



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](mailto: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.

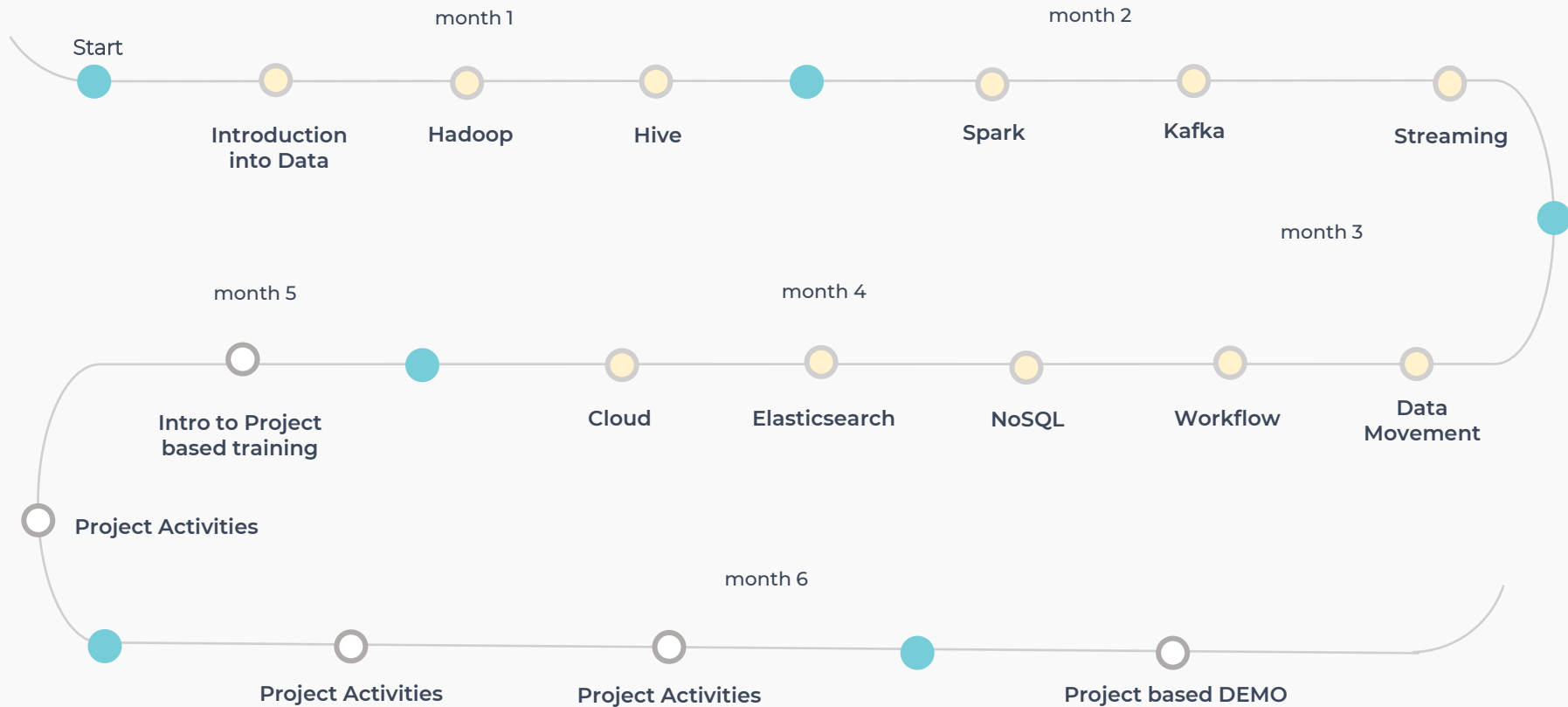




# **Program Curriculum**

# 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	
		21-Nov (Tue) 7 - 9 PM	27-Nov

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

Org Masters Program KZ  
To  
Cc  
Bcc

Reply Reply All Forward ...  
Cp 01.11.2023 12:26



Dear Students,  
On **November 1<sup>st</sup>, 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	17-Jan, 2024	

10. Elasticsearch	2-Feb		
11. Big Data in Clouds	16-Feb		
Final Certification	29-Feb		

ET General Posts **Files** Notes +

EPAM Confidential... Meet

Welcome to the team!

Here are some things to get going...

Add more people
Create more channels
Open the FAQ

24-Jan (We) 6 - 8 PM	29-Jan 5-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	



# 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**



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



WELCOME TO  
THE TEAM!



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).





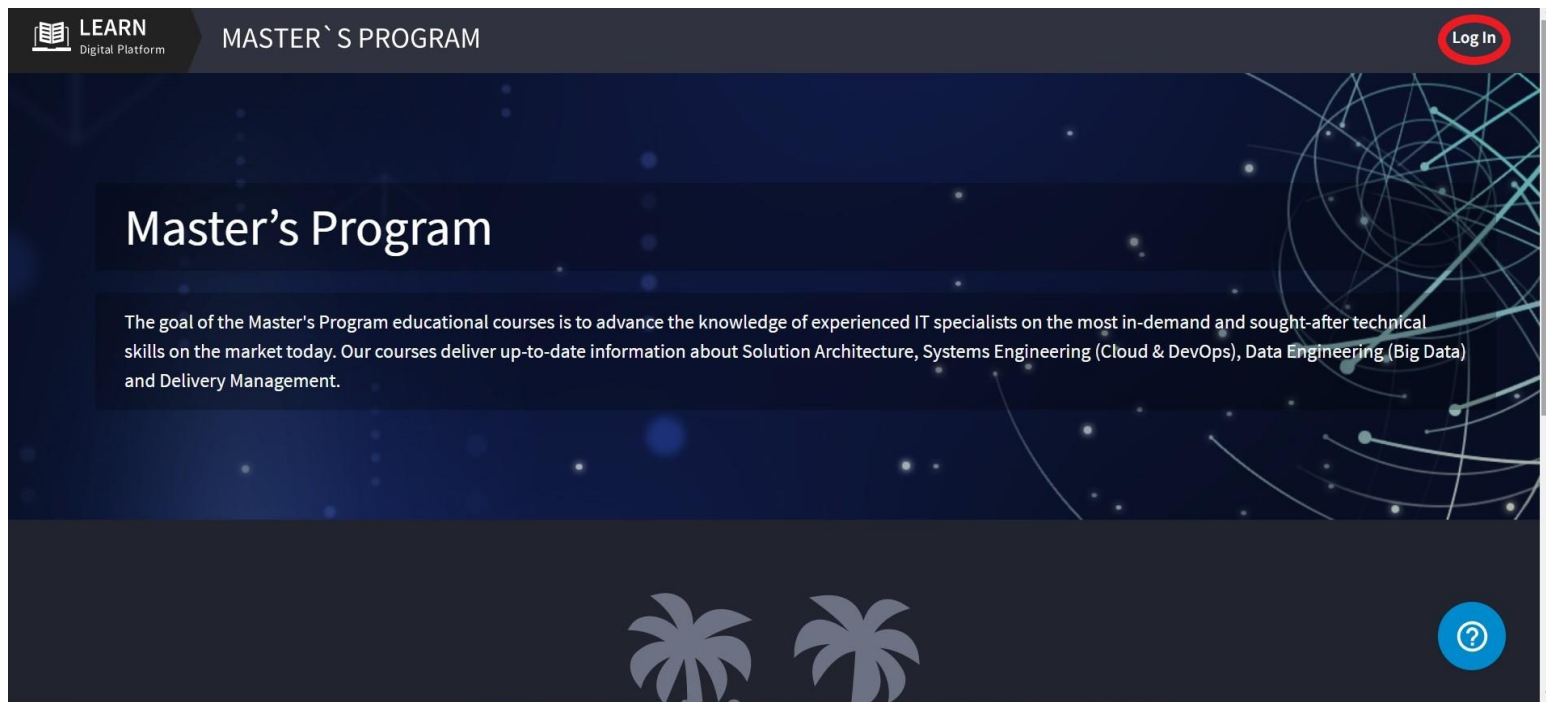
# ELEARN DIGITAL PLATFORM

Please visit the page:

<https://university.epam.com/>

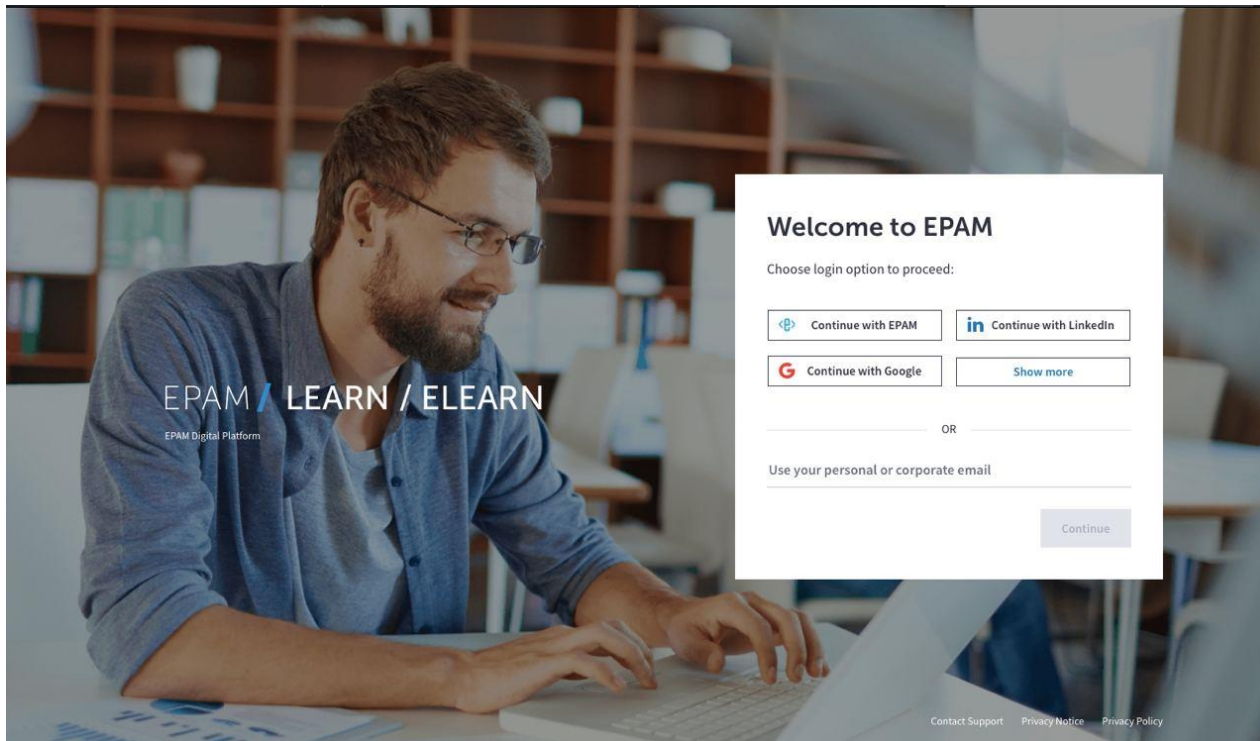
# LOG IN TO THE LEARNING PLATFORM STEP 1

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



## Sign in

user\_name@epam.com


---


[Can't access your account?](#)

[Next](#)

To change or reset your password or MFA, please use [password.epam.com](https://password.epam.com).

If you require any assistance, please contact [EPAM Support](#).

 Sign-in options



← [krystyna\\_popkova@epam.com](#)

## Enter password

Password

---

[Forgot my password](#)

[Sign in with Windows Hello or a security key](#)

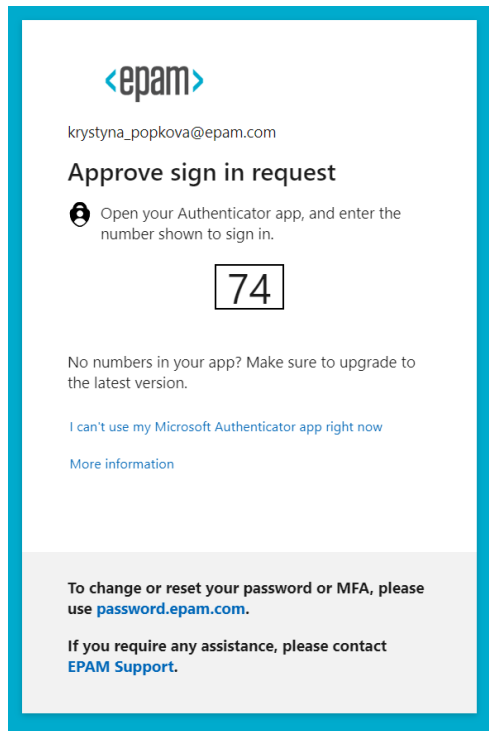
[Sign in](#)

To change or reset your password or MFA, please use [password.epam.com](https://password.epam.com).

If you require any assistance, please contact [EPAM Support](#).

# LOGIN TO THE LEARNING PLATFORM STEP 4

OPEN AUTHENTICATOR APP ON YOUR PHONE AND ENTER THE NUMBER.



The screenshot shows a mobile app interface for EPAM. At the top is the EPAM logo. Below it is the email address 'krystyna\_popkova@epam.com'. The main heading is 'Approve sign in request'. Below this is a list item with a circular icon containing the number 8, followed by the text 'Open your Authenticator app, and enter the number shown to sign in.' In the center, the number '74' is displayed inside a square box. Below the box, there is a message: 'No numbers in your app? Make sure to upgrade to the latest version.' followed by a link 'I can't use my Microsoft Authenticator app right now' and another link 'More information'. At the bottom, there is a grey footer area with the text 'To change or reset your password or MFA, please use [password.epam.com](https://password.epam.com).' and 'If you require any assistance, please contact [EPAM Support](#)'.

<epam>

krystyna\_popkova@epam.com

**Approve sign in request**

8 Open your Authenticator app, and enter the number shown to sign in.

74

No numbers in your app? Make sure to upgrade to the latest version.

[I can't use my Microsoft Authenticator app right now](#)

[More information](#)

To change or reset your password or MFA, please use [password.epam.com](https://password.epam.com).

If you require any assistance, please contact [EPAM Support](#).

# PLATFORM

The screenshot displays the EPAM School of Digital Engineering platform. The top navigation bar is dark blue and contains the EPAM School logo, an 'Admin' dropdown menu, a globe icon, and a user profile icon. Below this, a secondary navigation bar features a language selector set to 'English' and a menu with links to 'Courseware', 'Schedule', 'Dashboard', 'Settings', 'Contributors', and 'Attributes'. The 'Courseware' link is highlighted with a blue underline. The main content area has a dark background and features a course card for 'Data Engineering 2023 TechOrda'. The card includes a blue arrow icon on the left, the course title in large white text, a clock icon indicating a duration of 540 hours, and a descriptive paragraph about the course's focus on engineering practices and Big Data.

EPAM SCHOOL  
of Digital Engineering Admin ▾

English

Courseware Schedule Dashboard Settings Contributors Attributes

▶

## Data Engineering 2023 TechOrda

🕒 540hr

Obtain the necessary theoretical and practical background that covers best engineering practices, resulting in productive software development process along with product quality and aimed to improve the expertise in Big Data and prepare for real-time Big Data projects.

# LEARNING PATH

The screenshot displays the EPAM SCHOOL of Digital Engineering interface. At the top, the header includes the logo, 'Admin' dropdown, a globe icon, and a user profile icon. Below the header, there's a language selector set to 'English' and tabs for 'Courseware' and 'Timeline'. The main content area features a course card for 'Data Engineering 2023 TechOrda' with a duration of 540hr. Below the course card, a description states: 'Obtain the necessary theoretical and practical background that covers best engineering practices, resulting in productive software development process along with product quality and aimed to improve the expertise in Big Data and prepare for real-time Big Data projects.' The learning path is shown as a list of two items: 'Big Data' (marked 'Not Started' with a progress bar) and 'Project-Based Training'. A 'View' button is present next to the 'Big Data' item.

EPAM SCHOOL  
of Digital Engineering

Admin

English

Courseware Timeline

## Data Engineering 2023 TechOrda

540hr

Obtain the necessary theoretical and practical background that covers best engineering practices, resulting in productive software development process along with product quality and aimed to improve the expertise in Big Data and prepare for real-time Big Data projects.

- 1 Big Data  
Not Started View
- 2 Project-Based Training



# COURSE NAVIGATION

< Back to workspace

Data Engineering 2023 TechOrda

Home > Course > Big Data > Introduction into Data > L0: Submodule Introduction

COURSE NAME PROGRESS DISCUSSION INSTRUCTOR

Navigation

Search

- Big Data **1**
- Big Data Activities **2**
- Module Review **3**

< PREVIOUS

NEXT >

## L0: Submodule Introduction

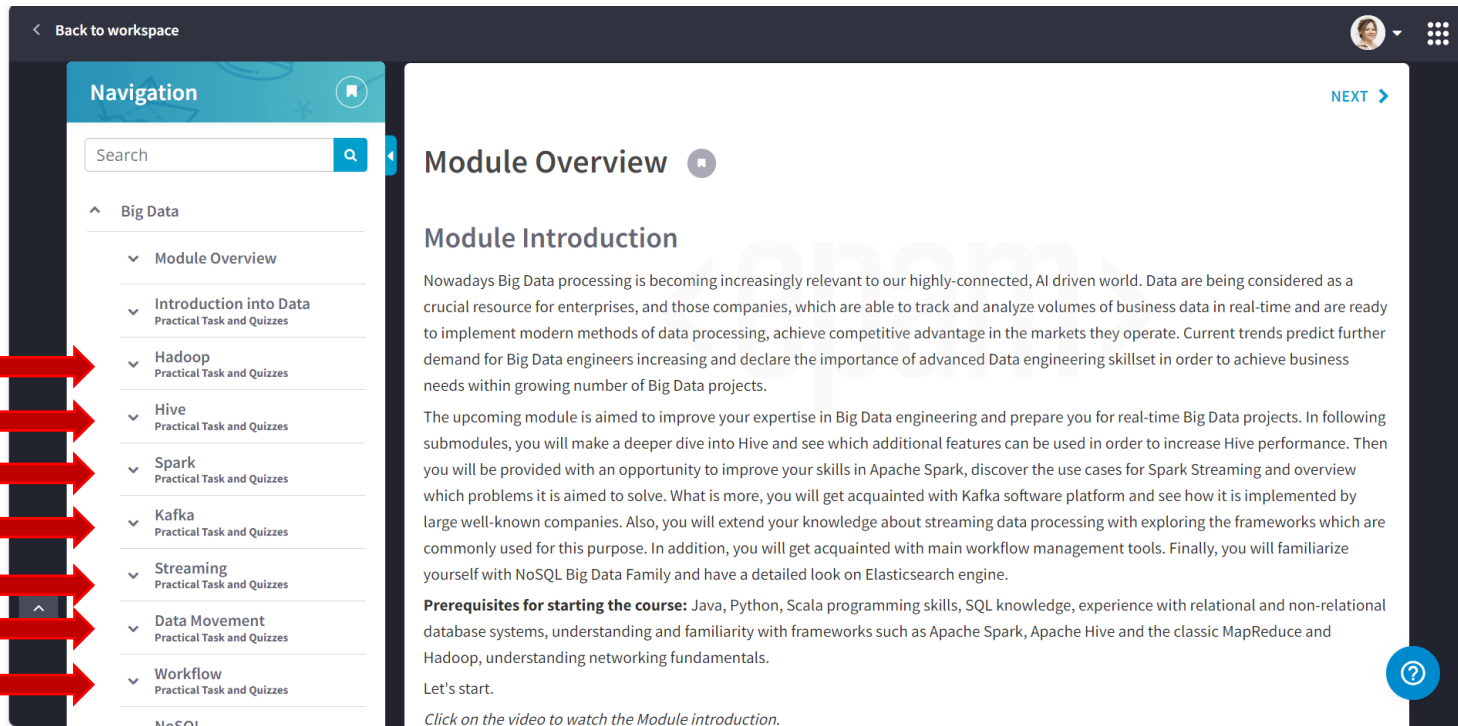
### Submodule Introduction

Big Data and analytics are essential for businesses to survive in a highly competitive environment. This is why big data technologies and trends greatly impact the IT landscape. In the following submodule, you will be introduced to the world of big data. You will get acquainted with different roles in data engineering, the data engineering lifecycle, and data product stages. Also, you will review the latest trends in data engineering alongside the key big data tools and applications. You will look at the characteristics of successful big data solutions, general architecture, data governance, and security issues. Moreover, you will see a comparison of the major cloud service providers.

Click on the video to watch the Submodule Introduction or click [here](#) to download a text script to find out what you will learn.

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

# COURSE NAVIGATION



The screenshot shows a course navigation interface. On the left is a 'Navigation' sidebar with a search bar and a list of items under the 'Big Data' category. Red arrows point to each item in this list. The main content area on the right shows the 'Module Overview' page, which includes a 'Module Introduction' section with text about Big Data processing and a 'Prerequisites for starting the course' section. A 'NEXT' button is visible in the top right of the main content area.

Back to workspace

Navigation

Search

Big Data

- Module Overview
- Introduction into Data  
Practical Task and Quizzes
- Hadoop  
Practical Task and Quizzes
- Hive  
Practical Task and Quizzes
- Spark  
Practical Task and Quizzes
- Kafka  
Practical Task and Quizzes
- Streaming  
Practical Task and Quizzes
- Data Movement  
Practical Task and Quizzes
- Workflow  
Practical Task and Quizzes

Module Overview

## Module Introduction

Nowadays Big Data processing is becoming increasingly relevant to our highly-connected, AI driven world. Data are being considered as a crucial resource for enterprises, and those companies, which are able to track and analyze volumes of business data in real-time and are ready to implement modern methods of data processing, achieve competitive advantage in the markets they operate. Current trends predict further demand for Big Data engineers increasing and declare the importance of advanced Data engineering skillset in order to achieve business needs within growing number of Big Data projects.

The upcoming module is aimed to improve your expertise in Big Data engineering and prepare you for real-time Big Data projects. In following submodules, you will make a deeper dive into Hive and see which additional features can be used in order to increase Hive performance. Then you will be provided with an opportunity to improve your skills in Apache Spark, discover the use cases for Spark Streaming and overview which problems it is aimed to solve. What is more, you will get acquainted with Kafka software platform and see how it is implemented by large well-known companies. Also, you will extend your knowledge about streaming data processing with exploring the frameworks which are commonly used for this purpose. In addition, you will get acquainted with main workflow management tools. Finally, you will familiarize yourself with NoSQL Big Data Family and have a detailed look on Elasticsearch engine.

**Prerequisites for starting the course:** Java, Python, Scala programming skills, SQL knowledge, experience with relational and non-relational database systems, understanding and familiarity with frameworks such as Apache Spark, Apache Hive and the classic MapReduce and Hadoop, understanding networking fundamentals.

Let's start.

Click on the video to watch the Module introduction.

NEXT

Not available  
before the  
START DATE

# DISCUSSION FORUM

[COURSEWARE](#) [PROGRESS](#) [DISCUSSION](#) [INSTRUCTOR](#)

## Filtration

All Discussions ▾

Search all posts

**Rules** 1

Dear students, All technical r...

## Rules

0

discussion posted a day ago by anonymous

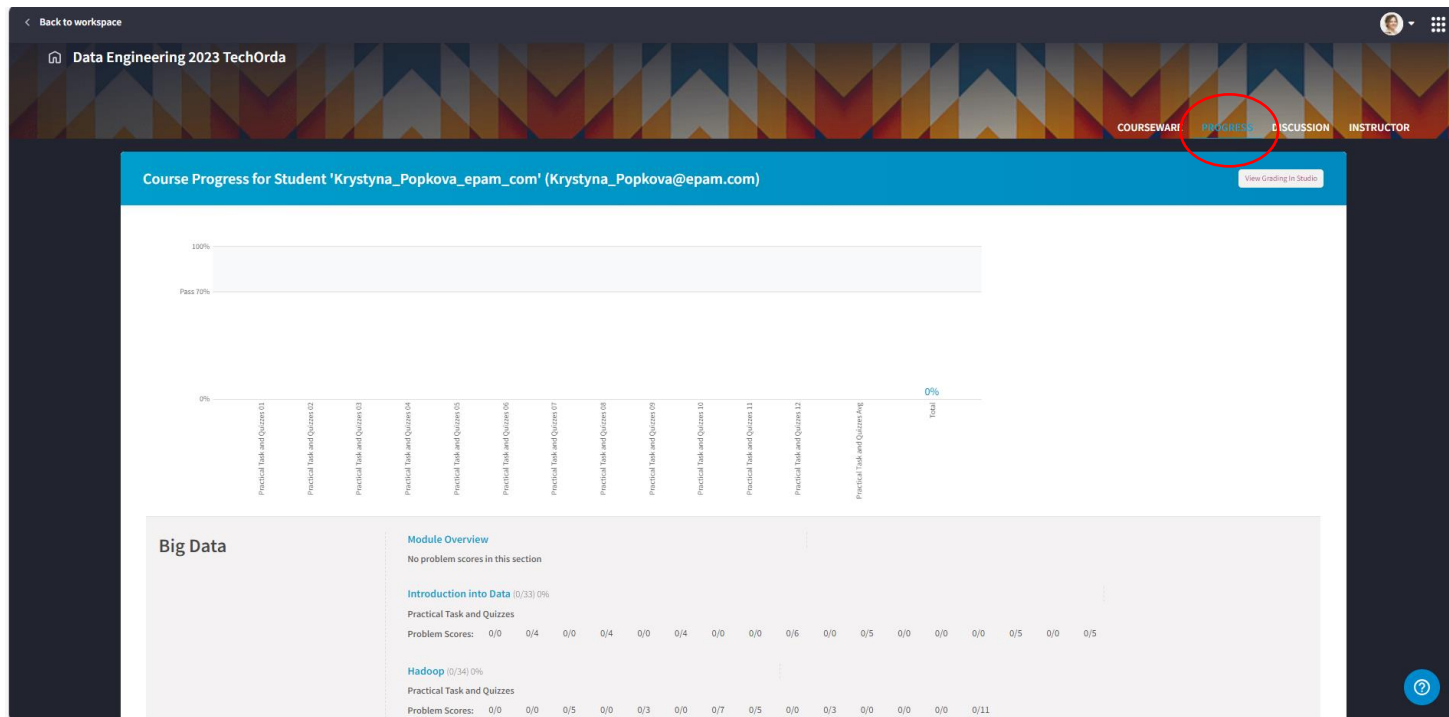
Dear students,

All technical requests should be submitted through support button on the right down corner.

All questions related to the specific topic should be provided in the formal chat with experts in Teams.

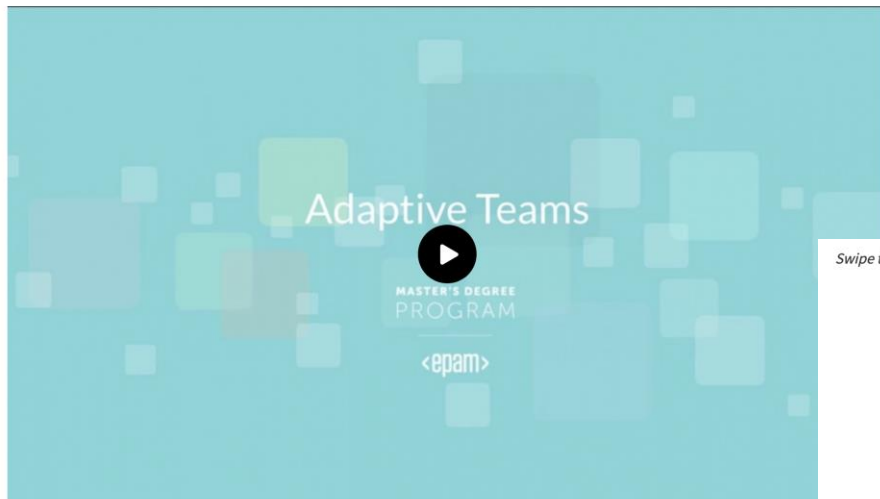
All personal requests should be submitted individually to the group manager.

# PROGRESS

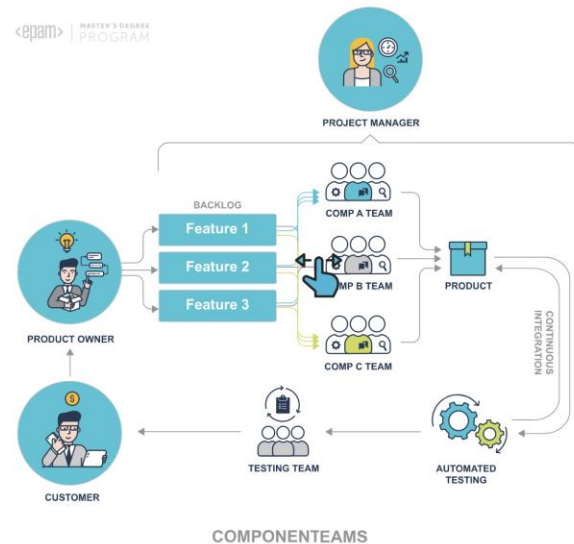


# PLATFORM INTERACTIONS

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

## Test Your Knowledge

0 points possible (ungraded)

*This is a non-graded quiz. When you have selected an answer, click the "Submit" button below.*

Which of the following best describes a user journey?

- ☒ A visualization of the process that a person goes through in order to accomplish a goal ✓
- ☐ A written case study about one person's interaction with a website
- ☐ A group of user complaints regarding a single topic
- ☐ An infographic showing the personal characteristics of a typical user

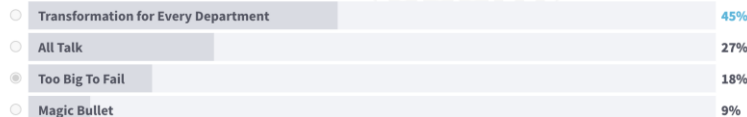
Submit

## Poll

Which of these mindsets do you think is the most problematic for your organization?

*This is a poll. When you have selected an answer, click the "Submit" button below. Once your answer is submitted, you see the percentage of respondents who chose the same answer.*

### Results



Submit

Results gathered from 11 respondents.

## Quiz

*These questions are designed to check your understanding of the key concepts of the lesson. This is a graded quiz. The result of this quiz is going to be included in your cumulative score for the Engineering Excellence 360 module. **Remember** that to complete the course successfully you need to have cumulatively at least 70% of correct answers. You have two attempts at the quiz, so please answer carefully!*

0/5 points (graded)

1. How adaptable organizations result on people?

- ☐ producing the same results every time and decreasing flexibility
- ☐ empowering to be creative and care about delivering quality work

2. How a modern IT leader delivers value?

- ☐ Aligns technology with business
- ☐ Takes the responsibility for growth
- ☐ Drives innovations and agility
- ☐ Prioritizes sharing and collaboration

# 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 5<sup>th</sup>**

# Customer Support

# CUSTOMER SUPPORT

EPAM SCHOOL  
of Digital Engineering

Admin

English

Courses Timeline GPT Copilot

## Business Analysis and Business Intelligence

The Program will provide you with the cutting edge knowledge and hands-on experience needed to start your career in IT as a Business Analyst. Here you will gain the required skills and become familiar with commonly used tools and approaches. After successful program completion, you will find yourself totally ready to start working on IT projects as

### Business Analysis

1. Welcome to the BA Module  
Release Date: Jun 9, 2023
1. Welcome to the BA Module  
Release Date: Jun 9, 2023
2. BA Fundamentals. Welcome to BA Fundamentals  
Release Date: Jun 9, 2023
1. Self-Study Materials  
Release Date: Jun 9, 2023
3. BA Entry Level. Big Picture of BA  
Release Date: Jun 9, 2023
1. Self-Study Materials  
Release Date: Jun 9, 2023
2. BA Entry Level. Big Picture of BA  
Release Date: Jun 9, 2023
1. Self-Study Materials  
Release Date: Jun 9, 2023

You are reviewing the Learning Path as participant [View Dashboard](#)

EPAM SCHOOL  
of Digital Engineering

Admin

English

## Business Analysis and Business Intelligence

The Program will provide you with the cutting edge knowledge and hands-on experience needed to start your career in IT as a Business Analyst. Here you will gain the required skills and become familiar with commonly used tools and approaches. After successful program completion, you will find yourself totally ready to start working on IT projects as

### Business Analysis

1. Welcome to the BA Module  
Release Date: Jun 9, 2023
1. Welcome to the BA Module  
Release Date: Jun 9, 2023
2. BA Fundamentals. Welcome to BA Fundamentals  
Release Date: Jun 9, 2023
1. Self-Study Materials  
Release Date: Jun 9, 2023
3. BA Entry Level. Big Picture of BA  
Release Date: Jun 9, 2023
1. Self-Study Materials  
Release Date: Jun 9, 2023
2. BA Entry Level. Big Picture of BA  
Release Date: Jun 9, 2023
1. Self-Study Materials  
Release Date: Jun 9, 2023

#### Help & Support

Home / [Contact Support](#)

Leave us your problem or idea. Sharing some screenshots to reproduce the corresponding steps would be helpful.

6 / 2000

Attach file

Privacy & Terms

Enter text from image

Powered by HTCaptha™

Send

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

# SUPPORT REQUEST


**Help & Support**

Home / [Contact Support](#)

Leave us your problem or idea.  
Sharing some screenshots to reproduce the corresponding steps would be helpful.

0 / 2000

Attach file

☐ I'm not a robot   
reCAPTCHA  
Privacy - Terms

Send

"Help & Support" pop-up opens.

**Help & Support**


Home / [Contact Support](#)

Leave us your problem or idea.  
Sharing some screenshots to reproduce the corresponding steps would be helpful.

Indicate module and lesson,  
describe your issue and share links  
here

0 / 2000

Attach file

☐ I'm not a robot   
reCAPTCHA  
Privacy - Terms

Send

Describe your issue or idea.  
Share links.


**Help & Support**

Home / [Contact Support](#)

Leave us your problem or idea.  
Sharing some screenshots to reproduce the corresponding steps would be helpful.

0 / 2000

Attach file

☐ I'm not a robot   
reCAPTCHA  
Privacy - Terms

Send

Attach screenshots to replay  
related steps


**Help & Support**

Home / [Contact Support](#)

The issue

9 / 2000

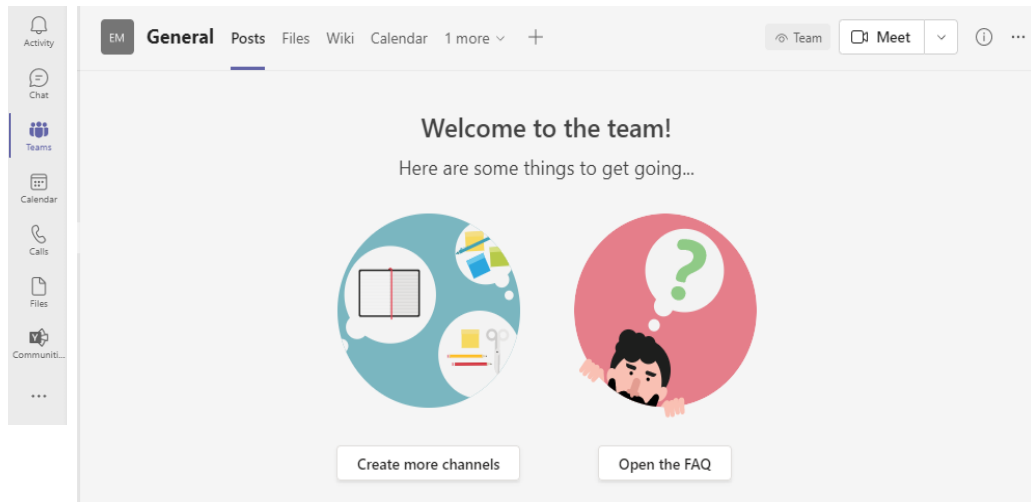
Attach file

☒ I'm not a robot   
reCAPTCHA  
Privacy - Terms

Send

Click the "Send" button to submit  
your request.

# COMMUNICATION WITHIN THE PROGRAM



## 1. Emails from [orgmastersprogramkz@epam.com](mailto: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

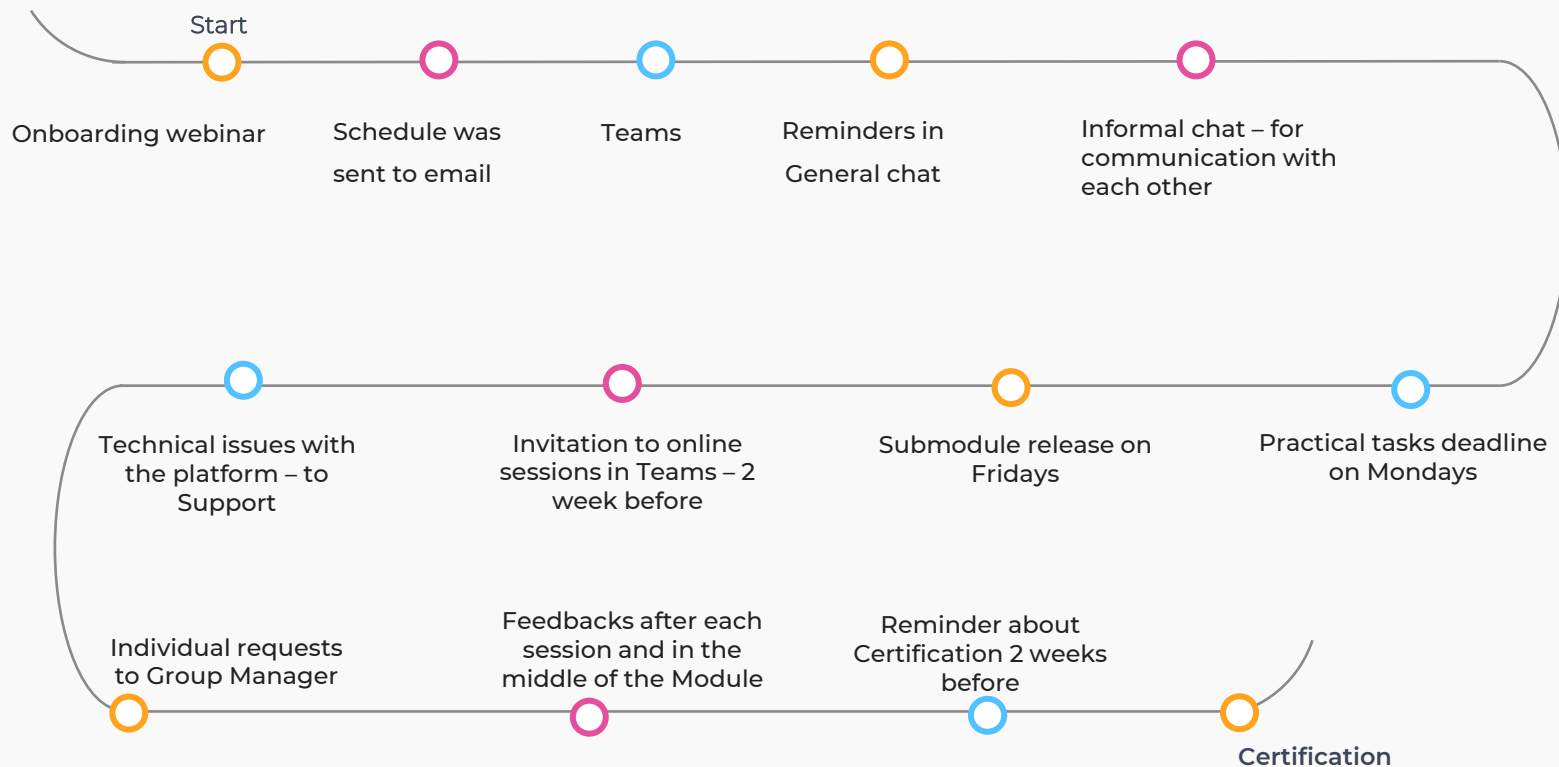
## Phase 2: Project work (weeks 2-8)

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**