Deepthi

Data Engineer

Mail: jobdeepthi6@gmail.com

Mobile: 6892201235

Summary:

- 8+ years of IT experience in implementing the end-to-end Data Warehouse and Data engineering solutions.
- Knowledge of **ETL** methods for data extraction, transformation and loading in corporate-wide ETL Solutions and **Data Warehouse tools** for reporting and data analysis.
- Intelligent in writing **Spark** transformations and actions using Spark SQL in **Scala**.
- Hands-on experience in developing SPARK applications using Spark tools like RDD transformations, Spark core, Spark MLlib, Spark Streaming and Spark SQL.
- Worked on **AWS** Cloud – used **S3** buckets for storing the data, **EMR** – for spinning up the cluster and running the spark jobs, **Athena** – for creating external tables, **IAM** – for authentication, **AWS GLUE** – to get data from sources.
- Hands-on experience working Amazon Web Services (AWS) using Elastic Map Reduce (EMR), **Redshift**, and **EC2** for data processing.
- Experienced on Migrating SQL database to **Azure** data Lake storage, Azure **Data** Factory (**ADF**), Azure data lake Analytics, Azure SQL Database, **Data Bricks** and Azure SQL **Data** warehouse and controlling and granting database access and migrating on premise databases to Azure **Data** Lake store using Azure **Data** factory.
- Strong working experience with **SQL** and **NoSQL** databases, data modeling and data pipelines. Involved in end-to-end development and automation of ETL pipelines using SQL and **Python**.
- Involved in setting up **Jenkins** Master and multiple slaves for the entire team as a CI tool as part of Continuous development and deployment process.
- Experience working on creating and running **Docker** images with multiple micro services and worked on **terraform** scripts to deploy the resources.
- Responsible for **GIT** version control to commit the code developer which further used for deployment using build and release tool Jenkins.
- Experienced in Optimizing the Pyspark jobs to run on **Kubernetes** Cluster for faster data processing.

- Proficiency in SQL and Python for data analysis and manipulation, including experience with popular libraries such as Spark and PySpark.
- Having good experience on Python scripting language to use in Spark development life cycle.
- Excellent in data manipulation using python for loading and extraction as well as with python libraries such as NumPy, SciPy and Pandas for data analysis and numerical computations.
- Solid experience in developing and designing data integration solutions using ETL tools such as Informatica PowerCenter, Dell Boomi, Teradata Utilities for handling large volumes of data.
- Hands-on experience in SQL and NOSQL databases such as Snowflake, HBase, Cassandra and MongoDB.
- Performed Data Profiling and Data Analysis using SQL on different extracts.
- Managed multiple tasks and worked under tight deadlines and in a fast paced environment.
- Excellent analytical, communication skills which helps to understand the business logics and develop a good relation between stakeholders and team members.

Education:
- Bachelors in Computer Science Engineering from KL University, India.
- Masters in University of Texas at Arlington, TX, USA.

Technical Skills:

Databases	Spark SQL, Teradata, Oracle, MySQL, Microsoft SQL, Postgrad SQL.
NoSQL Databases	MongoDB, Hadoop HBase, and Apache Cassandra.
Programming Languages	Python, SQL, Scala
Cloud Technologies	AWS, Azure, Kubernetes, and Docker
Data Formats	CSV, JSON, XML
Querying Languages	SQL, NO SQL, PostgreSQL, MySQL, Microsoft SQL, PL/SQL
Integration Tools	Jenkins, Informatica, Dell Boomi
Scalable Data Tools	Hadoop, Hive, Apache Spark, Map Reduce.

Professional Experience:

Client: Verizon-Tampa, FL   **October 2021 to Till Date**
Role: Data Engineer

Responsibilities:
- Implemented Installation and configuration of multi-node cluster on Cloud using Amazon Web Services (AWS) on EC2.
- Handled AWS Management Tools Cloud watch and Cloud Trail.

- Stored the log files in AWS S3. Used versioning in S3 buckets where the highly sensitive information is stored.
- Analyzed the SQL scripts and designed the solution to implement using Scala.
- Optimized the PySpark jobs to run on Kubernetes Cluster for faster data processing.
- Designed and developed AWS architecture, Cloud migration, AWS EMR, DynamoDB, Redshift and event processing using lambda function.
- Conducted ETL Data Integration, Cleansing, and Transformations using AWS glue Spark script.
- Analyzed the SQL scripts and designed it by using PySpark SQL for faster performance.
- Created PySpark code that uses Spark SQL to generate data frames from raw formatted layers and writing them to data service layer internal tables as orc format.
- Worked on terraform scripts to deploy the resources.
- Experience in developing scalable & secure data pipelines for large datasets.
- Migrated and developed daily running jobs from on-premises to EMR spark, Scala.
- Primarily Responsible for converting Manual Report system to fully automated CI/CD Data Pipeline that ingests data from different Marketing platforms to AWS S3 data lake.
- Used Spark Streaming to receive real time data from the Kafka and store the stream data to HDFS using Python and NoSQL databases such as DynamoDB and MongoDB.
- Performing ETL testing activities like running the Jobs, extracting data using necessary queries from database transform, and uploading data into Data warehouse servers.
- Developed and implemented core API services using Scala and Spark.
- Implemented a CI/ CD pipeline with Docker, Jenkins, and GitHub by virtualizing the servers using Docker for the Dev and Test environments by achieving needs through configuring automation using Containerization.
- Design Develop and test ETL Processes in AWS Glue to migrate Campaign data from external sources like S3, Parquet/Text Files into AWS Redshift.
- Hands on experience in working with Continuous Integration and Deployment (CI/CD) using Jenkins, Docker.
- Involved in developing applications in Python language for multiple platforms and good experience in handling data manipulation using python Scripts.
- Developed ETL jobs using PySpark, Data Lineage in which the data has been transformed in multiple stages and actions like aggregations are performed.
- Created scripts to read CSV, JSON, and parquet files from S3 buckets in Python and load into AWS S3, DynamoDB and Snowflake.
- Worked on Snowflake Schemas and Data Warehousing and processed batch and streaming data load pipeline using Snow Pipe and Matillion from data lake Confidential AWS S3 bucket.
- Worked on Dockers containers by combining them with the workflow to make them lightweight.
- Worked on MongoDB by using CRUD (Create, Read, Update and Delete), Indexing, Replication and Shading features.
- Worked with AWS Cloud Formation templates for reusable code deployment across environments.
- On call support during weekend installations.
- Handled production job failures and given permanent fixes.
- Worked on SQL queries in dimensional data warehouses and relational data warehouses. Performed Data Analysis and Data Profiling using Complex SQL queries on various systems.
- Enhanced and optimized Spark/Scala/pyspark jobs to aggregate, group and run data mining tasks using the Spark framework.
- Well versed with REST APIs and CRUD operations to access and transform analytic data.
- Facilitated the maintenance of the project by performance tuning Python apps.

Environment: AWS, API's, Cloud watch & trail, SQL, Scala, Spark, Redshift, NOSQL, Snowflake, MongoDB, AWS CloudFormation, Docker, DynamoDB and Python.

Client: American Express, Phoenix, AZ   **Feb 2019 - Sep 2021**
Role: Data Engineer

Responsibilities:
- Worked as Data Engineer to review business requirements and compose sources to target data mapping documents.
- Developed spark applications for performing large scale transformations and denormalization of relational datasets.
- Loading the data into AWS S3 using Databricks.
- Used Azure Databricks for performing data transformations and loading data into Azure Data Lake.
- Designed and implemented a wide range of ETL pipelines using Azure Databricks to extract, transform, and load data from various sources, including flat files, databases, and APIs.
- Utilized Snowflake as the data warehousing solution for storing and querying large amounts of data.
- Closely work with stakeholders to recommend and design Azure SQL database and Azure Data Lake storage solutions.
- Writing SQL stored procedures to transform the table into the target data model and load it into target tables.
- Implemented Kafka for real-time data streaming and handling high volume, high throughput, and low latency data.
- Working with Azure Data Factory Data transformations such as Stored Procedure, Databricks.
- Deploying ADF pipelines into different environments like QA, UAT and PROD using Azure DevOps pipelines.
- Extensively worked with various triggers and implanted as per the requirements.
- Deployment and monitoring of Pipelines & Scheduling and executing SSIS packages in ADF.
- Utilized Azure Cloud Technologies to store and manage data, including Azure Data Lake and Azure SQL Database.
- Worked on performance tuning for existing stored procedures, Functions.
- Developed and maintained data pipelines to support business intelligence and reporting needs, including creating/scheduling jobs, integrating with Azure DevOps, and utilizing notebooks and clusters.
- Conducted data quality checks and audits to ensure data accuracy and completeness, implemented processes for ongoing data monitoring and maintenance.
- Set up Azure Repo and Pipelines for CI/CD deployment of objects.
- Create Hive queries within Databricks to help market analysts spot emerging trends by comparing incremental data with Teradata reference tables and historical metrics.
- Develop normalized Logical and Physical database models for designing an OLTP application.
- Worked closely with data analysts and business users to understand data needs and requirements and provided technical guidance and support as needed.
- Written Python scripts to parse JSON documents and load the data in the database.
- Generating various capacity planning reports (graphical) using Python packages like NumPy, matplotlib.
- Worked with Azure DevOps to ensure standardization and automation of software build & release processes.

- Architected and implemented ETL and data movement solutions using Azure Data Factory.
- Analyzing various logs that have been generating and predicting/forecasting the next occurrence of events with various Python libraries.
- ETL pipelines in and out of the data warehouse using a combination of Python and Snowflakes Snow SQL Writing SQL queries against Snowflake.
- Experience in ETL processes involving migrations and in sync processes between two databases.
- Worked with the SCRUM team in delivering agreed user stories on time for every Sprint.
- Worked on analyzing and resolving production job failures in several scenarios.
- Created Dell Boomi interfaces based on interface specification and in conformance with Dell Boomi coding best practices.
- Build APIs according to the design specification, using programming languages and tools such as RESTful API, SOAP, and GraphQL. Orchestrated and deployed complex integrations using APIs, middleware, and databases. Ensuring that the APIs and integrations are properly documented and communicated to clients.
- Experienced working in document caching, document properties. Good knowledge on Build, Deploy and Manage in Boomi Atmosphere.
- Good Knowledge on detailed configuration of Dell Boomi Atom setup.
- Experienced working with connectors like FTP, Database, Mail, Disk, Http, Web Services server, JMS…etc.
- Configuring and customizing Dell Boomi connectors and components.
- Developing and maintaining integration workflows using Boomi and technical
- documentation related to Dell Boomi solutions.
- Developing custom scripts and integrations using Dell Boomi. Participating in the deployment and maintenance of Boomi solutions and in the evaluation and selection of new integration technologies and platforms.
- Proficient in XML, DATABASE, JSON, Flat file profiles in Dell Boomi.

Environment: AWS, Databricks, API's, ETL, Azure, Data Lake, Python, Dell boomi, Azure SQL Database and Kafka.

Client: Lumen Tech Inc, Chicago, IL   **Aug 2017 - Jan 2019**
Role: Data Engineer

Responsibilities:
- Built and maintained a wide range of ETL pipelines to extract, transform, and load data from various sources, including flat files, databases, and APIs.
- Developed ETL as per mapping requirements to load data into staging from multiple sources such as csv, xml, xlsx, etc.
- Utilized SQL and Python for data manipulation and analysis, including experience with popular libraries such as Spark and PySpark.
- Worked on developing ETL pipelines on S3 parquet files on data lake using AWS Glue.
- Wrote lambda functions using python for custom data transformation and ETL.
- Worked with AWS Cloud Formation templates for reusable code deployment across environments.
- Updated data mapping documents based on the actual development.
- Worked closely with data analysts and business users to understand data needs and requirements and provided technical guidance and support as needed.
- Devised PL/SQL Stored Procedures, Functions, Triggers, Views, and packages.

- Made use of Indexing, Aggregation and Materialized views to optimize query performance.
- Involved in developing and documenting the ETL (Extract, Transformation and Load) strategy to populate the Data Warehouse from various source systems.
- Design and develop spark job with Scala to implement end to end data pipeline for batch processing.
- Implemented Spark using Python and Spark SQL for faster processing of data.
- Developed a fully automated continuous integration system using Git, Jenkins, MySQL, and custom tools developed in Python and Bash.
- Developed complex mappings, Mapplets in Informatica workflow designer to integrate data from varied sources like Teradata, Oracle, SQL Server, Flat files and loaded into target.
- Worked on performance tuning by identifying the bottlenecks in Sources, Targets, and Mapping. Enhanced Performance on Informatica sessions using large data files by using partitions.
- Created Unix Shell Scripts for ETL jobs, session log cleanup, dynamic parameter and maintained shell scripts for data conversion.
- Validate the data in source and target systems using PL/SQL queries.
- Prepared Unit Test plan and efficient unit test documentation was created along with Unit test cases for the developed code.

Environment: AWS, Spark, Oracle, Python, PL/SQL, SQL, Informatica, Toad, ETL, AWS CloudFormation, AWS Glue, Jenkins, Bash, SQL Server, and Scala.

Client: Radiant soft Pvt Ltd, Hyderabad, India　　　　　　　　**June 2014 – Dec 2016**
Role: Data Engineer

Responsibilities:
- Normalized existing OLTP systems to speed-up the execution time in data modeling logical statements.
- Developed multitude of database objects such as tables, views, stored procedures, user-defined functions to support various database applications per business requirements.
- Gathered requirements and specifications to create detailed Functional and Technical Design Documents.
- Used Python scripts to update content in the database and manipulate files.
- Designed and implemented tables, views, functions, indexes in support of a multitude of databases and ETL applications.
- Designed & maintained databases using Python, troubleshot, fixed, and deployed many Python bug fixes successfully.
- Created SQL stored procedures and scheduled them in Azure DevOps.
- Worked with Azure DevOps to ensure standardization and automation of software build & release processes.
- Created stored procedures with error handling techniques, transactions, and performance tuning features.
- Participated in design and implementation of cloud data platforms including ETL, data processing, orchestration, and related applications.
- Designed ETL packages on SSIS to migrate data and make it available for the reporting applications.
- Wrote Python script for manipulating and looping through different user defined objects.
- Implemented Agile for frequent changes to client requirements and following parallel development and testing.
- Hands-on experience on developing SQL Scripts for automation purposes.
- Writing SQL Scripts to extract the data from Database and for Testing Purposes.

- Developed test cases, performed unit testing and integration testing.
- Involved in Daily scrum calls, sprint planning and sprint demos.
- Handle the migration process from Development, Test and Production Environments.
- Developed ETL jobs to extract information for Enterprise Data Warehouse.
- Extensive use of ETL process to load data from different RDBMS, XML, and flat files.
- Experience in Performance tuning of SQL queries and mainframe applications.
- Involved in Data Extraction from Oracle, Flat files, Mainframe files using Informatica.
- Developed PL/SQL Procedures at the database level that were used in the mappings through Stored Procedure Transformation.
- Created various UNIX shell scripts for Job automation of data loads.
- Writing Complex queries, stored procedures, Views, Cursors, SQL Joins.

Environment: Python, SQL, PL/SQL, Unix, Informatica, SSIS, RDBMS and Toad.

Deepthi J

- Arlington, TX, US

Contact Information

- w06-937-95u@mail.dice.com

- 6892201235

Work History

Total Work Experience: 10 years

- **Data Engineer Verizon-Tampa, FL**
  Oct 01, 2021

- **Data Engineer American Express, Phoenix, AZ**
  Feb 01, 2019

- **Data Engineer Lumen Tech Inc, Chicago, IL**
  Aug 01, 2017

- **Data Engineer Radiant Soft Pvt Ltd, Hyderabad, India**
  Jun 01, 2014

Education

- **Masters** | Masters in University of Texas at Arlington, TX, USA.

- **Bachelors** | Bachelors in Computer Science Engineering from KL University, India.

Skills

- **cloud** - 6 years

- **data** engineering - 6 years
- **data** warehouse - 6 years
- **database** - 6 years
- **etl** - 6 years
- **performance tuning** - 6 years
- **python** - 6 years
- **sql** - 6 years
- **automation** - 4 years
- **data** processing - 4 years
- **scripting** - 4 years
- **business requirements** - 5 years
- **documentation** - 5 years
- **flat file** - 5 years
- **stored procedures** - 5 years
- **transformation** - 5 years
- **xml** - 5 years
- **informatica** - 4 years
- **mapping** - 4 years
- **unix** - 4 years

Work Preferences

- Desired Work Settings: Remote or On-Site or Hybrid
- Likely to Switch: False
- Willing to Relocate: True
- Travel Preference: 0%
- Work Authorization:
  - US
- Work Documents:
  - Have H1 Visa
- Security Clearance: False
- Third Party: True
- Employment Type:
  - Contract - Corp-to-Corp

Profile Sources

- Dice:
  https://www.dice.com/employer/talent/profile/581ecb8c0e5c4e8c2bdc3120c13d25f8