

Современные подходы к управлению данными

Код: EAS-011

Длительность: 16 ч.

Описание:

При проектировании приложений одним из важных решений является выбор способа хранения данных. Несколько десятилетий реляционные СУБД были первым и единственным вариантом, проекты отличались только степенью нормализации, местом размещения бизнес логики и т.д. Последние же десять-пятнадцать лет бурно расцвели альтернативные системы – от объектно-ориентированных и документно-ориентированных СУБД до распределенных файловых систем и систем обработки потоков данных. В курсе рассматривается спектр современных решений, позволяющих долгосрочно надежно хранить данные, причины возникновения решений разных классов, их достоинства, недостатки и предпочтительные способы использования.

Цели:

После завершения обучения слушатели смогут:

- понимать, какие характеристики данных и запросов необходимо учитывать на этапе анализа требований и выбора систем управления данными;
- знать возможности и ограничения современных реляционных и нереляционных систем управления данными;
- уметь анализировать требования при выборе систем управления данными.

Разбираемые темы:

1. Эволюция подходов к хранению данных (теория – 2 часа).

- Базы данных, хранилища данных, машины баз данных, массово-параллельные архитектуры, гиперконвергенция.

2. Реляционная модель (теория – 2 часа).

- Какие проблемы решает, за счет чего.
- Репликация, шардинг, распределенные транзакции.

3. Минимальная модель "Ключ-значение" (теория – 1 час, практика – 1 час).

- Варианты структуры ключа, варианты структуры значения, программные интерфейсы.
- Эффективность применения нереляционных баз данных: необходимые и достаточные условия [Cassandra, HBase].

4. Документно-ориентированная модель [MongoDB] (теория – 0,5 часа,

практика – 0,5 часа).

5. Распределенные файловые системы вместо моделей данных: архитектура кластера [HDFS] (теория – 1 час, практика – 1 час).

6. SQL поверх распределенных файловых систем (теория – 1 час, практика – 2 часа).

- Варианты архитектур, форматы файлов, ограничения, транзакции [Hive, Spark, Spark SQL, Parquet, ORC].

7. Распределенные системы хранения данных в оперативной памяти [Hazelcast, Ignite, Tarantool] (теория – 1 час).

8. Распределенные OLAP-системы [Clickhouse, Druid] (теория – 1 час).

9. Обработка потоков данных [Spark Streaming] (теория – 1 час).

10. Самонастраиваемые и автономные базы данных (теория – 1 час).

Целевая аудитория:

Архитекторы, разработчики приложений, аналитики, администраторы баз данных.