# Seminar Nr. 1. Introducere în Conceptele și Ecosistemul Big Data

**Obiectiv:**

Prezentarea conceptelor fundamentale ale Big Data și a ecosistemului tehnologic asociat, în scopul oferirii unei baze pentru înțelegerea ulterioară a tehnologiilor și aplicațiilor specifice.

**Durata: 60 de minute**

**Structura seminarului:**

**1. Introducere în Big Data (10 minute)**

* **Definirea Big Data**:
  + Explicarea celor **5V**: volum, viteză, varietate, veridicitate, valoare.
* **De ce este important Big Data?**
  + Impactul datelor masive asupra afacerilor și societății.

**Activitate:** Întrebare interactivă: „Care sunt exemplele de Big Data pe care le întâlniți în viața de zi cu zi?” Studenții răspund cu exemple precum rețele sociale, aplicații de navigare sau platforme de streaming.

**2. Componentele Ecosistemului Big Data (25 minute)**

* **Stocarea datelor**:
  + Introducere în **Data Lakes** și **Data Warehouses**.
* **Procesarea datelor**:
  + Explicarea conceptelor de **procesare distribuită**.
  + Tehnologii majore: **Hadoop** și **MapReduce** (explicare succintă).
* **Instrumente de analiză**:
  + Menționarea pe scurt a **NoSQL databases** și **Apache Spark**.

**Activitate:** Prezentarea unui ecosistem simplu de Big Data – un exemplu vizual care arată fluxul de date de la stocare la analiză.

**3. Utilizări Big Data în practică (15 minute)**

* **Big Data în afaceri și decizii**:
  + Un exemplu rapid de utilizare a Big Data în analiza comportamentului utilizatorilor pe un website.
* **Big Data și sisteme suport decizii**:
  + Prezentarea unui caz simplu unde datele sunt folosite pentru optimizarea lanțului de aprovizionare.

**Activitate:** Mini-exercițiu de brainstorming: „Cum credeți că poate fi aplicat Big Data în domeniul vostru de studiu sau de interes?”

**4. Q&A și Concluzii (10 minute)**

* Recapitulare rapidă a conceptelor cheie.
* Răspunsuri la întrebările studenților.

**Resurse:**

* Prezentare PowerPoint concisă.
* Grafice/diagrame pentru vizualizarea ecosistemului.

# Семинар No 1. Введение в концепции и экосистему Big Data

**Цель:**

Представить основные концепции Big Data и связанные с ним технологические экосистемы, чтобы заложить основу для дальнейшего понимания специфических технологий и приложений.

**Длительность: 60 минут**

**Структура семинара:**

**1. Введение в Big Data (10 минут)**

* **Определение Big Data**:
  + Объяснение 5V: объем (volume), скорость (velocity), разнообразие (variety), достоверность (veracity), ценость (value).
* **Почему Big Data важны?**
  + Влияние больших данных на бизнес и общество.

**Активность:** Вопрос студентам: «Какие примеры Big Data вы встречаете в повседневной жизни?» Студенты приводят примеры, такие как социальные сети, навигационные приложения или стриминговые платформы.

**2. Компоненты экосистемы Big Data (25 минут)**

* **Хранение данных**:
  + Введение в **Data Lakes** и **Data Warehouses**.
* **Обработка данных**:
  + Объяснение концепций **распределенной обработки**.
  + Основные технологии: **Hadoop** и **MapReduce** (краткое объяснение).
* **Инструменты анализа**:
  + Краткое упоминание **NoSQL баз данных** и **Apache Spark**.

**Активность:** Презентация простой экосистемы Big Data – визуальный пример, показывающий поток данных от хранения до анализа.

**3. Применение Big Data на практике (15 минут)**

* **Big Data в бизнесе и принятии решений**:
  + Простой пример использования Big Data для анализа поведения пользователей на веб-сайте.
* **Big Data и системы поддержки принятия решений**:
  + Презентация простого случая, когда данные используются для оптимизации цепочки поставок.

**Активность:** Мозговой штурм: «Как, по вашему мнению, Big Data могут быть применены в вашей области обучения или интересов?»

**4. Вопросы и ответы, заключение (10 минут)**

* Быстрое повторение ключевых концепций.
* Ответы на вопросы студентов.

**Ресурсы:**

* Краткая презентация PowerPoint.
* Графики/диаграммы для визуализации экосистемы.

**Seminar 2: Tehnologii specifice și procesare avansată, optimizări în Big Data**

**Forme de organizare**: discuție ghidată, activitate comparativă, aplicare practică pe teme individuale de cercetare, materializare în tr-o prezentare PPTX.

**Obiective de învățare**

* Compararea principalelor tehnologii Big Data (Hadoop, Spark, Hive, Pig, NoSQL) în funcție de scopul și specificul utilizării.
* Explicarea principiilor de procesare distribuită și a modului în care RDD și DataFrames contribuie la optimizarea performanței.
* Identificarea tehnologiilor potrivite pentru prelucrarea și analiza datelor specifice fiecărei teme de cercetare.
* Aplicarea conceptelor discutate pentru proiectarea unui flux de prelucrare Big Data individualizat.
* Argumentarea alegerilor tehnologice în funcție de volum, viteză, varietate și valoarea datelor.

**Cerințe**:

* Scopul principal este dezvoltarea unei gândiri analitice și aplicative.
* Gândire comparativă, aplicare conceptelor în cercetarea proprie și înțelegerea criteriilor tehnice ale alegerii unei tehnologii Big Data.

**Desfășurarea seminarului**

Introducere sumară în ecosistemul Big Data. Explicarea rolului tehnologiilor principale în procesarea volumelor mari de date și se clarifică diferențele dintre procesarea batch și procesarea în timp real.

Conceptele-cheie: distribuirea datelor, stocarea distribuită, procesarea paralelă, optimizarea prin structuri eficiente de date.

Explicarea tehnologiilor de bază:

* Apache Hadoop — platformă pentru procesare batch, bazată pe modelul MapReduce și sistemul distribuit HDFS; potrivită pentru volume mari de date istorice.
* Apache Pig și Apache Hive — instrumente pentru prelucrare și interogare de date pe Hadoop; Pig folosește scripturi de procesare, iar Hive oferă o interfață SQL-like ușor de folosit.
* Apache Spark — motor de procesare in-memory, care permite calcule rapide, optimizări avansate și procesare în timp real. Se explică conceptele de RDD (Resilient Distributed Dataset) și DataFrames, precum și modul în care acestea reduc timpul de execuție.
* Tehnologiile NoSQL (ex. MongoDB, Apache Cassandra, Neo4j) — sisteme de stocare flexibile și scalabile, adaptate pentru date eterogene sau nestructurate. Se prezintă modelele Key-Value, Document, Columnar și Graph, precum și avantajele distribuției prin sharding și replicare.

Activitatea comparativă. Discutarea și formularea diferențelor esențiale dintre aceste tehnologii: scopul principal, tipul de date potrivit, avantajele și limitările, modul de utilizare în scenarii reale. Accentul cade pe înțelegerea logicii din spatele alegerii unei tehnologii, nu pe memorarea definițiilor.

Noțiunile studiate în contextul propriei teme de cercetare. Pe baza seturilor de date cu care se lucrează, se proiectează un mic flux de prelucrare Big Data: se aleg tehnologiile adecvate pentru stocare, procesare și analiză, justificând deciziile luate. Se pot combina mai multe instrumente — de exemplu, utilizarea unei baze NoSQL pentru stocare și a Spark pentru analiză rapidă.

**Conținut tematic**

* Distribuirea și procesarea paralelă a datelor mari.
* Arhitectura și principiile Hadoop: HDFS, MapReduce.
* Pig și Hive ca instrumente pentru transformare și interogare pe Hadoop.
* Spark: RDD, DataFrames, Lazy Evaluation, Catalyst Optimizer.
* Modele NoSQL: Key-Value, Document, Columnar, Graph.
* Sharding, replicare, partitioning, optimizări pentru reducerea timpilor de procesare.
* Selectarea tehnologiilor în funcție de tipul de date și obiectivele analizei.

**Exemplu orientativ:**

* Stocarea datelor brute într-o bază NoSQL de tip Big Column (ex. Casandra).
* Procesarea și curățarea datelor cu Spark DataFrames.
* Analiza temporală și agregările cu Hive pentru interogări rapide.
* Rezultatele finale folosite pentru modele predictive sau vizualizări interactive.

**Bibliografie recomandată**

* Tom White — *“Hadoop: The Definitive Guide”*
* Jules S. Damji et al. — *“Learning Spark”*
* Nathan Marz — *“Big Data: Principles and Best Practices of Scalable Realtime Data Systems”*

Documentații oficiale:

* <https://spark.apache.org>
* <https://hadoop.apache.org>
* <https://hive.apache.org>
* <https://cassandra.apache.org>

**Семинар 2: Специфические технологии и продвинутая обработка, оптимизации в Big Data**  
Формы организации: направленная дискуссия, сравнительная aктивность, практическое применение на индивидуальных исследовательских темах, подготовка презентации PPTX.

**Цели обучения**  
• Сравнение основных технологий Big Data (Hadoop, Spark, Hive, Pig, NoSQL) в зависимости от цели и специфики использования.  
• Объяснение принципов распределённой обработки и роли RDD и DataFrames в оптимизации производительности.  
• Идентификация подходящих технологий для обработки и анализа данных, соответствующих каждой исследовательской теме.  
• Применение изученных концепций для проектирования индивидуального потока обработки Big Data.

• Аргументация технологического выбора с учётом объёма, скорости, разнообразия и ценности данных.

**Требования**  
• Главная цель — развитие аналитического и прикладного мышления.  
• Формирование сравнительного подхода, применение концепций в собственном исследовании и понимание технических критериев выбора технологий Big Data.

**Ход семинара**

Краткое введение в экосистему Big Data. Объяснение роли основных технологий в обработке больших объёмов данных и различий между пакетной (batch) и потоковой (real-time) обработкой.  
Ключевые понятия: распределение данных, распределённое хранение, параллельная обработка, оптимизация с использованием эффективных структур данных.

**Объяснение базовых технологий**

• **Apache Hadoop** — платформа для пакетной обработки, основанная на модели MapReduce и распределённой файловой системе HDFS; подходит для обработки больших исторических объёмов данных.  
• **Apache Pig** и **Apache Hive** — инструменты для обработки и запросов данных на Hadoop; Pig использует скрипты обработки, а Hive предлагает SQL-подобный интерфейс, удобный для анализа.  
• **Apache Spark** — вычислительный движок с обработкой в оперативной памяти, обеспечивающий высокую скорость, продвинутые оптимизации и поддержку потоковой обработки. Рассматриваются понятия RDD (Resilient Distributed Dataset) и DataFrames, а также способы сокращения времени выполнения задач.  
• **NoSQL-технологии** (например, MongoDB, Apache Cassandra, Neo4j) — гибкие и масштабируемые системы хранения, предназначенные для работы с гетерогенными или неструктурированными данными. Представлены модели Key-Value, Document, Columnar и Graph, а также преимущества шардирования и репликации.

**Сравнительная активность**

Обсуждение и формулировка ключевых различий между этими технологиями: основное назначение, тип подходящих данных, преимущества и ограничения, способы применения в реальных сценариях. Особое внимание уделяется логике выбора технологии, а не запоминанию определений.

**Применение изученных понятий к собственной теме исследования**  
На основе доступных наборов данных проектируется небольшой поток обработки Big Data: выбираются подходящие технологии для хранения, обработки и анализа, обосновываются принятые решения. Возможна комбинация инструментов — например, использование NoSQL для хранения и Spark для быстрой аналитики.

**Тематическое содержание**  
• Распределение и параллельная обработка больших данных.  
• Архитектура и принципы Hadoop: HDFS, MapReduce.  
• Pig и Hive как инструменты для трансформации и запросов в Hadoop.  
• Spark: RDD, DataFrames, Lazy Evaluation, Catalyst Optimizer.  
• Модели NoSQL: Key-Value, Document, Columnar, Graph.  
• Шардирование, репликация, партиционирование, оптимизация времени обработки.  
• Выбор технологий в зависимости от типа данных и целей анализа.

**Пример**  
• Хранение исходных данных в NoSQL базе типа Big Column (например, Cassandra).  
• Обработка и очистка данных с использованием Spark DataFrames.  
• Временной анализ и агрегации с помощью Hive для быстрых запросов.  
• Использование результатов для предиктивных моделей или интерактивной визуализации.

**Рекомендуемая литература**

* Tom White — *“Hadoop: The Definitive Guide”*
* Jules S. Damji et al. — *“Learning Spark”*
* Nathan Marz — *“Big Data: Principles and Best Practices of Scalable Realtime Data Systems”*

Documentații oficiale:

* <https://spark.apache.org>
* <https://hadoop.apache.org>
* <https://hive.apache.org>
* <https://cassandra.apache.org>

**SEMINAR Nr. 3. Analiza Big Data în contextul Economiei 4.0 și 5.0**

#### **Obiective**:

* Înțelegerea rolului Big Data în tranziția către Economia 4.0 și 5.0.
* Explorarea cazurilor de utilizare a Big Data în Industria 4.0 (automatizare, IoT, predictive analytics) și Economia 5.0 (societate digitală, inteligență artificială, soluții orientate spre om).
* Aplicarea tehnicilor de analiză Big Data pentru a explora date relevante din domeniul industriei sau societății.

#### **Subiecte**:

1. **Introducere în Economia 4.0 și 5.0**:
   * Definirea conceptelor de **Industria 4.0** (automatizare, IoT, fabrici inteligente) și **Economia 5.0** (societate inteligentă, colaborare om-mașină, inteligență artificială).
   * Impactul Big Data în automatizarea industrială, optimizarea proceselor și luarea deciziilor inteligente.
2. **Cazuri de utilizare Big Data în Industria 4.0**:
   * Exemple de aplicații Big Data în automatizarea producției, întreținerea predictivă și monitorizarea proceselor prin IoT.
   * Explorarea unui set de date din domeniul manufacturii sau IoT (de exemplu, date despre performanța utilajelor, monitorizarea senzorială a producției).
3. **Cazuri de utilizare Big Data în Economia 5.0**:
   * Exemple de utilizare a Big Data în societatea inteligentă, interacțiunea om-mașină și deciziile centrate pe utilizator (de exemplu, date din sănătate, orașe inteligente).
   * Utilizarea unui set de date legate de societatea digitală sau smart cities (de exemplu, monitorizarea traficului urban, gestionarea energiei în orașe).
4. **Lucrare practică: Analiza datelor pentru optimizarea proceselor sau a serviciilor**:
   * Utilizarea unui instrument de analiză Big Data (Python, Pandas, Spark) pentru a analiza un set de date mare.
   * Scrierea unui cod simplu pentru a identifica tipare, anomalii sau tendințe în datele analizate.
   * Generarea de **vizualizări** (grafice, diagrame) pentru a prezenta concluziile analizelor.

#### **Exemplu de studiu de caz**:

* **Industria 4.0**: Analizarea unui set de date privind senzori industriali, pentru a detecta probleme predictive în utilaje (predictive maintenance).
* **Economia 5.0**: Analizarea datelor de la un oraș inteligent (trafic, consum energetic) pentru a identifica soluții de optimizare și eficientizare a resurselor.

#### **Cerințe de livrare**:

* Un raport care să includă:
  + O scurtă prezentare a conceptelor de **Economia 4.0** și **Economia 5.0** și rolul Big Data în acestea.
  + Prezentarea cazului de utilizare ales și descrierea setului de date analizat.
  + Codul sursă utilizat pentru analiză și vizualizările generate.
  + Concluzii bazate pe date și recomandări pentru optimizarea proceselor sau serviciilor.

#### **Durata: 4 ore**

#### **Evaluare**:

* Explicarea clară a conceptelor de **Economia 4.0** și **Economia 5.0**: 20%.
* Selectarea și explorarea adecvată a unui set de date relevant: 30%.
* Corectitudinea și relevanța analizei datelor (codul și vizualizările): 30%.
* Calitatea interpretării și concluziilor: 20%.

**Семинар Nr. 3. Экономика 4.0 и 5.0**

**Цель семинара:**

Целью данного семинара является углубленное понимание концепций Экономики 4.0 и 5.0, анализ влияния новейших технологий, таких как Big Data, IoT и искусственный интеллект (ИИ), на экономические и социальные преобразования. Студенты исследуют применение этих технологий в умных городах и отраслях, а также обсудят возможности и вызовы цифровой и устойчивой экономики.

**Проведение семинара**

**1. Введение и презентация концепций (15 минут)**

* Понятия Экономики 4.0 и 5.0, подчеркивая основные различия между ними.

**Вопросы для обсуждения:**

* Какие основные различия между Экономикой 4.0 и Экономикой 5.0?
* Как Big Data может поддержать переход от Экономики 4.0 к Экономике 5.0?

**2. Казусы:**

* **Казус 1:** Промышленный производитель внедряет IoT и Big Data для оптимизации производства и сокращения потерь сырья. Обсуждение: Какие экономические и экологические выгоды получены?
* **Казус 2:** Умный город интегрирует датчики и ИИ для улучшения городской мобильности и сокращения выбросов CO2. Обсуждение: Как эти решения способствуют развитию устойчивого города в контексте Экономики 5.0?
* **Казус 3:** Финансовая компания использует Big Data и машинное обучение для предоставления персонализированных финансовых услуг. Обсуждение: Как эта стратегия может способствовать более инклюзивной и справедливой экономике?

**Задания для студентов:**

* Проанализируйте кейс и обсудите, как внедренные технологические решения вписываются в парадигму Экономики 4.0 или 5.0.
* Определите вызовы, с которыми сталкиваются организации или города при использовании Big Data и ИИ, и предложите решения для их преодоления.

**3. Презентация и обсуждение результатов**

Каждая группа представляет свои выводы, подчеркивая, как обсуждаемые технологии способствуют экономическому развитию и их влияние на общество и окружающую среду.

**Вопросы для обсуждения:**

* Каковы риски и этические проблемы, связанные с массовым использованием Big Data и ИИ в экономике?
* Как общество может гарантировать, что новейшие технологии принесут пользу всем, а не только крупным корпорациям?

**4. Заключение и общая дискуссия**

* Преподаватель подводит итоги обсуждения, выделяя, как Big Data и другие новейшие технологии трансформируют глобальную экономику.
* Открытая дискуссия о будущем цифровой экономики и роли технологий в устойчивом развитии.
  + Насколько вы считаете, что Экономика 5.0 станет реальностью в ближайшие годы?
  + Какие первоочередные меры необходимы для перехода от Экономики 4.0 к экономике, ориентированной на людей и устойчивость?