|  |  |  |  |
| --- | --- | --- | --- |
| МИНОБРНАУКИ РОССИИ | **09.03.04 Программная инженерия** | | |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА–Российский технологический университет»** | Дисциплина: | | |
| **«Введение в профессиональную деятельность»** | | |
| Форма обучения: очная | | |
| Институт информационных технологий | Курс 1 | Семестр 1 | 2023/2024 учебный год |
| Группа: **ИКБО-74-23** | | |
| Кафедра корпоративных информационных систем |
| **ФИО: Зернов Никита Андреевич** | | | |
| Big data как инструменты и методы обработки данных для конкретных задач и целей.  В современном мире данные стали ключевым ресурсом, определяющим успех в различных областях. Однако объемы информации, генерируемой и собираемой каждый день, становятся настолько огромными, что традиционные методы обработки данных не способны эффективно справляться с таким потоком. В этом контексте, концепция Big Data, или больших данных, приходит на помощь, предоставляя инструменты и методы обработки данных, необходимые для анализа и извлечения ценной информации из огромных объемов данных.  Big Data представляет собой понятие, описывающее избыточные объемы данных, характеризующиеся высокой степенью сложности и разнообразия. Эти информационные массы могут оказаться слишком громоздкими и сложными для обработки с использованием традиционных методов баз данных или стандартных инструментов анализа данных.  Инструменты Big Data:  - Apache Hadoop является основополагающим элементом в области Big Data, представляя собой фреймворк для распределенного хранения и обработки данных, построенный на принципах MapReduce.  - Spark предоставляет более быстрые и гибкие возможности обработки данных по сравнению с Hadoop. Он поддерживает in-memory обработку данных, что повышает производительность.  - Для работы с неструктурированными данными часто используются NoSQL базы данных, такие как MongoDB, Cassandra, и Couchbase.  - Flink предлагает эффективную и масштабируемую обработку потоков данных, что делает его отличным выбором для проведения анализа данных в режиме реального времени.  Методы Обработки Данных в контексте Big Data:  - С использованием машинного обучение Big Data открывает новые горизонты для применения машинного обучения. Модели могут обучаться на огромных объемах данных, что улучшает их точность и способность к обобщению.  - Учитывая потребность в обработке данных в реальном времени, анализ потоков данных становится важным инструментом, обеспечивающим мгновенный отклик на динамические изменения и события.  - Для изучения взаимосвязей и связей в обширных объемах данных применяются графовые базы данных, вроде Neo4j.  Примеры Применения Big Data:  - Анализ данных в области медицины направлен на диагностику и прогнозирование заболеваний у пациентов.  - Обнаружение мошенничества, прогнозирование рыночных тенденций в реальном времени.  - Оптимизация сетей, путем прогнозирования загруженности сети.  Big Data становится интегральной частью современного общества, предоставляя мощные средства и технологии для обработки данных. Возможность проведения анализа обширных информационных потоков в режиме реального времени открывает новые перспективы для принятия решений и достижения стратегических целей. Тем не менее, для достижения максимальной эффективности необходимо не только использовать передовые инструменты, но и разрабатывать глубокие методологии анализа данных с целью извлечения максимальной ценности из обширных объемов информации, доступных в наше время. | | | |