



**Bikash Jha**  
Data Engineer

Leverkusen, Germany 51381  
+49 (0) 157 50325815  
[jhabikash2829@gmail.com](mailto:jhabikash2829@gmail.com)  
[linkedin.com/in/bikash-jha/](https://linkedin.com/in/bikash-jha/)

---

## Profile

I'm a driven data engineer with 5+ years of experience playing with big datasets. I have experience and knowledge in a diverse set of disciplines, technologies and tools like Data Migration, Cloud and data pipelines, and data architecture.

Looking for a career opportunity to apply my skills/experience on challenging projects. Eager to build robust databases that lay the groundwork for revealing game-changing insights for the business. Involving extensive use of current IT techniques to contribute productively towards the growth of the company thus growing professionally.

Skills: Data Structures, Algorithm, Serverless, Data Pipelines  
Technologies: Python 2/3, Spark, Spark Streaming, Elastic, Kubernetes, Docker, Airflow  
Cloud: AWS, GCP  
Databases: MySQL, MongoDB, DynamoDB, PostGre, HBase  
Big Data: Kafka, Hive, Zookeeper, Oozie, HBase

---

## Experience

### Planet Labs, Germany

#### Senior Data Engineer

Feb 2022 – Current

- Creating infrastructure design to process geolocation data coming from 250+ satellite
- Implementing engine to process terabytes of geolocation data using Geospark analytics
- Designing DataLake (Snowflake) and loading the data
- Maintaining MySQL database on Cloud
- Storing Latitude & Longitude data to MySQL database as backend of one of our services
- Storing Spatial data type and schema design on MySQL
- Creating imagery APIs for google maps used by Google

Tech Stack: MySQL, AWS, CloudWatch, Kubernetes, Python, Kafka, PySpark

### Homelike Internet GmbH, Cologne, Germany

#### Senior Data Engineer

OCT 2021 – Feb 2022

- Managing 6 different ETL processes to transfer data from MySQL to MongoDB, BigQuery, Salesforce, Snowflake and several Marketing Channels using PySpark
- Migrating database to Cloud SQL for MySQL by using Database Migration Service
- Migrating MySQL server databases/ existing licenses to Azure SQL Database using serverless framework
- Designing Snowflake Schema and NoSQL Schema Design best practices.
- Designing serverless framework for real-time consumption of user tracking data using kinesis
- PySpark and Airflow autoscaling solution with Kubernetes (Helm)

- POC for a customer: Migration of MySQL to Amazon Cloud (RDS)
- Optimizing Schema for import/instance/monitoring on import process

Tech Stack: MySQL, GCP, BigQuery, AWS, Lambda, Kinesis, SQS, CloudWatch, Kubernetes, Python  
 Migration Tools: PySpark on Kubernetes, DMA, DMS, JDBC Spark Connector

## Aurigo Software Technologies, Bangalore, India

### Senior Data Engineer

JAN 2021 - SEP 2021

Company Description : Helps state agencies, cities, counties, water authorities, airports, and facility owners plan, build and maintain capital assets, infrastructure and facilities by combining Data Engineering, Democratized Data Science and Data Orchestration ([link](#))

Project: Serverless / Managed AWS services

- Implementation of Amazon Kinesis Firehose to collect Realtime, streaming data
- Designing AWS S3 as the Data Lake to store all Raw Data.
- Setting up AWS Athena to analyze data in Amazon S3 using standard SQL.
- Aggregated report ETL to MySQL from S3 for reporting dashboard
- Kubernetes on AWS EKS Cluster using Cloud Formation Template.
- EKS Cluster Auto-scaler and setting up Amazon ECS for container orchestration.

Project: Microservices Components on Kubernetes.

- Writing Connector Microservices to fetch data from different sources.
- Pyspark/PandasDF/Boto3 Connector/: Twitter, S3 bucket, DynamoDB, Filesystem.
- .
- Dag Engine Microservices: Integration engine for all microservices (ML + Connector)
- Airflow Microservices: programmatically author and schedule their workflows and monitor them via the built-in Airflow user interface.
- Attaching EFS volume mount/AWS RDS instance to Airflow Kubernetes Pods.
- Fluentd microservices and Elastic stack to fetch/store logs from Airflow.
- Implementing spark-on-k8s operator pod to run spark jobs on Kubernetes.
- Monitoring Kubernetes cluster health using K8s Dashboard, Prometheus and Grafana.
- Exploring the GCP Platform for future POC's and Kafka Microservice.

Tech Stack: Linux, AWS, Kubernetes, Docker, Python, Airflow, DynamoDB

## LTI (Larsen & Toubro Infotech), Pune, India

### Senior Product Engineer

JAN 2020 - DEC 2020

Company Description : LTI Mosaic Decisions Platform

Responsibilities:

- Creation of connector service to load data from and to MySQL/RDBMS from different sources
- Refactoring existing code and evolving new architecture fit for our existing product
- Writing Connector code using spark/ pySpark / PandasDF for Azure/S3/NO SQL etc
- Kubernetes setup on AMAZON EKS cluster and implementation of key vault (HashiCorp)
- Writing our own Kubernetes API to submit spark jobs on K8s and KubeSpawner APIs.
- Horizontal Autoscaling on Kubernetes Pods and optimization of existing spark Jobs.

Tech Stack: Python , Amazon S3, Azure Blob, Cosmos DB, MongoDB, Kafka, Kubernetes, PySpark

## AMDOCS, Pune, India

### Software Developer

AUG 2016 - DEC 2019

Company Description : Platform that implements business logic and Allows the marketer to inject machine learning logic run on a big data system for any decision that needs to be taken within the experience ( Email/Message text etc.)

Project: Auto-ML Product Recommendation

- Setup/Code for consumers/Producer and offset management in Kafka streaming (Listener balancer)
- Cleaning up data and data preparation for Model Creation using Pyspark
- Implementation of ML model (Random Forest/Linear Regression based on demographic data and supervised learning to predict potential customers
- Analysis of target customer feedback and provide insight to marketing team

Tech Stack: Pyspark, Python, MapR, Kafka, Hive, Random Forest

---

## Education

University Of Kalyani, West Bengal

Bachelor of Technology in 2016 Information Technology

2016

Awards: Awarded as Employee of the Month 2017, Awarded as Employee of the Quarter Q2 '18, On-Site Opportunities - Mexico for AT&T and Manila for PLDT.

---

## Other Interesting Facts

Languages: English (Professional), Hindi (Native), Bengali (C1)

Interests: contributing to open Source Platform , Volleyball, Cricket

---

## Reference

Available upon request