

# ANITHA SAI SRI BOKKA

+1(913) 407-6697



[anithassri@gmail.com](mailto:anithassri@gmail.com)

## Professional Summary

---

- 4+ years of professional experience in information technology as Data Engineer with an expert hand in the areas of Database Development, ETL Development, Data modelling, Report Development and Big Data Technologies.
- Experience in Data Integration and Data Warehousing using various ETL tools Informatica Power Center, AWS Glue, SQL Server Integration Services (SSIS), Talend.
- Proficient in Designing Business Intelligence Solutions with Microsoft SQL Server and using MS SQL Server Integration Services (SSIS), MS SQL Server Reporting Services (SSRS) and SQL Server Analysis Services (SSAS).
- Extensively used Informatica PowerCenter, Informatica Data Quality (IDQ) as ETL tool for extracting, transforming, loading and cleansing data from various source data inputs to various targets, in batch and real time.
- Experience working with Amazon Web Services (AWS) cloud and its services like Snowflake, EC2, S3, RDS, EMR, VPC, IAM, Elastic Load Balancing, Lambda, RedShift, Elastic Cache, Auto Scaling, Cloud Front, Cloud Watch, Data Pipeline, DMS, Aurora, ETL and other AWS Services.
- Strong expertise in Relational Data Base systems like Oracle, MS SQL Server, Teradata, MS Access, DB2 design and database development using SQL, PL/SQL, SQL PLUS, SQL - LOADER. Highly proficient in writing, testing and implementation of triggers, stored procedures, functions, packages, Cursors using PL/SQL.
- Hands on Experience with AWS Snowflake cloud data warehouse and AWS S3 bucket for integrating data from multiple source system which include loading nested JSON formatted data into Snowflake table.
- Experience in building and architecting multiple Data pipelines, end to end ETL and ELT process for Data ingestion and transformation in AWS.
- Extensive experience in Data Mining solutions to various business problems and generating data visualizations using Tableau, PowerBI and SSRS
- Well knowledge and experience in Cloudera ecosystem such as HDFS, Hive, SQOOP, HBASE, Kafka, Data pipeline, Data analysis and processing with Hive SQL, IMPALA, SPARK, SPARK SQL.
- Experienced in design, development, Unit testing, integration, debugging and implementation and production support, client interaction and understanding business application, business data flow and data relations.
- Using Flume, Kafka and Spark streaming to ingest real time or near real time data to HDFS.
- Worked on AWS Data Pipeline to configure data loads from S3 into Redshift.
- Developed Python scripts to parse the Flat Files, CSV, XML, JSON files and extract the data from various sources and load the data into data warehouse.
- Developed Automated scripts to do the migration using Unix shell scripting, Python, Oracle/TD SQL, TD Macros and Procedures.
- Good Knowledge on No SQL database like MongoDB
- Expert-level mastery in designing and developing complex mappings to extract data from diverse sources including flat files, RDBMS tables, legacy system files, XML files, Applications, COBOL Sources & Teradata.
- Worked on JIRA for defect/issues logging & tracking and documented all my work using CONFLUENCE.
- Experience with ETL workflow Management tools like Apache Airflow and have significant experience in writing the python scripts to implement the workflow.
- Experience in identifying Bottlenecks in ETL Processes and Performance tuning of the production applications using Database Tuning, Partitioning, Index Usage, Aggregate Tables, Session partitioning, Load strategies, commit intervals and transformation tuning.
- Worked on performance tuning of user queries by analyzing the explain plans, recreating the user driver tables by right Primary Index, scheduled collection of statistics, secondary or various join indexes.

## TECHNICAL SKILLS

---

<b>RDBMS/Databases</b>	MSSQL Server, Oracle 9i/10g, MS Access, MYSQL, MongoDB, Redshift, Teradata, PostgreSQL
<b>ETL Tools</b>	AWS Glue, Talend, SQL Server Integration Services (SSIS), Informatica
<b>Reporting Tools</b>	Tableau, Power BI, SQL Server Reporting services (SSRS)
<b>Operating systems</b>	Windows 7/8/10, MS - Dos, Unix
<b>Big Data Technologies:</b>	Spark, Pyspark, Hadoop, Hive, Kafka, Databricks, Spark-SQL
<b>Scripts</b>	T SQL, Python, VB, JavaScript, PHP, PL/SQL

## Experience

---

<b>Data Engineer</b>	Jan'24 –Present
<b>DirecTV</b>	Charlotte, NC

- Designed and developed ETL processes in AWS glue to migrate campaign data from external sources like S3, ORC/Parquet/Text files to AWS redshift
- Monitored and maintained Hadoop cluster performance, ensuring high availability and reliability.
- Integrated PySpark with big data technologies like Hadoop, Hive, and Kafka to streamline data ingestion and real-time data processing.
- Demonstrated experience in design and implementation of Statistical models, Predictive models, enterprise data model, metadata solution and data lifecycle management in both RDBMS, Big Data environments.
- Utilized domain knowledge and application portfolio knowledge to play a key role in defining the future state of large, business technology programs.
- 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 SQL and also performed dataset sorting and merging.
- Tested Complex ETL Mappings and Sessions based on business user requirements and business. rules to load data from source flat files and RDBMS tables to target tables.
- Hands-on database design, relational integrity constraints, OLAP, OLTP, Cubes and Normalization (3NF) and De-normalization of the database.
- Developed MapReduce/Spark Python modules for machine learning & predictive analytics in Hadoop on AWS.
- Worked on customer segmentation using an unsupervised learning technique - clustering.
- Worked with various Teradata 15 tools and utilities like Teradata Viewpoint, Multi-Load, ARC, Teradata Administrator, BTEQ and other Teradata Utilities.
- Worked on EMR, S3 and EC2 services in AWS cloud and migrating servers, databases, applications from on premises to AWS
- Worked on centralized logging solutions using tools like Grafana Loki, consolidating logs from various sources for efficient analysis and troubleshooting.
- Devised PL/SQL Stored Procedures, Functions, Triggers, Views and packages. Made use of Indexing, Aggregation and Materialized views to optimize query performance.
- Implemented FTP operations using Talend Studio to transfer files in between network folders as well as to FTP server using components like tFileCopy, TFileArchive, tFileDelete, tCreateTemporaryFile, tFTPDelete, tFTPCopy, tFTPRename, tFTPput, EFTP Get etc.
- Ingested data from disparate sources using SQL and Google Analytics API to construct data views in Tableau and created Tableau dashboards/reports for data visualization, Reporting and Analysis and presented it to Business.
- Created Data Connections, published on Tableau Server for usage with Operational or Monitoring Dashboards.
- Integrated Docker and Kubernetes with CI/CD pipelines, automating build, test, and deployment processes.

- Responsible for daily communications to management and internal organizations regarding status of all assigned projects and tasks.
- Participated in Agile/SAFE planning for 60+ data feature requests.

**Environment:** AWS Glue, S3, Redshift, Hadoop, PySpark, Hive, Kafka, SQL, PL/SQL, Talend, Tableau, Google Analytics API, EMR, EC2, Teradata, Talend, FTP Server

**Data Engineer**  
**BMO**

Nov'20-Dec'22  
Chicago, IL

- Orchestrated the seamless migration of legacy data systems to MongoDB on AWS, ensuring minimal downtime and preserving data integrity throughout the process.
- Utilized AWS Glue for ETL (Extract, Transform, Load) processes, ensuring data consistency and integrity across various CRM and sales systems.
- Employed AWS Lambda for serverless execution of data validation scripts, allowing for efficient scaling and cost-effective resource utilization during data processing.
- Configured AWS Step Functions to orchestrate complex data workflows, ensuring error-free, sequential, and parallel execution of data validation and quality management tasks.
- Implemented Python scripts for orchestration of batch processes, ensuring robust data validation and ETL workflows.
- Conducted meticulous end-to-end testing of SQL processes post-migration to AWS Redshift, utilizing the ELK stack, DVC, and Jenkins to guarantee data integrity and functional parity with Teradata systems.
- Utilized Spark, PySpark, and Scala for high-performance data processing and transformation tasks, improving processing times and scalability.
- Leveraged PySpark extensively for data processing tasks within Databricks environments, developing and optimizing PySpark scripts to perform complex data transformations, aggregations, and analytics on large datasets.
- Utilized Spark-SQL to read and process data in Parquet format, facilitating efficient data analysis and reporting.
- Implemented advanced data manipulation techniques such as pivot queries and hierarchical queries using PL/SQL.
- Created a defect tracker using JIRA and integrated it with Jenkins to monitor and resolve data migration issues, establishing a transparent and effective issue resolution protocol.
- Enhanced data migration security and governance using AWS IAM, automated infrastructure setup with AWS CloudFormation, and facilitated real-time streaming analytics with Amazon Kinesis.
- Troubleshoot and resolved data process issues by analyzing logs and backtracking errors, ensuring high data quality and system reliability using Talend, Hadoop, and the ELK stack.
- Collaborated with DevOps teams to automate deployment and configuration management of MongoDB instances on AWS using Infrastructure as Code (IaC) tools like AWS CloudFormation and Terraform.
- Implemented CI/CD pipelines for AWS Glue jobs using services like AWS CodePipeline and AWS CodeBuild, enabling automated testing, deployment, and versioning of ETL code.
- Built a program with Python and Apache Beam, executing it in cloud Dataflow to run data validation between raw source files and BigQuery tables.
- Applied required transformations using AWS Glue and loaded data back to Redshift and S3.
- Extensively worked on making REST API calls to get data as JSON responses and parse it.
- Experience in analyzing and writing SQL queries to extract data in JSON format through REST API calls with API Keys, ADMIN Keys, and Query Keys, and loading the data into data warehouses.
- Extensively worked on Informatica tools like Source Analyzer, Mapping Designer, Workflow Manager, Workflow Monitor, Maplets, Worklets, and Repository Manager.
- Designed and implemented ETL pipelines from various relational databases to the data warehouse using Apache Airflow.
- Worked on data extraction, aggregations, and consolidation of Adobe data within AWS Glue using PySpark.
- Built a Scala and Spark-based configurable framework to connect common data sources like MySQL, Oracle, and Postgres.

- Configured row-level security (RLS) in Power BI to ensure that sensitive data is accessible only to authorized users, maintaining data confidentiality and compliance.
- Designed and developed interactive dashboards and reports in Power BI to provide actionable insights and support data-driven decision-making across various business units.
- Strong background in data warehousing, business intelligence, and ETL processes (Informatica, AWS Glue) with expertise in working on large data sets and analysis.

**Environment:** AWS Glue, MongoDB, AWS Lambda, Step Functions, Python, Redshift, ELK Stack, Jenkins, Spark, PySpark, Scala, Databricks, Parquet, PL/SQL, JIRA, AWS IAM, CloudFormation, Kinesis, Talend, Hadoop, Airflow

## **SQL Developer**

May'19 - Nov'20

### **United Health care**

Minnetonka, MN

- Generated SQL and PL/SQL scripts to install create and drop database objects, including tables, views, primary keys, indexes, constraints, packages, sequences, grants and synonyms.
- Maintained a series of UNIX Shell Scripts to manage batch order processing.
- Developed dynamic SQL scripts to handle diverse data processing tasks, including data cleansing, aggregation, and transformation.
- Utilized tools such as TKPROF, SQL Profiler, and execution plans to identify and resolve performance bottlenecks.
- Tune SQL statements using hints for maximum efficiency and performance, create and maintain/modify PL/SQL packages, mentor others with the creation of complex SQL Statements, perform data modeling and create/maintain and modify complex database Triggers and Data Migration scripts.
- Developed and maintained reports using SQL Server Reporting Services (SSRS), providing users with timely and accurate business intelligence.
- Implemented and enforced database security measures, including user roles, permissions, and encryption, to protect sensitive data and ensure compliance with regulatory requirements.
- Created, debugged, and modified Stored Procedures, Triggers, Tables, Views and User-Defined Functions.

**Environment:** Oracle 11g, Oracle 9i, SQL Developer, SQL Loader, TOAD, UNIX, Sun Solaris, Server Reporting Services (SSRS)

---