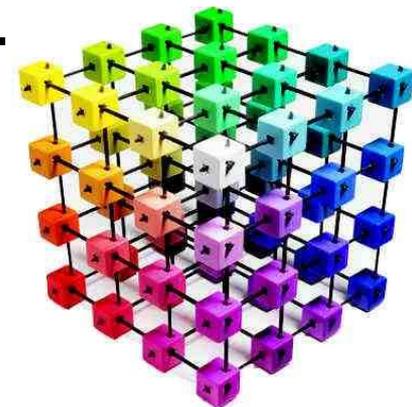


Хранилища данных



Хранилища данных – это процесс сбора, отсеивания и предварительной обработки данных с целью представления результирующей информации пользователям для статистического анализа и аналитических отчетов.

Ральф Кинболл (автор концепции хранилищ данных) описывал хранилища данных как «место, где люди могут получить доступ к своим данным».



Основные требования к хранилищам данных

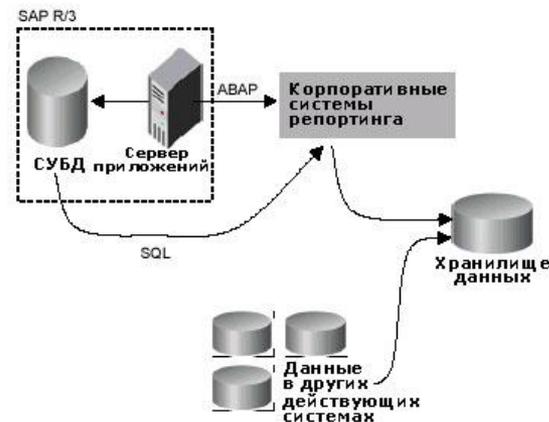
- поддержка высокой скорости данных из хранилища;
- поддержка внутренней непротиворечивости данных;
- возможность получения и сравнения данных;
- наличие удобных утилит просмотра данных хранилища;
- полнота и достоверность хранимых данных;
- поддержка качественного процесса пополнения данных.

Всем перечисленным требованиям удовлетворять зачастую не удается, поэтому **для реализации хранилищ данных используют несколько продуктов:**

1- средства хранения данных;

2–средства извлечения и просмотра данных;

3–средства пополнения хранилищ данных;

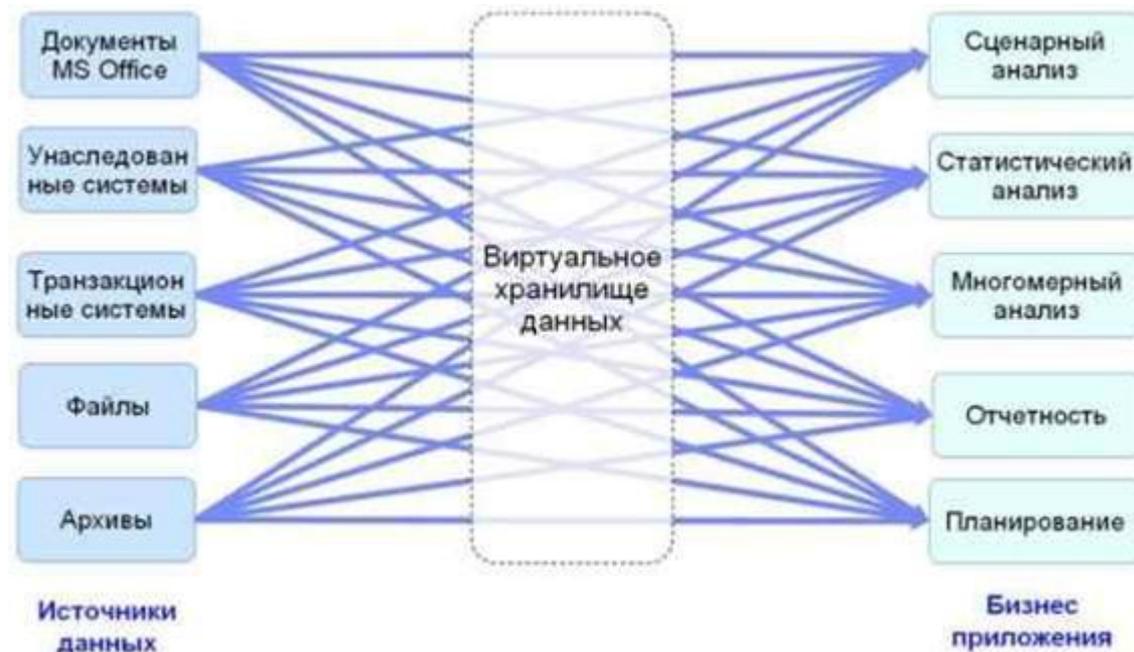


Типичное хранилище данных как правило отличается от реляционной базы данных:

1) Обычная база данных предназначена для того, чтобы помочь пользователям выполнять повседневную работу, тогда как хранилища данных предназначены для принятия решений;

2) Обычная база данных подвержена постоянным изменениям в процессе работы пользователей, а хранилища данных относительно стабильно; данные в нем обновляются согласно расписанию (например, ежечасно, ежедневно, ежемесячно), в идеале, процесс пополнения данными за определенный период времени без изменения прежней информации находящейся уже в хранилище.

з) Обычная база данных чаще всего является источником данных попадающих в хранилище, кроме того хранилище может пополняться за счет внешних источников (например, сжатия данных).



Принципы построения ХД

Информация, которая загружается в хранилище, должна интегрироваться в целостную структуру, отвечающую целям анализа данных.

При этом минимизируются несоответствия между данными из различных оперативных систем, в хранилище именуются и выражаются единым образом.

Данные интегрированы на множестве уровней: на уровне ключа, атрибута, на описательном, структурном уровне и так далее.

Общие данные и общая обработка данных консолидированы и являются единообразным для всех данных, которые подобны или схожи в хранилище данных.

При этом информация структурируется по разным уровням детализации:

- высокая степень суммаризации;
- низкая степень суммаризации;
- текущая детальная информация.

Хранилища можно рассматривать как набор моментальных снимков состояния данных: можно восстановить картинку на любой момент времени.

Атрибут времени всегда явно присутствует в структурах данных хранилища.

Попав однажды в хранилище, данные уже никогда не изменяются, а только пополняются новыми данными из оперативных систем, где данные постоянно меняются.

Новые данные по мере поступления обобщаются с уже накопленной информацией в хранилище данных.

Основные компоненты хранилища данных

Использование технологии хранилищ данных предполагает наличие в системе **следующих компонентов:**

- оперативных источников данных;
- средств переноса и трансформации данных;
- метаданных – включают каталог хранилища и правила преобразования данных при загрузке их из оперативных баз данных;
- реляционного хранилища;
- OLAP-хранилища;
- средств доступа и анализа данных.

Оперативные данные собираются из различных источников.

Поступившие оперативные данные очищаются, интегрируются и складываются в реляционные хранилище.

Они уже доступны для анализа при помощи средств построения отчетов.

Затем данные (полностью или частично) подготавливаются с использованием средств переноса и трансформации данных для OLAP-анализа, который реализуется применением средств доступа и анализа данных.

При этом они могут быть загружены в специальную базу данных OLAP или оставаться в реляционном хранилище.

Важнейшим элементом хранилища являются **метаданные**, т.е. данные о структуре, размещении, трансформации данных, которые используются любыми процессами хранилища.

Метаданные могут быть востребованы для различных целей, например: извлечения и загрузки данных; обслуживании хранилища и запросов.

Метаданные для различных процессов могут иметь различную структуру, т.е. для одного и того же элемента данных может существовать несколько вариантов метаданных.

Итак, хранилища данных являются структурированными.

Они содержат базовые данные, которые образуют единый источник для обработки данных во всех системах поддержки принятия решений.

Элементарные данные, присутствующие в хранилище, могут быть представлены в различной форме.



Хранилища данных исключительно велики, поскольку в них содержатся интегрированные и детализированные данные.

Эти характеристики являются общими для всех хранилищ данных. Но, несмотря на то что хранилища обладают общими свойствами, разные типы хранилищ имеют свои индивидуальные особенности.