

Инструментальная Система управления потоками данных (ИСУПД). Некоторые области применения Системы

1. Построение информационных систем анализа данных.

Для выполнения анализа больших объёмов данных с применением специальных алгоритмов, как правило, используются специально разработанные базы данных (OLAP-хранилища), которые загружаются данными из оперативных БД периодически по установленному регламенту. Такой подход снижает нагрузку на оперативные OLTP-системы, исключает коллизии доступа к данным при их оперативной обработке, улучшает характеристики процесса анализа за счёт оптимизации состава и структуры хранения данных, подлежащих анализу.

Примером такой системы может служить система анализа результатов хозяйственной (финансовой, производственной, научной и т.д.) деятельности организации за отчётный период, выполняемой на основе данных из БД различных эксплуатируемых в организации систем: бухгалтерского учёта, финансовой системы, депозитарной системы, торговой системы, CRM-системы, кадровой системы и т.д.

Для создания подобной OLAP-системы с использованием ИСУПД необходимо:

- Спроектировать и описать Операции ИСУПД, моделирующие хранилище «срезов» данных определённой структуры и степени обобщения. Данные в Хранилище могут быть представлены в виде «кубов» данных (трёхмерная структура).
- Разработать Регламент загрузки данных из оперативных OLTP-систем в Операции ИСУПД. В рамках Регламента разрабатывается технологическая карта загрузки, содержащая:
 - Источники исходных данных, график/расписание их использования;
 - Структуру исходных данных и способы их получения для каждого источника;
 - Алгоритмы тестирования, очистки, преобразования и обобщения исходных данных;
 - Схему записи данных в Операции ИСУПД;
 - Порядок удаления/очистки «устаревших» данных.
- Спроектировать и разработать Обработчики импорта данных, выполняющие сбор, очистку, проверку, преобразование и обобщение исходных данных.
- Спроектировать и разработать Обработчики анализа данных Операций ИСУПД.
- Спроектировать и разработать Обработчики экспорта результатов анализа данных Операций ИСУПД (при необходимости).
- Спроектировать и описать Задания и Процессы ИСУПД, в рамках которых активизируются Обработчики ИСУПД. Задания активизируются по установленному регламенту, например, один раз в сутки после завершения основного цикла обработки данных в оперативных системах или по событиям встроеной подсистемы событий.

В общем случае Система после загрузки исходных данных в хранилище (Операции ИСУПД) выполняет определённые действия (запросы) по их отбору и анализу. Результаты анализа (тестирования) данных могут использоваться для формирования различной отчётности или передаваться (экспорт) в другие системы-потребители данных с помощью Обработчиков экспорта данных. Данные также могут быть переданы в другие экземпляры ИСУПД, функционирующие на других узлах локальной или глобальной сети. В случае необходимости хранилище очищается от уже не нужных данных.

2. Построение информационных систем мониторинга данных.

В процессе функционирования информационных систем как правило выполняются действия по вводу и корректировке данных в их БД. В некоторых случаях может потребоваться выполнение анализа определённых данных (параметров) без вмешательства в работу этих систем, например, с целью выявления каких-то тенденций или отклонений в изменении или состоянии данных. По

результатам такого анализа вырабатываются какие-то оповещения, рекомендации или управляющие воздействия.

Примером такой системы может служить система «Противодействие отмыванию доходов и финансированию терроризма» (ПОД ФТ). В ней осуществляется анализ операций (сделок), выполняемых в рамках различных систем, эксплуатируемых в организации, с целью выявления «подозрительных» действий или участия в сделках установленных ранее нарушителей (физических или юридических лиц, разыскиваемых органами или взятых на особый контроль). В случае обнаружения таких фактов формируются и передаются определённые сведения в соответствующие организации.

Для создания подобной системы с использованием ИСУПД необходимо:

- Спроектировать и описать Операции ИСУПД, содержащие в своей структуре контролируемые данные (атрибуты), соответствующие контролируемым действиям (процессам) информационных систем организации.
- Разработать Регламент загрузки данных из оперативных OLTP-систем в Операции ИСУПД. В рамках Регламента разрабатывается технологическая карта загрузки, содержащая:
 - Источники исходных данных, график/расписание их использования;
 - Структуру исходных данных и способы их получения для каждого источника;
 - Алгоритмы тестирования, очистки, преобразования и обобщения исходных данных;
 - Схему записи данных в Операции ИСУПД;
 - Порядок удаления/очистки «устаревших» данных.
- Спроектировать и разработать Обработчики импорта данных, выполняющие сбор, очистку, проверку, преобразование и обобщение исходных данных.
- Спроектировать и разработать Обработчики анализа (тестирования) данных Операций ИСУПД.
- Спроектировать и разработать Обработчики экспорта сведений о выявленных отклонениях (нарушениях) в данных Операций ИСУПД.
- Спроектировать и описать Задания и Процессы ИСУПД, в рамках которых активизируются Обработчики ИСУПД. Задания активизируются по установленному регламенту, например, один раз в сутки после завершения основного цикла обработки данных в оперативных системах или по событиям встроенной подсистемы событий.

В общем случае Система «следит» за процессами обработки данных в OLTP-системах путём копирования и упорядочивания контролируемых данных из БД OLTP-систем в свою БД в соответствие с созданными объектами типа «Операция» с последующим анализом этих данных. Данные Операций тестируются специально разрабатываемыми Обработчиками анализа (тестирования) данных. Результаты тестирования используются для формирования специальных донесений или управляющих воздействий для передачи (экспорт) в другие системы-потребители данных Обработчиками экспорта данных. Данные также могут быть переданы в другие экземпляры ИСУПД, функционирующие на других узлах локальной или глобальной сети. В случае необходимости проверенные данные Операций могут быть удалены из БД или переписаны в архивную БД.

3. Построение информационных систем интегрированных хранилищ данных.

В случае, когда необходимо организовать стандартную обработку данных, находящихся в различных источниках (БД, файлы, сайты и т.д.), в которых они имеют различное устройство и структуру, а также могут быть не формализованы (тексты), создаётся специальная комплексная БД, интегрирующая в себе все такие данные или ссылки на такие данные (ETL-система). За счёт строгой типизации и структуризации данных в такой БД их обработка значительно упрощается.

Примером такой системы может служить поисковая система, которая осуществляет поиск объектов, хранящихся в различных источниках, по заданным поисковым критериям.

Для создания подобной системы с использованием ИСУПД необходимо:

- Спроектировать и описать Операции ИСУПД, содержащие в своей структуре обрабатываемые данные (атрибуты) или их фрагменты. Кроме самих данных в Операциях могут содержаться ссылки на данные, используемые для доступа к данным при их обработке.
- Разработать Регламент загрузки данных из источников в Операции ИСУПД. В рамках Регламента разрабатывается технологическая карта загрузки, содержащая:
 - Источники исходных данных, график/расписание их использования;
 - Структуру исходных данных и способы их получения для каждого источника;
 - Алгоритмы тестирования, очистки, преобразования и обобщения исходных данных;
 - Схему записи данных в Операции ИСУПД;
 - Порядок удаления/очистки «устаревших» данных.
- Спроектировать и разработать Обработчики импорта данных, выполняющие сбор, очистку, проверку, преобразование и обобщение исходных данных.
- Спроектировать и разработать Обработчики анализа (тестирования) данных Операций ИСУПД.
- Спроектировать и разработать Обработчики экспорта (передачи) отобранных данных Операций ИСУПД (при необходимости).
- Спроектировать и описать Задания и Процессы ИСУПД, в рамках которых активизируются Обработчики ИСУПД. Задания активизируются по установленному регламенту или по событиям встроенной подсистемы событий.

В данном варианте внутренняя БД ИСУПД используется как интегрированное хранилище данных, получаемых из различных источников. Обработчики импорта данных по установленному регламенту проводят анализ исходных данных из этих источников, осуществляют отбор требуемых данных, их контроль, очистку и преобразование с занесением их во внутреннюю БД Системы. Данные в БД могут как накапливаться, так и заменять уже существующие. В последующем внутренняя БД Системы служит источником для выдачи данных в виде установленной отчётности или источником для поиска и отбора требуемых суммарных данных.

4. Построение подсистем архивирования данных.

В некоторых предметных областях ИС функционируют в рамках установленных отчётных или регламентных периодов, например, год, квартал, месяц и т.д., когда по истечении такого периода данные могут быть удалены из БД ИС или обобщаются с преобразованием в итоговые (суммарные) данные истёкшего периода (суммы, остатки и т.д.). Такой подход значительно «облегчает» БД ИС. В то же время, если имеется вероятность, что в будущем могут всё же потребоваться исходные данные предыдущих периодов, то необходимо использовать архивную БД, в которую переписываются удаляемые данные из основной БД. В этом случае можно говорить о хранении поколений данных. Оперативная работа ведётся в рамках основной оперативной БД, а «старые» данные доступны в архивной БД в течение заданного периода хранения.

Для создания подобной системы с использованием ИСУПД необходимо:

- Спроектировать и описать Операции ИСУПД, содержащие в своей структуре обрабатываемые данные (атрибуты) с привязкой к отчётным периодам для моделирования поколений данных.
- Разработать Регламент загрузки данных из источников в Операции ИСУПД. В рамках Регламента разрабатывается технологическая карта загрузки, содержащая:
 - Источники исходных данных, график/расписание их использования;
 - Структуру исходных данных и способы их получения для каждого источника;
 - Алгоритмы тестирования, очистки, преобразования и обобщения исходных данных;
 - Схему записи данных в Операции ИСУПД;

- Порядок удаления/очистки «устаревших» данных.
- Спроектировать и разработать Обработчики импорта данных, выполняющие сбор, очистку, проверку, преобразование и обобщение исходных данных.
- Спроектировать и разработать Обработчики анализа архивных данных Операций ИСУПД (при необходимости).
- Спроектировать и разработать Обработчики экспорта (восстановления) данных Операций ИСУПД в основную БД (при необходимости).
- Спроектировать и описать Задания и Процессы ИСУПД, в рамках которых активизируются Обработчики ИСУПД. Задания активизируются по установленному регламенту или по событиям встроенной подсистемы событий.

В данном варианте внутренняя БД ИСУПД используется как архив «устаревших» данных из оперативной БД ИС. Обработчики импорта данных по установленному регламенту проводят анализ оперативной БД ИС, осуществляют отбор и изъятие «устаревших» данных с занесением их (после требуемых преобразований) во внутреннюю БД Системы. В последующем внутренняя БД Системы служит источником для выдачи архивных данных в виде установленной отчётности или источником для восстановления изъятых данных в оперативную БД ИС.

5. Построение подсистем обеспечения репликации данных между БД или ИС.

В случае многозвенной структуры ИС, когда в ней функционируют несколько БД и/или подсистем различного уровня, ИСУПД может использоваться для организации обмена данными между различными узлами такой ИС, например, в рамках филиальной сети одной организации или между ИС различных организаций.

В данном варианте внутренняя БД Системы используется как временное хранилище данных, подлежащих передаче из одной БД в другую или из одного узла сети на другой узел (фактически система выступает посредником при экспорте и импорте данных). Обработчики импорта (получения) данных выполняют приём исходных данных из заданного источника данных (или выполняют непосредственный доступ к данным), их проверку и преобразование. Система организует временное хранение данных в своей БД и по установленному регламенту при доступности системы-приёмника выполняет занесение этих данных в заданном виде в другую БД или осуществляет их передачу на другой узел сети. После доведения до системы-приёмника данные удаляются из внутренней БД или архивируются в специальной архивной БД.

Возможно создание сложных многоуровневых линейных или древовидных «конвейеров передачи данных» от системы к системе и/или от узла к узлу с анализом и фильтрацией требуемых данных.

6. Комплексование (интеграция) программных комплексов или ИС.

Как правило, в организациях эксплуатируются несколько самостоятельных ИС с целью автоматизации (информатизации) деятельности различных предметных областей, например, кадровая система, бухгалтерская и финансовая системы, торговая система, депозитарная система, электронный документооборот и т.д. К тому же они могут быть разработаны на различных средствах с использованием различных СУБД и форматов данных. В то же время, имеется реальная необходимость использовать данные одних систем в других без их повторного ввода (принцип ручного ввода данных только в той системе, где по регламенту функционирования организации они порождаются). Кроме того, аналогичные потребности возникают при функционировании многоуровневых систем, например, филиальной структуры организации. Т.е. необходимо говорить об интеграции различных систем в единый комплекс как по данным так и по управлению ими.

Фактически это более сложный вариант предыдущего применения ИСУПД. В данном варианте создаётся комплекс из существующих самостоятельных ПК и/или ИС. Система в этом случае выступает как объединяющая и управляющая «шина», организующая синхронное или

асинхронное взаимодействие различных систем между собой. Управляющие воздействия и потоки данных передаются от одной системы к другой через ИСУПД по событиям и/или инициативе самих систем (с использованием API ИСУПД и интегрируемых систем). При этом ИСУПД выполняя функции посредника, учитывает и «выравнивает» требуемый перечень и форматы данных для каждой из взаимодействующих систем, использует механизмы доступа (API, протоколы и т.д.), предусмотренные каждой системой.

7. Доступ к данным в БД из программных комплексов или ИС.

Использование COM/OLE объектов, реализованных в Сервере ИСУПД, напрямую через Сервер ИСУПД или через Агента удалённого доступа ИСУПД позволяет абстрагироваться от конкретных СУБД и, используя методы работы с данными, осуществлять доступ к данным с последующей их обработкой. Особенно это актуально в Web-приложениях, когда в коде обработки событий, написанных на скрипт-языках (Java, VB, ...), создаётся объект внешнего или удалённого управления с дальнейшим использованием их методов для доступа к данным.