

Корпоративное хранилище данных (Enterprise Data Warehouse)

Синонимы: Хранилище данных предприятия, Corporate Data Warehouse, CDW, EDW

Разделы: [Источники данных](#)

Корпоративным называют хранилище данных, которое консолидирует информацию и поддерживает бизнес-аналитику по всей компании, а не отдельных ее подразделений и направлений деятельности.

Основное различие между корпоративным хранилищем данных (КХД) и обычным хранилищем данных (ХД) заключается в масштабе, а также в архитектурных и функциональных особенностях. ХД создается и используется на уровне отдельного подразделения компании. Оно хранит данные, связанные с конкретным направлением деятельности, за которое отвечает это подразделение, и поддерживает аналитику исключительно для его нужд.

КХД обеспечивает сбор, хранение и анализ всех данных компании, по всем направлениям деятельности, подразделениям и бизнес-процессам. Обычное ХД масштаба подразделения или направления деятельности, отражает только их специфику. Напротив, КХД должно быть единым репозиторием для всей организации и централизованно обеспечивать все потребности в аналитической обработке информации.

Масштаб КХД предъявляет повышенные требования к его структуре и семантической модели, которые значительно сложнее. Кроме того, КХД ориентировано на поддержку управленческих решений стратегического уровня предприятия, в отличие от локальных ХД подразделений, которые используются для тактических задач.

Это, в свою очередь предъявляет повышенные требования к уровню качества и безопасности данных, а также надежности процедур их сбора и интеграции. Поскольку КХД хранит и поддерживает аналитику большего объема данных, чем локальные ХД, оно должно обеспечивать и более высокую скорость работы с ними.

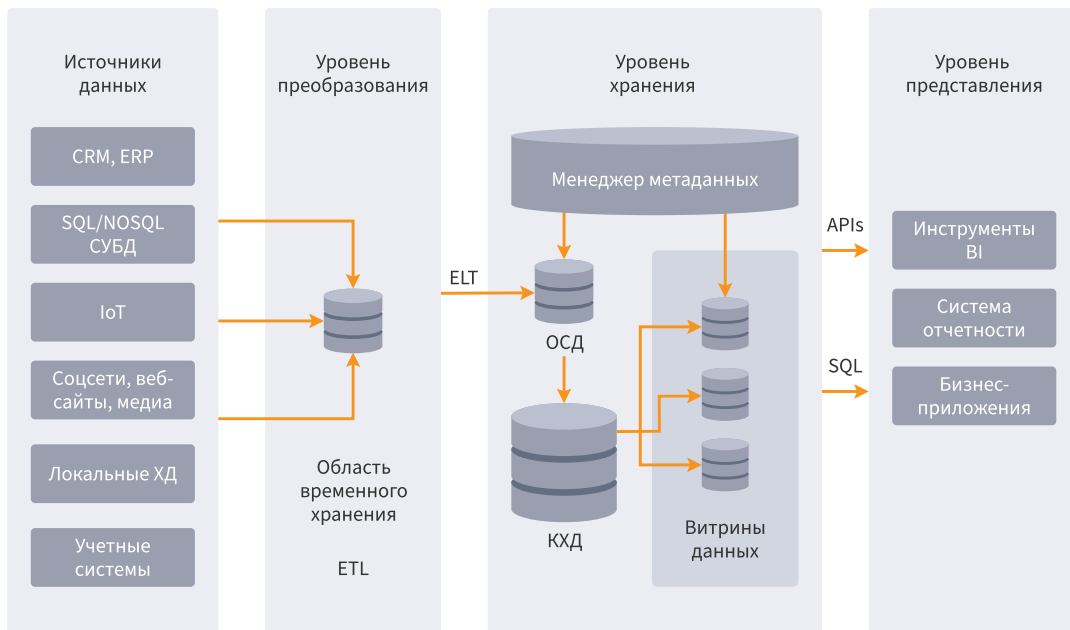
Еще одним отличием КХД от обычного ХД является большой интервал накопленных исторических данных. Это связано с тем, что в корпоративном хранилище накапливаются данные за весь период деятельности компании, в то время как в обычном хранилище данные собираются только за время существования конкретного подразделения или направления деятельности, которые могут создаваться и ликвидироваться.

Параметр	Обычное ХД	Корпоративное ХД
----------	------------	------------------

Параметр	Обычное ХД	Корпоративное ХД
Масштаб	Локальный (подразделение, направление деятельности)	Глобальный (компания, организация)
Сложность	Обычная	Повышенная
Уровень поддерживаемых решений	Тактический	Стратегический
Требования к качеству данных	Стандартные	Повышенные
Требования к безопасности данных	Стандартные	Повышенные
Требования к надежности	Стандартные	Повышенные
Интервал исторических данных	Время существования подразделения (направления)	Время существования компании

И, наконец, что самое важное, корпоративное хранилище данных (КХД) имеет более сложную и, соответственно, дорогостоящую поддерживающую инфраструктуру, некоторые элементы которой могут быть избыточными для обычных хранилищ данных (например, оперативный склад данных). Более того, обычные локальные хранилища могут быть частью инфраструктуры централизованного корпоративного хранилища и даже служить источниками данных для него.

На рисунке представлена типичная инфраструктура КХД. Она содержит следующие элементы, объединенные в уровни.



- 1. Источники данных.** На этом уровне находятся как сырые данные, так и данные, прошедшие некоторую предобработку. К таким источникам могут относиться производственные и управленческие системы (CRM, ERP), системы управления базами данных (СУБД), устройства интернета вещей, социальные сети, веб-сайты, учетные системы и локальные хранилища данных подразделений компании.
- 2. Уровень преобразования.** Содержит область временного хранения, где сырые данные из источников преобразуются перед загрузкой в КХД: производится их очистка, согласование, интеграция и преобразование в единый формат, соответствующий модели данных хранилища. При этом могут быть реализованы две стратегии: ETL (extract, transform, load) и ELT (extract, load, transform). Различие между ними заключается в порядке событий. В ETL преобразование происходит в области временного хранения (виртуальных таблицах), перед тем, как данные попадут в КХД. ELT — это более современный подход, в нем все задачи по преобразованию выполняются внутри хранилища, а промежуточный этап отсутствует.
- 3. Уровень хранения.** Содержит собственно инфраструктуру хранения данных и метаданных. При использовании ELT данные здесь могут подвергаться преобразованиям. В случае ETL все необходимые преобразования данных уже выполнены и они загружаются в свои целевые структуры хранения. На этом уровне в специальной репозитории хранятся технические и бизнес метаданные, управление которыми производится с помощью специальной системы (менеджера метаданных). Здесь же могут содержаться оперативный склад данных и витрины данных (при наличии).
- 4. Уровень представления.** Содержит средства, обеспечивающие конечному пользователю доступ к данным и инструменты их визуализации, анализа и формирования отчетности. Этот уровень также называют интерфейсом бизнес-аналитики (BI interface).

Таким образом, различие между обычным и корпоративным хранилищем данных заключается не только в масштабе, но и в более широком архитектурном разнообразии и функциональности. Одним из ключевых принципов, заложенных в функциональность КХД, является его **конечность** — оно служит конечной точкой для всех потоков данных в организации. В то время как локальные хранилища данных подразделений могут быть источниками для более централизованных систем.

Несмотря на то, что развертывание и поддержка КХД приводит к дополнительным затратам, их использование обеспечивает ряд важных преимуществ:

- 1. Повышается согласованность данных.** Без КХД каждое подразделение отвечает за свою аналитику. Данные могут быть корректными, но несовместимыми с данными других отделов, что затрудняет объединение аналитики для использования на уровне всей компании. С КХД данные и результаты их анализа становятся согласованными, что позволяет увидеть общую картину бизнеса. Аналитика из разных отделов может быть объединена по всей компании, что улучшает координацию и повышает качество управленческих решений.
- 2. Улучшенное обслуживание клиентов.** КХД часто является единственным и наиболее точным источником для описания каждого клиента и всей истории его взаимодействия с компанией, включая заказы, историю покупок и поведенческие модели. Анализ истории клиентов помогает улучшить качество обслуживания на всех этапах, что способствует повышению их удовлетворенности и лояльности. Это, в свою очередь, помогает увеличить доходы и прибыльность на одного клиента.
- 3. Точные данные по всем локациям.** Если бизнес компании географически распределен, КХД не только агрегирует данные в реальном времени из различных мест, но и стандартизирует их в соответствии с установленными бизнес-правилами, обеспечивая доступ к единому источнику правды.

Таким образом, несмотря на более сложную и дорогостоящую организацию, КХД является предпочтительной информационной архитектурой для крупных компаний, имеющих большое количество подразделений и направлений деятельности.