MASTER`S PROGRAM / EPAM SCHOOL of Digital Engineering

Data Engineering Specialization Onboarding webinar

November 2, 2023

‹epam›

AGENDA

1	Master’s Program Team Introduction
2	Program Curriculum
3	Digital Platforms
4	Customer Support
5	Project-based Training
6	Q & A

Master’s Program Team Introduction

MASTER'S PROGRAM TEAM

Head of the Program
Natalia Ostashko

Group Manager
Krystyna Popkova

Support Specialist
Darya Ruban

Project-based Training Manager
Dmytro Shlepakov

Administrative Specialists
Tetiana Koshelenko

Daryna Pihareva

* if any questions: orgmastersprogramkz@epam.com

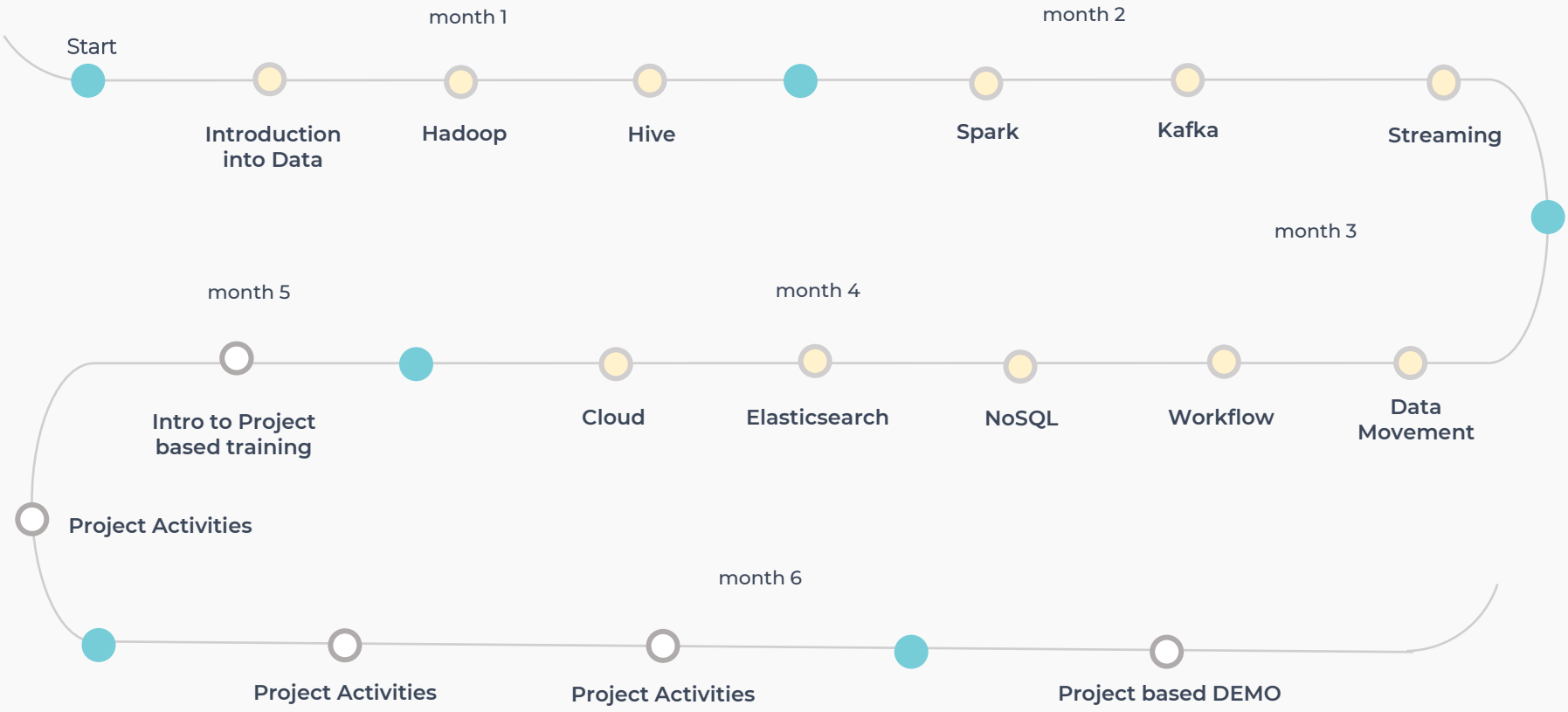
MASTER'S PROGRAM MISSION

Our **mission** is to equip you with advanced knowledge combined with hands on skills taught by experts who have solved complex real-world problems for clients.

Program Curriculum

PROGRAM TIMELINE

Data Engineering (Big Data)



BIG DATA SCHEDULE

Topic	Start Date	Session date (KZ time)	Practical Task Deadline
0. Module Overview	1-Nov		
1. Introduction into Data	1-Nov	8-Nov (Wed) 6 - 8 PM	
2. Hadoop	10-Nov	14-Nov (Tue) 6 - 8 PM 15-Nov (Wed) 6 - 8 PM	
3. Hive	17-Nov	21-Nov (Tue) 7 - 9 PM	27-Nov
4. Spark	24-Nov		
5. Kafka	1-Dec		
6. Streaming	8-Dec		
7. Data movement	12-Jan	24-Jan (We) 6 - 8 PM	29-Jan 5-Feb
8. Data Lakes	19-Jan		
9. Data Warehouses	26-Jan		
10. Elasticsearch	2-Feb	6-Feb (Tue) 6 - 8 PM 8-Feb (Thu) 6 - 8 PM 13-Feb (Tue) 6 - 9 PM 15-Feb (Thu) 6 - 9 PM	
11. Big Data in Clouds	16-Feb		
Final Certification	29-Feb		

[EPAM School of Digital Engineering] [Data Engineering] Module schedule

OM Org Masters Program KZ
To
Cc
Bcc

Reply | Reply All | Forward

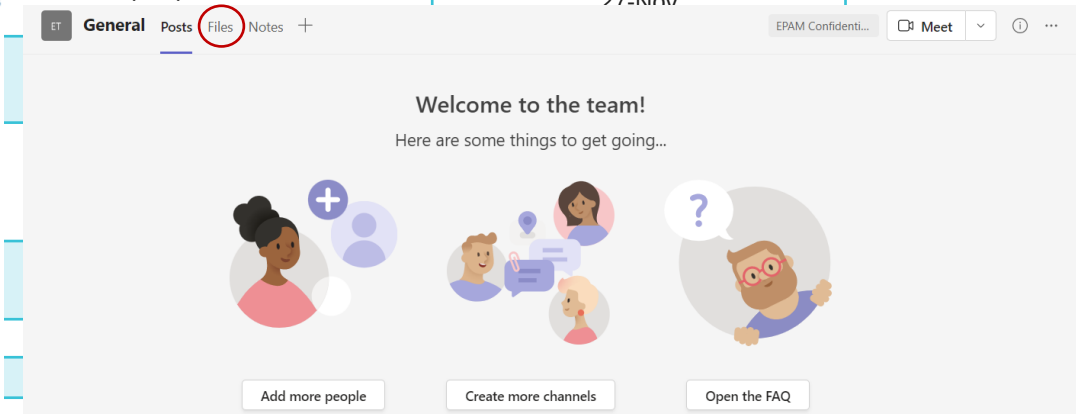
Ср 01.11.2023 12:26

astana hub | <epam> | Tech Orda

Dear Students,
On **November 1st, 2023** the module on **Data Engineering** within EPAM School of Digital Engineering is starting.
Data Engineering Module consists of 11 submodules and 5 practical tasks.

The submodules, deadlines and online session will come into effect step by step on the dates scheduled below:

Submodule	Start date (available on the platform)	Task should be uploaded by
1. Introduction into Data	1-Nov, 2023	
2. Hadoop	10-Nov, 2023	
3. Hive	17-Nov, 2023	27-Nov, 2023
4. Spark	24-Nov, 2023	19-Dec, 2023
5. Kafka	1-Dec, 2023	8-Jan, 2024
6. Streaming	8-Dec, 2023	
7. Data movement	12-Jan, 2024	



BIG DATA MODULE

The Big Data course includes 11 modules designed to provide robust information about data engineering, machine learning and cloud computing.

Introduction into Data

- You will be introduced to the world of big data
- Get acquainted with the data engineering workflow and data product stages.
- Overview of the latest trends in data engineering and key big data tools and applications.

Hadoop

- Will introduce you to the world of Hadoop
- You will overview its ecosystem, get acquainted with the main features, and figure out its possibilities.

Hive

- You will make a deep dive into Hive features
- Will introduce you to User Defined Functions
- You will get acquainted with Hive statistics
- You will have a detailed look at Hive optimization techniques

Spark

- You will familiarize with the basics of Spark, key components, architecture, and various applications
- You will get knowledge on Catalyst
- You will use Spark Streaming in real-time analysis

Kafka

- You will get all the information about Apache Kafka, its capabilities, advantages and drawbacks.
- Will also explore Kafka architecture, components and configuration
- You will get acquainted with Kafka Connect framework and learn how to optimize Kafka

BIG DATA MODULE

Streaming

- You will understand the data streaming approaches, use-cases of streaming data and stream processing concepts
- You will become familiar with more in-depth aspects of streaming with Spark and Kafka

Data Movement

- You will become familiar with Flow-Based
- You will explore about Apache NiFi framework
- You will get acquainted with StreamSets Data Collector

Workflow

- You will become familiar with Workflow tools to manage Big Data Workflows

NoSQL

- You will discover NoSQL databases, its advantages, disadvantages and types
- You will explore MongoDB, HBase, and Cassandra, and how to use them

Elasticsearch

- You will be introduced to Elasticsearch, specifically designed to solve a common problem in software development

Cloud

- You will learn ways to extract data on a much larger scale with better methods to process and analyze this data.
- You will learn how to merge big data with cloud computing

LEARNING APPROACH AND TOOLS

BLENDED APPROACH

Combination of self-study and online sessions.

Theoretical part will be provided through self-study, while live sessions with trainers are focused on collaboration and practice.

SELF-STUDY

Platform

PRACTICAL TASKS & QUIZZES

Platform

ONLINE SESSIONS

Teams

COMMUNICATION WITH TRAINERS

Teams

COMPLETION CRITERIA

What should be done to complete the program successfully?

1. The course includes the following graded activities:

   - Practical tasks
   - Quizzes

2. Each practical task can bring 100 points max.
   Each quiz question counts as 1 point.

3. The points gained by a student for practical tasks and quizzes through the course sum up into the resulting score. The maximum possible score is equivalent to 100%.

4. The final grade for the course is calculated in percent as follows:

$$\frac{\text{total score gained by student through the course}}{\text{maximum possible score}} \times 100\%$$

Success criteria for the program:
final grade is no less than 70%

* The access to final Certification is granted under condition of attendance 70% of online sessions.

CERTIFICATE

After successful completion of the program students receive a certificate.

Certificates include:

- information about courses
- ECTS (hours)
- learning outcomes
- student’s grade

eLearn Digital Platform

EPAM EMAIL

EPAM user account credentials

WS

WFT Support EPAM

To

Cc Anastasiia Bohush

Reply | Reply All | Forward

Thu 14/09/2023 09:20

If there are problems with how this message is displayed, click here to view it in a web browser.

Set_Up_Microsoft_Multi-Factor_Authenticator.pdf
666 KB

We are glad to have you as part of EPAM! Your corporate user account has been created for you to provide access to EPAM resources. Please follow the steps below to complete the account setup.

IMPORTANT:

1. **Before you start setting up your EPAM account, please make sure that your mobile device supports the installation of the Microsoft Authenticator app.**

Minimum system requirements:
- iOS version - 12.0 and above.
- Android version - 6.0 and above, with support of Google services (most of Huawei mobile devices are not supported).

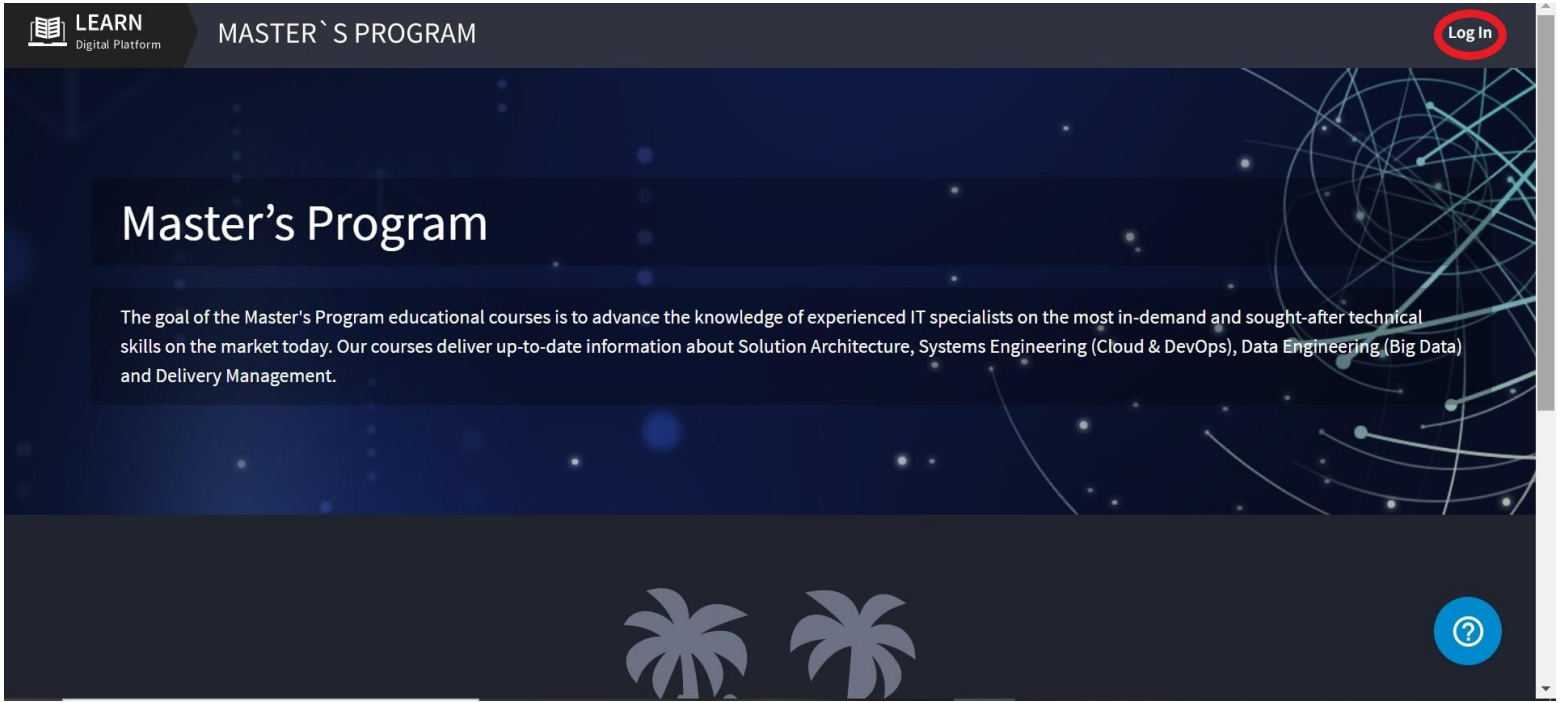
ELEARN DIGITAL PLATFORM

Please visit the page:

https://university.epam.com/

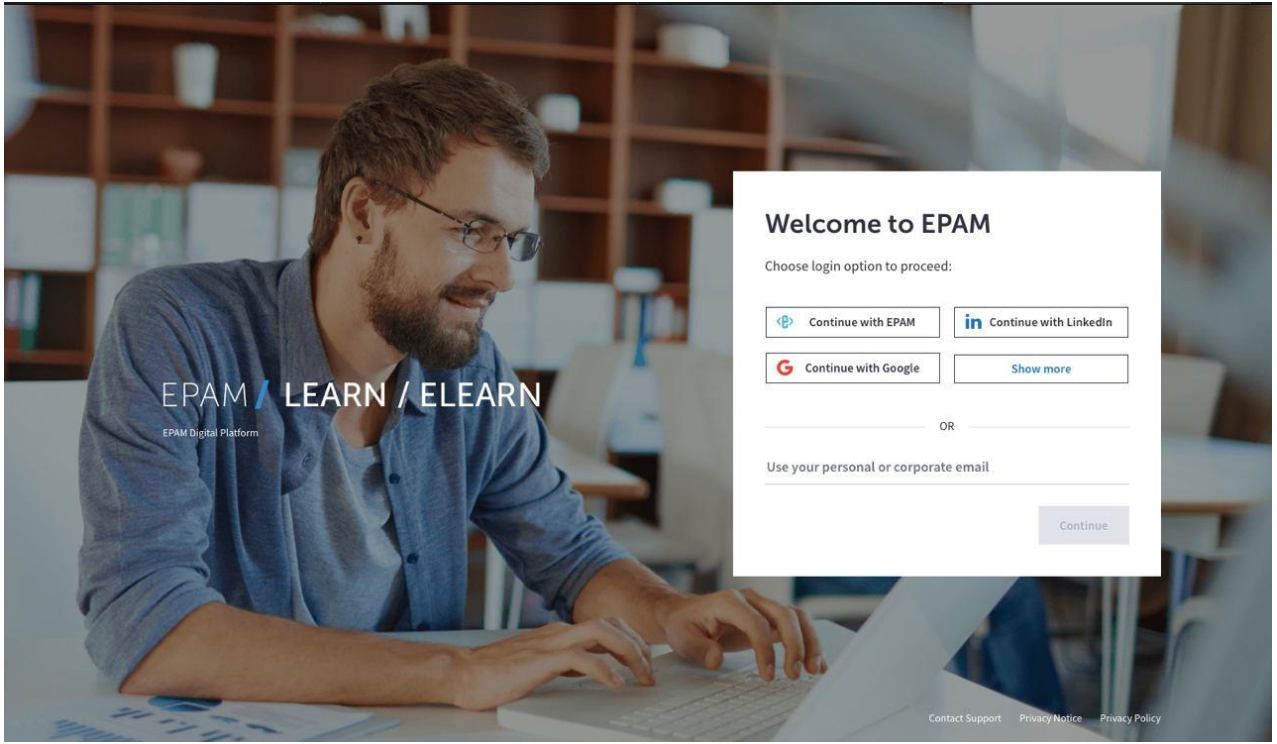
LOG IN TO THE LEARNING PLATFORM STEP 1

CLICK LOG IN ICON IN THE UPPER RIGHT CORNER



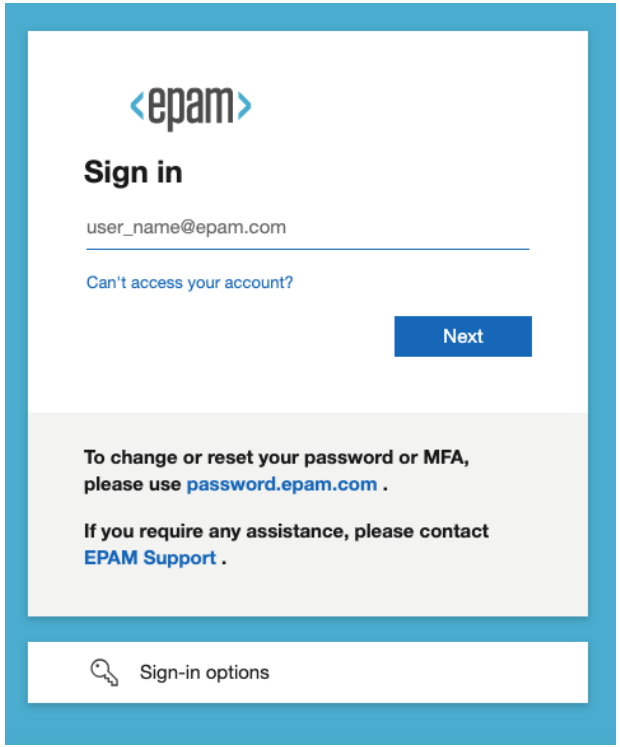
LOGIN TO THE LEARNING PLATFORM STEP 2

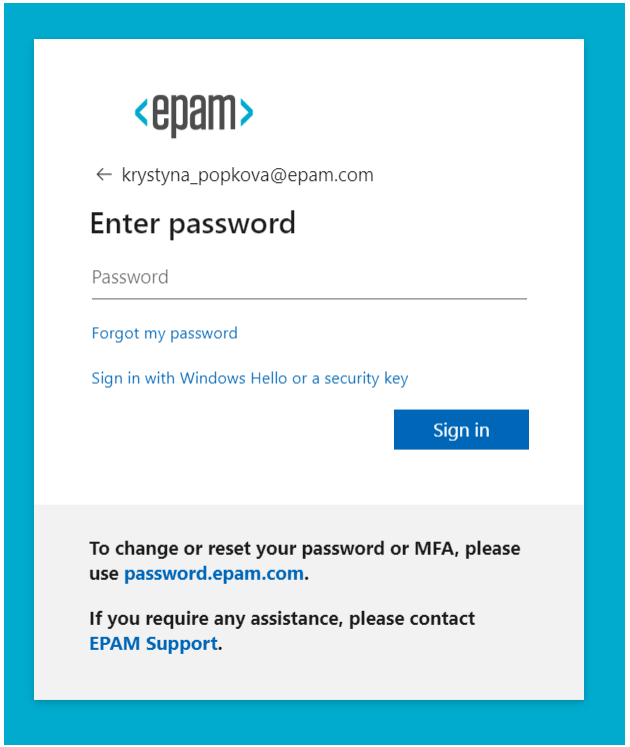
USE YOUR EPAM EMAIL



LOGIN TO THE LEARNING PLATFORM STEP 3

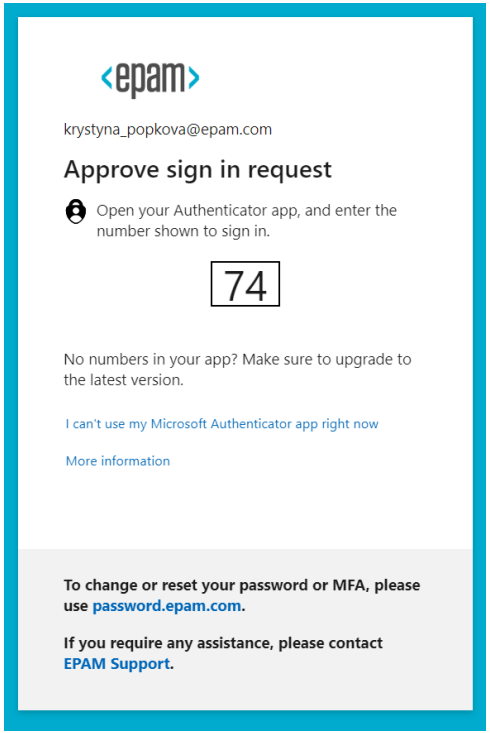
ENTER YOUR EMAIL AND PASSWORD. CLICK SIGN IN





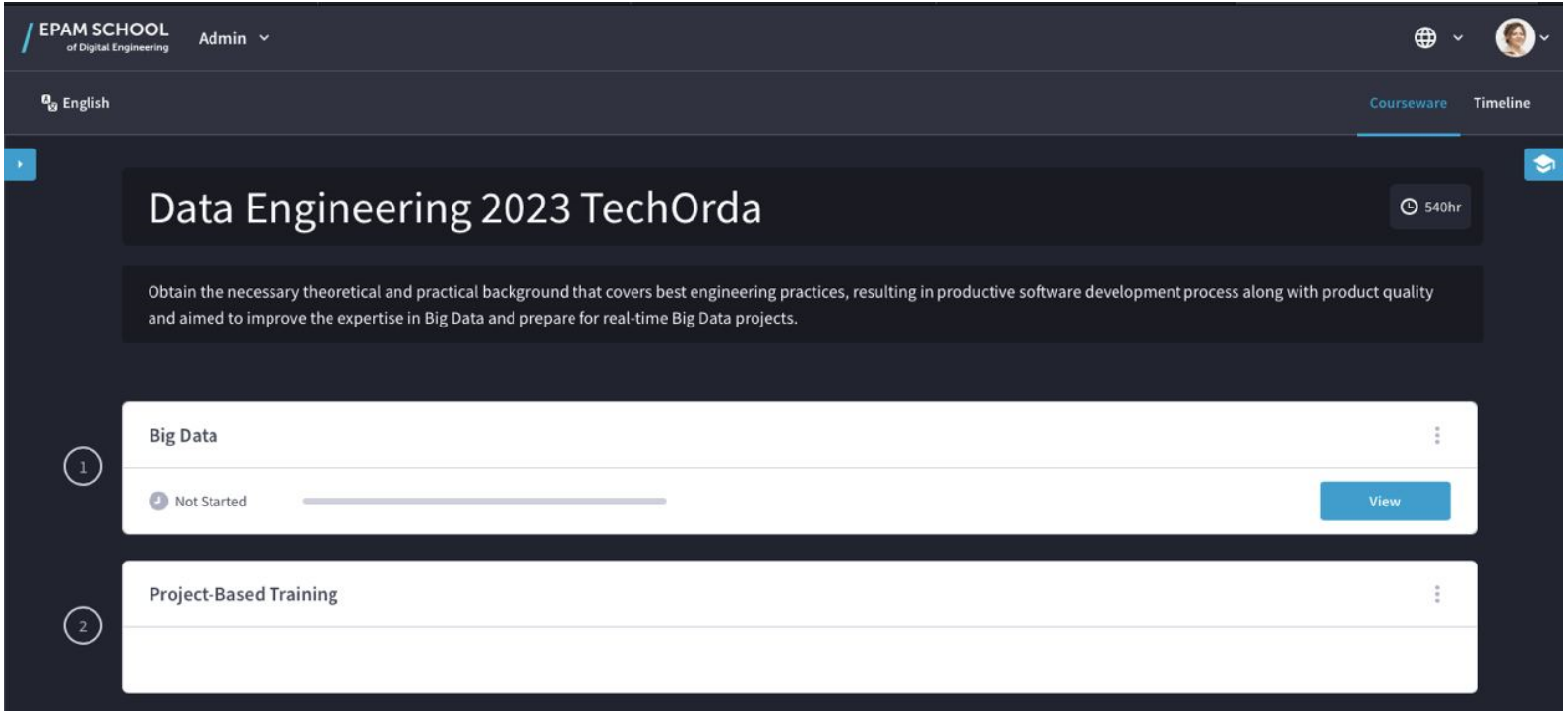
LOGIN TO THE LEARNING PLATFORM STEP 4

OPEN AUTHENTICATOR APP ON YOUR PHONE AND ENTER THE NUMBER.

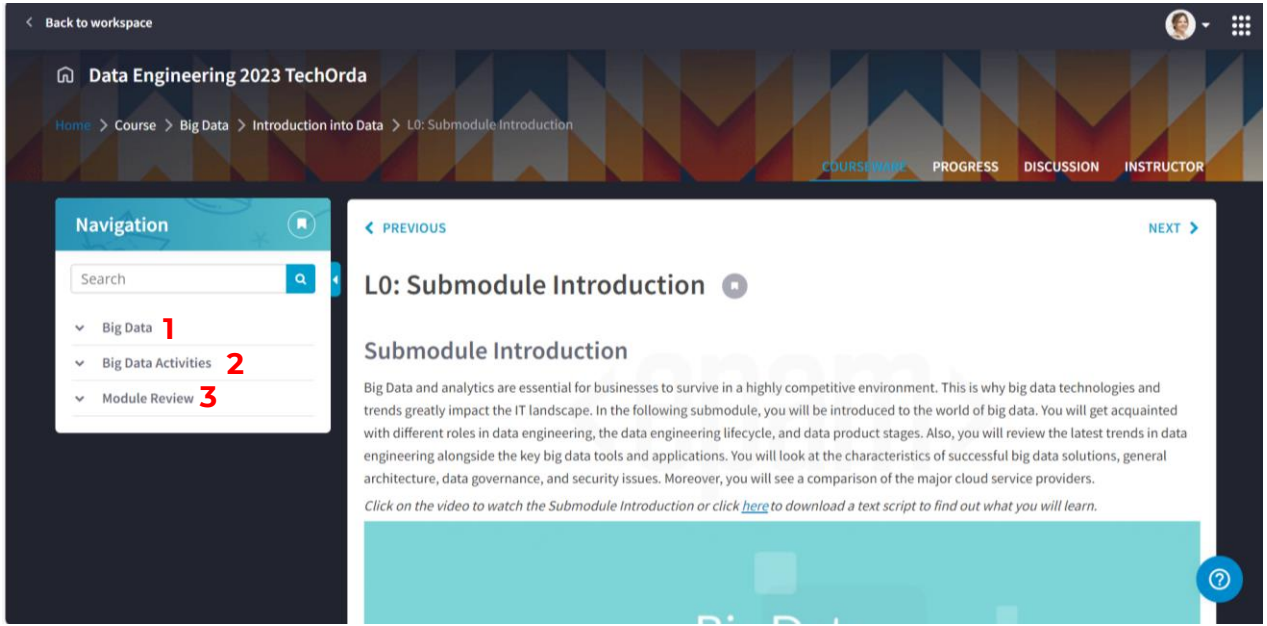


PLATFORM

LEARNING PATH



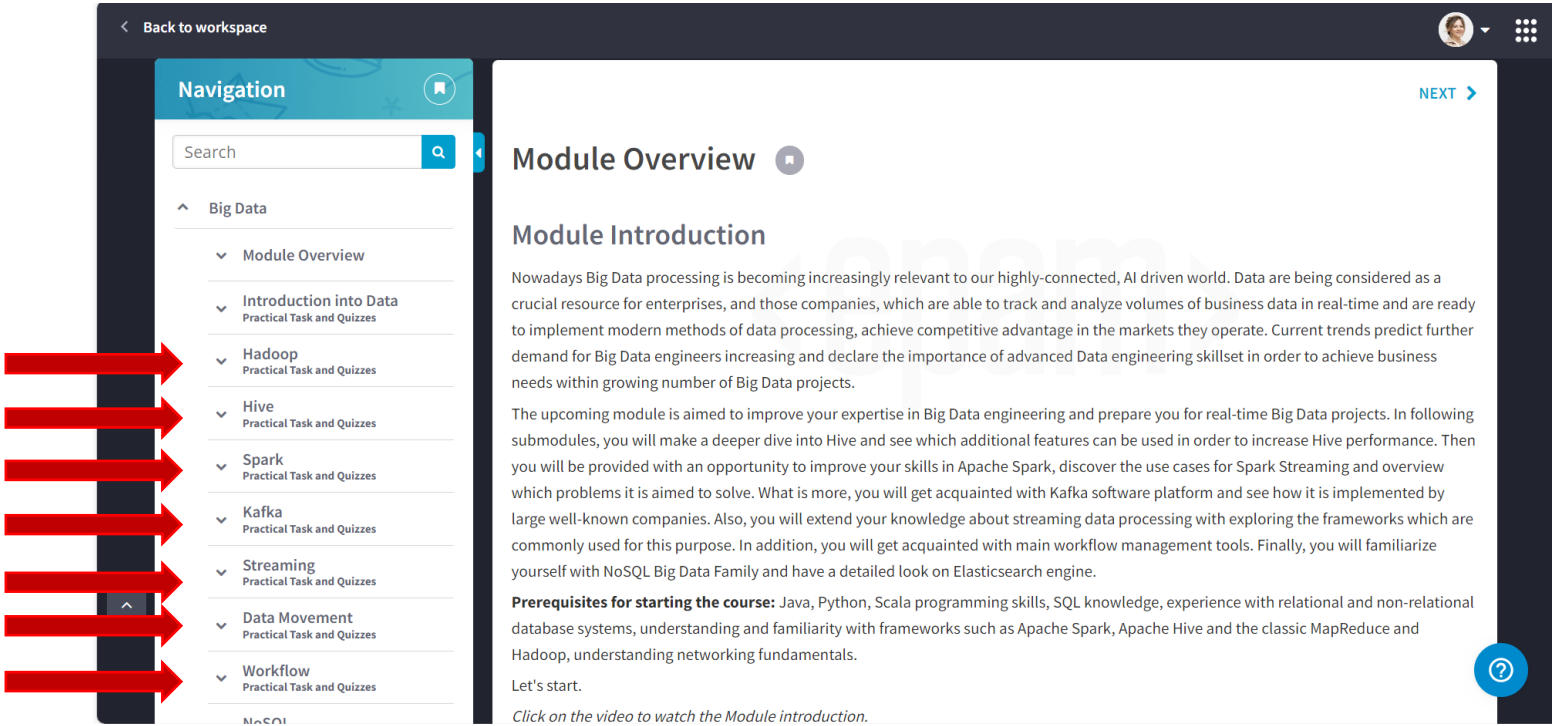
COURSE NAVIGATION



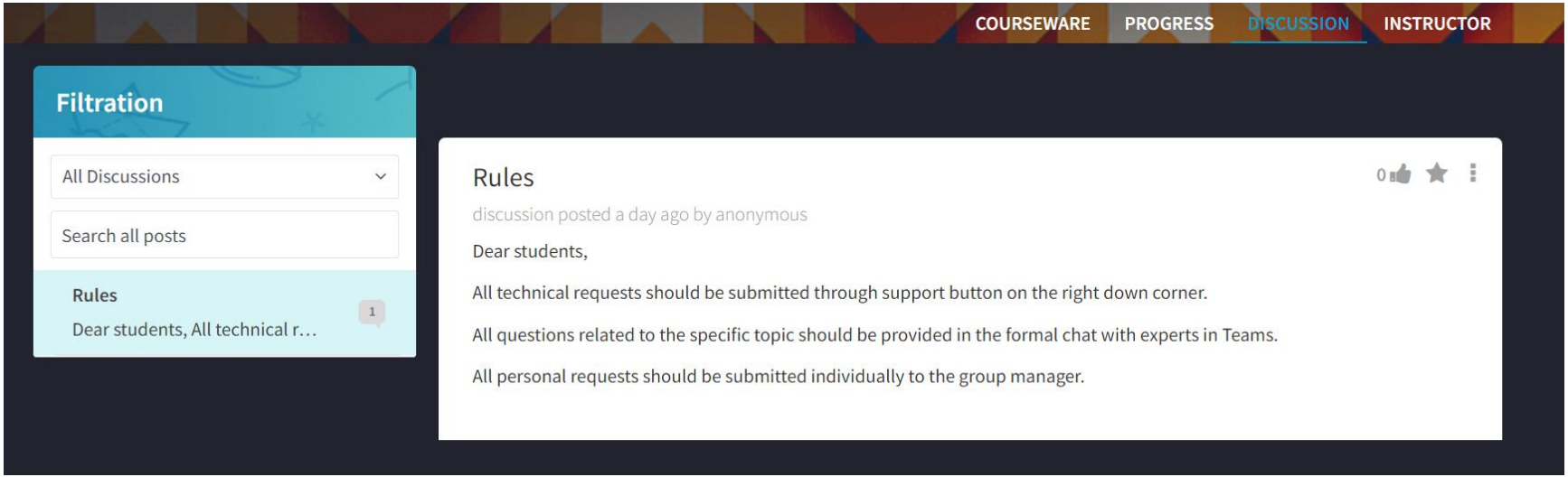
1. ALL THE EDUCATIONAL MATERIALS
2. THE PRESENTATIONS AND RECORDINGS OF THE SESSIONS
3. FINAL ASSESSMENT

COURSE NAVIGATION

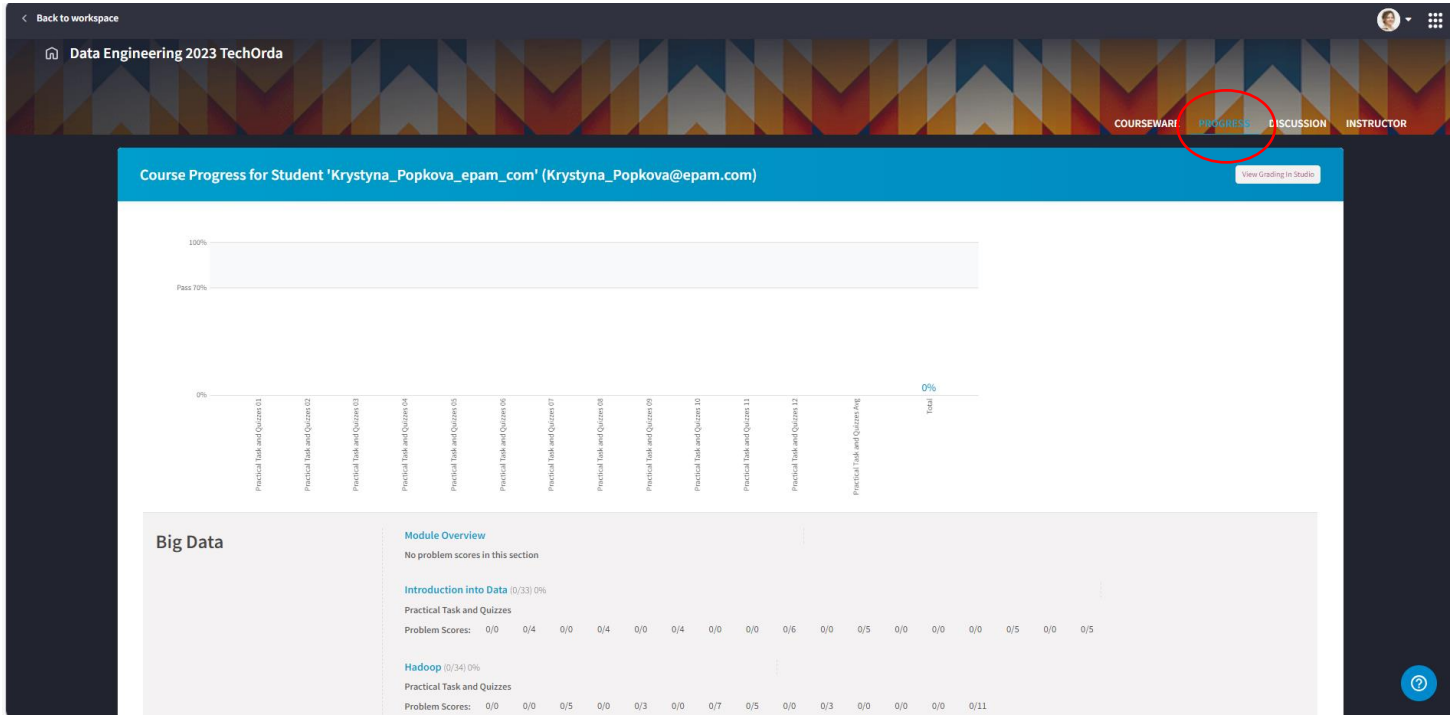
Not available before the START DATE



DISCUSSION FORUM



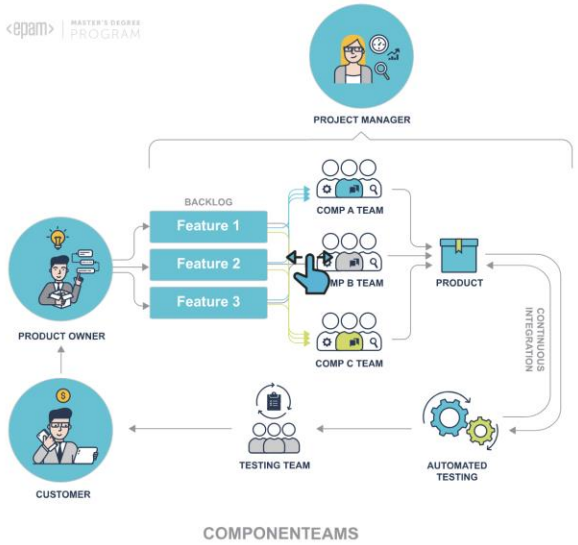
PROGRESS



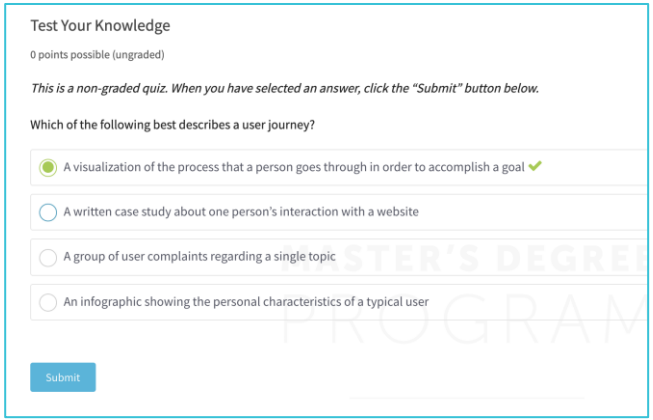
PLATFORM INTERACTIONS

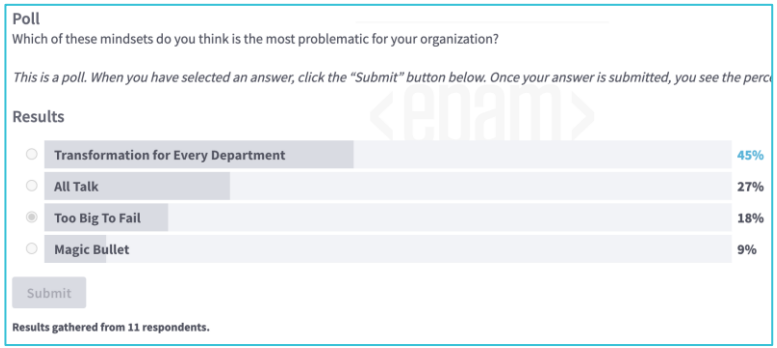
Watch the video to learn about team-oriented adaptable organizations. Alternatively, you may click here to download a text.

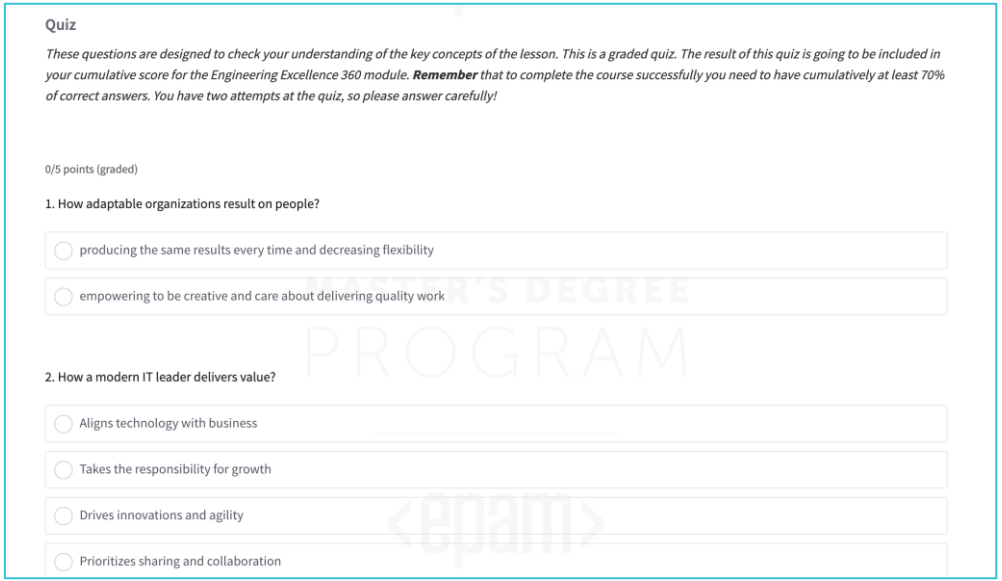
Swipe the illustration to see differences between feature and component teams.



PLATFORM INTERACTIONS







Onboarding Digital Platform

ONBOARDING DIGITAL PLATFORM

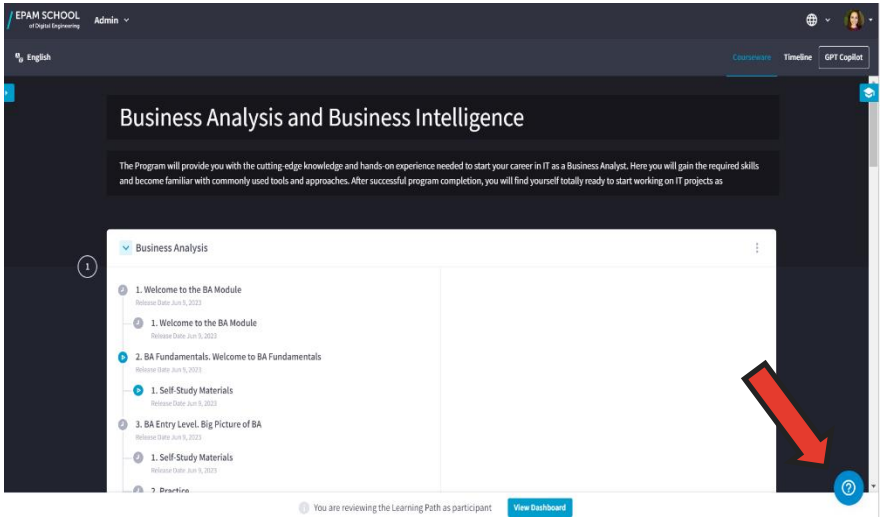
Please visit the page:

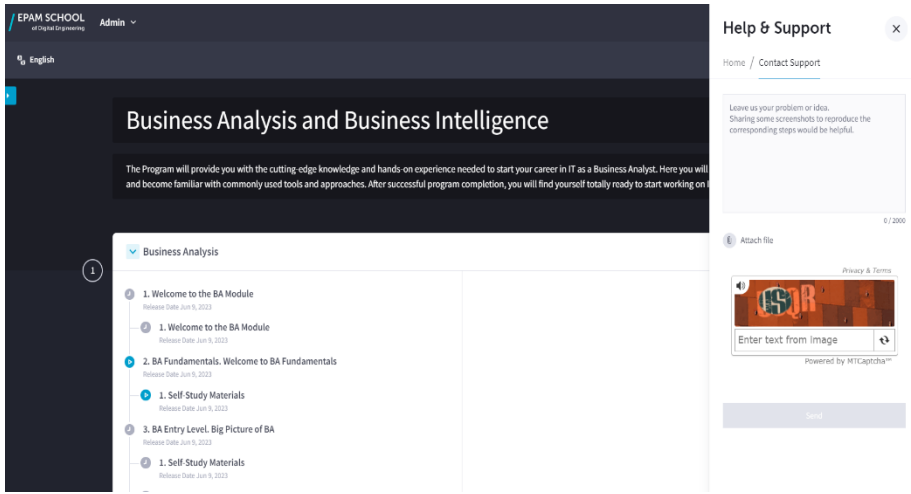
https://onboarding.epam.com/

The modules within the **Onboarding** portal should be done until **November 5th**

Customer Support

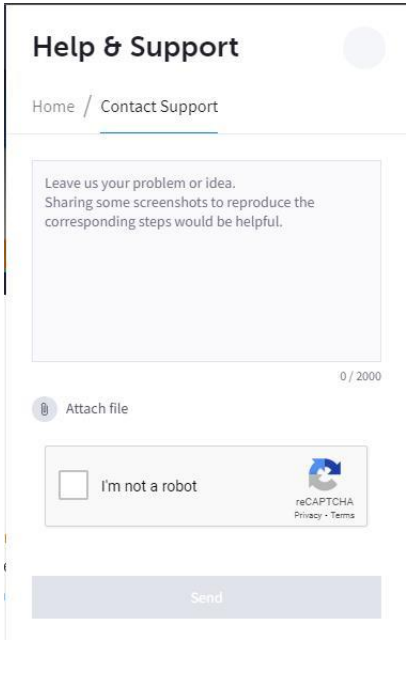
CUSTOMER SUPPORT



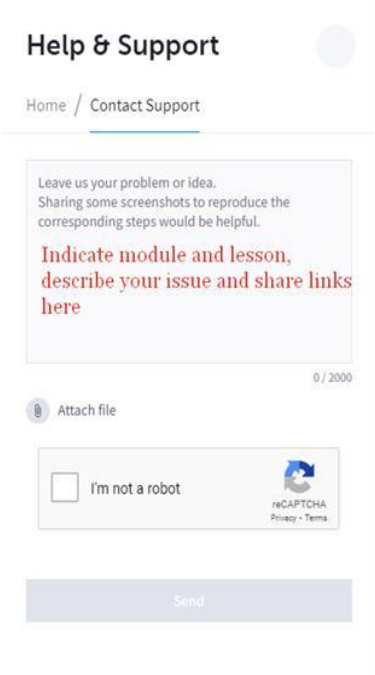


Important! Please keep the request number until it is resolved.

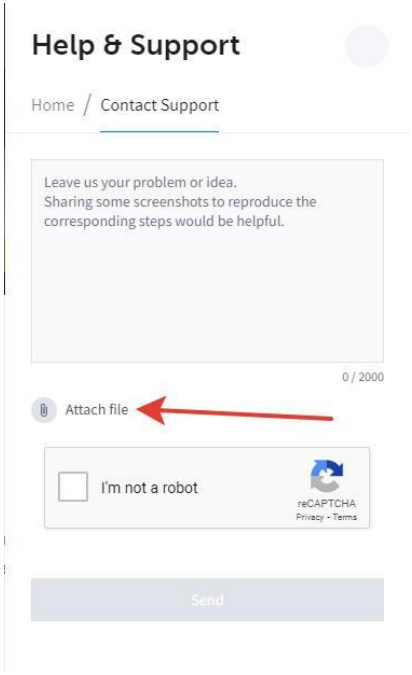
SUPPORT REQUEST



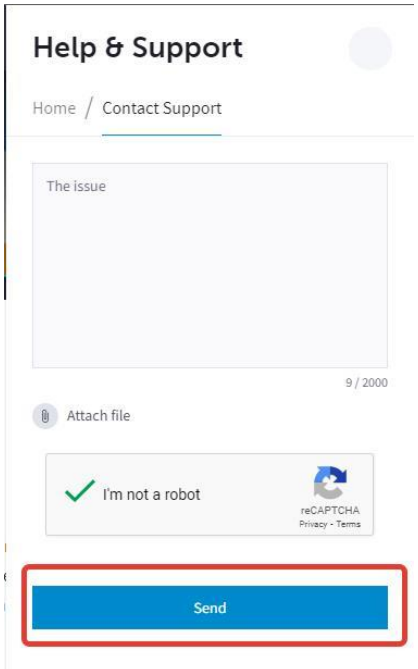
"Help & Support" pop-up opens.



Describe your issue or idea. Share links.

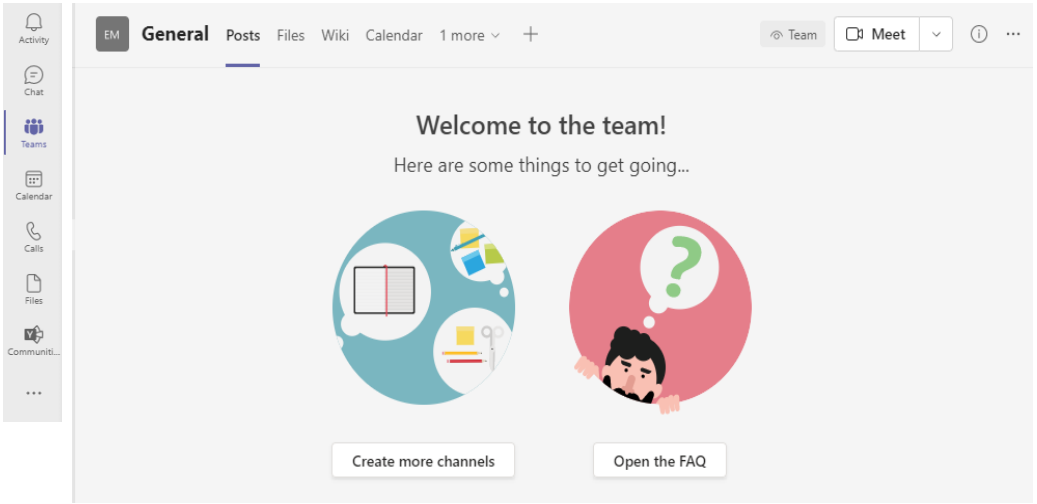


Attach screenshots to replay related steps



Click the "Send" button to submit your request.

COMMUNICATION WITHIN THE PROGRAM



1. Emails from orgmastersprogramkz@epam.com:

- General info
- Invitations to the online sessions
- Certificate of transfer and acceptance of services

2. MS Teams Group:

EPAM Tech Orda Big Data 2023

- Posts
- Reminders
- News on the program
- Communication with trainers
- Feedback requests
- Links to Recordings/ Presentations
- Informal chat
- Online sessions
- Files: Folder shared materials

PROJECT-BASED TRAINING

PROJECT-BASED TRAINING

GOAL

Familiarize participants with the real industry standards of IT projects in the software development enterprise, gain valuable practical experience and apply theoretical knowledge in real life.

COURSE MODEL

In the scope of the Project based training students will participate in the internal and educational EPAM projects as Data engineers.

DURATION

9 weeks

RESULT

As a result, students will be able to gain production experience and add it to their portfolio

PROJECT-BASED TRAINING

- Project work in EPAM for 9 weeks
- Working with quasi-Production data mimicking real EPAM Data Engineering projects
- Workload is spread between synchronous teamwork, mentoring sessions, and self-study

PROJECT-BASED TRAINING PHASES

Phase 1: Introduction (week 1)

Project teams' formation

Project onboarding

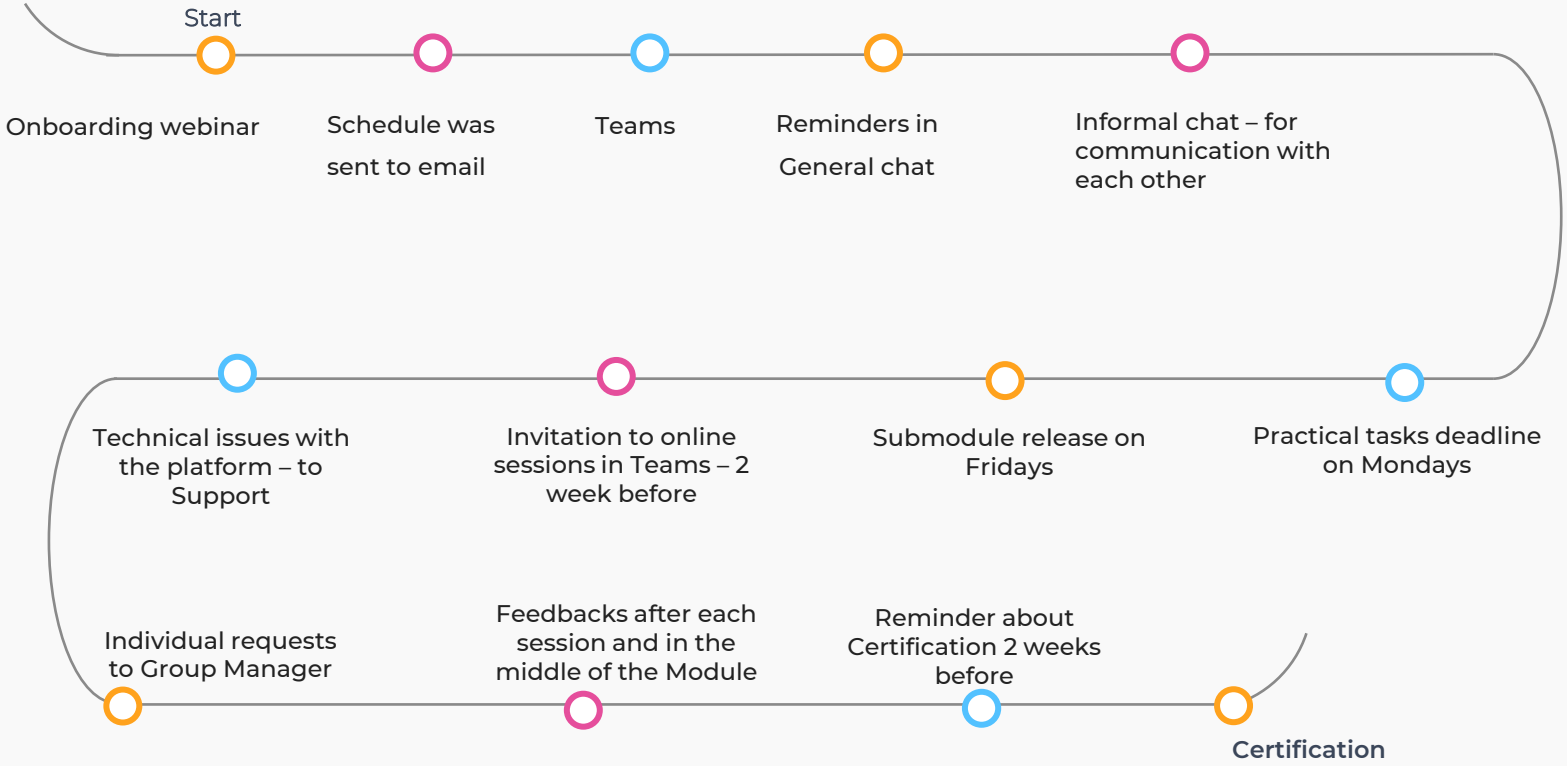
Phase 2: Project work (weeks 2-8)

Team work on projects supporting by EPAM mentors

Phase 3: Assessment (week 9)

Presentation and Code assessment. Final interview mentors. Demo

STUDYING FLOW SUMMARY



Q & A

Thank you!

Let’s get to know each other