### Вариант 1 1. В чем разница плотного и разреженного индекса? Плотный индекс хранит в своём поле “значение” указатель на запись в сортированном файле данных, в то время как разреженный индекс хранит в себе указатель на блок данных (как правило на первый элемент этих данных). 2. Перечислите недостатки представлений. Ограниченные возможности обновлений. Необходимость контроля архитектуры. Снижение производительности. 3. Дайте определение транзакции. Транзакция - последовательность действий с БД, в которой либо все действия выполняются успешно, либо не выполняется ни одно из них. 4. Какие системы попадают в класс безопасности “D” в соответствии с “оранжевой” книгой? Системы, которые не соответствуют всем остальным классам (C, B, A). 5. Перечислите типы идентификаторов (типы аутентификации). То, что субъект знает. То, что ему принадлежит. То, что является его неотъемлемой характеристикой. 6. В чем разница горизонтального и вертикального фрагментирования в распределенной БД? Вертикальное фрагментирование - пытаемся хранить таблицу не как кортежи со всеми атрибутами, а отдельно храним атрибуты. Горизонтальное (шардорование) - выделение подмножеств строк. 7. Перечислите виды прозрачности в распределенных СУБД. Прозрачность фрагментации, расположения объекта, кол-ва реплик, контроля доступа 8. Перечислите предпосылки к возникновению NoSQL решений. Увеличение объёма хранимых данных. Данные становятся всё более взаимосвязными. Слабоструктурированная информация(невозможность использования каких-то “жёстких” структур для её хранения). Изменения в архитектуре информационных систем(туманные, облачные вычисления). **9. Опишите принцип хранения данных в базах “ключ-значение”.** Не храним отношения, храним множества пар ключ-значение. **10. Опишите принцип построения логической модели базы знаний. Логическая модель** - совокупность из четырех множеств M = <T, P, A, B> T - базовые элементы, алфавит, словарь. P - множество формул, синтаксических правил. A - аксиомы. B - множество правил вывода.

### **Вариант 2** 1. В чем разница первичного индекса и кластерного индекса? **Первичный индекс** - файл упорядочен по первичному ключу, и по нему же построен индекс, так что гарантируется уникальность каждой записи. **Индекс кластеризации** - файл упорядочен по ключевому или неключевому аттрибуту по которому построен индекс, при этом несколько кортежей соответствует одному значению индекса. 2. В чем разница представления замены и материализованного представления? **Материализованное представление** - непосредственно в памяти хранятся дубликаты данных, полученные в результате выполнения некоторого запроса **Представление замены** - хранится только подзапрос, тогда когда нужно выполнить родительский запрос. 3. Перечислите свойства транзакции Атомарность - транзакции неделимы - либо выполняются все действия, либо не выполняются никакие. Самосогласованность - перевод БД из одного согласованного состояние в другое, а между - несогласованные состояния. Изоляция - если запущено несколько конкурирующих транзакций, то любое обновление состояния БД некоторой транзакцией должно быть сокрыто от других транзакций до момента её исполнения Долговечность - если транзакция завершилась корректно, то она будет сохранена, даже если в следующий момент произойдет сбой. 4. Какие системы попадают в класс безопасности “C” в соответствии с “оранжевой” книгой? Системы, обладающие подсистемой идентификации и аутентификации, подсистемы учёта событий, связанных с безопасностью и дискреционным(избирательным) контролем доступа 5. Что такое дискреционный (избирательный) контроль доступа? Для каждой пары субъект-объект должно быть явно задано перечисление комбинаций доступа. 6. В чем суть стратегии раздельного (фрагментного) размещения данных в распределенной БД? БД разделена на некоторые фрагменты, которые хранятся ровно на одном узле 7. В чем различие гомогенной и гетерогенной распределенной БД? **Гомогенные** - на всех узлах одна и та же СУБД. **Гетерогенные** - на разных узлах могут быть разные СУБД 8. Перечислите общие характеристики NoSQL решений Отказ от SQL стандарта. Shemaless - не должно быть четкой структуры данных. Aggregates - представление данных в виде агрегатов. Weak ACID (Atomicity+Consistency+Isolation+Durability) = BASE = Basic Availability + Soft State + Eventual Consistency. BA - любой запрос завершится в течение какого-то времени. S - гибкое состояние хранилища данных. E - согласованность в конечном счёте. 9. Опишите принцип хранения данных в документоориентированных БД Пары Ключ-значение (ключ - число, значение - любая сложная структура) 10. Опишите принцип построения сетевой модели базы данных Формально может быть записана в виде: H = <I, C1, C2, … Cn, G>, где **I** — информационные единицы (множество фактов, значений, слов, типа алфавит кароче). **С1 … Сn** — множество типов связей (красим дуги графа) **G** — множество связей, каждый элемент которого говорит о наличии связи между какими-то элементами из **I**.

### Вариант 3 1. За счет чего использование индекса повышает скорость выполнения запроса? За счёт того, что при необходимости найти какое-то поле/кортеж больше нет необходимости пробегаться по всем значениям, сравнивая, как правило, строки. С помощью индекса мы ставим каждой строке в соответствие числовой индекс и ищем уже по нему. В итоге мы сравниваем не строки, а числа, что гораздо быстрее 2. Перечислите преимущества представлений. Независимость от данных. Повышение защищённости данных. Снижение сложности запросов 3. Перечислите проблемы конкурирующих транзакций. Проблема потерянного обновления — несколько транзакций меняют один и тот же кортеж, в результате “засчитывается” только последний. Проблема грязного чтения — чтение одной транзакцией кортежа, который уже изменён другой транзакцией, которая поздее будет отменена (откатится). Проблема неповторямеого чтения — при повторном чтении ранее считанных данных, транзакция обнаруживает модифицированные данные. Проблема фантомного чтения — первая конкурирующая транзакция добавляет кортежи между двумя актами считывания данных из БД второй конкурирующей транзакцией. В результате чего получаем несогласованность данных. 4. Какие системы попадают в класс безопасности “B” в соответствии с “оранжевой” книгой? Всё из С, + мандатный доступ к данным. Чтение: субъект может читать объекты со своим уровнем или ниже. Запись: субъект может записывать в объекты своего уровня или выше. 5. Что такое мандатный контроль доступа? **Мандатный контроль доступа** — каждому объекту и субъекту ставится в соответствие своя метка секретности, исходя из которой в дальнейшем контролируется доступ. 6. В чем суть стратегии размещения данных с полной репликацией в распределенных БД? На каждом узле есть своя реплика (копия) БД. 7. Перечислите не мнее пяти правил К. Дейта распределенных БД. Локальной автономности. Отсутствие опоры на центральный узел. Непрерывное функционирование. Независимость от расположения. Независимость от фрагментации 8. В чем разница ACID и BASE? В ACID любой запрос должен гарантированно успешно завершаться, тогда как в BASE запрос должен гарантированно выполняться вне зависимости от его успеха. В ACID БД может переходить только из одного согласованного состояния в другое, в промежутке он недоступна. В BASE гарантируется согласованность в конечном счёте, что значит, что в какой-то момент времени БД может находиться в несогласованном состоянии, и при обращении к ней мы можем получить неконсиситентные данные. 9. Опишите принцип хранения данных в колоночных БД. Данные каждого столбца хранятся отдельно от других столбцов 10. Опишите принцип построения фреймовой модели базы данных. Вводится понятие “фрейм”. **Фрейм** — структура для описание стереотипной ситуации, состоящая из характеристик этой ситуации (слотов) и их значений. Значением слота может быть константа или процедура вычисления значения. Множество фреймов — фреймовая модель.

### Вариант 4 1. Что такое индекс, какая у него структура? Индекс — это метод, который позволяет получить структуру данных, предназначенную для оптимизации поиска. 2. Дайте определение представления Это динамически сформированный результат одной или нескольких реляционных операций, выполненных над некоторым отношением из БД для получения новых отношений. 3. Перечислите уровни изоляции транзакций Незавершённое чтение - лишь одна транзакция может изменять данные Завершённое чтение - если одна транзакция изменяет данные, то никакая другая транзакция не сможет читать и изменять данные до момента её завершения Воспроизводимое чтение - если одна транзакция начала считывать данные, то никакая другая транзакция не может их изменить до завершения предыдущей. Сериализуемость - если некоторая транзакция обращается к данным, то никакая другая транзакция не сможет добавить новые строки или изменить существующие, которые уже могли бы быть считаны при выполнении текущей транзакции. 4. Какие системы попадают в класс безопасности “А” в соответствии с “оранжевой” книгой? Проверенный дизайн - по сути B3, но весь процесс создания контролируется 5. Что такое прозрачное шифрование БД (Transparent Database Encryption)? При записи на диск файлы шифруются, при прочтении – дешифруются 6. В чем суть размещения данных с выборочной репликацией в распределенных БД? Репликация - поддержка синхронизированных физических копий некоторго объекта БД Разделяем БД тем или иным способом, и для каждого фрагмента решаем 2 задачи: Сколько копий фрагмента сделать Где их расположить СУБД должна принимать данные решения 7. В чем суть механизма двухфазной транзакции в распределенных БД? В первой фазе транзакции диспетчер транзакций опрашивает каждый ресурс, чтобы определить, следует ли выполнить фиксацию или откат транзакции Во второй фазе транзакции диспетчер транзакций уведомляет каждый ресурс о результате опроса, позволяя ресурсам выполнить необходимые операции очистки 8. Сформулируйте CAP теорему В распределённой информационной системы возможно обеспечить не более двух из перечисленных свойств: Consistency — Сложность обеспечивания консистентности данных (согласованность данных). Availability — Доступность. Partition tolerance — Устойчивость к распределению 9. Опишите принцип хранения данных в графовых БД У узлов есть атрибуты, у связей тоже есть атрибуты 10. Что отличает базу знаний от базы данных? Считается что знания присущи только человеку. Данные - это факт, а знания - обработка этих данных. Данные не врут, а знания врут