KANNOJI ARUN

**Data Engineer**

# CAREER OBJECTIVE



A passionate & Experience Data Engineer with an overall 2.5 years of experience in designing & building ETL pipelines for data processing and implementing and processing large datasets of different types it’s including structured, semi-structured and unstructured data. 7981968681

|  |
| --- |
| **PROFESSIONAL EXPERIENCE** |
| **LOGRAN TECHNOLOGIES PRIVATE LIMITED (Bangalore)**  **(Sep 2020 to Present)**   * Experienced in working with different file formats which includes CSV, Parquet,Avro, JSON etc. * Automate the **ETL** process * Experienced in the Azure ecosystem such as Azure Data Factory (ADF), Azure Data Pipeline, Data Flows, Azure Synapse analytics & Data Bricks,Azure data lake,Blob storage. * Doing with Best practices of ADF implementation and performance tuning techniques. * Created mount point for data lake storage gen2 from Databricks notebook. * Migrated data from Amazon S3 Bucket to Azure Synapse analytics. * Loading data from Amazon S3 bucket To AWS Redshift. * Knowledge on **Spark /Pyspark** and worked on Spark Data Frames/Dataset * Experienced in **Spark** using optimization techniques (join, Partitioning, Bucketing etc.) * Worked and exposure on **AWS services** such as S3,Data Brew,EMR, Glue, Athena Redshift,EC2,VPC,Cloud Load Balancing. * Worked on python for **pre-processing** and **Transformation**. * Experienced in working **SQL** (joins, CTE, sub-queries,etc) * Experienced in implementing data integration standards. * Excellent communication, interpersonal, analytical skills, and strong ability to perform in a team * Experience in end-to-end implementation of the Project including requirement mapping, finalization of specifications, design, development, testing and implementation. * Have knowledge of data warehousing processes such as extracting, transforming, and loading. * Good understanding star Schema and Snowflake Schema. |

arun.karun2568@gmail.com

http://linkedin.com/in/arun-kannoji-20a330250

# TECHNICAL SKILLS

## BIG Data Tools:

* Apache Spark

## RDBMS:

* SSMS

## 

## Programming Languages:

* Python
* Pyspark

## Operating Systems:

* Linux
* MS Windows

## Cloud computing:

* Azure
* AWS compute services
* Google Cloud Platform

**Responsibilities**

As a Team member took responsibilities of:

**Soft skills**

* Management
* Communications
* Presentation skills
* Leadership

**Education**

* B. Tech in KESHAV MEMORIAL INSTITUTE OF TECHNOLOGY.

* Managed data coming from different sources.
* Transferred data from a Data Source to RDBS to perform analysis.
* Loaded and transformed large sets of structured, and semi structured data.
* Load data into the target table by normalizing all the data.

# Projects

## Publicis (Sep 20 –Present)

**Objective:** It was automated Data Management Monitoring process of customer. reducing the Manual effort. It consisted of various modules such as data collection, data cleaning, Transformation, segmentation, and reporting.

## Responsibilities:

* Experiencing with core Azure data components & very strong adjacent experience from other platforms.
* Working on copy data process for extracting data from different

sources and storing into Azure Blob storage.

* Created the mappings using transformations such as the Derived Column,

Aggregate, Merge, Joins, Conditional Split and Lookup.

* Orchestrated data integration pipelines in ADF using various Activities like

Get Metadata, Lookup, For Each, Wait, Execute Pipeline, Set Variable, Filter, until, etc.

* Worked with SCD Type 0, Type 1 & Type 2.
* Experience building and optimizing big data pipelines, architectures, and data sets.
* Built new end to end integration’swhich was integrated with various servers. Understood the system requirements of different targets and transformed the data accordingly.
* Worked closely with team on design, implementation, and performance issues.
* Managed data accordingly. Define and implement solution to minimize efforts and speed up the deliverables.
* Created External tables in Azure SQL Datawarehouse

for data visualization and reporting purpose for business users.

* Created Data Marts for different use cases.
* Configured Email alerts for success and failures of ADF Pipelines by using Logic apps.
* Configured Triggers for the synapse pipelines.
* Created GitHub repositories and specified.

**Tech Stack** Azure, ADF, Synapse, Data Bricks, GitHub.

**DECLARATION:**

I hereby declare that the information given above is true and correct to the best of my knowledge.

**Place: Bangalore, India.** **Name: Kannoji Arun**