

Корпоративные хранилища данных. Интеграция систем. Проектная документация.

Проектный опыт. Аналитика и статьи. Примеры и шаблоны документов. ГОСТы.

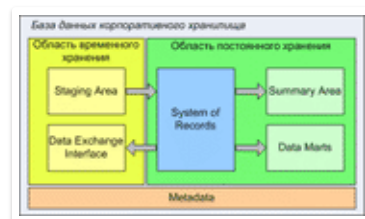
[О проекте](#) | [Добавить статью](#) | [Рассылка](#)

Тематика

[Data Warehouse](#)[Модели данных](#)[DSS & BI](#)[Управление данными](#)[ETL](#)[DW Appliance](#)[Интеграция](#)[Управление проектом](#)[Oracle Tech](#)[Cloud Tech](#)[Видео материал](#)

[Главная](#) » [Хранилище данных](#) » Архитектура корпоративного хранилища данных

Архитектура корпоративного хранилища данных



Основными компонентами корпоративного хранилища данных являются:

Модель данных;

База данных;

ETL-приложение;

BI-приложение.

Архитектура области хранения данных базы данных корпоративного хранилища, как правило, состоит из следующих областей:

Популярные:

- **область временного хранения данных (Staging Area)** – предназначена для временного хранения данных, извлеченных из систем-источников; является промежуточным слоем между операционными системами компании и хранилищем данных;

ГОСТы

[ГОСТ 19](#)[ГОСТ 34](#)[Другие ГОСТы](#)[Примеры по ГОСТ](#)

Главная

Хранилище данных

Интеграция систем

Примеры документов

ГОСТы

Автоматизированные системы. Термины и определения

За всё время:

ГОСТ 34.602-89

Техническое задание на создание автоматизированной системы (пример)

РД 50-34.698-90

Пояснительная записка к эскизному проекту на создание автоматизированной системы (пример эскизного проекта)

Этапы разработки и внедрения

информационно-аналитической системы

РД 50-34.698-90

Пояснительная записка к

техническому проекту

• область постоянного хранения данных, которая включает:

— **детальные данные (System of records)** – область хранения детальных данных, приведенных к структуре **модели данных** корпоративного хранилища, прошедших очистку и обогащение;

— **агрегаты (Summary area)** – сгруппированные по времени (чаще просуммированные) детальные данные;

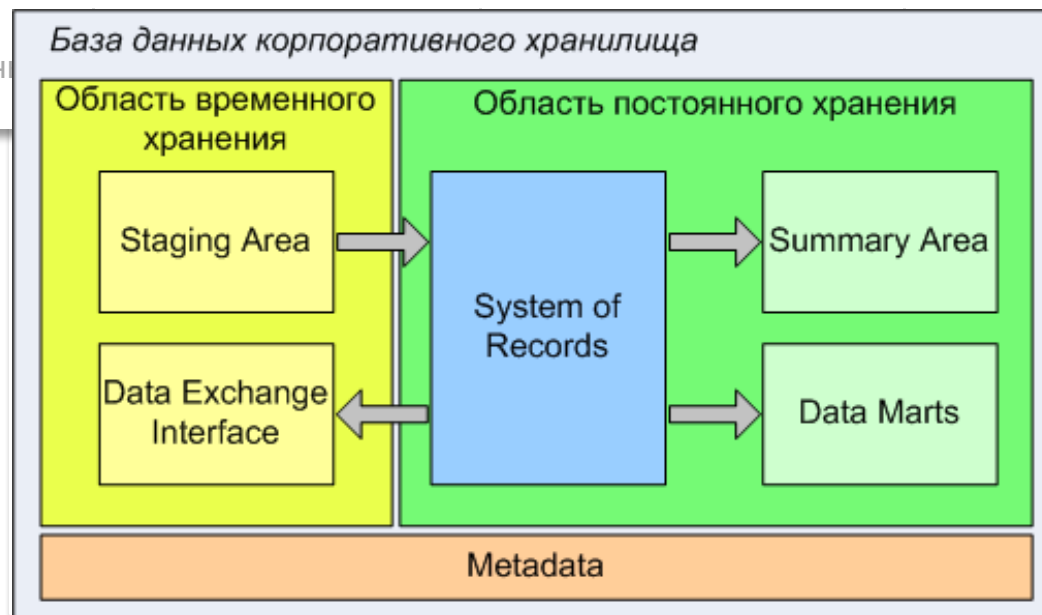
— **витрины данных (Data Marts)** – тематические наборы данных, хранящиеся в виде пригодном для их анализа (например, схема «звезда»); ориентированны на поддержку конкретных бизнес-процессов, приложений, подразделений компании, бизнес-целей;

интерфейсы обмена данными с другими системами (*Data Exchange Interface* или *Feedback Area*) – таблицы БД, в которых хранятся подготовленные для передачи в другие информационные системы компании данные из области постоянного хранения данных;

метаданные (Metadata) – являются важной частью архитектуры хранилища данных. Метаданные - это данные, описывающие правила, по которым «живет» хранилище. Например, с точки зрения базы данных хранилища, метаданными является описание структур таблиц, взаимосвязей между ними, правил секционирования, описание витрин данных и т.п. С точки зрения ETL, метаданными являются описания правил извлечения и преобразования данных, периодичность выполнения ETL-процессов и т.п.

Обычно приведенные выше области хранения данных реализуются в виде отдельных схем одной или нескольких баз данных.

Ниже представлена общая схема организации областей хранения данных.



Область временного хранения данных (Staging Area)

Область временного хранения данных является **промежуточным слоем между источниками данных и областью постоянного хранения**. В данной области сохраняются извлеченные из операционных систем-источников (СУБД, csv, dbf, xml файлов, web-сервисов и т.д.) данные, производится их очистка, трансформация, обогащение, подготовка к загрузке в область постоянного хранения. Зачастую очередной цикл обработки и загрузки данных в хранилище не может быть начат пока не будут извлечены все необходимые данные из различных систем-источников, а в силу ряда причин (географической распределенности, разных циклов функционирования систем и т.п.) данные в источниках могут быть доступны в разные моменты по времени. **Область временного хранения служит для сбора всех необходимых данных перед началом трансформации.**

Одной из наиболее важных задач при построении хранилища данных является определение соответствия (mapping) сущностей систем-источников данных и сущностей модели хранилища данных.

источников данных и сущностей модели хранилища данных. Обычно подобное соответствие представляет собой отношение десятков (а иногда и сотен) таблиц систем-источников к десяткам таблиц области постоянного хранения данных. Правильно организованная область временного хранения данных позволяет значительно упростить организацию процессов загрузки данных из области временного в область постоянного хранения.

Ниже представлены основные принципы формирования области временного хранения.

1. В области временного хранения данных должно быть относительно небольшое количество сущностей - до 20, в которые сохраняются все необходимые данные, извлеченные из систем-источников.
2. Основой для проектирования состава сущностей области временного хранения должны являться предметные области (Subject Area) модели данных.
3. При извлечении данных из систем-источников сами данные и их типы не должны принципиально изменяться.

Детальные данные (System of records)

Данная область является основной хранилища данных. В этой области хранятся преобразованные и очищенные детальные данные, полученные из систем-источников, и основные классификаторы. Хорошо спроектированная модель данной области является залогом дальнейшего успешного функционирования базы данных и BI-приложения.

Данная область содержит следующие типы сущностей:

- › справочники и классификаторы;
- › сущности, содержащие фактические значения;
- › сущности, описывающие связи.

Справочники и классификаторы определяют:

участников основных бизнес-процессов – клиентов, поставщиков, филиалы, услуги, продукты и т.п.

базовые справочники – дата и время, валюта, страны и т.п.

прочие справочники – отражающие потребности бизнеса в необходимой аналитике данных, определяющие в разрезе каких справочников необходимо анализировать фактические данные.

Сущности, содержащие фактические значения – транзакционные данные из систем источников. Например, информация о совершенных телефонных звонках, выставленных счетах, проводках, проданных товарах и т.п.

Сущности, содержащие связи определяют взаимосвязи между остальными сущностями. Например, Клиент-Услуга.

Область детальных данных не содержит никаких агрегатов. Только детальные, очищенные и структурированные в соответствии с моделью данные.

Агрегаты (Summary area)

В данной области хранятся агрегаты данных, которые в основном строятся для сущностей, описывающих участников бизнес-процессов. Например, агрегаты строятся для данных по продажам товаров, оказанию услуг, клиентам и т.п. Данные агрегируются в разрезе времени – от часа, дня к неделе, месяцу. Для каждого агрегата может быть определена своя степень агрегации данных.

информации, нацеленными на поддержку конкретных бизнес-функций, конкретных подразделений компании. На уровне базы данных витрины обычно реализуются по схеме «звезда» или «снежинка» и содержат данные из области детальных данных (*System of records*). Также могут быть реализованы в виде многомерного OLAP-куба. Витрины данных являются основой, обеспечивающей возможность проведения многомерного анализа (OLAP) данных.

Ниже представлены основные принципы проектирования витрин данных.

1. Витрины данных ориентированы на бизнес и при их проектировании необходимо учесть все измерения, показатели и иерархии, необходимые пользователям.
2. При проектировании витрин данных необходимо учитывать особенности BI-приложения, используемого на проекте. Например, в Oracle Discoverer нет возможности создавать несбалансированные иерархии и это нужно учитывать.

Интерфейсы обмена данными (Data Exchange Interface)

Хранилище обычно строится с целью консолидации в нем данных компании, и поэтому оно зачастую является источником данных для других информационных систем. Для обмена данными создаются интерфейсы обмена (обычно это таблицы базы данных), в которых и хранятся специально подготовленные (возможно, перед передачей данных потребуется их предобработка) для передачи данные. Интерфейсы обмена желательно создавать как можно более универсальными.

описанными метаданными является более простой задачей, нежели при отсутствии таковых. Метаданные хранилища включают:

- информацию о данных, их бизнес-описание и структуру хранения;
- описание структур источников данных, их доступности;
- информацию о структуре процессов ETL, периодичности их выполнения, применяемых правил очистки и преобразования данных;
- описание бизнес-представления данных, помогающее пользователю работать с BI-приложением;
- информацию о настройках безопасности, правил аутентификации и назначенных прав доступа;
- статистику утилизации ресурсов, обращений к данным и др., которая помогает администратору оптимизировать работу базы данных хранилища.

Обычно управление метаданными осуществляется отдельными инструментами для каждого из компонентов хранилища. Например, для базы данных Oracle, метаданные которой хранятся в системных таблицах и настроечных файлах, это будет Oracle Enterprise Manager.

Ковтун М.В. Апрель 2011

Источники:

<http://www.globaldataconsulting.net/articles/theory/inside-data-warehouse-model>

<http://www.linktionary.com/d/datawarehouse.html>

[Главная](#)[Хранилище данных](#)[Интеграция систем](#)[Примеры документов](#)[ГОСТы](#)[Хранилище данных, Data Warehouse](#)

Предыдущая:

[Типы моделей данных
корпоративного хранилища
данных](#)

Следующая:

[Oracle Exadata](#)[Хранилище данных](#)[Интеграция систем](#)[Примеры документов](#)[Госстандарты](#)Допускается перепечатка материала с обязательным указанием [источника](#).Адрес для связи admin@prj-exp.ru.

© 2010-2021 Project Experience.

