

ОРГАНИЗАЦИЯ МЕТАДАНЫХ В ХРАНИЛИЩЕ ДАННЫХ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Ильючик В.А.

Николаев А.Ю. магистр техники и технологии, аспирант

Популярность хранилищ данных в последние годы существенно возросла. Конкурентоспособные организации находятся на пути построения хранилищ данных, либо расширения, перепроектирования и усовершенствования уже имеющихся. Метаданные считаются ключевым фактором успеха в проектах по внедрению хранилищ. Они содержат всю информацию, необходимую для извлечения, преобразования и загрузки данных из исходных систем, а также для последующего использования и интерпретации содержимого [1].

Хранение и обработка данных в компьютерных системах, обмен данными между ними и доступ к ним пользователей невозможны без явно представленных описаний свойств этих данных. Эти описания необходимы программным средствам, выполняющим указанные функции, а также пользователям для формулировки запросов, анализа данных и интерпретации их содержания. Описания такого рода называются метаданными и являются особым видом информационных ресурсов.

В случае информационных систем метаданные особенно упрощают управление, создание запросов, полноценное использование и понимание данных. Генерирование, хранение и управление метаданными помогают в поддержке использования огромных объемов информации, доступных в наши дни в любой электронной форме. Интерес к метаданным существенно возрос в связи с интенсивным развитием сетевых технологий, которые предполагают формирование и существование многочисленных сообществ, в которых взаимодействуют друг с другом люди с различными уровнями знаний и интересов, а границы между традиционными ролями, например, издатель – автор – библиотека, довольно размыты.

Типы метаданных:

1. Метаданные исходной системы:

- спецификации источников данных, таких как репозитории;
- описательная информация, например, частота обновления, юридические ограничения и методы доступа;
- информация о процессах, таких как график заданий и коды извлечения.

2. Метаданные преобразования данных:

- информация о получении данных (например, планирование передачи данных и результатов, а также сведения об использовании файлов);
- управление таблицами измерений, например, определения измерений и присвоения суррогатных ключей;
- преобразование и агрегирование, например, расширение и отображение данных, программы, скрипты загрузки СУБД, определения агрегатов данных;
- документирование проверок, работ и журналов, например, журналов преобразования данных и записей слежения за происхождением данных.

3. Метаданные СУБД:

- содержание системных таблиц СУБД;
- рекомендации по обработке.

Метаданные можно использовать тремя способами:

1. Пассивно, обеспечивая четкую документацию о структуре, процессе разработки и использовании системы Хранилища данных. Доступная документация необходима всем участникам, т.е. конечным пользователям, системным администраторам, а также разработчикам приложений.

2. Активно, путем хранения конкретных семантических аспектов (например, правил преобразования) в виде метаданных, которые можно интерпретировать и использовать во время исполнения. В этом случае процессы Хранилища данных управляются метаданными. Активные метаданные и дополнительная документация согласованно и унифицировано управляются в одном репозитории, при этом актуальность документации возрастает.

3. Полуактивно, за счет хранения статической информации, например, определений структур, спецификаций конфигураций, которую будет считывать другой программный компонент во время выполнения.

Суть метаданных можно представить, описывая их роль и назначение в реализации процессов хранилища данных (рис. 1).

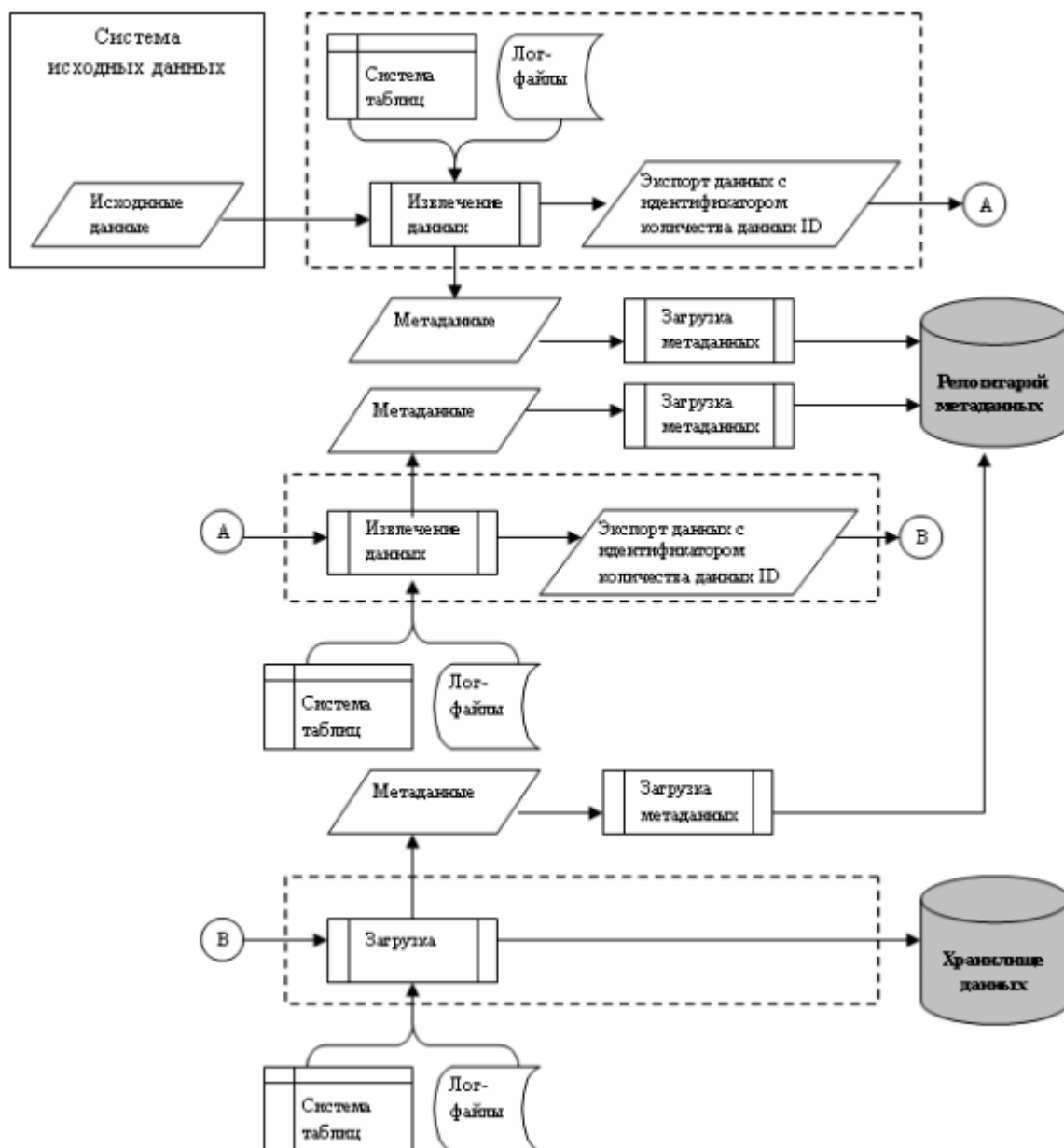


Рис. 1. Выработка метаданных в процессах хранилища данных

Метаданные обеспечивают аутентичность, надежность, применимость и целостность файлов, обеспечивают удобное управление ими и понимание информационных объектов как физических, так и аналоговых, и цифровых; обеспечивают эффективную организацию работ по разработке и администрированию БД, более быстрому извлечению информации из БД (повышение качества информационных услуг) и предполагают грамотное стратегическое управление данными и получение автоматизированной отчетности по управлению данными, включая мониторинг состояния БД.

Метаданные необходимы для анализа, проектирования, построения, внедрения и применения практически любой компьютерной системы. Так как все, с чем работает компьютер, по сути, является данными, а метаданные сопровождают любые данные, то это понятие имеет место в любой предметной области.

Список использованных источников:

1. Метаданные и их место в хранилище данных [Электронный ресурс]. Режим доступа – <http://bourabai.kz/tpoi/metadata.html>