Задание 1. СУБД

Кейс

Крупная строительная компания, которая так же занимается проектированием и девелопментом, решила создать правильную архитектуру для работы с данными. Ниже представлены задачи, которые необходимо решить для каждой предметной области.

Какие типы СУБД, на Ваш взгляд, лучше всего подойдут для решения этих задач и почему?

1.1 Бюджетирование проектов с дальнейшим формированием финансовых аналитических отчетов и прогнозирования рисков (СУБД должна гарантировать целостность и четкую структуру данных).

**РСУБД, MySQL, MS SQL, PostgreSQL**

**Отчеты и прогнозы, как правило, имеют четкую структуру, которую мы знаем до начала проектирования. Поэтому реляционная база данных вполне подойдет.**

1.1\* Хеширование стало занимать длительное время, какое API можно использовать для ускорения работы?

1.2 Под каждый девелоперский проект создается отдельный лендинг и все данные по лидам стекаются в CRM к маркетологам и менеджерам по продажам. Какой тип СУБД лучше использовать для лендингов и для CRM? (СУБД должны быть гибкими и быстрыми).

**В обоих случаях - документоориентированная СУБД. Например, MongoDB**

**Набор значений которые будут отражаться на лендингах может различаться от проекта к проекту. И не обязательно будет пересекаться между ними.**

**В части CRM, если речь идет о контактах клиентов, также может быть большая вариативность. Особенно при заполнении контактов на сайте.**

1.2\* Можно ли данную задачу закрыть одной СУБД? И если да, то какой именно СУБД и какой реализацией?

1.3 Отдел контроля качества решил создать базу по корпоративным нормам и правилам, обучающему материалу и так далее, сформированную согласно структуры компании (СУБД должна иметь простую и понятную структуру).

**Иерархическая СУБД**

**Поскольку нужно отразить структуру компании - можно использовать иерархическую СУБД.**

1.3\* Можно ли под эту задачу использовать уже существующую СУБД из задач выше и если да, то как лучше это реализовать?

1.4 Департамент логистики нуждается в решении задач по быстрому формированию маршрутов доставки материалов по объектам и распределению курьеров по маршрутам с доставкой документов (СУБД должна уметь быстро работать со связями).

**Графовая СУБД**

**Из задания следует четкое выделение структуры данных - узлы и связи между ними.**

1.4\* Можно ли к этой СУБД подключить отдел Закупок или для них лучше сформировать свою СУБД в связке с СУБД логистов?

1.5\* Можно ли все перечисленные выше задачи решить используя одну СУБД? Если да, то какой именно?

Приведите ответ в свободной форме.

**Можно в РСУБД, но для каждой задачи возникнут свои сложности в реализации и выполнении запросов…  
У меня был опыт, когда набор полей заменялся атрибутом типа json (контакты пользователя, ответ API-службы об описании публичных компаний)**

**Также была задача маршрутизации, где мы хранили сотни тысяч пар “откуда-куда” с параметрами времени, цены, расстояния. Выбор СУБД был продиктован корпоративным стандартом.**

Задание 2. Транзакции.

2.1 Пользователь пополняет баланс счета телефона, распишите пошагово какие действия должны произойти для того, чтобы транзакция завершилась успешно (ориентируйтесь на 6 действий).

**Получен запрос на увеличение баланса**

**Сохранить запрос в БД (статус - в работе)**

**Запросить текущий баланс**

**Рассчитать новый баланс**

**Записать баланс в таблицу**

**Запросить текущий баланс (после обновления)**

**Обновить статус запроса (статус - выполнен)**

2.1\* Какие действия должны произойти, если пополнение счета телефона происходило бы через автоплатеж?

Приведите ответ в свободной форме.

Задание 3. SQL vs NoSQL

3.1 Напишите 5 преимуществ SQL систем по отношению к NoSQL.

**Четкая структура данных**

**Управление доступом и защита данных**

**Поддержка ACID-транзакций**

**Создание представление и процедур**

**Универсальность и распространенность SQL**

3.1\* Какие, на Ваш взгляд, преимущества у NewSQL систем перед SQL и NoSQL.

Приведите ответ в свободной форме.

Задание 4. Кластеры

Необходимо производить большое количество вычислений при работе с огромным количеством данных, под эту задачу выделено 1000 машин.

На основе какого критерия будете выбирать тип СУБД и какая модель распределенных вычислений здесь справится лучше всего и почему?

Приведите ответ в свободной форме.

**Честно говоря, первым в голову пришла идея использовать колоночную СУБД. Однако, в этом случае, если не ошибаюсь, 999 машин будут стоять без дела.**

**Метод вычислений - MapReduce, т.к. предполагаем что данные легко разделяются на части для вычислений и также легко собираются в итоговый ответ.**

**Если массив данных статичен и расчеты проводятся не часто, то можно использовать PostgreSQL для хранения данных, а расчет производить через, например hadoop.**

**С другой стороны, если частота обновления данных высокая, то возможно подойдет MongoDB или другая NoSQL СУБД. Если не ошибаюсь, в MongoDB нативно реализован метод для использования MapReduce**