**Санкт-Петербургский государственный университет**

**Р А Б О Ч А Я П Р О Г Р А М М А**

**УЧЕБНОЙ ДИСЦИПЛИНЫ**

Распределенная обработка информации и NoSQL базы данных

Distributed Information Processing and NoSQL DBMSs

**Язык(и) обучения**

русский

Трудоемкость в зачетных единицах: 2

Регистрационный номер рабочей программы: 010075

Санкт-Петербург

2020

## Раздел 1. Характеристики учебных занятий

## 1.1. Цели и задачи учебных занятий

Сформировать у участников семинара общее представление о распределённых вычислительных системах, предназначенных для хранения и обработки значительных объёмов информации.

Рассмотреть ряд примеров исторических, современных и перспективных технологий распределённой обработки данных.

Опробовать несколько из рассмотренных подходов распределённой обработки информации, реализованный самостоятельно или на базе общедоступного программного обеспечения, дать количественную и качественную оценки полученным результатам.

На основании у слушателей общее представление о содержании, задачах и методах современного системного программирования как самостоятельной научной и инженерной дисциплины, о диапазоне и разнообразии ее типичных приложений.

Поставленные цели достигаются путём подготовки и слушания и обсуждения докладов по тематике семинара и самостоятельного решения практических задач.

## 1.2. Требования подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)

Знание основ информатики, программирования и математики в пределах бакалаврской подготовки.

Требуется предварительно или параллельно с семинаром слушать общий вводный теоретический курс по СУБД.

Желательно иметь представление о сетевых технологиях.

## 1.3. Перечень результатов обучения (learning outcomes)

В процессе участия в семинаре «Распределённая обработка данных и NoSQL СУБД» обучаемые приобретают следующие знания, умения и навыки.

Знания:

* сущности и значения информации в развитии общества, основных методов, способов и средств получения, хранения, переработки информации;
* исторических, современных и перспективных тенденций развития программного обеспечения распределённых вычислительных систем;
* методов проектирования моделей данных, предназначенных для обработки при помощи распределённых СУБД;
* современных парадигм и языков распределённого программирования, языков программирования и базовых парадигм и алгоритмов;
* принципов организации современных распределённых СУБД, информационных систем, принципов взаимодействия их внутренних механизмов.

Умения:

* использовать в научной и познавательной деятельности, а также в социальной сфере профессиональные навыки работы с информационными и компьютерными технологиями;
* готовить и делать доклады, сопровождаемые презентациями и демонстрациями программного обеспечения;
* критически переосмысливать свой опыт, адаптироваться к различным ситуациям, проявлять творческий подход, инициативу и настойчивость в достижении целей профессиональной деятельности;
* делать анализ и грамотную оценку эффективности применяемых и разрабатываемых подходов распределённого программирования;
* реализовывать решения, направленные на поддержку социально значимых проектов, на повышение электронной грамотности населения, обеспечения общедоступности информационных услуг;
* формализовывать поставленные задачи и реализовывать сложные программные комплексы с точки зрения грамотной профессиональной разработки различного рода проектов для достижения эффективного результата.

Навыки:

* работы с информацией из различных источников, включая сетевые ресурсы сети Интернет, для решения профессиональных задач;
* осуществления целенаправленного поиска информации о технологических достижениях в сети Интернет и из других источников;
* применения в профессиональной деятельности современных языков программирования и языков баз данных, СУБД, операционных систем, электронных библиотек и пакетов программ, сетевых технологий;
* взаимодействия с коллегами, работы в коллективе.

Компетенции, формируемые дисциплиной:

* ОПК-1 – способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональной деятельности;
* ОПК-3 – способен применять современные информационные технологии, в том числе отечественные, при создании программных продуктов и программных комплексов различного назначения;
* ОПК-4 – способен участвовать в разработке технической документации программных продуктов и программных комплексов;
* ОПК-5 – способен инсталлировать и сопровождать программное обеспечение для информационных систем и баз данных, в том числе отечественного производства;
* ПКА-1 – способен демонстрировать базовые знания математических и естественных наук, программирования и информационных технологий;
* ПКП-1 – способность проводить под научным руководством исследование на основе существующих методов в конкретной области профессиональной деятельности;
* ПКП-2 – способен решать задачи в области развития науки, техники и технологии с учетом нормативного правового регулирования в сфере интеллектуальной собственности;
* ПКП-4 – способен применять современные информационные технологии при проектировании, реализации, оценке качества и анализа эффективности программного обеспечения для решения задач в различных предметных областях;
* ПКП-5 – способен использовать основные методы и средства автоматизации проектирования, реализации, испытаний и оценки качества при создании конкурентоспособного программного продукта и программных комплексов, а также способен использовать методы и средства автоматизации, связанные с сопровождением, администрированием и модернизацией программных продуктов и программных комплексов;
* ПКП-6 – способен использовать знания направлений развития компьютеров с традиционной (нетрадиционной) архитектурой; современных системных программных средств: операционных систем, операционных и сетевых оболочек, сервисных программ; тенденции развития функций и архитектур проблемно-ориентированных программных систем и комплексов в профессиональной деятельности;
* ПКП-8 – способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования;
* УКБ-3 – способен понимать сущность и значение информации в развитии общества, использовать основные методы получения и работы с информацией с учетом современных технологий цифровой экономики и информационной безопасности.

## 1.4. Перечень и объём активных и интерактивных форм учебных занятий

Обучение полностью (30 ак. часов) проводится в активных (доклады на семинаре) и интерактивных (последующая научная дискуссия) формах.

## Раздел 2. Организация, структура и содержание учебных занятий

## 2.1. Организация учебных занятий

## 2.1.1 Основной курс

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся | | | | | | | | | | | | | | | | | | |
| Код модуля в составе дисциплины,  практики и т.п. | Контактная работа обучающихся с преподавателем | | | | | | | | | | | | Самостоятельная работа | | | | Объём активных и интерактивных  форм учебных занятий | Трудоёмкость |
| лекции | семинары | консультации | практические  занятия | лабораторные работы | контрольные работы | коллоквиумы | текущий контроль | промежуточная  аттестация | итоговая аттестация | под руководством преподавателя | в присутствии  преподавателя | сам. раб. с использованием  методических материалов | текущий контроль (сам.раб.) | промежуточная аттестация (сам.раб.) | итоговая аттестация  (сам.раб.) |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | | | | | | | | | | | | | |
| Форма обучения: очная | | | | | | | | | | | | | | | | | | |
| Семестр 7 |  | 30 |  |  |  |  |  |  | 2 |  |  |  | 34 |  | 6 |  | 30 | 2 |
|  |  | 5-  25 |  |  |  |  |  |  | 5-  25 |  |  |  | 1-1 |  | 1-1 |  |  |  |
| ИТОГО |  | 30 |  |  |  |  |  |  | 2 |  |  |  | 34 |  | 6 |  |  | 2 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Виды, формы и сроки текущего контроля успеваемости и промежуточной аттестации | | | | | | |
| Код модуля в составе дисциплины, практики и т.п. | Формы текущего контроля успеваемости | | Виды промежуточной аттестации | | Виды итоговой аттестации  (только для программ итоговой аттестации и дополнительных образовательных программ) | |
| Формы | Сроки | Виды | Сроки | Виды | Сроки |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | |
| Форма обучения: очная | | | | | | |
| Семестр 7 |  |  | зачёт, устно, традиционная форма | по графику промежуточной аттестации |  |  |

## 2.2. Структура и содержание учебных занятий

Период обучения: семестр 7

| № п.п. | Наименование темы (раздела, части) | Вид учебных занятий | Кол-во часов |
| --- | --- | --- | --- |
| 1 | Тема 1. Отличные от реляционных методологии хранения и обработки данных и реализующие их технологии. | семинары | 4 |
| с исп. мет. материалов | 4 |
| 2 | Тема 2. Исторический обзор распределённых вычислительных сред и систем хранения данных. | семинары | 4 |
| с исп. мет. материалов | 4 |
| 3 | Тема 3. Специфические парадигмы организации распределённых вычислений. | семинары | 6 |
| с исп. мет. материалов | 6 |
| 4 | Тема 4. Современные распределённые, в т.ч. NoSQL, СУБД. | семинары | 6 |
| с исп. мет. материалов | 8 |
| 5 | Тема 5. Применения децентрализованных вычислительных сред | семинары | 6 |
| с исп. мет. материалов | 6 |
| 6 | Тема 6. Решение практических задач параллельного и распределённого программирования | семинары | 4 |
| с исп. мет. материалов | 6 |
| 7 | Промежуточная аттестация | самостоятельная работа | 6 |
| промежуточная аттестация | 2 |
| **Итого** | | | **72** |

**Тема 1. Отличные от реляционных методологии хранения и обработки данных и реализующие их технологии.**

Парадигмы моделирования и структурирования данных. Объектно-ориентированные, слабоструктурированные, сетевые, функциональные и логические модели данных. Автономные не реляционные СУБД и системы отображения не реляционных моделей данных в реляционные для обработки данных реляционными СУБД.

**Тема 2. Исторический обзор распределённых вычислительных сред и систем хранения данных.**

Исторический обзор распределённых вычислительных сред, файловых систем, распределённых операционных систем, специфических для кластеров систем хранения данных.

**Тема 3. Специфические парадигмы организации распределённых вычислений.**

Организация распределённых вычислений при помощи вычислительных кластеров, GRID, систем, реализующих подход Map-Reduce. Особенности индексирования распределённых данных. Языки и парадигмы распределённого программирования.

**Тема 4. Современные распределённые, в т.ч. NoSQL, СУБД.**

NoSQL Movement. Рассмотрение ряда современных NoSQL СУБД, их сравнение, выделение их сильных и слабых сторон.

**Тема 5. Применения децентрализованных вычислительных сред.**

Распределённые и децентрализованные системы передачи и хранения данных. Устойчивость к централизованным атакам. Децентрализованные системы проводки финансовых транзакций. Децентрализованные распределённые вычислительные сети с динамической топологией.

**Тема 6. Решение практических задач параллельного и распределённого программирования.**

Выполняется сформированными из состава обучающихся рабочими группами дома либо в компьютерном классе по выбору участников. Темы и подходы к решению предварительно согласуются с преподавателем. Результаты представляются на семинарских занятиях.

## Раздел 3. Обеспечение учебных занятий

## 3.1. Методическое обеспечение

## 3.1.1 Методические указания по освоению дисциплины

Успешное освоение дисциплины возможно благодаря посещению лекций, участию в обсуждении вопросов, подготовленных к занятию, самостоятельной работе, включающей в себя чтение специальной литературы по разделам темы, подготовка презентаций по тематике курса.

## 3.1.2 Методическое обеспечение самостоятельной работы

Настоящей программой предусмотрены формы самостоятельной работы с использованием научной и популярной литературы по тематике дисциплины. Основная форма самостоятельной работы – подготовка презентаций и сообщений по тематике курса и источникам, указанным в списке литературы.

## 3.1.3 Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания

Текущий контроль сводится к проверке того, насколько обучающиеся учитывают замечания преподавателя по сделанным докладам.

При промежуточной аттестации используются следующие критерии.

Для положительной аттестации в любом случае требуется сделать минимум один доклад.

Один сделанный доклад оценивается по шкале от 0 до 40 баллов:

* от 0 до 10 за полноту доклада в плане собственно предмета доклада,
* от 0 до 10 за полноту доклада в плане основ и контекста (background и related work),
* от 0 до 10 за речь, оформление слайдов,
* от 0 до 10 за научную дискуссию (ответы на вопросы).

Работа на семинаре, за исключением собственных докладов, оценивается по шкале от 0 до 20 баллов:

* от 0 до 10 оценивается посещаемость семинара (доля посещённых докладов).
* от 0 до 10 оценивается активность участия в научной дискуссии (удвоенная доля докладов, по которым обучающийся участвовал в дискуссии, но не более 1).

В случае, если обучающийся сделал лишь один доклад или недостаточно активно участвовал в работе семинара, по инициативе обучающегося на зачёте может быть проведена отдельная беседа без подготовки по материалам одного из сделанных в семестре докладов на выбор преподавателя. Полнота знаний обучающегося по выбранной теме оценивается по шкале от 0 до 40 баллов.

Сделав в течение семестра два доклада без нареканий, хорошо посещая семинар и участвуя в дискуссии, обучающийся может получить в сумме 100 баллов за работу в семестре. Баллы при аттестации интерпретируются следующим образом.

| **Баллы, 0–100** | **Оценка по шкале ECTS** | **Аттестация СПбГУ** |
| --- | --- | --- |
| 90–100 | A | зачтено |
| 80–89 | B |
| 70–79 | C |
| 60–69 | D |
| 50–59 | E |
| Менее 50 | F | не зачтено |

## 3.1.4 Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)

*Примеры тем докладов:*

1. Object-Relational Mapping
2. Object-XML Mapping
3. «Урождённые» ОО СУБД
4. «Урождённые» XML СУБД
5. XQuery, XUpdate
6. Дедуктивные СУБД
7. Графовые БД
8. Паттерн «Map-Reduce»
9. GRID, очереди заданий для распределённых вычислителей
10. Архитектура Google Pregel
11. Протоколы распределённого консенсуса на примере Raft
12. Распределённые хэш-таблицы
13. Архитектура типичных криптовалют
14. Сети с недетерминированной топологией сетевого уровня (самоорганизующиеся Mesh)
15. Цензуроустойчивые сети и хранилища; атаки на них
16. MapReduce: внутри, снаружи или сбоку от параллельных СУБД?
17. Транзакционные параллельные СУБД: новая волна
18. Дискуссия по поводу "NoSQL" не имеет никакого отношения к SQL
19. "Моя десятка основных утверждений о хранилищах данных"
20. Уточнения по поводу теоремы CAP и ошибок, связанных с данными
21. Ошибки в системах баз данных, согласованность "в конечном счете" и теорема CAP
22. Обзор и сравнительный анализ систем управления не реляционными базами данных
23. Критика CAP, теорема PACELC
24. NoSQL и Agile: а это вообще связано?
25. The NoSQL movementt
26. Consisency, partitioning
27. Storage Layout, Query Models
28. Amazon’s Dynamo
29. Project Voldemort, Other Key-/Value-Stores
30. Apache CouchDB
31. MongoDB
32. Joins and Other Aggregation Enhancements Coming in MongoDB 3.2
33. Google BigTable
34. Bigtable Derivatives, Cassandra
35. Google Spanner
36. Apache Hadoop
37. Apache Spark
38. Яндекс Clickhouse
39. Google Firebase
40. Архитектура Tarantool

## 3.1.5 Методические материалы для оценки обучающимися содержания и качества учебного процесса

Для оценки обучающимися содержания и качества учебного процесса применяется анкетирование в соответствии с методикой и графиком, утвержденными в установленном порядке.

## 3.2. Кадровое обеспечение

## 3.2.1 Образование и (или) квалификация штатных преподавателей и иных лиц, допущенных к проведению учебных занятий

К чтению лекций привлекаются преподаватели, имеющие базовое образование и/или ученую степень, соответствующую профилю преподаваемой дисциплины.

## 3.2.2 Обеспечение учебно-вспомогательным и (или) иным персоналом

Не требуется.

## 3.3. Материально-техническое обеспечение

## 3.3.1 Характеристики аудиторий (помещений, мест) для проведения занятий

Аудитории должны соответствовать требованиям СанПиН 2.2.2/2.4.1340-03 в редакции от 21 июня 2016 года.

## 3.3.2 Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования

Стандартно оборудованные лекционные аудитории для проведения интерактивных лекций: один компьютер, видеопроектор, экран, маркерная доска. Программное обеспечение для демонстрации слайдов в форматах PDF и PowerPoint. Любой Веб-браузер актуальной на момент реализации дисциплины версии.

## 3.3.3 Характеристики специализированного оборудования

Не предусмотрено.

## 3.3.4 Характеристики специализированного программного обеспечения

Не предусмотрено.

## 3.3.5 Перечень и объёмы требуемых расходных материалов

Маркеры цветные, губки для очистки доски.

## 3.4. Информационное обеспечение

## 3.4.1 Список обязательной литературы

Литература по курсу находится в свободном доступе и представлет собой общедоступные технические отчёты, научные и популярные статьи.

1. Christof Strauch. NoSQL Databases. 2011. <http://www.christof-strauch.de/nosqldbs.pdf>
2. Craig S. Mullins. DRDA. 2002. <http://www.craigsmullins.com/drda.htm>
3. Michael Stonebraker. Errors in Database Systems, Eventual Consistency, and the CAP Theorem. BLOG@CACM, April 5, 2010. <http://cacm.acm.org/blogs/blog-cacm/83396-errors-in-database-systems-eventual-consistency-and-the-cap-theorem/fulltext>; Перевод С.Д. Кузнецова, ИСП. РАН <http://citforum.ru/gazeta/154/>
4. Кузнецов С. Д. Транзакционные параллельные СУБД: новая волна // Труды ИСП РАН. 2011. №. URL: <https://cyberleninka.ru/article/n/tranzaktsionnye-parallelnye-subd-novaya-volna>
5. Stonebraker M. Clarifications on the CAP theorem and data-related errors // VoltDB. com, October. – 2010. – Т. 21. <https://web.archive.org/web/20201104182853/https://www.voltdb.com/blog/2010/10/clarifications-cap-theorem-data-related-errors/>
6. Michael Stonebraker. The "NoSQL" Discussion has Nothing to Do With SQL. BLOG@CACM, November 4, 2009 <https://web.archive.org/web/20200811223507/https://cacm.acm.org/blogs/blog-cacm/50678-the-no-sql-discussion-has-nothing-to-do-with-sql/fulltext>
7. Michael Stonebraker. My Top 10 Assertions About Data Warehouses. BLOG@CACM, August 26, 2010 <https://web.archive.org/web/20200110144547/https://m-cacm.acm.org/blogs/blog-cacm/98136-my-top-10-assertions-about-data-warehouses/fulltext>
8. Кузнецов С. Д. MapReduce: внутри, снаружи или сбоку от параллельных СУБД? // Труды ИСП РАН. 2010. №. URL: <https://cyberleninka.ru/article/n/mapreduce-vnutri-snaruzhi-ili-sboku-ot-parallelnyh-subd>
9. Kleppmann M. A Critique of the CAP Theorem //arXiv preprint arXiv:1509.05393. – 2015 <https://arxiv.org/pdf/1509.05393.pdf>
10. Abadi D. Consistency tradeoffs in modern distributed database system design: CAP is only part of the story //Computer. – 2012. – Т. 45. – №. 2. – С. 37-42. <http://www.cs.umd.edu/~abadi/papers/abadi-pacelc.pdf>
11. O'Reilly Graph Databases The Definitive Book of Graph Databases <https://graphdatabases.com/>

## 3.4.2 Список дополнительной литературы

Не предусмотрено.

## 3.4.3 Перечень иных информационных источников

Не предусмотрено.

## Раздел 4. Разработчики программы

Луцив Дмитрий Вадимович, доцент кафедры системного программирования СПбГУ. d.lutsiv@spbu.ru +79119101002.