# Amar Kumar.

(HIVE, SQL, python,PySpark Control-M, ABINITIO, Unix, Azure DataFactory, AWS ,Azure Data Lake)

Professional Summary

* Have 3 years of IT experience.
* Having good knowledge on **Hive**, Data warehouse, **SQL**, , **PySpark**, **DataFactory** , and Unix.
* Having good knowledge on tools like **HiveQL**, **PIG**, **Spark SQL**, **Spark RDD**, **Azure DataFactory**
* Involve in create Hive tables, loading with data and writing Hive queries which will run internally in MapReduce way.
* Experience in Pyspark to transform the data as per business logics and processing facts & dimensions.
* Created reusable function using PySpark to reduce the code redundancy in multiple notebook
* Worked on improving the performance of Pyspark code by using broadcast Join , persist and changing the shuffling and splitting the join into multiple dataframe.
* Extracting, transforming and loading data (ETL) between various Homogenous and Heterogeneous data sources using Abinitio & Azure DataFactory and worked on improving ETL performance.
* Good Knowledge on **Hadoop** data storage system (Mapreduce).
* Experience in **Sqoop** for loading data into HDFS.
* Experience in **Control-M** tool for scheduling.
* Created metadata driven Azure Data Factory pipeline to load data from various source by implementing Data Quality checks & re-start from point of failure.
* Excellent communication skills, Goal oriented and highly motivated organized team player with high aptitude for analytical, technical and problem solving skills.
* Worked on **Delta lake.**
* Experience building web portals using Services Oriented Architecture and web services and able to consume/create if required. Proficient with knowledge of Python, Django, Mysql, Js, HTML, Extensive experience in designing web sites. .Planning, design and implemen tation of complete website and Admin Panel. .Strong logical and analytical skills. .

|  |  |
| --- | --- |
|  |  |

Work Experience

Associate Data Engineer April 2020-Present

Project Profile

Project: Modern Data Platform

**Working Company: Futurense Technologies**

**Client: The Smartcube**

* Understand data requirements from the business for the sources to onboard and to discuss on the logics to implement for getting the desired metrics.
* Created reusable function using PySpark to reduce the code redundancy in multiple notebook
* Worked on improving the performance of Pyspark code by using broadcast Join , persist and changing the shuffling and splitting the join into multiple dataframe
* Created metadata driven Azure Data Factory pipeline to load data from various source by implementing Data Quality checks & re-start from point of failure.
* Mounted ADLS in Azure Data bricks and created Notebook using PySpark to clean the data and apply business logic.
* Created reusable function to handle basic operations like duplicate check, creation of default value for Dimension & Facts etc.
* Used ADF for running Azure Databricks notebook by creating dependency with all the stage load.
* Moved final facts and dimension into Azure Data Warehouse for downstream to consume.
* Environment: SQL Server 2016, Azure Data Factory (V1/v2), Azure DataBricks, Azure DataLake, Azure DataWarehouse

Project: TA\_Associates

**Working Company: Futurense Technologies**

**Client: The Smartcube**

**Description:** TA\_ Associates platform for portfolio company. Porfolio is a collection finance investment like stocks, bonds, commodities and cash and cash equivalents including close-end fund. They comes and upload file for quarterly and yearly basis.

**Responsibilities:**

* Understand data requirements from the business for new sources and for the existing sources analyzed legacy system.
* Writing APi’s
* Creating Schema’s for database
* Upload file on AWS S3
* Storing data on AWS RDS

Project: Biocine

**Working Company: Oodles Technologies**

**Client: Biocine**

**Description:** Biocine Entertainment was established as a partnership firm in 2017. It provide entertainment based mobile and web application service.

**Responsibilities:**

* Understand data requirements from the business for new sources and for the existing sources analyzed legacy system.
* Creating some module and design.
* Perform testing

Project: Planetshare

**Working Company: Planetcast Media Service Pvt Ltd.**

**Client: Planetcast**

**Description:** Planetcast platform integrate different products and services under one unique platform to foster profitability for your business. With our platform, you can connect with buyer, seller and vendor(service provider) from all over the world to monetize your content and services. This system provide servicing like dubbing, transcoding etc.

**Responsibilities:**

* Understand data requirements from the business for new sources and for the existing sources analyzed legacy system.
* Writing API’s
* Creating schema’s for database

Trainings

* Futurense Training in **Big Data** stream attended for 45 Days

Education