

Jithendra Kanna

AWS Data Engineer

Email : jithendrakanna95@gmail.com

Contact No.: +91 8142000703

CAREER OBJECTIVE

Seeking responsibility and challenging opportunity that would utilize and encourage my abilities through dedicated hard work and resources of the company to achieve combined growth.

WORK SUMMARY

- Holding 3 +years of experience in **Data Engineering** with sound experience in **BigData Ecosystem**.
- Good hands-on experience on BigData/Cloud tools like AWS Glue, Redshift, Athena, Azure Databricks, Azure Data Factory, Azure Synapse Analytics, Azure SQL Database, SparkCore, Spark SQL using language Python, Hive and HDFS.
- Experience in developing Spark jobs to process data and storing into Delta lake, Azure Data lake storage, AWS S3, Hive tables and databases.
- Worked on Data movement and data transformation activities using compute services such as Databricks, Azure Synapse Analytics, AWS Glue.
- Excellent knowledge with distributed storages (HDFS) and Spark on Standalone and Cluster mode
- Having good knowledge on YARN and various HADOOP daemons such as Name Node, Data Node, Secondary Name Node, Resource Manager, and Node Manager.
- Usage of Import and export data from different databases like Oracle, Teradata into HDFS and vice versa using Sqoop.
- Experience in working with the different file formats like TEXTFILE, ORC,AVRO and PARQUET files.
- Well experienced in team co-ordination, taking ownership of responsibilities, analytical, problem solving and negotiation skills with a high degree of comfort in dealing with management and vendors.

TECHNICAL SKILLS

BigData Ecosystem : Spark, Hive, Sqoop

Cloud Technologies : Azure, AWS

Languages :Python

Database : Teradata, Snowflake

WORK EXPERIENCE

2.Company : Blazeclan Technologies Pvt Ltd-IT Company

Designation :Data Engineer

Duration :Aug'21 till date

Project #3

Name: Data warehouse implementation

Duration: Jun 2022 - present

Client: CamillaAustralia

Team Size: 5

Role: Developer

Technologies used: AWS Glue, Redshift, Athena, S3, Python, Apache Spark

Description:

Camilla Australia is a designer wear company and has boutiques in Australia and US and online business across the world.

Roles & Responsibilities:

- Create ETL framework in AWS Glue using Pyspark.
- Extract data from API using Python.
- Normalise JSON data and store it in AWS S3 using Pandas.
- Process/transform JSON data in S3 using Pyspark.
- Load processed/transformed data into AWS Redshift tables using Pyspark.
- Implement SCD logics to upsert data in Redshift tables.
- Built generic data quality framework using Pyspark to perform data quality check.
- Create audit framework to capture run logs.
- Reports Daily/weekly updates to customer.

Project #2

Name: Cloud Migration

Duration: Jan 2022 –Jun 2022

Client: Pelabuhan Tanjung Pelepas

Team Size: 4

Role: Developer

Technologies: Azure Synapse Analytics, Azure SQL Database, PySpark

Description:

The Port of Tanjung Pelepas (PTP) is Malaysia's largest container terminal. The port delivers reliable, efficient and advanced services to major shipping lines and box operators, providing shippers in Malaysia and abroad extensive connectivity to the global market.

As part of this project we migrated the functional logic of on-premise ETL operations into Azure cloud with the help of Azure Synapse Analytics to load data into Azure Data Lake Storage Gen2 and Azure SQL Database.

Roles & Responsibilities:

- Build reliable ETL/ELT pipeline using Azure Synapse Analytics.
- Copying data from on-premise database into data lake store using Azure Synapse Analytics.
- Transforming the data using DataFlow and loading into Azure SQL DB in Synapse Analytics
- Implementing SCD type 1 in Data Flow and updating the target database.
- Transforming data into multiple tables using Pyspark programming and writing data into target tables.
- Handling scrum for the team
- POC on Azure Synapse Analytics to ingest, load and transform data from multiple sources.

Project #1

Name: Delta Lake Implementation

Duration: Aug 2021 – Jan 2022

Client: Slintel

Team Size: 4

Role: Developer

Technologies: Databricks, PySpark, Python, SQL

Description:

Slintel evaluates over 100 billion data points and analyzes factors such as buyer journeys, technology adoption patterns, and other digital footprints to deliver market & sales intelligence.

Slintel integrates with the technology that matters the most to you - CRMs like Salesforce, HubSpot and Pipedrive, and sales tools like Outreach.

As part of this project we handle slintel's data using Databricks. By making use of Databricks capabilities, reliable ETL/ELT pipeline were built and stores data in AWS S3 in delta format. Provides the refined data to the customer which helps in efficient data insights.

Roles & Responsibilities:

- Create ETL/ELT pipelines using pyspark on Databricks platform.
- Read data from external storage like AWS S3 and converting into delta format.

- Minimize job failures incase of schema mismatch by utilising schema enforcement and schema evolution techniques.
- Optimize jobs for better performance in data processing and maintain data quality.
- Implement SCD logics to update and insert data in delta lake.
- Integrate Databricks with AWS Quicksight to generate reports.
- Handling scrum call daily for the updates and reportings.
-

WORK EXPERIENCE

1.Company : Cognizant Technology Solutions

Designation :Data Engineer

Duration :Marc'19 to Aug'21

Project #2 – Cognizant Technology Solutions

Name: Terminus Logistics – RFO

Duration: Jan 2021 – Aug 2021

Client: Apple

TeamSize: 6

Role : Developer

Technologies: Teradata, Snowflake, Spark, Autosys

Description:

Terminus (Teradata Minus) is a program to reduce the dependency on Teradata and host the data and reports on cloud Datawarehouse systems like Snowflake. As part of this project we migrated the functional logic with the help of Keystone Framework, built on Spark & Scala. Keystone framework is a set of applications with which we create data pipelines & schedule these pipelines using the framework to load the data into various databases like Teradata, Vertica & Snowflake etc., These data pipelines load the data into Teradata & replicate the data to Snowflake simultaneously. Also, the reports are re-pointed to data available on Snowflake with which users are also migrated to cloud based LDAP authentication.

Roles and Responsibilities:

- Have played a key role in Design, Development & Testing phases.
- Created data pipelines to load the data into Core & Semantic layers.
- Constantly worked with stake holders to evaluate on the risk and migrate the interface to Keystone.
- Worked on creating the data pipelines to load the data into Snowflake.
- Have done end to end testing of data pipelines.
- Involved in resolving the issues in production.

- Provided warranty support on the interfaces migrated to new process.
- Validated the data loaded between legacy & current process.

Project #1 – Cognizant Technology Solutions

Project: Walt Disney Parks and Resorts

March 2019 – Jan 2021

Title : WDPR IM SS

Client : Walt Disney Parks and Resorts

Role : Data Engineer

TeamSize : 15

Technologies : Teradata, Spark, Hive, Python

Client Description:

Walt Disney is one of the important clients in the Cognizant that comes under IME (Information Media and Entertainment) vertical, this account deals with all the Disney theme park related information into a different data warehouses based on its type of its details respectively.

Project Objective:

To facilitate the Walt Disney resort's Guest and Labor related data. We are responsible for the resort's day to day transaction, offers by analyzing the guest data based on the seasons and the response over offers and packages.

Roles and Responsibilities:

- Creating Spark jobs to process data and store into Hive tables.
- Used Spark-SQL to process the data and to run on Spark engine
- Worked on creating Hive managed and external tables based on the requirement.
- Implemented Partitioning and Bucketing on Hive tables for better performance.
- Effectively coordinated with Business for Ad-hoc queries raised in the form of incidents.
- Monitoring and checking the jobs
- Responsible as an offshore Team Person to provide regular updates/interactions with the Business on the application ongoing activities.

CERTIFICATIONS

- Azure Data Fundamentals – DP 900
- Azure Data Engineering Associate – DP 203

EDUCATIONAL DETAILS

B.Tech from Madanapalle Institute of Technology and Science with 7.85 CGPA

DECLARATION

I hereby declare that the above mentioned details are true to the best of my knowledge.