



Name

Data Engineer

[email](#)

Phone

Overview

Radha has 4+ years of experience in developing data pipelines and migration routines for Top 5 pharmaceutical company leveraging Azure Data Factory, Azure Databricks and StreamSets. Her deep technology expertise has enabled to craft innovative technology solutions that now enable a business to visualize their dashboards in the almost real-time interval. Radha holds a bachelor's degree in Information Technology, has a wide repertoire of technology skills, and is also a **certified Azure Data Engineer**.

Highlights

- Experience as a developer in cross-functional teams to deliver end-to-end 'data solutions' using various frameworks, and tools both on-prem and cloud
- Experience handling both structured and semi-structured data, appropriate storage formats, and various processing engines such as Spark and Azure Data Lake
- Implemented multiple pipelines with Change Data Capture (CDC), and developed approach to have 'chain of trust' on ingested data
- Experience in Apache Kafka including broker configurations, partitioning of Kafka messages, and setting up the replication factors in Kafka Cluster.
- Experience handling message-based data using Azure Data Factory
- Experience in optimizing queries, and selecting appropriate storage formats
- Experience in developing the CI/CD for pipelines, applying DevOps practices using Azure DevOps

Technical Skill Set

- | | |
|--------------------------------|---|
| • Data Ingestion Tools | : Azure Data Factory, StreamSets |
| • Cloud Computing | : Azure Platform, GCP |
| • Relational Databases | : SQL Server, PostgreSQL, Oracle, MySQL |
| • Programming Languages | : PL/SQL, Python, Java |
| • Analytics Tool | : Power BI, Qlik View |
| • Hadoop Ecosystems | : Kafka |

Certification

- Microsoft Azure Fundamentals
- Microsoft Azure Data Fundamentals
- Microsoft Azure Data Engineer
- Microsoft Azure Data Analyst

Projects

Project Aurora

Client : Top 5 Life Sciences Company

Role : Developer

Company :

Duration : Jan 2019 – Nov 2020

Description

Project Aurora was complete on the existing Customer Relations System (CRS). The client wanted to build new and improved CRS using a cloud solution to improve customer requests and create efficiency in the existing customer service processes. The system primarily contained Salesforce as a source system and built on Azure technology. As an agile team, we were working with primarily developing the Analytics components of CRMS along with required dashboards and actions. This effort involved the integration of data from 125+ source objects like Social Media, Emails, Phone, etc., and Agent Call Detail Records (CDR) into cloud storage using Azure Data Factory. A lot of intermediate transformations were involved. The effort included designing of multiple alert-based mechanisms for intimation of process halts in production.

Accountabilities

- One of the key participants in the design discussions
- Since this was a ground-up implementation, there were multiple changes and iterations in the source systems which meant the pipelines had to be modified as the data model changes. Given the tight timelines, we automated the pipelines to alert if the incoming data has changed
- Designed the data de-duplication mechanisms for most of the data pipelines using the MD5 hashing
- Audited the data using Change Data Capture
- Designed and developed the data profiling of incoming data so that 'chain of trust' can be established

ChEMBL

Role : Data Engineer

Company :

Duration : Sept 2018 – Dec 2018

Description

ChEMBL is a manually curated database of bioactive molecules with drug-like properties. It brings together chemical, bioactivity, and genomic data to aid the translation of genomic information into effective new drugs. It is an open-source database that can provide cross-references between chemical structure from various open source datasets. This data is available on FTP sites for download and new versions will be released either quarterly. As new versions are released regularly, users need access to the updated data frequently. Using our data model, we were able to send a notification to the user whenever a new version was released. We have designed and developed a process where we were able to automate this whole process and fetch data from this FTP site, orchestrate data pipelines to ingest data into the platform.

Accountabilities

- Designed StreamSets pipelines to ingest data into platform

- Triggered and configured a BOT workflow using Kafka messages based on availability of new version

Accolade

Client : Top 5 Life Sciences Company

Role : Data Engineer

Company :

Duration : Sept 2020 – Sep 2021

Description

Project Accolade was completely based on insurance data. This project is Extract Transform Load process (ETL) where we need to extract fixed width format files from SFTP and perform necessary transformations as per the project rules using streamsets (ETL tool) and loading it into various destinations like Mongo, Kafka and GCS. This process involved designing of multiple pipelines for performing logging, validations, error handling etc. It is an automated process where all the pipelines will be in running state and picks the files whenever the files are available on SFTP server.

Accountabilities

- Developed pipelines for converting incoming fixed width data files to JSON records.
- Automated the logging, validation process and Error handling for the data by using MongoDB to store data.
- Publishing the data to Kafka topic

CR Analytics

Client : Top 5 Life Sciences Company

Role : Data Analyst

Company :

Duration : Apr 2020 – Sept 2021

Description

The primary goal of the CR Analytics project was to build complete cloud technology for Consumer Healthcare to build the future state analytical reporting in the Power BI platform which enables self-service analytics and to drive improvements across decision-making capabilities of business operations. The current system was outsourced to an on-premise environment and the client wanted the same solution replicated in the cloud environment. The client categorizes its products into various categories in the Consumer Health care division with around 363 products in 120 operation centers around the world.

Business logic and necessary complex calculations (say day light savings, local time zone conversions, excluding the weekends/ holidays, etc.,) for 50+ KPIs are implemented on the SQL Server database engine for processing the data. Implemented data integration process between the older and latest versions of data sources. Azure cognitive services are used to extract the keyword, determine the sentiment, and analyze the verbatim provided by the consumer. Implemented models using Bert Algorithm for text classification and sentiment analysis. Key insights are generated for the stakeholders to determine the business decisions.

Accountabilities

- As a Developer, participated in weekly workshops with client SMEs for the requirements and business operations. The Key Performance Indicators (KPI) were to be developed with business rules which requires a

lot of insights into the business and data capture operations

- Designed an efficient workflow in ADF to ingest data into specific destinations
- Integrated data from multiple data sources across various geographies into a single format for further processing
- Implemented convoluted business logics using SQL queries, PLSQL in Azure SQL Database
- Created interactive and comprehensively illustrated dashboards, reports, Apps to visualize the data in Power BI
- Implemented Azure cognitive services to evaluate sentiment and keywords from free text
- Automated the ad hoc requests Monthly, Weekly, and Daily into SharePoint using Power Automate flows

Goldmine Project (POC)

Client : Top 5 Life Sciences Company

Role : Developer

Company :

Duration : Nov 2020 – Feb 2021

Description

Analyzing the free text to uncover temporal trends in consumer relations data. We use natural language processing to understand how consumers feel so that we create emotional triggers that have high impact marketing activations with consumers. We perform longitudinal studies that reveal how consumers' behaviors, perceptions change over time. We use consumer data to get insights for value proposition canvas i.e., gains, pains, customer jobs, etc.

Accountabilities

- As a Developer, participated in weekly workshops with client SMEs for the requirements and business operations
- Analyzed the free text to tag the emotional triggers

Quantum Analytics

Client : Top 5 Financial Cooperation

Role : BI Engineer

Company :

Duration : Oct 2021 (Current Project)

Description

Migration of On-Premises data server to cloud server using Azure cloud service. As an agile team, we were working with primarily understanding the transformations and architecture of the current existing system and trying to implement on the migration process.

Accountabilities

- Key participant in creating Architecture designs for Migration process
- Migrated On-Premises warehouse and BI Solutions to Azure
- Implemented change management, data deduplication, indexing, process automation through stored procedures and automated the output data consumption patterns (Data extracts, BI Tools).
- Migrated the existing CDIC tracking dashboards from Cognos, Excel, and Report Server to Power BI
- Publishing downstream data and making it accessible to different teams.

Bimbo Project

Client : Top 5 Consumer Good Company in US

Role : Data Engineer

Company :

Duration : Mar 2022 (Current Project)

Description

Migration Grupo Bimbo s worldwide leader in the baking and confectionery industry. They are looking the extract value from the data by standardizing the data models designing a process to model, extract and adapt the data into standardize model. The Solution will focus on building cloud-based intelligence system based on data models that enable faster insights obtention.

Accountabilities

- Key participant in creating Architecture designs for Migration process
- Used serverless SQL Lake Database presented in synapse workspace.
- Used Map tool using Industry Data Model(IDM) we will be loading the data.
- Created a testing framework for dynamically calling the azure pipelines and running the test cases related to the data ingestion.

Education

- Bachelor of Technology 2014 – 2018, Information Technology, College Name, Affiliated to JNTU Hyderabad