

KARTHIK THANDAPANI

Enterprise Data Architect

+91 89399 79393

advant.analytics@gmail.com

India



EXPERIENCE

Cloud Architect

Indium Software Pvt.Ltd

06/2021 - 05/2022 Remote

- Designed & built an ETL pipeline to ingest data from multiple data sources **S3, Athena, Lambda, Glue & Redshift**. Fully automated it.
- Worked with pre-sales in various tech-solutioning activities. Brought design architecture maturity to their data-driven activities.
- Designed modern data warehouse on **Snowflake**. Built quick prototypes/POCs. Built serverless applications on **AWS cloud**.

Lead Business Consultant (AWS)

Tech Mahindra

12/2018 - 05/2021 Hyderabad, Canada, Mexico

- Developed ETL process using **BigData** technologies like **Hadoop, Spark** to process ~~terabytes of data~~ and its downstream activities.
- Built data pipelines and analysis tools using **Python, PySpark & Scala**.
- Designed and developed real time streaming pipelines for sourcing data from *IOT devices*, defining strategy for data lakes, data warehouse, retention, aggregation, summarisation for optimising the performance of analytics products using **Snowflake**.
- Created **Scala** pipeline activities to bring data from **RDBMS** to Amazon **S3**. Ingestion pipeline using **Lambda & Glue ETL**.

Technical Lead

Franklin Templeton (Innova Solutions)

05/2018 - 11/2018 Vizag

- Designed & built flawless ETL pipelines using **S3, AWS Glue, Lambda & Snowflake**.
- Designed & built real-time applications using **Apache Kafka & Spark Streaming**.
- Experience in *Data Warehousing applications*, responsible for the Extraction, Transformation and Loading ETL of data from multiple sources into Data Warehouse. 85+ data sources.
- Provided guidance to development team working on **Scala** as ETL platform.

Cloud BPI Advisor

NTT Data

03/2016 - 04/2018 Chennai, OKC, California

- Migrated data from clients **HDFS cluster** to the cloud platform by building ETL process using **BigData** tools like **Sqoop, Hive, Spark, Scala** and **HBase**.
- Experience in optimising Hive **SQL** queries and Spark Jobs.
- Optimized the **Pyspark jobs** to run on **Kubernetes Cluster** for faster data processing.
- Designed & build modern **data-warehouse in Snowflake**.

Project Manager

Encore Software & IT Services

10/2015 - 02/2016 Chennai

- Built **PySpark** based applications for both *batch and data-streaming* requirements. Optimised existing data pipelines.
- Led **database migration** across 3 continents.

TECHNOLOGY STACK

AWS Azure Snowflake Spark

S3 Athena Lambda Glue

Redshift Kafka Python Spark

Scala Talend Cribl DBT

Airbyte Kinesis SQL Spark

HIVE EMR DynamoDB Pyspark

R Hubspot Google BigQuery

EC2 RDS Dremio

Machine Learning AI

AWS IOT Core Apache Airflow

Terraform IT-Automation

Docker Kubernetes Devops

Jenkins AWS Codepipeline

Azure Devops

PowerBI Tableau RShiny

Quicksight

JIRA Confluence Git

EDUCATION

Bachelor of Science

Acharya Arts & Science College

2007

EXPERIENCE

Senior Consultant

Accenture

📅 03/2014 - 10/2015 📍 Chennai

- *Optimised performance* for data access requirements by choosing the appropriate native **Hadoop** file formats **Avro, Parquet, ORC** and compression techniques, respectively.
 - Worked on various Hadoop file formats like Parquet, ORC & AVRO.
 - Migrated On-prem **informatica** ETL process to **AWS cloud. Airflow**.
 - Promoted the components to production environment through CICD process by using **Jenkins, Github and Kubernetes**.
-

Sales Analyst

NetApp

📅 06/2010 - 10/2013 📍 Bengaluru

- Build integrated solutions leveraging **Unix shell scripting, RDBMS, Hive, Hadoop, HDFS File System, HDFS File Types, HDFS compression codec**.
 - Created and managed schemas and instances for **SQL** and **noSQL** databases.
 - Excellent knowledge of Machine Learning, Mathematical data Modelling and operations Research. Comfortable with **R, Python, SAS** and **Relational databases**. Deep understanding & exposure of **Big Data Ecosystem**.
 - Experience implementing operational best practices such as **monitoring, alerting, catalog and metadata management**.
 - *Optimisation of Hive queries* using best practices and right parameters . *Automated devOps using ADO pipelines*.
-

Process Executive

Infosys

📅 06/2007 - 05/2010 📍 Bengaluru

- Built analytics tools that utilise the **data pipeline** to provide *actionable insights* into customer acquisition, operational efficiency and other key business performance metrics.
- Helped with **intelligent data partitioning and bucketing techniques**.
- Designed & developed ETL proceses on **AWS cloud** and *automated event driven activities*.
- Helped to build modern data lake using **AWS S3**.