**1) Search Engines. Основные принципы работы и общие черты. Наиболее распространённые Search Engines. Взаимодействие краулеров и Search Engines. Могут ли Search Engines работать без краулеров? Роль Search Engines в корпоративных системах масштаба предприятия.**

Search Engines - это системы управления базами данных NoSQL, предназначенные для поиска контента данных. В дополнение к общей оптимизации для этого типа приложений, специализация обычно заключается в предложении следующих функций:

* Поддержка сложных поисковых выражений
* Полный текстовый поиск
* Стемминг (сокращение склоняемых слов до их основы)
* Ранжирование и группировка результатов поиска
* Распределенный поиск для высокой масштабируемости

Наиболее распространённые приложения:

* Elasticsearch
* Splunk
* Apache Solr
* OpenSearch
* Sphinx

Краулеры собирают данные из веб-сайтов, после чего передают в Search Engines. Search Engines можно использовать без краулеров, но из-за этого объём данных для обработки будет ограничен тем, что задано разработчиком.

Роль Search Engines в корпоративных системах масштаба предприятия исходит из его функций. Благодаря Search Engines можно значительно сократили количество серверов, необходимых для поддержки поисковой инфраструктуры. Search Engines может помочь с обработка структурированных и неструктурированных данных. Так же помогает повысить производительность работников при работе с Big Data.

**2) Сильные и слабые стороны SAP HANA Direct Extractor Connection (DXC). Типичные варианты его применения.**

SAP HANA Direct Extractor Connection (DXC) — это средство предоставления готовых базовых моделей данных для SAP HANA, которые основаны на сущностях SAP Business Suite. DXC также является методом сбора данных для SAP HANA. Обоснованием DXC является простота и низкая совокупная стоимость владения (TCO) для сбора данных для SAP HANA с использованием существующих поставляемых моделей данных. DXC – прямое подключение экстрактора. Большой стимул для использования DXC возникает, когда вы хотите использовать сложные экстракторы из систем SAP для загрузки данных в SAP HANA. Особенно когда экстракторы требуют активации, службы данных могут не иметь возможности их использовать.

Проекты клиентов могут сталкиваться со значительными сложностями в моделировании сущностей в системах SAP Business Suite. Во многих случаях данные из различных областей в системах SAP Business Suite требуют прикладной логики для надлежащего представления состояния бизнес-документов.

Основной бизнес-кейс для этого – когда вы хотите получить некоторые данные из системы SAP Business Suite, например, финансовые данные. Вы можете реплицировать несколько финансовых таблиц из исходных систем SAP с помощью SLT, но в итоге у вас будет более 10 таблиц внутри SAP HANA. Затем вам нужно создать сложную модель данных, чтобы объединить все таблицы и манипулировать данными. Используя экстракторы через DXC, вы получаете финансовые данные, доставленные вам в SAP HANA в том виде, в котором они готовы к немедленному использованию, и вам не нужно сначала создавать сложную модель данных.

**3) Основные игроки на рынке облачных платформ, их сильные и слабые стороны?**

1) Amazon Web Service (AWS)

Сильные стороны: Самая распространённая платформа; самая большое количество сервисов.

Слабые стороны: Высокие цены; сложность для новичков.

2) Google Cloud Platform (GCP)

Сильные стороны: Обслуживание и использование обходится дешевле, чем другие платформы; кроссплатформенность.

Слабые стороны: Меньше сервисов, чем у AWS; меньше регионов (влияет на скорость доступа к данным).

3) SAP Cloud Platform

Сильные стороны: Глубокая интеграция с SAP-экосистемой; мощные аналитические возможности.

Слабые стороны: Узкая специализация; необходимость лицензирования.

4) Microsoft Azure

Сильные стороны: Интеграция с продуктами Microsoft; развитые enterprise-решения.

Слабые стороны: Сложности с не-Windows окружениями; менее гибкий, чем AWS.