to target **Oracle Data Warehouse** database.
* Worked with business analysts to identify appropriate sources for **Data** warehouse and to document
* business needs for decision support data.
* Led successful migration from Informatica to modularized **GCP** BigQuery code, adept in

Informatica and skilled in managing data ingestion using **FiveTran**, Workday, and

scheduling tools like **Control M**

* Possesses adeptness in rapid **SQL** query execution, data warehousing, modeling, and machine

learning enablement within Google Cloud Platform. Developing comprehensive skills in Google data structures, while preferring experience in Amex and familiarity with Cornerstone

DAM, Automated Case Workflow Management, ISP, and Inventory Management tools.

* Performed data **manipulations** using various Informatica **Transformations** like Joiner Expression **Lookup** **Aggregate** Filter Update Strategy and **Sequence Generator** etc.
* Involved in creating logical and physical data modeling with **STAR** and **SNOWFLAKE** schema techniques using **Erwin** in Data warehouse as well as in Data Mart.
* Created sessions and workflow for designed **mappings**.
* Written per session and post session scripts in mappings.
* Extensively worked in the performance tuning of the programs **ETL** Procedures and processes.
* **Migrated** the **mapping** to the testing and production department and introduced the concepts of **Informatica** to the people in the testing department.
* Developed **PL**/**SQL** procedures for **processing** business logic in the database.

# Environment: **PowerCenter, Informatica power center**, **Data Quality, Master Data Management (MDM),** **Informatica Data Integration Hub**

# EDUCATION

**Master of Engineering in Computer Science:** Trine University, Angola, Indiana

**Bachelor of Technology in Computer Science:** Prist University, Thanjavur, Tamil Nadu, India