

<b>Course intended for when</b>	Comparative analysis of data processing tools
<b># of credits</b>	3 course, Data Science and Business Analytics programme
<b>Total # of hours</b>	3-4 module (march-june 2022)
<b>incl. # classroom hours</b>	5
<b>incl. # independent study hours</b>	190
	30
	160
<b>Teachers</b>	<p>* Elena Tsybak - Head of Division in Glowbyte. {Over 10 years experience with Data Warehousing and Data Integration projects in Banking. Last 3 years working on Data Integration projects with Otkritie Bank.}</p> <p>* Alexander Borodin, Glowbyte</p> <p>* Alexander Sergenko, Neoflex</p> <p>* Oleg Myagkov, Neoflex</p>
<b>Annotation</b>	<p>Over the past few years, the amount of data that humanity generates has been growing very rapidly. It's important that not only the volume of data is increasing, but also its variety. Usage of all possible data is critical to the competitiveness of enterprises in the digital transformation era. This led to the creation of new technologies and approaches to collecting, storing, processing, and analyzing data, such as Big Data, Hadoop, MPP, and NoSQL databases, and others.</p> <p>The purpose of this course is to review the most modern tools and methods and how they are applied in one of the top 10 major banks in Russia - Otkritie Bank.</p> <p>We will focus on the most popular architecture patterns and frameworks for batch and real-time stream processing, as well as best practices for automating the data processing lifecycle (DataOps). Modern data processing tasks require the use of ML, which cannot be implemented without the specific tools and processes (MLOps). During our course, we will learn what data is being used in real business processes in Otkritie Bank and which tools and techniques are required to obtain, transform and analyze this data.</p> <p>In our practical sessions, we will work on building an analytical CRM platform for Retail Banking and design data models for real business cases.</p>
<b>Audience coverage</b>	non inter-campuse course
<b>Format</b>	non blended, non online course. Course is offline
<b>Maximum audience</b>	30 students
<b>Learning Objectives</b>	<p>- Familiarity with the most modern data processing, such as Hadoop, Spark, Hive, Impala, Flink, Akka, taking into account the specifics of their field of application</p> <p>- Understanding DataOps and MLOps tools and processes (MLFlow, Kubeflow)</p> <p>- Understanding real data flows in Banking and tools and techniques that are being used to process this data.</p> <p>Current control: (0-10%) attendance check, 3 practical sessions with assignments</p>
<b>Forms of control</b>	<p>Intermediate control: (0-20%) Online-test with choice of answer options at the end of the 3rd module</p> <p>Final control: (0-70%) presentation of the result of the practical assignments</p>
<b>Assesment rules</b>	100% = attendance check (0-10%)+ online test (0-20%) + graduating of the practical ssignment results (0-70%)
<b>Contact</b>	Oxana Gaydash, Deputy Head of Joint Department with Otkritie Financial Group
<b>Email</b>	oxana.gaydash@open.ru, ogaidash@hse.ru