Analysis and visualization of complex software system traces

FINAL EVALUATION

Group 7

QA for Fintech Applications using Machine Learning

PRODUCT OWNERS

Dr. Asitha Bandaranayake

Senior Lecturer
Department of Computer Engineering
Faculty of Engineering
University of Peradeniya

Prof. Rostislav Yavorskiy

Head of Research at Exactpro Systems

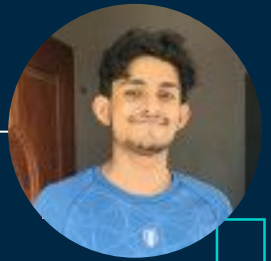
TEAM MEMBERS (Group 07)

ISHAN FERNANDO

E/18/098

ADEEPA FERNANDO

E/18/100



RIDMA JAYASUNDARA

E/18/155

PART 01

INTRODUCTION

Understanding the Problem

1.1

Understanding the Problem

- About 15 years ago log files were simple and human readable
- Now, log files extends upto GigaBytes while selections containing the word ‘error’ is about 200 -300 MegaBytes
- Typically, log files can contain millions of lines per working day, making manual analysis impractical and error-prone.
- QA Engineers nightmare …

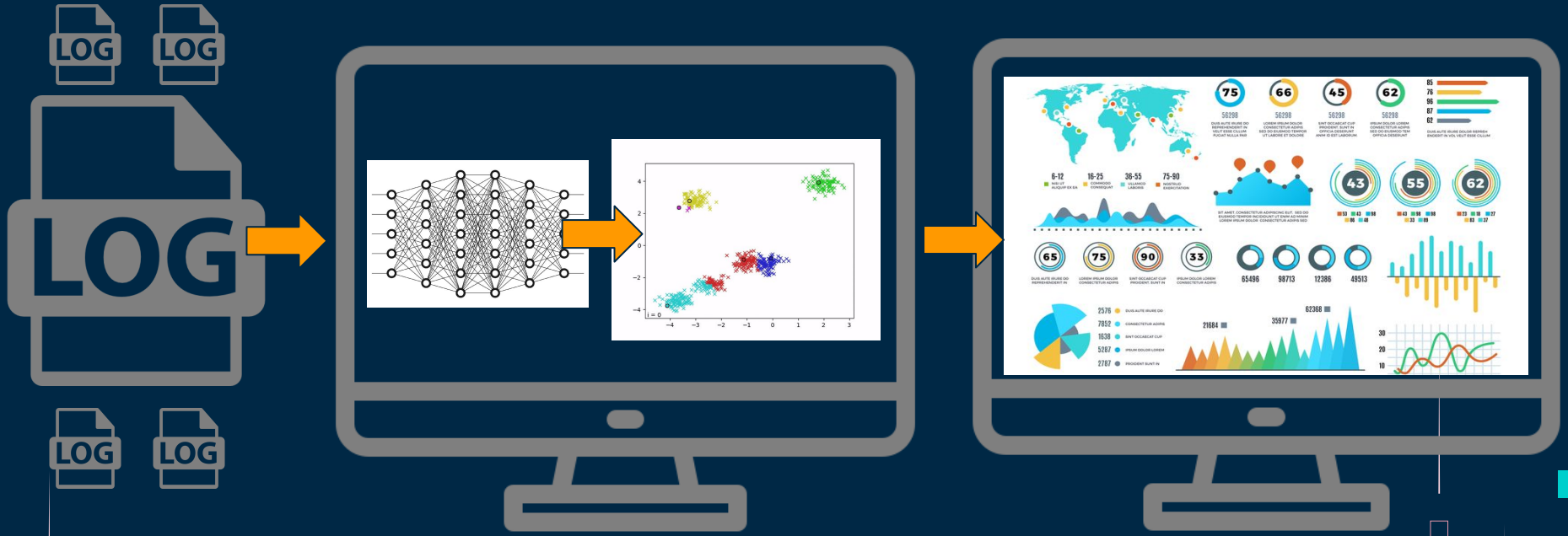
Solution

1.2

Solution

- To develop machine learning models for anomaly detection and issue identification in system traces (in form of log files)
- To visualize the system traces using interactive dashboards to provide a human understanding of what is happening in the system
- To evaluate the effectiveness of the proposed approach on real-world FinTech applications

Solution



Process 1.3

PROCESS

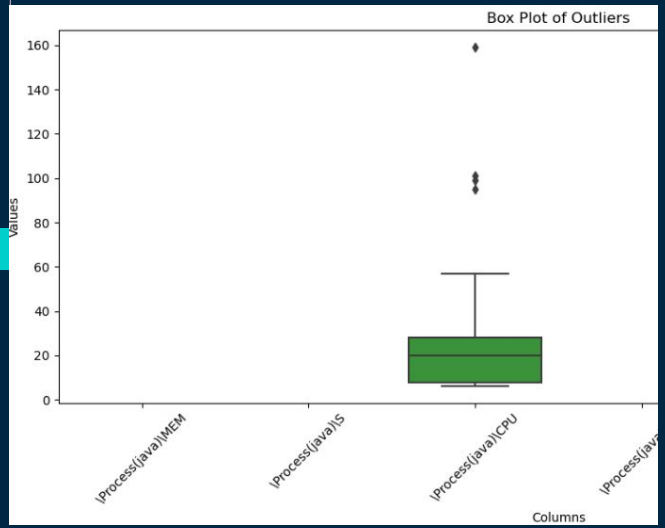
1. **Data Collection**
    - Size of data set = 45 GB
    - No. of log files = More than 27,000
2. **Exploratory Data Analysis (EDA)**

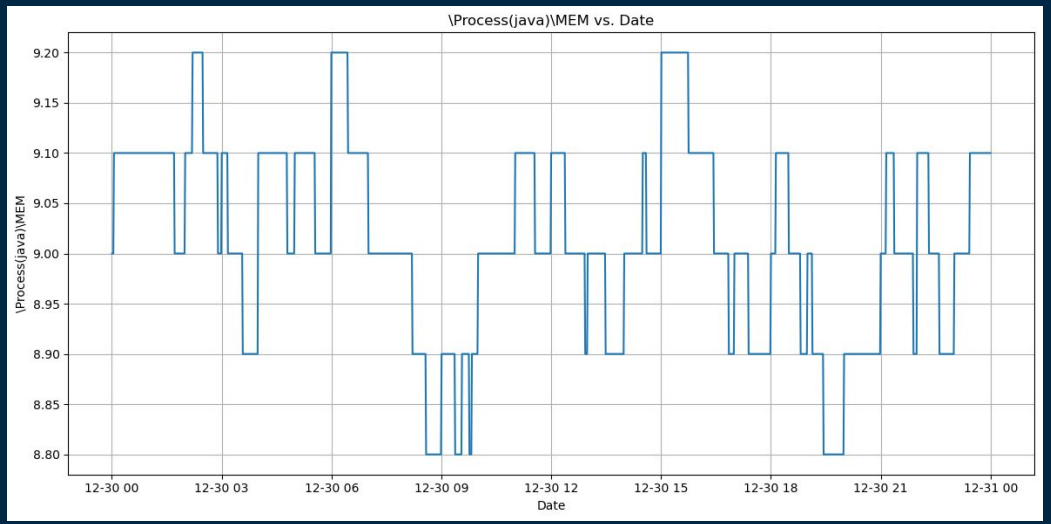

```
[
    {
        "name": "2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.03.01_00.00.01.csv",
        "columns": 13,
        "size": "156K",
        "line_count": "1419"
    },
    {
        "name": "2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.02.28_00.00.01.csv",
        "columns": 13,
        "size": "156K",
        "line_count": "1418"
    },
    {
        "name": "2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.02.27_00.00.00.csv",
        "columns": 13,
        "size": "156K",
        "line_count": "1414"
    },
    {
        "name": "2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.02.26_00.00.00.csv",
        "columns": 13,
        "size": "152K",
        "line_count": "1417"
    },
    {
        "name": "2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.02.25_00.00.01.csv",
        "columns": 13,
        "size": "152K",
```


File Name	Number of Columns	Size	Line Count
2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.03.01_00.00.01.csv	13	156K	1419
2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.02.28_00.00.01.csv	13	156K	1418
2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.02.27_00.00.00.csv	13	156K	1414
2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.02.26_00.00.00.csv	13	152K	1417
2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.02.25_00.00.01.csv	13	152K	1417
2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.02.24_00.00.01.csv	13	156K	1416
2015/2015-CW-9/lphost18_wls2_unix-process_prod-appserver_2015.02.23_00.00.00.csv	13	156K	1416
2015/2015-CW-9/lphost18_wls2_log_prod-appserver-log_2015.03.01_00.00.01.csv	3	52K	1441
2015/2015-CW-9/lphost18_wls2_log_prod-appserver-log_2015.02.28_00.00.01.csv	3	52K	1441
2015/2015-CW-9/lphost18_wls2_log_prod-appserver-log_2015.02.27_00.00.00.csv	3	52K	1441

PROCESS

3. **Statistical Log File Analysis**

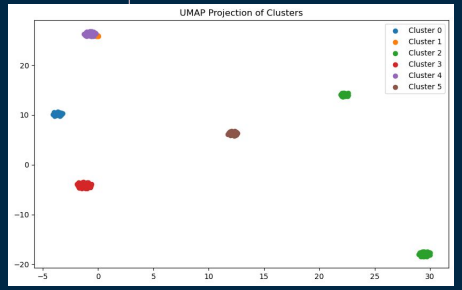


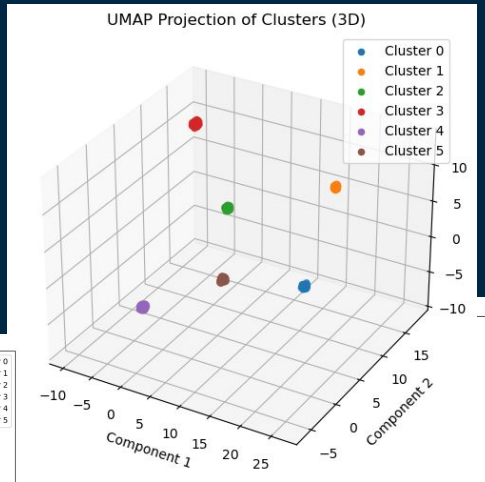


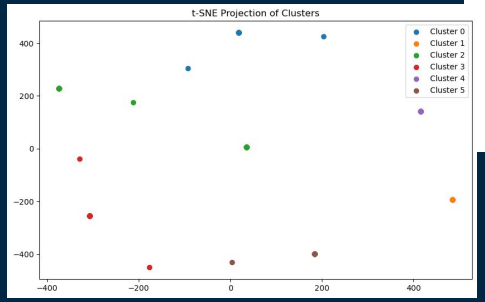
PROCESS

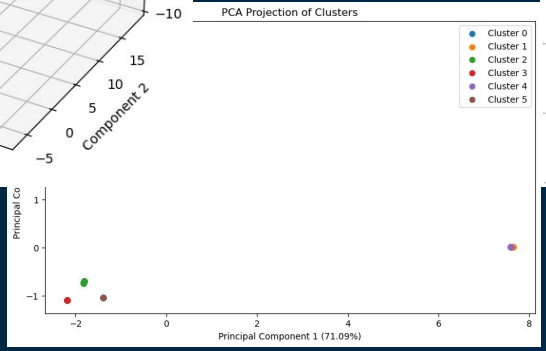
4. Raw Log Clustering Approach

K-Means Clustering

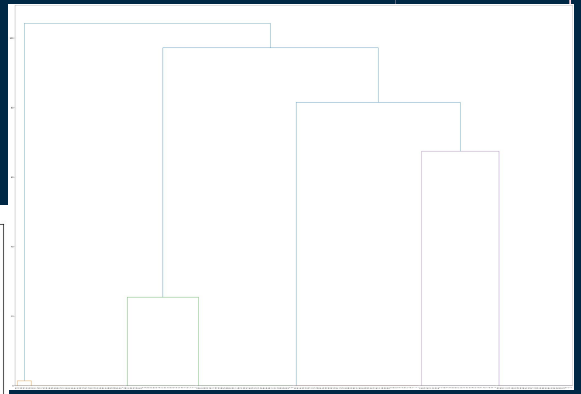






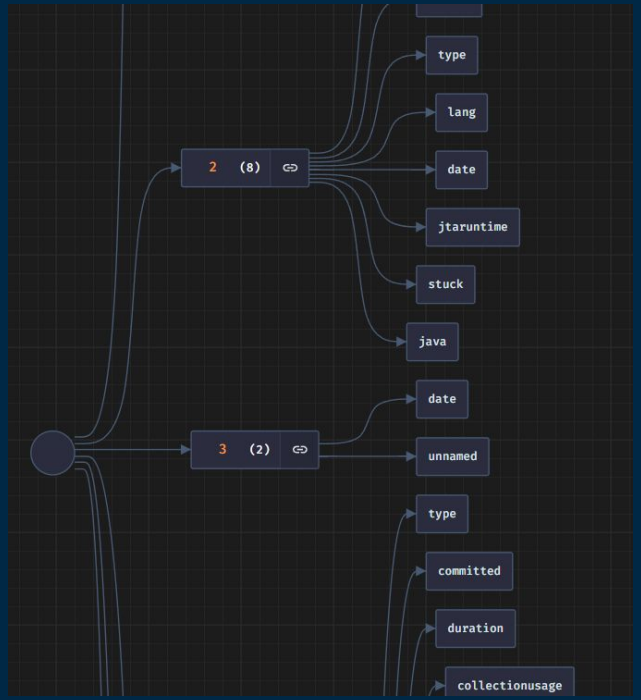


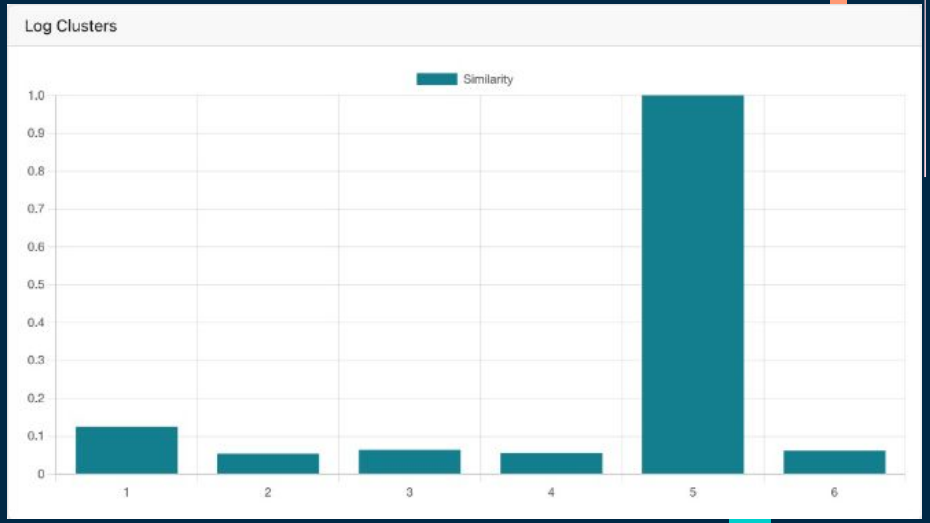
DBSCAN Clustering



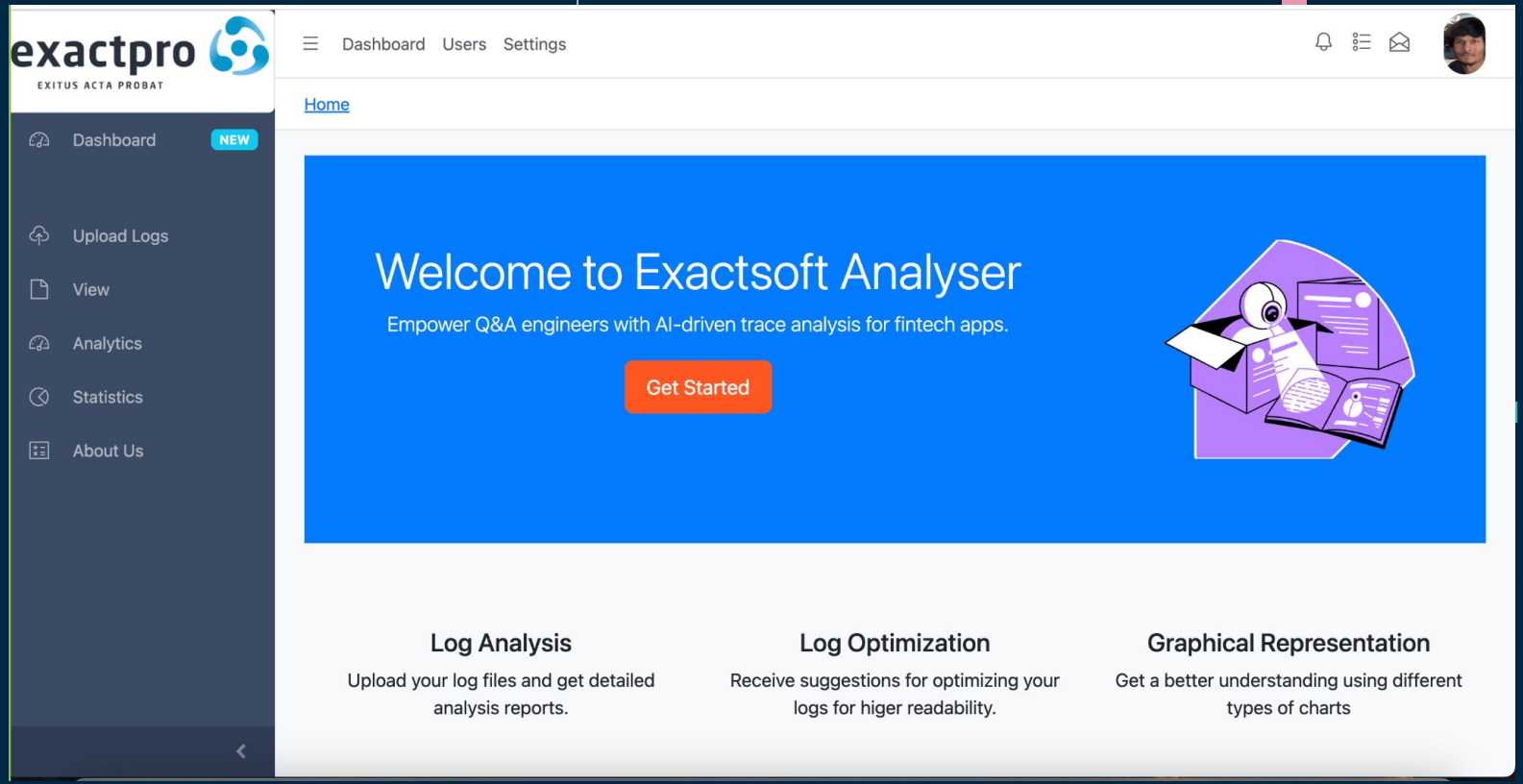
PROCESS

5. Template-Based Approach

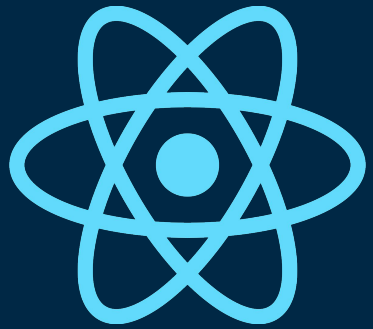


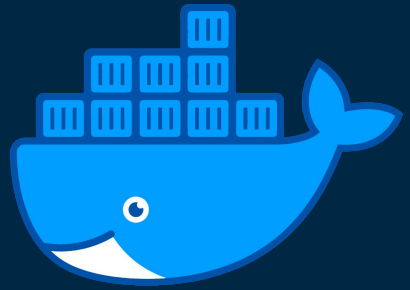


Website Development

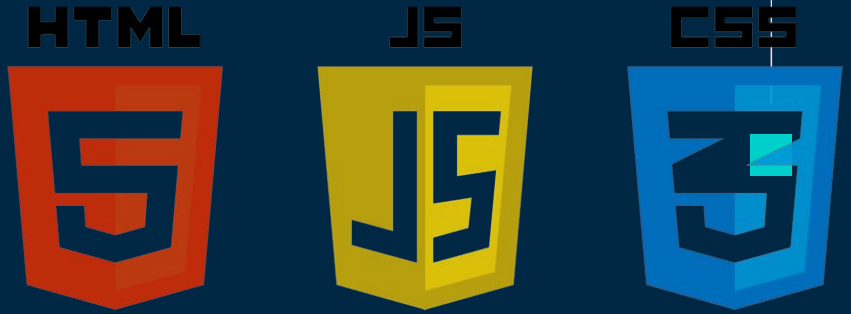


python™





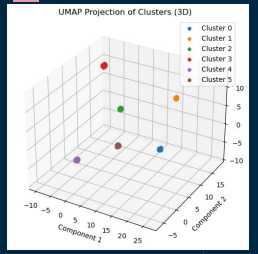
docker



System Architecture

LOG

FRONTEND (REACT)



BACKEND (PYTHON)

python

MongoDB

DATABASE (MONGODB)

MACHINE LEARNING MODEL (PYTHON)

NumPy

pandas

seaborn

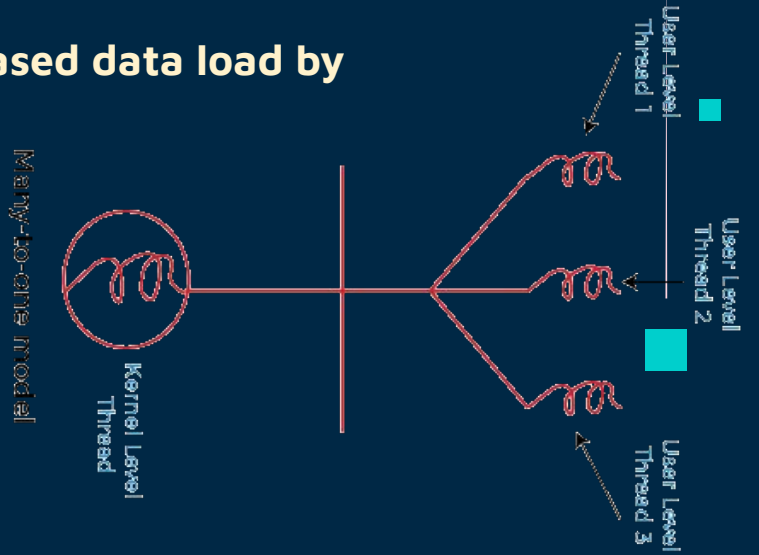
scikit learn

PyTorch

TensorFlow

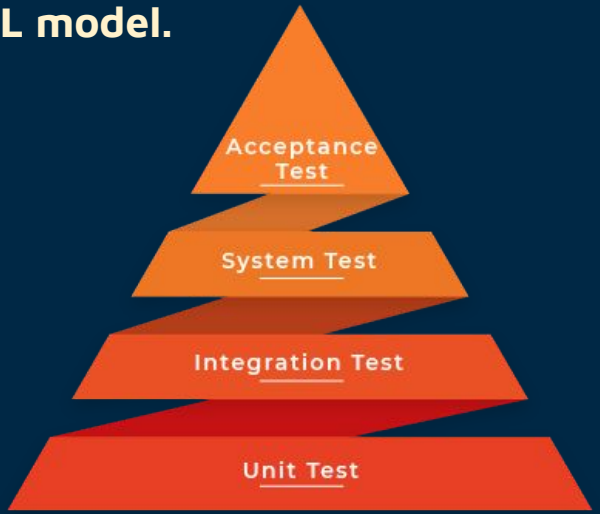
Multithreading

- Concurrent handling of requests and ML computations for faster data evaluation.
- Optimal utilization of system resources like CPU cores for improved performance.
- Easy expansion to handle increased data load by distributing tasks efficiently.



Testing

- Unit tests for the backend API's
- Regression testing for changes and updates in the code
- Integration testing to ensure dataflow from frontend to backend to ML model.



PART 02

DEMONSTRATION

THANK YOU!