

**Федеральное государственное образовательное бюджетное
учреждение высшего образования
«Финансовый университет при Правительстве Российской Федерации»
(Финансовый университет)**

Кафедра информационных технологий
Факультет информационных технологий и анализа больших данных

Документ подписан усиленной неквалифицированной электронной подписью

Организация: Финансовый университет при Правительстве РФ

Сертификат: e2mrKgbyndwJ+2YLLPXW5ItVL3kSxWRV

Утверждено: Проректор по учебной и методической работе Е.А. Каменева

Дата: 09.02.2026 г.

А.В. Антипов

Разработка архитектуры информационно-аналитических систем

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки:

09.03.04 - Программная инженерия,

Образовательная программа «Разработка и внедрение информационно - аналитических систем»

Профиль:

«Разработка и внедрение информационно - аналитических систем»

Рекомендовано

Факультет информационных технологий и анализа больших данных

(протокол № 06 от 17.03.2026 г.)

Одобрено

Кафедра информационных технологий

(протокол № 1 от 27.02.2026 г.)

Содержание

1. Наименование дисциплины	4
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	4
3. Место дисциплины в структуре образовательной программы	6
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	6
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий . . .	7
5.1. Содержание дисциплины	7
5.2. Учебно-тематический план	10
5.3. Содержание семинаров	11
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине	15
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	15
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю . . .	18
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	20
8. Перечень основной и дополнительной литературы, необходимой для освоения дисциплины	28
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	29
10. Методические указания для обучающихся по освоению дисциплины	30
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	32

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	33
--	----

1. Наименование дисциплины

«Разработка архитектуры информационно-аналитических систем».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Код компетенции	Наименование компетенции	
ПКП-4	Способность разрабатывать и предлагать архитектурные решения для информационно-аналитических систем	
Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	
Умеет выбирать архитектурные решения для разрабатываемых информационных систем	знать: основные архитектурные паттерны ИАС (монолит, микросервисы, EDA), принципы проектирования DWH и Data Lake, жизненный цикл ИАС уметь: анализировать требования и выбирать оптимальную архитектуру хранения и обработки данных для конкретного бизнес-сценария	
Способен выбирать современные технологии, методы и подходы к реализации архитектурных решений	знать: современные СУБД (реляционные, NoSQL, NewSQL), ETL/ELT инструменты, технологии Big Data (Hadoop, Spark), BI-платформы, потоковая обработка уметь: подбирать технологический стек под заданные нефункциональные требования, проектировать ETL-процессы и конвейеры данных	
Способен проводить интеграцию архитектурных и проектных решений для информационных систем с существующими решениями	знать: стратегии интеграции с Legacy-системами, паттерны интеграции (ACL, CDC, Strangler Fig), принципы Data Quality и MDM уметь: разрабатывать планы интеграции новых систем с унаследованными, проектировать слои изоляции и миграции данных	

Умеет оформлять проектную документацию в соответствии с действующими стандартами	знать: стандарты проектной документации ИАС, формат Architecture Decision Records (ADR), структура технического задания, требования к документированию ETL-процессов уметь: составлять архитектурные решения, технические задания и эксплуатационную документацию в соответствии с корпоративными стандартами
---	---

3. Место дисциплины в структуре образовательной программы

Дисциплина «Разработка архитектуры информационно-аналитических систем» относится к «Циклу профиля».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Очная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 6 (в часах)
Общая трудоёмкость дисциплины	5/180	180
Контактная работа – Аудиторные занятия	50	50
Лекции	16	16
Семинары, практические занятия	34	34
Самостоятельная работа	130	130
Вид текущего контроля	Проектная работа	Проектная работа
Вид промежуточной аттестации	Экзамен	Экзамен

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Введение в информационно-аналитические системы

Определение ИАС и ее назначение для поддержки принятия управленческих решений. Фундаментальное различие OLTP и OLAP систем. Пирамида DIKW—трансформация данных в знания. Классификация ИАС по масштабу (DesktopBI, EnterpriseBI) и функциональному назначению (Reporting, Ad-hoc анализ, DataMining, DSS). Типовая слоистая архитектура ИАС: источники данных, ETL/ELT, хранилище данных (DWH), витрины данных (DataMarts), BI-инструменты. Жизненный цикл ИАС: анализ требований и обследование источников, проектирование архитектуры, реализация, тестирование, внедрение и обучение, эксплуатация и развитие. Роль архитектора ИАС как моста между бизнесом и технологиями.

Тема 2. Архитектурные требования и паттерны проектирования

Функциональные требования: источники данных (список, типы, методы захвата), преобразование и логика (правила очистки, бизнес-правила, историчность SCD), потребление данных (форматы отчетности, типы запросов). Нефункциональные требования: производительность (Latency, Throughput, DataFreshness), масштабируемость (вертикальная и горизонтальная), доступность (MTBF, MTTR, RPO, RTO), качество данных, безопасность. Методы сбора требований: интервью, воркшопы, анализ документации, прототипирование. Architecture Decision Records (ADR). Компромиссы и CAP-теорема. Архитектурные стили: монолитная архитектура, сервис-ориентированная архитектура (SOA) и ESB, микросервисная архитектура и DataMesh, событийно-ориентированная архитектура (EDA). Закон Амдала и закон Литтла. Гибридные подходы.

Тема 3. Проектирование слоя хранения данных

CAP-теорема и компромиссы в распределенных системах. Реляционные СУБД:

строковые (PostgreSQL, Oracle) vs колоночные (ClickHouse, Vertica), эффективность сжатия, MPP-системы (Greenplum, Teradata). NoSQL: документоориентированные (MongoDB), ключ-значение (Redis), wide-column (Cassandra), графовые (Neo4j). NewSQL и HTAP (CockroachDB, TiDB). Полиглотное хранение (Polyglot Persistence). Data Warehouse (DWH): принцип Schema-on-Write, архитектурные слои (Staging Area, Core/Integration Layer, Data Marts), модели «Звезда» и «Снежинка». Data Lake: принцип Schema-on-Read, зональная архитектура Medallion (Bronze, Silver, Gold). Математика стоимости хранения. Риск Data Swamp. Data Lakehouse (Delta Lake, Apache Iceberg, Apache Hudi). Критерии выбора архитектуры хранения.

Тема 4. Обработка и интеграция данных

Экосистема Hadoop: HDFS и MapReduce, математическая модель времени выполнения. Apache Spark: RDD, DAG, Lazy Evaluation, DataFrames, Spark SQL. Поточковая обработка: оконные функции (Tumbling, Sliding, Session), семантика обработки (At-most-once, At-least-once, Exactly-once). Apache Kafka, Spark Structured Streaming, Apache Flink. Архитектурные паттерны Lambda и Kappa. ETL vs ELT: области применения. Стратегии извлечения: Full Load, Incremental Load, CDC (Change Data Capture). Стратегии загрузки: Append, Truncate/Insert, Merge/Upsert. Data Quality: блокирующие и предупреждающие тесты. Оркестрация: DAG, критический путь, идемпотентность. Инструментарий: Informatica, Talend, NiFi, dbt, Apache Airflow.

Тема 5. Архитектура аналитического слоя

OLAP: концепция многомерного куба, измерения и меры, операции (Slice, Dice, Drill-down, Roll-up, Pivot). Архитектуры OLAP: MOLAP, ROLAP, HOLAP. In-Memory Tabular модели. Data Mining и Advanced Analytics: классификация, кластеризация, регрессия, ассоциативные правила, методология CRISP-DM. Интеграция ML в архитектуру (In-Database, отдельный сервис, Batch-scoring). BI-платформы: семантический слой, слой визуализации. Self-Service BI против Enterprise Reporting. Headless BI и Metrics Store. Современные BI-инструменты: Power BI, Tableau, Apache Superset.

Тема 6. Интеграция с унаследованными системами

Legacy-системы: определение, причины невозможности переписывания, главные вызовы интеграции (хрупкость, экзотические форматы, отсутствие API). Стратегии интеграции: прямой доступ к БД (Read-OnlyReplica), файловый обмен (CSV, XML, FTP), антикоррупционный слой (ACL), CDC на уровне логов (OracleGoldenGate, Debezium), экранный скрапинг и RPA. Оценка нагрузки при интеграции: формулы для Polling и инкрементальной загрузки. Паттерн «Душитель» (StranglerFigPattern). Матрица выбора решения. Работа с качеством данных из Legacy: StagingArea, очистка и валидация.

Тема 7. Управление качеством данных и мастер-данными

Принцип GIGO и правило 1-10-100. Измерения качества данных (DQDimensions): полнота, точность, согласованность, своевременность, уникальность, валидность. DataQualityIndex(DQI). Архитектура проверок качества: AtSource, DataQualityFirewall(жесткая блокировка, карантин), DataObservability. Master Data Management (MDM): мастер-данные, Golden Record. Процесс создания: Standardization, Matching, Merging. Алгоритмы матчинга: детерминистический, вероятностный (расстояние Левенштейна, коэффициент Жаккара). Архитектурные стили MDM: Registry, Consolidation, Coexistence, Centralized. DataGovernance: DataStewards, DataGovernanceCouncil.

Тема 8. Информационная безопасность в аналитических системах

Триада CIA: конфиденциальность, целостность, доступность. Оценка рисков: $R = P \times I$. Зонирование сети (DMZ): PublicZone, DMZ, TrustedZone. Bastion Host. Аутентификация: SSO, LDAP, Kerberos, SAML, OIDC. Авторизация: RBAC (Role-Based Access Control), RLS (Row-Level Security), CLS (Column-Level Security). Шифрование: atRest(TDE, FDE), inTransit(TLS/SSL). Обезличивание данных: статическое маскирование (SDM), динамическое маскирование (DDM), k-анонимность. Аудит и мониторинг: логирование доступа, SIEM.

5.2. Учебно-тематический план

№ п/п	Наименование тем(разделов) дисциплины	Трудоемкость в часах				
		Всего	Контактная работа - Аудиторная работа			Самостоя- тельная работа
			Общая, в т.ч.:	Лекции	Семинары, практическ- ие занятия	
1	Введение в информационно-аналитические системы	22	6	2	4	16
2	Архитектурные требования и паттерны проектирования	22	6	2	4	16
3	Проектирование слоя хранения данных	24	8	2	6	16
4	Обработка и интеграция данных	22	6	2	4	16
5	Архитектура аналитического слоя	22	6	2	4	16
6	Интеграция с унаследованными системами	24	6	2	4	18
7	Управление качеством данных и мастер-данными	22	6	2	4	16
8	Информационная безопасность в аналитических системах	22	6	2	4	16
	Итого	180	50	16	34	130

Формы текущего контроля успеваемости согласно учебному плану: **Проектная работа.**

5.3. Содержание семинаров

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарах, практических занятиях	Формы проведения занятий
Введение в информационно-аналитические системы	В чем фундаментальное различие между OLTP и OLAP системами и почему их нельзя объединить в одну систему. Как пирамида DIKW объясняет трансформацию данных в знания на примере конкретного бизнес-процесса. Какие классы задач решают Desktop BI и Enterprise BI системы. Из каких слоев состоит типовая архитектура ИАС и каковы их функции. Какова роль архитектора ИАС в жизненном цикле системы. Какие этапы жизненного цикла ИАС являются критическими с точки зрения рисков.	Интерактивная форма, практикум по решению задач по тематике занятия и коллективное обсуждение решений
Архитектурные требования и паттерны проектирования	Какие функциональные требования определяют выбор архитектуры хранения данных. Как измерить и обеспечить нефункциональные требования: Latency, Throughput, Data Freshness. Как применить CAP-теорему для выбора между консистентностью и доступностью в распределенной системе. В чем различие между вертикальным и горизонтальным масштабированием, когда применять каждый подход. Какие компромиссы возникают при выборе между монолитной, SOA и микросервисной архитектурой. Как закон Амдала ограничивает ускорение системы при добавлении ресурсов. Как оформить Architecture Decision Record (ADR) для фиксации архитектурного решения.	Интерактивная форма, практикум по решению задач по тематике занятия и коллективное обсуждение решений

Проектирование слоя хранения данных	Как CAP-теорема влияет на выбор типа СУБД для аналитической системы. В чем преимущества колоночных СУБД перед строковыми для аналитических нагрузок. Какие задачи решают MPP-системы и в чем их отличие от традиционных СУБД. Когда выбрать NoSQL решение (документоориентированное, ключ-значение, wide-column, графовое). Что такое полиглотное хранение (Polyglot Persistence) и когда оно оправдано. Какие архитектурные слои выделяют в Data Warehouse и какова их специфика. В чем различие между моделями «Звезда» и «Снежинка», какие компромиссы они предполагают. Как устроена зональная архитектура Data Lake (Medallion: Bronze, Silver, Gold). Что такое Data Swamp и какие меры предотвращают его образование. Как Data Lakehouse совмещает преимущества DWH и Data Lake.	Интерактивная форма, практикум по решению задач по тематике занятия и коллективное обсуждение решений
Обработка и интеграция данных	Как математическая модель MapReduce определяет время выполнения распределенных вычислений. Что такое Lazy Evaluation в Apache Spark и как это влияет на производительность. Какие семантики обработки существуют (At-most-once, At-least-once, Exactly-once) и как обеспечить Exactly-once. В чем различие между архитектурными паттернами Lambda и Кappa. Когда выбрать ETL, а когда ELT подход. Какие стратегии извлечения данных применяются (Full Load, Incremental Load, CDC). Как обеспечить идемпотентность в ETL-процессах и почему это критически важно. Какие инструменты оркестрации (Airflow, NiFi, dbt) подходят для разных сценариев.	Интерактивная форма, практикум по решению задач по тематике занятия и коллективное обсуждение решений

Архитектура аналитического слоя	Как устроена концепция многомерного OLAP-куба, что такое измерения и меры. Как выполняются операции Slice, Dice, Drill-down, Roll-up, Pivot в многомерном анализе. В чем различие между MOLAP, ROLAP и HOLAP архитектурами. Как интегрировать Data Mining и ML-модели в архитектуру ИАС (In-Database, отдельный сервис, Batch-scoring). Что такое семантический слой в BI-платформе и какова его роль. В чем различие между Self-Service BI и Enterprise Reporting, какие компромиссы они предполагают. Что такое Headless BI и Metrics Store, как они меняют архитектуру аналитических систем.	Интерактивная форма, практикум по решению задач по тематике занятия и коллективное обсуждение решений
Интеграция с унаследованными системами	Почему Legacy-системы невозможно или нецелесообразно переписывать. Какие главные вызовы возникают при интеграции с Legacy (хрупкость, экзотические форматы, отсутствие API). Какие стратегии интеграции существуют и когда применять каждую (Read-Only Replica, файловый обмен, CDC, RPA). Что такое антикоррупционный слой (ACL) и как он изолирует новую систему от Legacy. Как оценить нагрузку на источник при интеграции (формулы для Polling и инкрементальной загрузки). Как применить паттерн «Душитель» (Strangler Fig Pattern) для постепенной миграции от Legacy. Как обеспечить качество данных при интеграции с Legacy-системами.	Интерактивная форма, практикум по решению задач по тематике занятия и коллективное обсуждение решений
Управление качеством данных и мастер-данными	Как правило 1-10-100 объясняет экономическую целесообразность инвестиций в качество данных. Какие измерения качества данных (DQ Dimensions) критичны для аналитических систем. Как рассчитать Data Quality Index (DQI) для оценки качества данных. Что такое Data Quality Firewall и как он работает (жесткая блокировка vs карантин). Что такое мастер-данные и Golden Record в контексте MDM. Какие алгоритмы матчинга применяются для создания Golden Record (детерминистический, вероятностный). В чем различие между архитектурными стилями MDM: Registry, Consolidation, Coexistence, Centralized. Какова роль Data Stewards и Data Governance Council в управлении данными.	Интерактивная форма, практикум по решению задач по тематике занятия и коллективное обсуждение решений

Информационная безопасность в аналитических системах	Как триада CIA (конфиденциальность, целостность, доступность) применяется к аналитическим системам. Как рассчитать риск информационной безопасности по формуле $R = P \times I$. Как зонирование сети (DMZ) защищает аналитическую систему. В чем различие между аутентификацией (SSO, LDAP, Kerberos, SAML, OIDC) и когда применять каждый протокол. Как обеспечить авторизацию на уровне строк (RLS) и колонок (CLS) в аналитической СУБД. Какие методы шифрования применяются для данных at Rest и in Transit. Что такое k-анонимность и как она применяется для обезличивания данных. Как организовать аудит и мониторинг доступа к данным с использованием SIEM.	Интерактивная форма, практикум по решению задач по тематике занятия и коллективное обсуждение решений
---	---	--

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Введение в информационно-аналитические системы	История развития DSS-систем и эволюция от EIS к современным BI-платформам. Сравнительный анализ ведущих вендоров BI-рынка (Microsoft, Tableau, Qlik, SAP). Кейсы провалов внедрения ИАС и их причины. Роль Data Literacy в успешности внедрения аналитических систем.	Работа с учебной литературой. Разбор вопросов по теме занятия.
Архитектурные требования и паттерны проектирования	Закон Густафсона и его применение для оценки масштабируемости систем. Практические кейсы применения CAP-теоремы в распределенных системах (Dynamo, Cassandra, MongoDB). Сравнение Service Mesh и традиционной микросервисной архитектуры. Эволюция от ESB к API Gateway и Service Mesh. Паттерны устойчивости распределенных систем (Circuit Breaker, Bulkhead, Retry).	Работа с учебной литературой. Разбор вопросов по теме занятия.
Проектирование слоя хранения данных	Внутреннее устройство колоночных СУБД (ClickHouse, Vertica): сжатие данных, векторизация запросов. Сравнительный анализ производительности PostgreSQL vs ClickHouse для аналитических нагрузок. Архитектура распределенных файловых систем (HDFS, Ceph, MinIO). Эволюция форматов хранения данных (CSV, Parquet, ORC, Arrow): сравнение по эффективности сжатия и скорости чтения. Устройство индексов в LSM-tree (RocksDB, LevelDB) и B-tree.	Работа с учебной литературой. Разбор вопросов по теме занятия.

Обработка и интеграция данных	Внутреннее устройство Apache Kafka: лог-структурированное хранение, репликация, ISR. Сравнение производительности Spark vs Flink для потоковой обработки. Устройство и оптимизация DAG в Apache Airflow. Практические подходы к реализации идемпотентности в распределенных системах. Change Data Capture: сравнение Debezium, Maxwell, Canal. Форматы сериализации для потоковых данных (Avro, Protobuf, JSON Schema).	Работа с учебной литературой. Разбор вопросов по теме занятия.
Архитектура аналитического слоя	Внутреннее устройство In-Memory движков (Power BI VertiPaq, Tableau Hyper, ClickHouse). Сравнение производительности MOLAP vs ROLAP на практических наборах данных. Методология CRISP-DM: применение на реальных проектах Data Mining. Feature Store как элемент архитектуры ML-систем. Сравнение современных BI-инструментов по критериям: производительность, стоимость владения, Self-Service возможности.	Работа с учебной литературой. Разбор вопросов по теме занятия.
Интеграция с унаследованными системами	Кейсы успешной интеграции с мейнфреймами и AS/400. Практическое применение RPA-инструментов (UiPath, Automation Anywhere) для интеграции с Legacy. Паттерн SAGA для управления распределенными транзакциями при интеграции. API Gateway как единая точка входа для гибридных систем. Практики работы с COBOL-кодом и его интеграции с современными системами.	Работа с учебной литературой. Разбор вопросов по теме занятия.
Управление качеством данных и мастер-данными	Инструменты Data Quality (Great Expectations, Deequ, Soda Core): сравнительный анализ. Практическое применение алгоритмов нечеткого сравнения строк (Levenshtein, Jaro-Winkler, Soundex). Реализация Data Lineage в современных инструментах (Apache Atlas, DataHub, OpenLineage). Кейсы внедрения MDM-систем в крупных организациях. Data Catalog как элемент Data Governance: сравнение Collibra, Alation, Amundsen.	Работа с учебной литературой. Разбор вопросов по теме занятия.

Информационная безопасность в аналитических системах	Практические атаки на аналитические системы и методы защиты. Реализация дифференциальной приватности (Differential Privacy) для аналитических данных. Сравнение решений для маскирования данных (Delphix, IBM InfoSphere, Informatica). Zero Trust архитектура для аналитических систем. Практики безопасной работы с облачными хранилищами данных (AWS, Azure, GCP). Соответствие требованиям регуляторов (GDPR, 152-ФЗ, PCI DSS) при построении ИАС.	Работа с учебной литературой. Разбор вопросов по теме занятия.
---	---	---

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерные задания проектной работы

1. На основе заданных требований (объем данных 500 ТБ, latency < 100 мс, тип нагрузки — потоковая обработка событий IoT) выбрать и обосновать архитектуру хранения (DWH/Data Lake/Lakehouse). Указать технологический стек для каждого слоя. Применить CAP-теорему для обоснования выбора СУБД. Оценить стоимость хранения данных.
2. На основе заданных требований (объем данных 5 ТБ, latency < 1 сек, тип нагрузки — транзакционные операции с элементами аналитики HTAP) выбрать и обосновать архитектуру хранения (DWH/Data Lake/Lakehouse). Указать технологический стек для каждого слоя. Применить CAP-теорему для обоснования выбора СУБД. Оценить стоимость хранения данных.
3. На основе заданных требований (объем данных 5 ПБ, latency < 10 сек, тип нагрузки — ML-обучение на неструктурированных данных) выбрать и обосновать архитектуру хранения (DWH/Data Lake/Lakehouse). Указать технологический стек для каждого слоя. Применить CAP-теорему для обоснования выбора СУБД. Оценить стоимость хранения данных.
4. Спроектировать схему данных для хранилища телеком-оператора (звонки, абоненты, тарифы, базовые станции, время). Построить модель «Снежинка» с обоснованием выбора измерений и мер. Рассчитать примерный объем хранения при 100 млн звонков в сутки. Описать процесс инкрементальной загрузки фактов.
5. Спроектировать схему данных для хранилища медицинской клиники (приемы, пациенты, врачи, диагнозы, время). Построить модель «Звезда» с обоснованием выбора измерений и мер. Рассчитать примерный объем хранения при 1 млн приемов. Описать процесс инкрементальной загрузки фактов с учетом версионирования данных (SCD Type 2).
6. Спроектировать схему данных для хранилища финтех-платформы (транзакции, пользователи, мерчанты, категории операций, время). Построить модель «Конstellляция» (набор связанных звезд) с обоснованием выбора измерений и мер. Рассчитать примерный объем хранения при 50 млн транзакций. Описать процесс инкрементальной загрузки фактов.
7. Для корпоративного хранилища с требованиями (10 млрд строк, отклик < 5 сек, загрузка каждые 5 мин) подобрать: СУБД для хранения, инструмент ETL/ELT, BI-

- платформу. Обосновать выбор по каждому компоненту с учетом NFR. Рассчитать требуемые вычислительные ресурсы.
8. Для корпоративного хранилища с требованиями (100 млн строк, отклик < 0.5 сек, загрузка раз в сутки) подобрать: СУБД для хранения, инструмент ETL/ELT, BI-платформу. Обосновать выбор по каждому компоненту с учетом NFR. Рассчитать требуемые вычислительные ресурсы.
 9. Для корпоративного хранилища с требованиями (50 млрд строк, отклик < 10 сек, загрузка в реальном времени streaming) подобрать: СУБД для хранения, инструмент ETL/ELT, BI-платформу. Обосновать выбор по каждому компоненту с учетом NFR. Рассчитать требуемые вычислительные ресурсы.

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Кафедры информационных технологий Факультета информационных технологий и анализа больших данных.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. *Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине*

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

ПКП-4 Способность разрабатывать и предлагать архитектурные решения для информационно-аналитических систем

1) Умеет выбирать архитектурные решения для разрабатываемых информационных систем

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: основные архитектурные паттерны ИАС (монолит, микросервисы, EDA), принципы проектирования DWH и Data Lake, жизненный цикл ИАС

Уметь: анализировать требования и выбирать оптимальную архитектуру хранения и обработки данных для конкретного бизнес-сценария

Типовые контрольные задания

Задание 1. На основе заданных требований (объем данных, latency, тип нагрузки) выбрать подходящую архитектуру хранения (DWH/Data Lake/Lakehouse) и обосновать выбор.

Задание 2. Спроектировать архитектуру ИАС для интернет-магазина с учетом интеграции с существующей CRM и ERP системами.

Задание 3. Сравнить монолитную и микросервисную архитектуру для заданного сценария и обосновать выбор.

2) Способен выбирать современные технологии, методы и подходы к реализации архитектурных решений

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: современные СУБД (реляционные, NoSQL, NewSQL), ETL/ELT инструменты, технологии Big Data (Hadoop, Spark), BI-платформы, потоковая обработка

Уметь: подбирать технологический стек под заданные нефункциональные требования, проектировать ETL-процессы и конвейеры данных

Типовые контрольные задания

Задание 1. Выбрать стек технологий (СУБД, ETL-инструмент, BI-платформу) для корпоративного хранилища данных с заданными NFR.

Задание 2. Разработать схему ETL-процесса с использованием CDC для репликации данных из OLTP в OLAP.

Задание 3. Спроектировать потоковую обработку данных с использованием Kafka и Flink для системы real-time аналитики.

3) Способен проводить интеграцию архитектурных и проектных решений для информационных систем с существующими решениями

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: стратегии интеграции с Legacy-системами, паттерны интеграции (ACL, CDC, Strangler Fig), принципы Data Quality и MDM

Уметь: разрабатывать планы интеграции новых систем с унаследованными, проектировать слои изоляции и миграции данных

Типовые контрольные задания

Задание 1. Разработать стратегию интеграции новой аналитической платформы с унаследованной системой на базе Oracle Forms без прямого доступа к БД.

Задание 2. Спроектировать антикоррупционный слой (ACL) для изоляции новой ИАС от Legacy-системы с неконсистентными данными.

Задание 3. Разработать план миграции данных из старого DWH в новое Lakehouse-решение с минимальным downtime.

4) Умеет оформлять проектную документацию в соответствии с действующими стандартами

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: стандарты проектной документации ИАС, формат Architecture Decision Records (ADR), структура технического задания, требования к документированию ETL-процессов

Уметь: составлять архитектурные решения, технические задания и эксплуатационную документацию в соответствии с корпоративными стандартами

Типовые контрольные задания

Задание 1. Составить Architecture Decision Record (ADR) для выбора ClickHouse в качестве основной аналитической СУБД.

Задание 2. Разработать техническое задание на создание витрины данных для отдела маркетинга с описанием источников, бизнес-правил и метрик качества.

Задание 3. Оформить документацию на ETL-процесс: диаграмму DAG, описание зависимостей, процедуры восстановления после сбоев.

Примеры практико-ориентированных заданий

Задание 1.

На основе заданных требований (объем данных 50 ТБ, latency < 5 сек, тип нагрузки — аналитические запросы ad-hoc) выбрать и обосновать архитектуру хранения (DWH/Data Lake/Lakehouse). Указать технологический стек для каждого слоя. Применить CAP-теорему для обоснования выбора СУБД. Оценить стоимость хранения данных.

Задание 2.

Спроектировать схему данных для хранилища интернет-магазина (заказы, товары, клиенты, время). Построить модель «Звезда» с обоснованием выбора измерений и мер. Рассчитать примерный объем хранения при 10 млн заказов. Описать процесс инкрементальной загрузки фактов.

Задание 3.

Для корпоративного хранилища с требованиями (1 млрд строк, отклик < 2 сек, загрузка каждые 30 мин) подобрать: СУБД для хранения, инструмент ETL/ELT, BI-платформу. Обосновать выбор по каждому компоненту с учетом NFR. Рассчитать требуемые вычислительные ресурсы.

Задание 4.

Разработать схему репликации данных из PostgreSQL (OLTP) в ClickHouse (OLAP) с

использованием CDC. Выбрать инструмент CDC и обосновать выбор. Описать форматы данных на каждом этапе. Обеспечить идемпотентность загрузки. Нарисовать диаграмму потока данных.

Задание 5.

Спроектировать архитектуру real-time аналитики для системы мониторинга транзакций с использованием Kafka и Flink. Обеспечить семантику Exactly-once. Описать обработку оконных агрегаций (Tumbling window 1 минута). Спроектировать схему восстановления после сбоя.

Задание 6.

Сравнить монолитную и микросервисную архитектуру для корпоративного DWH. Применить закон Амдала для оценки потенциального ускорения при распараллеливании ETL-процессов. Описать критерии выбора архитектуры для заданного сценария.

Задание 7.

Система обработки данных содержит последовательную часть 15% и параллелизуемую 85%. Рассчитать теоретическое ускорение при увеличении числа процессоров с 4 до 16. Определить максимально достижимое ускорение.

Задание 8.

Объем сырых данных — 100 ТБ. Коэффициент сжатия в Parquet — 0.3, в ClickHouse — 0.15. Рассчитать стоимость хранения в облаке (AWS S3: \$0.023/ГБ, ClickHouse Cloud: \$0.04/ГБ) за год с учетом репликации 2х. Определить оптимальную стратегию размещения данных по зонам Bronze-Silver-Gold при соотношении объемов 100%-30%-10%.

Задание 9.

Источник (PostgreSQL) обрабатывает 1000 транзакций/сек. Необходимо реплицировать изменения в DWH. Рассчитать нагрузку при полной выгрузке (Full Load раз в сутки) vs CDC (поток изменений). Оценить задержку репликации при CDC при пропускной способности канала 100 Мбит/с и среднем размере изменения 1 КБ.

Задание 10.

Задача обработки 10 ТБ данных в Spark. Расчетное время выполнения на кластере: 100 исполнителей по 4 ядра, 16 ГБ RAM каждый. Оценить требуемый объем памяти с учетом 3х репликации HDFS. Рассчитать стоимость выполнения в AWS EMR (\$0.27/час за узел r5.xlarge). Оптимизировать конфигурацию для снижения стоимости на 30%

Примерные вопросы для подготовки к экзамену

1. Фундаментальное различие между OLTP и OLAP системами и обоснование их разделения в корпоративной архитектуре.
2. Пирамида DIKW: уровни, механизмы трансформации и практическое применение для проектирования аналитических решений.
3. Типовая слоистая архитектура ИАС: состав слоев, их функции и взаимодействие на примере конкретной системы.
4. Жизненный цикл ИАС: этапы, критические точки рисков и роль архитектора на каждом этапе.
5. Функциональные и нефункциональные требования к ИАС: метрики, методы сбора и формализации.
6. CAP-теорема: суть, компромиссы и практическое применение при выборе технологий хранения данных.
7. Сравнительный анализ архитектурных стилей: монолит, SOA, микросервисы, EDA, Data Mesh — критерии выбора.
8. Закон Амдала: расчет теоретического ускорения и ограничения масштабирования системы.
9. Структура и назначение Architecture Decision Records (ADR) в проектной документации.
10. Строковые vs колоночные СУБД: внутреннее устройство, сценарии применения и сравнение производительности.
11. MPP-системы: принципы работы, представители и критерии выбора для больших данных.
12. Классификация NoSQL баз данных: типы, особенности и задачи, которые они решают.
13. NewSQL и HTAP: концепция, примеры реализации и сценарии применения.
14. Полиглотное хранение (Polyglot Persistence): обоснование и архитектура гетерогенных систем хранения.
15. Архитектура Data Warehouse: слои Staging Area, Core/Integration Layer, Data Marts и их специфика.
16. Модели данных «Звезда» и «Снежинка»: структура, компромиссы и критерии выбора.
17. Data Lake: зональная архитектура Medallion (Bronze, Silver, Gold) и эволюция качества данных.
18. Data Swamp: причины возникновения, признаки и меры профилактики.
19. Data Lakehouse: концепция, технологии (Delta Lake, Apache Iceberg, Apache Hudi) и

преимущества.

20. Математическая модель MapReduce: расчет времени выполнения распределенных вычислений.

21. Apache Spark: RDD, DAG, Lazy Evaluation, DataFrames и оптимизация выполнения запросов.

22. Поточковая обработка данных: оконные функции (Tumbling, Sliding, Session) и их применение.

23. Семантики обработки сообщений: At-most-once, At-least-once, Exactly-once — реализация и гарантии.

24. Apache Kafka: архитектура, лог-структурированное хранение, репликация и гарантии доставки.

25. Паттерны потоковой обработки: Lambda и Карра архитектуры —сравнение и области применения.

26. ETL vs ELT: критерии выбора подхода в зависимости от архитектуры и требований.

27. Стратегии извлечения данных: Full Load, Incremental Load, CDC —сценарии и реализация.

28. Стратегии загрузки данных: Append, Truncate/Insert, Merge/Upsert —особенности и риски.

29. Идемпотентность в ETL-процессах: определение, значение и практические подходы к реализации.

30. Инструменты оркестрации данных: сравнение Apache Airflow, NiFi, Talend, dbt, Informatica.

31. OLAP: концепция многомерного куба, измерения, меры и операции (Slice, Dice, Drill-down, Roll-up, Pivot).

32. Архитектуры OLAP: MOLAP, ROLAP, HOLAP —сравнение и критерии выбора.

33. In-Memory Tabular модели: принципы работы и сценарии применения.

34. Интеграция Machine Learning в архитектуру ИАС: подходы In-Database, отдельный сервис, Batch-scoring.

35. Семантический слой в BI-платформах: назначение, реализация и влияние на архитектуру.

36. Self-Service BI vs Enterprise Reporting: компромиссы, риски и стратегии внедрения.

37. Headless BI и Metrics Store: концепция и трансформация архитектуры аналитических систем.

38. Legacy-системы: определение, причины сохранения и главные вызовы интеграции.

39. Стратегии интеграции с Legacy: прямой доступ, файловый обмен, CDC, API-фасад,

RPA —сравнение.

40. Антикоррупционный слой (ACL): назначение, архитектура и паттерны реализации.
41. Оценка нагрузки на источники при интеграции: формулы для Polling и инкрементальной загрузки.
42. Паттерн «Душитель» (Strangler Fig Pattern): применение для постепенной миграции от Legacy.
43. Принцип GIGO и правило 1-10-100: экономическое обоснование инвестиций в качество данных.
44. Измерения качества данных (DQ Dimensions): полнота, точность, согласованность, своевременность, уникальность, валидность.
45. Data Quality Index (DQI): методология расчета и применение для мониторинга.
46. Data Quality Firewall: архитектура, жесткая блокировка vs карантин и Data Observability.
47. Master Data Management (MDM): мастер-данные, Golden Record и процесс их создания.
48. Алгоритмы матчинга в MDM: детерминистический, вероятностный, расстояние Левенштейна, коэффициент Жаккара.
49. Архитектурные стили MDM: Registry, Consolidation, Coexistence, Centralized — сравнение и выбор.
50. Data Governance: роли Data Stewards, Data Governance Council и организация процессов.
51. Триада информационной безопасности CIA: конфиденциальность, целостность, доступность в контексте ИАС.
52. Оценка рисков информационной безопасности: формула $R = P \times I$ и методология применения.
53. Зонирование сети (DMZ): Public Zone, DMZ, Trusted Zone и размещение компонентов ИАС.
54. Протоколы аутентификации: SSO, LDAP, Kerberos, SAML, OIDC —сравнение и сценарии применения.
55. Авторизация в аналитических системах: RBAC, Row-Level Security (RLS), Column-Level Security (CLS).
56. Шифрование данных: at Rest (TDE, FDE) и in Transit (TLS/SSL) —реализация и управление ключами.
57. Обезличивание данных: статическое маскирование (SDM), динамическое маскирование (DDM), k-анонимность.

- 58. Аудит и мониторинг доступа к данным: логирование, SIEM и обнаружение аномалий.
- 59. Комплексное проектирование архитектуры ИАС: от требований до выбора технологического стека.
- 60. Разработка плана миграции данных с минимальным downtime: стратегии, риски и контрольные точки.

Пример экзаменационного билета

Экзаменационный билет №

- 1. Закон Амдала: расчет теоретического ускорения и ограничения масштабирования системы **(30 баллов)**
- 2. Для корпоративного хранилища с требованиями (1 млрд строк, отклик < 2 сек, загрузка каждые 30 мин) подобрать: СУБД для хранения, инструмент ETL/ELT, BI-платформу. Обосновать выбор по каждому компоненту с учетом NFR. Рассчитать требуемые вычислительные ресурсы **(30 баллов)**

8. Перечень основной и дополнительной литературы, необходимой для освоения дисциплины

Основная литература:

1. Парфенов, Юрий Павлович Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов ; под научной редакцией Н. В. Папуловской. Электрон. дан. Москва : Юрайт, 2025 97 с (Высшее образование) URL: <https://urait.ru/bcode/559502> (дата обращения: 08.04.2025). Режим доступа: Электронно-библиотечная система Юрайт, для авториз. пользователей <https://urait.ru/bcode/559502> ISBN 978-5-534-21173-3 : 439.00. [БИК ID: RU2fURAIT2f559502]
2. Бененсон, М. З. Информационные системы, базы и хранилища данных [Электронный ресурс] / Бененсон М. З., Сорокин С. А. Москва : РТУ МИРЭА, 2024 37 с. Книга из коллекции РТУ МИРЭА - Информатика <https://e.lanbook.com/book/405209>. [БИК ID: RU-LAN-BOOK-405209]

Дополнительная литература:

1. Середенко, Н. Н. Информационно-аналитические системы : Учебное пособие / Н. Н. Середенко, К. В. Чернышева, С. И. Афанасьева Электрон. дан. Москва : КноРус, 2025 200 с. Режим доступа: book.ru Internet access <https://book.ru/book/956557> ISBN 978-5-406-14030-7. [БИК ID: RU\{}bookru\{}bibl\{}956557]
2. Целых, Александр Николаевич Информационно-аналитические системы обеспечения информационной и финансовой безопасности : Учебное пособие / Южный федеральный университет 1 Ростов-на-Дону : Издательство Южного федерального университета (ЮФУ), 2024 123 с. ВО - Бакалавриат <https://znanium.ru/catalog/document?id=462801> ISBN 978-5-9275-4787-6. [БИК ID: RU\{}infra-m\{}znanium\{}bibl\{}2204521]
3. Балдин, К. В. Информационно-аналитические системы в управлении предприятием [Электронный ресурс] : учебное пособие / Балдин К. В., Фархшатова Г. Р. Москва : РТУ МИРЭА, 2023 78 с. Книга из коллекции РТУ МИРЭА - Информатика <https://e.lanbook.com/book/398138> ISBN 978-5-7339-2012-2. [БИК ID: RU-LAN-BOOK-398138]

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
(<http://library.fa.ru/files/elibfa.pdf>)
2. Электронно-библиотечная система BOOK. RU <http://www.book.ru>
3. Электронно-библиотечная система "Университетская библиотека
ОНЛАЙН" <http://biblioclub.ru/>
4. Образовательная платформа "ЮРАЙТ" <https://urait.ru/>
5. Электронно-библиотечная система издательства "Лань" <https://e.lanbook.com/>
6. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
7. Информационно-образовательный портал Финуниверситета: <https://org.fa.ru>

10. Методические указания для обучающихся по освоению дисциплины

Рекомендации по работе с учебным материалом:

1. Осознавайте наличный уровень полученных вами знаний.
2. В ситуации непонимания нужно выявить тот первичный уровень и факторы непонимания, которые стали препятствием понимания последующего.
3. Задавайте сами себе вопросы и пытайтесь ответить на них.

Рекомендации по работе на лекции и с лекционным материалом:

1. Основная задача на лекции – осмысление излагаемого в ней материала. Для этого необходимо слушать лекцию с самого начала, не упуская общих, ориентирующих в материале рассуждений и установок лектора.
2. Ведение записей на лекции важно и полезно для лучшего осмысливания материала, для сохранения информации, с целью ее дальнейшего использования.
3. Для облегчения записи рекомендуется применять сокращения повторяющихся терминов или хорошо известных понятий.

Рекомендации по работе с литературой:

1. Если возникли затруднения при разыскивании материала, по какому-либо конкретному вопросу, следует обратиться к предметному указателю, напечатанному, как правило, в конце каждого литературного источника.
2. Предметный указатель – это алфавитный список основных научных понятий (терминов), содержание которых раскрыто в книге, рядом с термином стоят числа, обозначающие номера страниц, на которых изложен материал, относящийся к данному понятию.

Рекомендации по выполнению проектной работы:

- Анализ исходных требований

Перед началом работы проведите декомпозицию исходных требований:

Параметр	Что определить	На что влияет
----------	----------------	---------------

Объем данных	Текущий и прогнозируемый объем, темпы роста	Выбор архитектуры, шардирование, масштабирование
Latency	Требования к времени отклика (сек/мс)	Выбор между OLAP/OLTP/HTAP, кэширование
Тип нагрузки	Транзакционная/аналитическая/архивная	Архитектура хранения, формат данных
Частота загрузки	Batch/real-time/near real-time	Инструменты интеграции, CDC-механизмы

- Структура решения проектной работы

Каждое задание должно содержать четкие разделы:

1. Обоснование — почему выбрано именно это решение
2. Альтернативы — какие варианты рассматривались и почему отвергнуты
3. Расчеты — количественные оценки с приведением формул
4. Ограничения — когда предложенное решение перестанет работать

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

Комплект лицензионного программного обеспечения:

1. Astra Linux
2. Libre Office
3. Реляционная СУБД
4. Kaspersky

Современные профессиональные базы данных и информационные справочные системы:

1. Информационно-правовая система «Гарант»
2. Информационно-правовая система «Консультант Плюс»
3. Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>

Сертифицированные программные и аппаратные средства защиты информации:

1. не предусмотрены

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

1. **Учебная аудитория** для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная оборудованием и техническими средствами обучения: мебель аудиторная (столы, стулья, доска аудиторная), набор демонстрационного оборудования (проектор, экран)
2. **Компьютерный класс** для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенный оборудованием и техническими средствами обучения: мебель аудиторная (столы, стулья, доска аудиторная), персональные компьютеры, набор демонстрационного оборудования (проектор, экран)