**SOWMYA EPPA**

**Ph: 717-455-0829 Email: sowmyaeppa98@gmail.com**

**Sr Data Engineer**

**PROFESSIONAL SUMMARY:**

* Around 9+ years of IT Experience in Database Design, Data Analyst and Business Analyst, Test and Production Environments.
* Strong expertise in using Hadoop big data technologies Apache Spark, Scala, python, Kafka, Cassandra, Jenkins Pipelines, Kubernetes, Kibana, Rancher, GITHUB, Rancher, Kibana, Hadoop HDFS, Hive, IntelliJ, Cassandra, SQL server etc.
* Experience working in Data warehouse Design, methodologies Extraction, Transformation and Loading, Data Analysis, Data migration, Data preparation, Data Validation, Data Quality, Data Profiling Reporting and Documentation.
* Developed data pipelines for moving data from on-premises (Oracle) to Google Cloud Platform, transforming data from in Google Big Query.
* Knowledge about setting up Python REST API Framework using Django.
* Worked with Google Cloud (GCP) Services like Compute Engine, Cloud Functions, Cloud DNS, Cloud Storage and Cloud Deployment Manager and SaaS, PaaS and IaaS concepts of Cloud Computing and Implementation using GCP.
* Manager and SaaS, PaaS and IaaS concepts of Cloud Computing and Implementation Worked with Google Cloud (GCP) Services like Compute Engine, Cloud Functions, Cloud DNS, Cloud Storage and Cloud Deployment using GCP.
* Designing and developing Spark jobs with data frames using different file formats like Text, Sequence, Xml, parquet, and Avro
* Design, implement, and support on-premises and cloud-based data infrastructure that is resilient to disruptions and failures.
* Participate in designing and building Azure data platforms supporting both batch and real-time (event-based) architecture.
* Have familiarity with the technology stack available in the industry for metadata management, including Data Governance, Data Quality, MDM, Lineage, and Data Catalog.
* Expertise in SQL Server Integration Services (SSIS) and SQL Server Reporting Services (SSRS) with good knowledge on SQL Server Analysis Services (SSAS).
* Proficient of AWS services like VPC, Glue Pipelines, Glue Crawler, Cloud front, EC2, ECS, EKS, Elastic beanstalk, Lambda, S3 Storage, RDBS, Dynamo db, Redshift, Elastic Cache, DMS, SMS, Data Pipeline, IAM, WAF, Artifacts, API gateway, SNS, SQS, SES, Auto Scaling, Cloud Formation, Cloud Watch and Cloud Trail
* Worked with various transformations like Normalizer, expression, rank, filter, group, aggregator, lookups, joiner, sequence generator, sorter, SQLT, stored procedure, Update strategy, Source Qualifier, Transaction Control, JAVA, Union, CDC etc.
* Used Python with OpenStack, Open ERP (now ODOO), SQL Alchemy, Django CMS.
* Excellent experience with Requests, NumPy, Matplotlib, SciPy, PySpark and Pandas python libraries during development lifecycle and experience in developing APIs for the application using Python, Django, MongoDB, Express, ReactJS, and NodeJS.
* Strong experience with spark real time streaming data using Kafka and Spring boot API.
* Handled operations and maintenance support for AWS cloud resources, including launching, maintaining, and troubleshooting EC2 instances, S3 buckets, Auto Scaling, DynamoDB, AWS IAM, and Elastic Load Balancers (ELB) and Relational Database Services (RDS). Also created snapshots for data to store in AWS S3.
* 2+ years of Strong expertise in using ETL Tool Informatica Power Center 10.x/9.x/8.x (Designer, Workflow Manager, Repository Manager, ETL and Data Warehouse.
* Experienced in using advanced concepts of Informatica like push down optimization (PDO).
* Experience in Performance Tuning and Debugging of existing ETL processes.

TECHNICAL SKILLS:

Languages: Core Java, Java 1.x to 1.8 SQL, PL/SQL, Python, Scala, PySpark

Databases: Oracle 9i/10g/11g/12, SQL Server 2000/2005, MS SQL, Mongo DB, MySQL, Cassandra, DynamoDB, PostgreSQL

Bigdata Technologies: Hadoop, Hive, Pig, Azure, GCP, Flume, Sqoop, Spark, MapReduce, Oozie

ETL Tools: Informatica Power Center 10.x/9.x/8.x, SQL Loader, Talend

Development IDE`s: Eclipse, Visual Studio Code, JetBrains, Toad, SQL Developer

Logging & Monitoring: Splunk, CloudWatch, Log4J, SLF4J, Zipkins, Grahana

Operating Systems: UNIX, Linux, Ubuntu, Windows XP/2000/VISTA

Unix & Linux Unix Grid computing and Shell Scripting

Cloud Technologies: AWS (Lambda, EC2, S3, SNS, CloudWatch, Cloud Formation Template, RDS, VPC, Auto Scaling, IAM, AWS Glue, AWS Batch, AWS DMS, Code Build, Code Deploy etc.,)

Version Control Tools: CVS, SVN, GitHub, and Bitbucket

Test Frameworks: Junit, Mockito

**PROFESSIONAL EXPERIENCE:**

**AT&T, Seattle, WA Aug 2022 - Present**

**Data Engineer**

**Roles and Responsibilities:**

* Worked extensively on AWS S3 data transfer and AWS Redshift was used for cloud data storage.
* Handled data extraction and data ingestion from different data sources into S3 by creating ETL pipelines using Spark.
* Extensively worked with Pyspark / Spark SQL for data cleansing and generating data frames and RDDs.
* Experience working in Multi cloud architecture designing, building multiple Data pipelines, end to end ETL from Data ingestion and transformation in GCP and AWS data pipelines.
* Implemented Big Query data process from GCP Pub/Subtopic to BigQuery using cloud Dataflow with python and also used rest API with python to ingest Data from other systems into BigQuery.
* Involved in executing cloud Dataflow to run Data Validation between raw source file and BigQuery and integrated monitoring BigQuery, Dataproc using Stack driver across environments.
* Implemented Data security and Access controls using GCP’s Identity and Access Management(IAM) and Cloud Security Command Center.
* Design and Implemented data pipelines using GCP services such as Dataflow, Dataproc, and Pub/Sub.
* Redesigned cloud-based data warehouse to enhance security and improve performance.
* Enhanced quality of data insights through implementation of automated data validation processes and improved access to data sources.
* Extensively worked on Data Modeling involving Dimensional Data modeling, Star Schema/Snowflake schema, FACT & Dimensions tables, Physical & logical data modeling.
* Implemented AWS solutions using E2C, S3, RDS, EBS, Elastic Load Balancer, Glue Pipelines, Glue Crawler, Auto scaling groups, Optimized volumes, and EC2 instances and created monitors, alarms, and notifications for EC2 hosts using Cloud Watch.
* Involved in Designing Snowflake Schema for Data Warehouse, ODS architecture by using tools like Data Model, Erwin.
* Developed data models and data migration strategies utilizing concepts of snowflake schema.
* Developed Kibana Dashboards based on the Log Stash data and integrated different source and target systems into Elastic Search for near real time log analysis of monitoring End to End transactions.
* Extensively worked on Informatica power center Mappings, Mapping Parameters, Workflows, Variables and Session Parameters.
* Worked on SQL tools like TOAD and SQL Developer to run SQL Queries and validate the data.
* Worked on various automation tools like GIT, Terraform, Ansible Playbooks and Dimensional Modeling Star Schema, Snowflake schema, transactional modeling.
* Study the existing system and conduct reviews to provide a unified review on jobs.
* Involving in testing the database using complex SQL scripts and handling the performance issues effectively.

**Environment**: Apache spark 2.4.5, Databricks, Scala2.1.1, Cassandra, HDFS, Hive, GitHub, Jenkins, Kafka, Informatica PowerCenter 10.x, SQL Server 2008, Salesforce Cloud, Visio, TOAD, Putty, Autosys Scheduler, UNIX, AWS, snowflake, GCP, CSV, WinSCP, Salesforce data loader, SFDC Developer console, python, Version One, Service Now etc.

**Lowes, Mooresville, NC** **November 2020 to July 2022**

**Role: Azure Data Engineer**

**Roles and Responsibilities:**

* Extensively worked on Spark Applications using Spark-SQL in Databricks for data extraction and also worked with event based distributed systems like Kafka (Distributed Stream Processing)
* Experience on SQL data migration to Azure Data Lake/Analytics and used Azure Data Warehouse.
* Worked on Informatica Power Center tools Designer, Repository Manager, Workflow Manager, and Workflow Monitor.
* Expert in developing Data bricks notebooks for extracting data from various source systems such as DB2, Teradata, and performing data cleansing, wrangling, ETL processing, and loading to Azure SQL DB.
* ETL operations were carried out in Azure Data Factory by connecting to various relational database source systems via JDBC connectors.
* Configured data pipelines using Azure Data Factory, and a custom alerts platform was built for monitoring. Ingested data in mini batches and performed RDD transformations using spark streaming analytics in Azure Data Bricks.
* Designed dimensional data models using Star and Snowflake Schemas.
* Worked on Snowflake environment to remove redundancy and load real time data from various data sources into HDFS using Kafka.
* Created Mappings using Designer and extracted data from various sources, transformed data according to the requirement.
* Parsed high-level design specification to simple ETL coding and mapping standards.
* Designed and customized data models for Data Warehouse supporting data from multiple sources on real time.
* Involved in building the ETL architecture and Source to Target mapping to load data into the Data warehouse.
* Maintained stored definitions, transformation rules and targets definitions using Informatica Repository Manager.
* Used various transformations like Filter, Expression, Sequence Generator, Update Strategy, Joiner, Stored Procedure, and Union to develop robust mappings in the Informatica Designer.
* Coordinated with the front-end design team to provide them with the necessary stored procedures and packages along with necessary insight into the data.
* Involved in the continuous enhancements and fixing production problems.
* Generated server-side PL/SQL scripts for data manipulation, validation and materialized views for remote instances.
* Developed PL/SQL triggers and master tables for automatic creation of primary keys.
* Created PL/SQL stored procedures, functions, and packages for moving the data from staging area to data mart.
* Created scripts to create new tables, views, queries for new enhancement in the application using TOAD.
* Created indexes on the tables for faster retrieval of the data to enhance database performance.
* Involved in data loading using PL/SQL and SQL Loader calling UNIX scripts to download and manipulate files.
* Performed SQL and PL/SQL tuning and Application tuning using various tools like EXPLAIN PLAN, SQL TRACE, and AUTO TRACE.
* Extensively involved in using hints to direct the optimizer to choose an optimum query execution plan.
* Used Bulk Collections for better performance and easy retrieval of data, by reducing context switching between SQL and PL/SQL engines.
* Creation of database objects like tables, views, materialized views, procedures and packages using oracle tools like Toad, PL/SQL Developer and SQL\* plus.
* Partitioned the fact tables and materialized views to enhance the performance.
* Extensively used bulk collection in PL/SQL objects for improving the performing.
* Created records, tables, collections (nested tables and arrays) for improving Query performance by reducing context switching.
* Used Pragma Autonomous Transaction to avoid mutating problem in database trigger.
* Extensively used the advanced features of PL/SQL like Records, Tables, Object types and Dynamic SQL.
* Handled errors using Exception Handling extensively for the ease of debugging and displaying the error messages in the application.
* Used Splunk for monitoring, setting up alerts and creating dashboards.

**Environment**: Oracle 11g, SQL Plus, TOAD, SQL Loader, SQL Developer, PL/SQL, Informatica Power Center 10.2.0, Informatica EDC, Designer, Workflow Manager, Workflow Monitor, Repository Manager, Shell Scripts, UNIX, Windows XP, Splunk.

**Ernst & Young, Chicago,** **IL Sep 2019 - Oct 2020**

**ETL/Data Engineer**

**Roles and Responsibilities:**

* Gathering specification, designing the system, development, testing and implementing the system. Collecting the requirements from end users.
* Developed Advance PL/SQL packages, procedures, triggers, functions, Indexes and Collections to implement business logic using SQL Navigator.
* Generated server-side PL/SQL scripts for data manipulation and validation and materialized views for remote instances.
* Created management analysis reporting using Parallel Queries, Java Stored Procedures. Participated in change and code reviews to understand the testing needs of the change components. Worked on troubleshooting defects in a timely manner.
* Involved in defragmentation of tables, partitioning, compressing and indexes for improved performance and efficiency. Involved in table redesigning with implementation of Partitions and Partition Indexes to make Database performant and easier to maintain.
* Experience in Database Application Development, Query Optimization, Performance Tuning and DBA solutions and implementation experience in complete System Development Life Cycle.
* Used Informatica Power Center Designer to analyze the source data to Extract & Transform from various source systems by incorporating business rules using different objects and functions that the tool supports.
* Used Power Center Designer to create mappings and mapplets to transform the data according to the business rules.
* Used various transformations like Source Qualifier, Joiner, Lookup, SQL, Router, Filter, Expression and Update Strategy etc.,
* Created and Configured Workflows and Sessions to transport the data to target Oracle tables using Informatica Workflow Manager.
* Implemented complex business rules in Informatica Power Center by creating re-usable transformations, and robust Mapplets.
* Implemented performance tuning of Sources, Targets, Mappings and Sessions by identifying bottlenecks and used debuggers to debug the complex mappings and fix them.
* Designed and Developed Informatica Workflow to extract data from XML files and loaded it into the database.
* Improved session Performance by enabling property incremental aggregation to load incremental data into the target table.
* Implemented performance tuning logic on Targets, Sources, Mappings and Sessions to provide maximum efficiency and performance.
* Worked along with the UNIX team for writing UNIX shell scripts to customize the server scheduling jobs.
* Used principles of Normalization to improve the performance. Involved in ETL code using PL/SQL in order to meet requirements for Extract, transformation, cleansing and loading of data from source to target data structures.
* Involved in the continuous enhancements and fixing of production problems. Designed, Implemented and Tuned interfaces and batch jobs using PL/SQL.
* Involved in data replication and high availability design scenarios with Oracle Streams. Developed UNIX Shell Scripts to automate repetitive database processes.
* Written documentation to describe program development, logic, coding, testing, changes and corrections.

**Environment**: Oracle 10g/11g, SQL Plus, TOAD, SQL Loader, SQL Developer, PL/SQL, Informatica Power Center 9.6, Designer, Workflow Manager, Workflow Monitor, Repository Manager, Shell Scripts, UNIX, Windows XP, Splunk, HTML, TOAD, XML.

**AmTrust Financial Services, Inc. New York July 2015 - Aug 2019**

**Data Analyst**

**Roles and Responsibilities:**

* Interacted with clients and business users. Involved in requirement gathering and impact analysis.
* As a Software Engineer, involved in designing business layer and data management components.
* Developed Weekly Billing Process (WBP) which runs every week which includes processes for creating customer invoices and usage reports.
* Designed, Developed and Supported Extraction, Transformation and Load Process (ETL) for data migration with Informatica 8.6.
* Developed various mappings using Mapping Designer and worked with Aggregator, Lookup, Filter, Router, Joiner, Source Qualifier, Expression, Stored Procedure, Sorter and Sequence Generator transformations.
* Worked extensively with the connected Lookup Transformations using dynamic cache.
* Created and scheduled Sessions, Jobs based on-demand, run on scheduled time, and run only once.
* Monitored Workflows and Sessions using Workflow Monitor.
* Responsible for design, development and testing of server-based software that provides secure mobile workforce solutions.
* Created conceptual, logical and physical data models.
* Written Procedures, Functions & Triggers for different operations.
* Performed performance tuning of query and scripts.
* Coordinated with the Quality Assurance/Testing team members to perform both SIT and UAT testing.

**Environment**: Python, Informatica 8.6, Oracle, SQL, Perl, PL/SQL, Shell scripting, Windows, UNIX, MSSQL.

**Avantel, India July 2013 to June 2015**

**Database Engineer**

**Responsibilities:**

* Experienced with PLSQL programming for developing applications on databases. Developed PLSQL procedures, functions, triggers, and packages as per application need.
* Automated database health checks and daily operations using Python programming and Unix Shell scripting.
* Developed UNIX Shell scripts for generating HTML email alerts for various performance and monitoring metrics of databases.
* Developed bash scripts to identify database growth periodically and took necessary actions by contacting application developers for either archiving data or purging unwanted data from production environment.
* As a part of performance management, identified long running jobs and user application queries and accordingly took necessary actions. Identified server and memory level bottlenecks. Helped application developers in resolving their database performance related issues.
* Experienced with identifying worst performing SQLs and fixed them with the help of development Teams. Periodically run automated scripts that shows long running queries’ report and thereby understand the performance issues in the databases. Discussed the same with the development teams and tool necessary actions such as tuning queries, changing queries, indexing tables, or purging unwanted data from tables.
* Worked on application developers’ change requests to roll out the development changes in production.
* Performed database installation, upgrade, PSU, CPU, Bugfix patching. Worked on managing disaster recovery (DR) databases and carried out DR drills.

**EDUCATION**

* **CHRISTIAN BROTHERS UNIVERSITY,** Memphis, Tennessee.

Masters in computer science, 2015-2017

* **DVR College of Engineering and Technology,** Kandi, Sanga Reddy.

Bachelors in computer science, 2013.