



Data Science at Scale



# Introduction

## **Shashank Agarwal**

Associate Director, Engineering

Founder,  THIRDWATCH (Acquired by Razorpay)



[www.linkedin.com/in/shashank734/](https://www.linkedin.com/in/shashank734/)

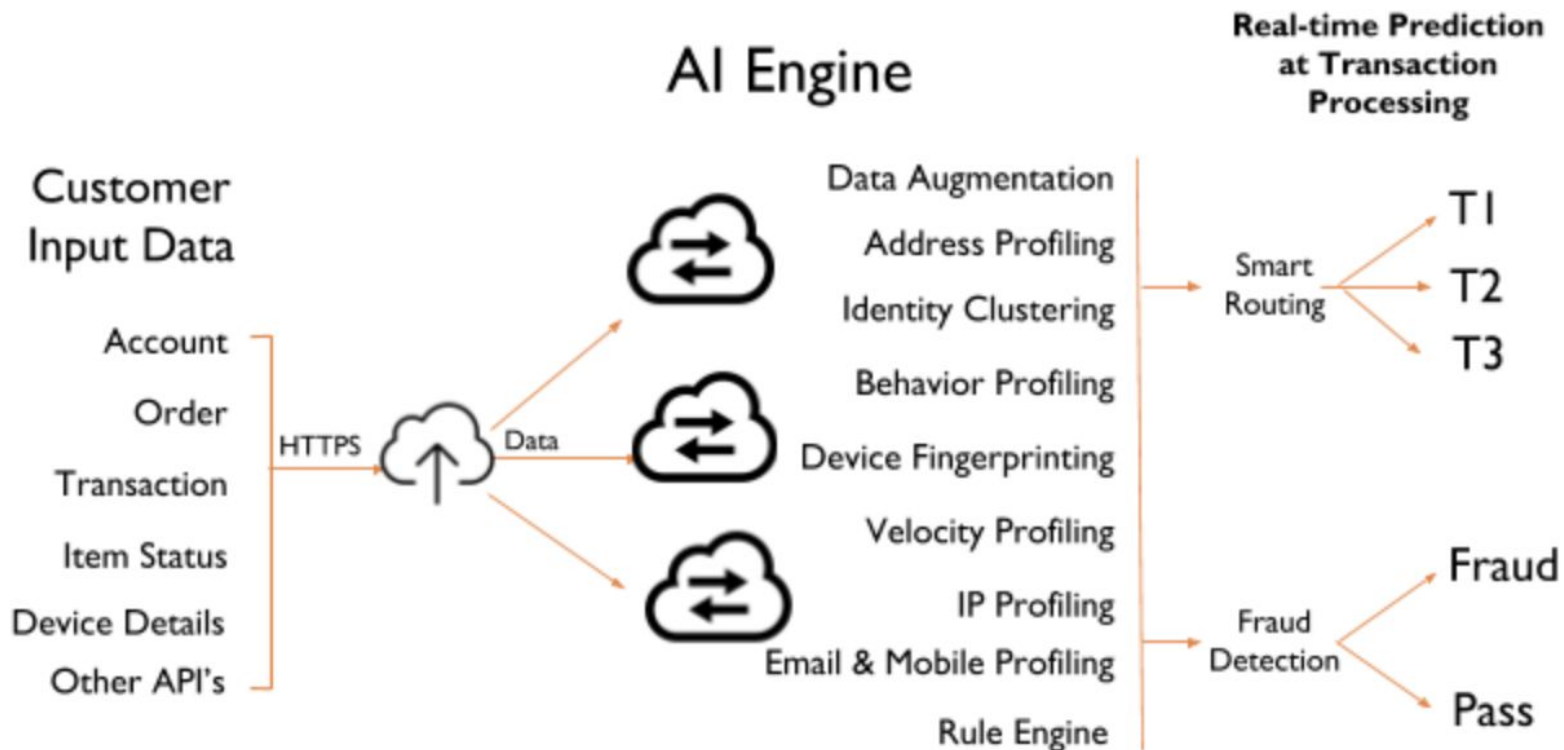


@shashank734

Razorpay processes a huge amount of data on a daily basis

- 5 Million merchants
- 25 Billion USD Annualized Payment volumes
- 100+ parameters for each transaction

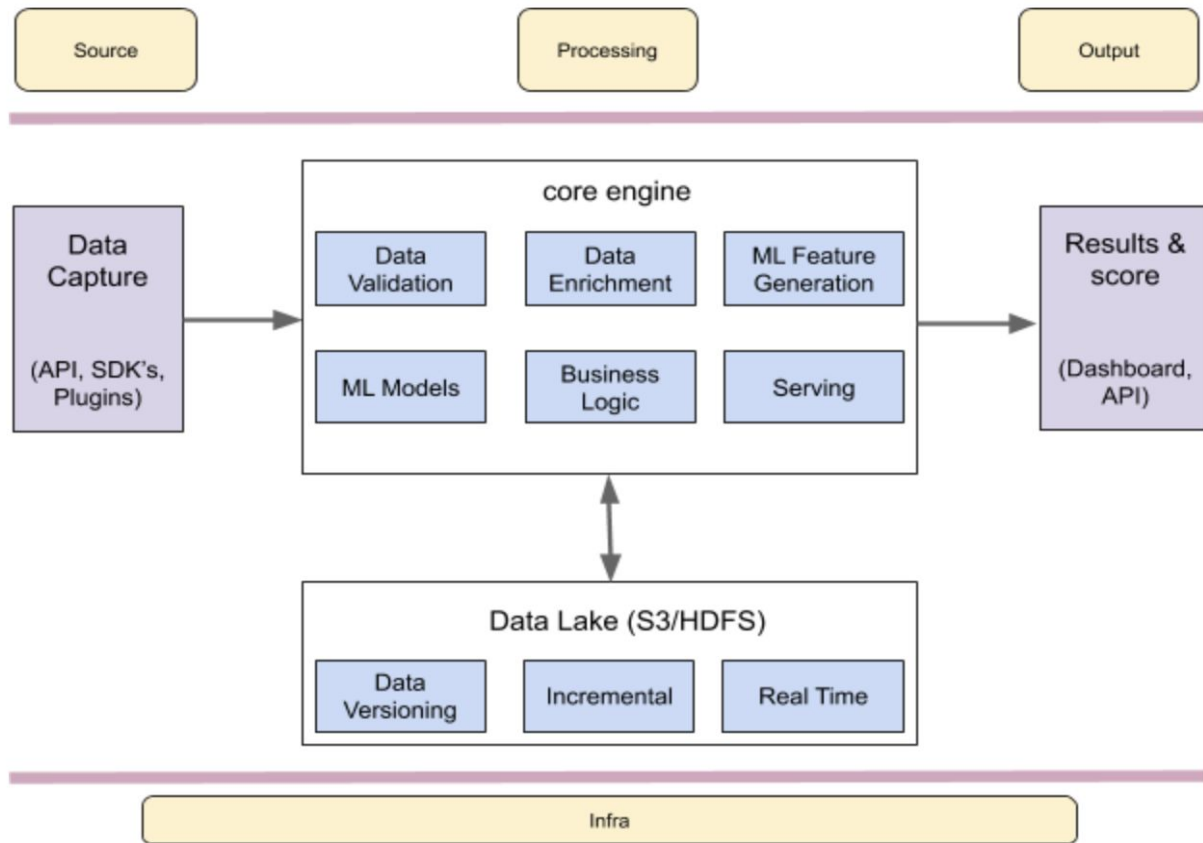
# Data Science @ Razorpay



# Why Apache Flink?

- Low latency streaming
- Flink's in-memory states
- CEP (Complex event processing) and event time streams
- Async IO
- Automatic Checkpointing for fault tolerance
- Active community support and active development. For example, Flink introduced the most awaited features schema evolution and State TTL.

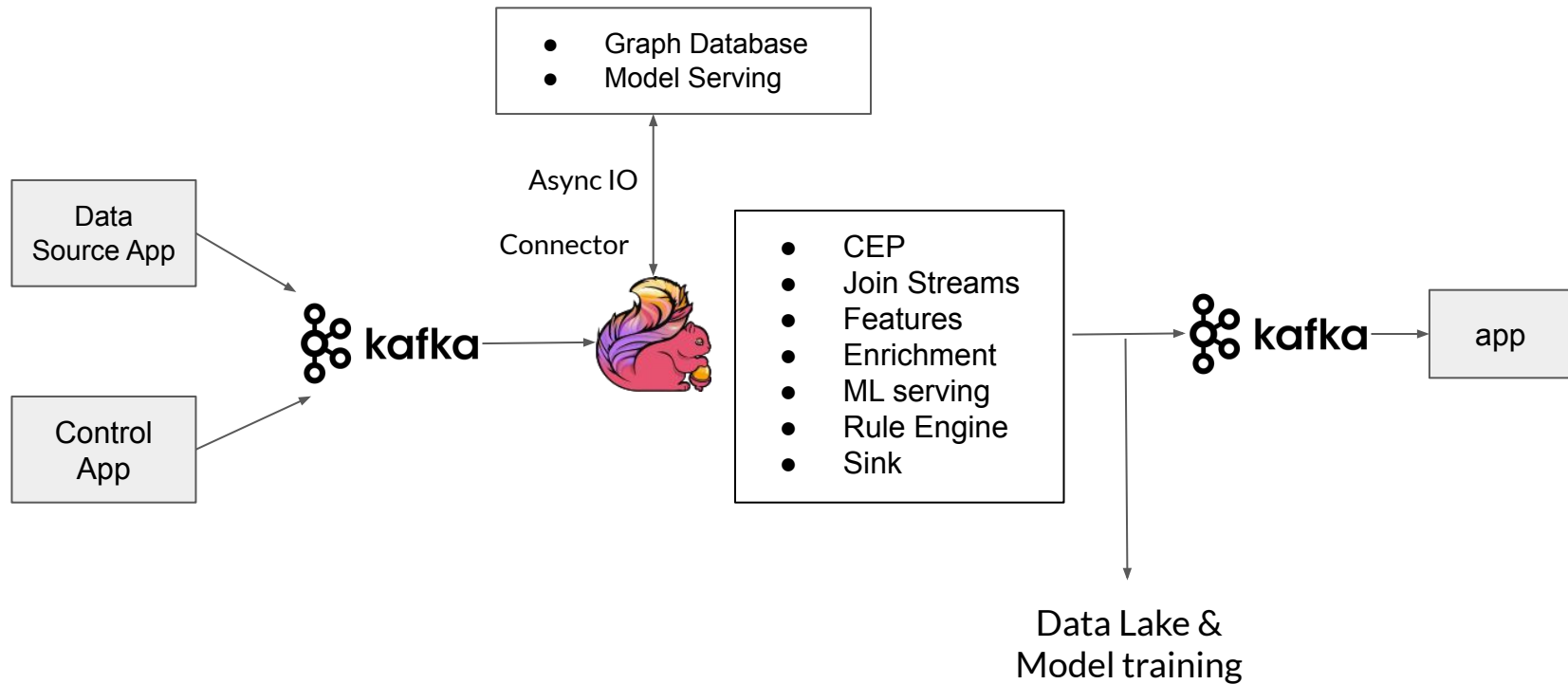
# Introduction to Mitra



# Key Features of Mitra

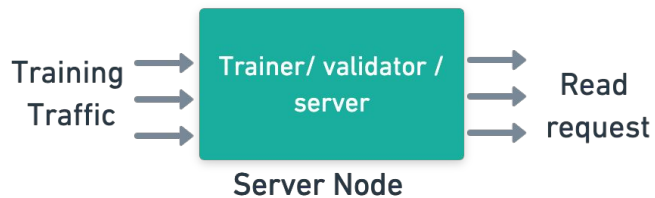
- Predict results within 200 milliseconds in the distributed environment
- Generate Hundreds of stateful features on the fly using 60+ operators
- Serve results from multiple deployed ML models
- Dynamic rule engine on Flink streams
- Heavily use Flink's in memory states for storing features and data
- Use CEP (Complex event processing) to handle async events

# Core Engine



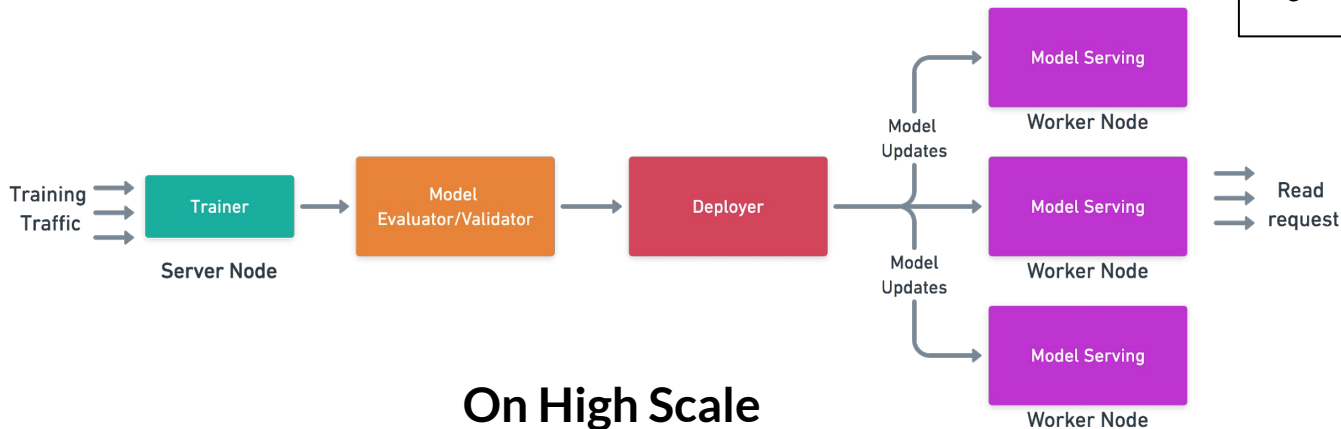


# Model Deployment



On Small Scale

- Network Load
- Resource allocation
- Limit traffic size





Thank You!

Ready for a new challenge? We are Hiring!  
Write to me at [shashank.a@razorpay.com](mailto:shashank.a@razorpay.com)



[www.linkedin.com/in/shashank734](https://www.linkedin.com/in/shashank734)



@shashank734

