

ОПЫТ ПОСТРОЕНИЯ ПРОМЫШЛЕННОЙ БАЗЫ ДАННЫХ С БОЛЬШИМИ ОБЪЕМАМИ ДАННЫХ

Антонов А.В.

*Институт проблем управления им. В.А. Трапезникова РАН,
Россия, г. Москва, ул. Профсоюзная д.65*

antonov@ipu.ru

Аннотация: Представлены результаты опыта построения базы данных хранения результатов стендовых испытаний различных новых видов образцов продукции отрасли

Ключевые слова: база данных хранения, большие объемы данных, опыт построения.

Введение

Промышленная отраслевая база данных испытаний включает базу данных хранения и базу данных представлений, как результатов испытаний, так и документации по самим объектам испытаний, и документации по подготовке, а также проведению их испытаний. В данном материале представлен опыт построения только базы данных хранения (БДХ) результатов испытаний новых и модернизированных промышленных образцов техники и относящихся к ним документации.

1 Постановка задачи создания БДХ

1.1 Инфраструктура испытаний

Заказчик имеет в своём распоряжении более десятка крупных стендов, предназначенных для всесторонних испытаний образцов продукции отрасли. Каждый стенд оборудован рабочими местами (5,...20 персоналок), на которых производится запись и хранение входных данных, настроечных данных и данных испытаний.

Каждый из стендов обслуживается одним или несколькими структурными подразделениями, которые позиционируют себя как «владельцы» данных (напрямую никаких данных взять нельзя – обмен данными только по служебным запросам), а также аналитическое подразделение, ответственное за выпуск итоговых отчётов по результатам испытаний. Данные испытаний хранятся в форматах либо текстовых таблиц, либо бинарных файлов, с информационными заголовочными разделами. Объём каждого файла от нескольких сотен байт до нескольких гигабайт. По каждому испытанию «больших» файлов (более 10 Мб, а в среднем 300,..600 Мб) – до нескольких десятков. Число регистрируемых при испытании параметров – до нескольких тысяч. Объекты испытаний, как правило, новые, поэтому списки регистрируемых параметров различны

1.2 Основные требования и ограничения на формирование БДХ

- Пользователи не должны иметь доступ ко всем объектам испытаний и общему списку объектов испытаний, а иметь доступ только к тем, к которым у них есть разрешение
- Должна быть возможность менять (увеличивать) число параметров по каждому объекту испытаний (это касается, как правило, файлов результатов обработки испытаний).
- По каждой записи должна быть информация о том кто, где и когда её внёс в БД или модифицировал
- Записи должны удаляться только в исключительных случаях
- В состав сохраняемых данных должны входить документы по подготовке и проведению испытаний (в электронных форматах – «docx», «txt», «xls», «jpg», «png», «mov»,...), программы и процедуры обработки данных; экранные формы представления данных пользователю в скриптовом или бинарном виде.
- **Одно из самых больших ограничений на формирование БД** – минимальное воздействие на существующие в подразделениях технологические процессы первичного сохранения и первичной обработки данных, за которые отвечают эти подразделения.

1.3 Общие сведения о задачах, решаемых БДХ

В БДХ решаются три основные группы задач:

- создание и модификация структуры и состава рабочих таблиц данных
- наполнение рабочих таблиц данными и поддержание данных в актуальном состоянии
- предоставление пользователям необходимых им данных в соответствии с их правами.

2 Результат построения БДХ

В качестве СУБД для БДХ была взята СУБД Oracle

2.1 Основные объекты БД

В БДХ представлены следующие объекты БД:

- табличные пространства (table space);
- пользователи БД Oracle (users);
- роли пользователей БД Oracle (roles);
- таблицы (tables);
- синонимы (synonyms);
- представления данных таблиц (views);

Ниже даны краткие характеристики основных объектов БДХ

2.2 Табличные пространства

При создании БДХ новые табличные пространства не создавались, а использовались стандартные: табличное пространство USER - для хранения пользовательских таблиц и табличное пространство TEMP - для работы со временными данными.

2.3. Пользователи

В БДХ представлены следующие пользователи

- Менеджер
- Редактор
- Регистратор
- не привилегированные пользователи (просто Пользователи)

Менеджер - создаёт и удаляет *таблицы* БДХ (задание паролей опущено). В единственном экземпляре

Редактор - создаёт, модифицирует, удаляет *виртуальные таблицы и папки*. (о виртуальных таблицах и папках – ниже). В единственном экземпляре.

Регистратор - добавляет, модифицирует, «удаляет» (условно) записи в таблицах, Три экземпляра

Не привилегированные пользователи (или просто «**Пользователи**») - запрашивают и читают данные таблиц в соответствии с запросами и правами. Несколько экземпляров – по числу объектов испытаний с различными правами доступа к их данным.

2.4 Роли пользователей

Роли используются для тиражирования специфических прав. В БДХ создана роль, для выполнения общих для всех пользователей задач, а также частные роли для Регистраторов и Пользователей.

2.5 Таблицы

Таблицы - именованные хранилища данных, закрепляемые внутри табличного пространства за конкретным пользователем. Доступ к данным таблицы по квалифицированному имени таблицы:

[квалифицированное имя таблицы] = [имя пользователя]. [имя таблицы].

Число таблиц в БДХ – фиксировано. Более подробно о таблицах - ниже

2.6 Синонимы

Синонимы - синонимы таблиц используются для сокрытия размещения и владельца таблицы и упрощения обращения к данным таблицы. Используется не полное квалифицированное имя таблицы, а зарегистрированное в СУБД имя её синонима.

2.7 Представления данных таблиц группам пользователей

Представления данных таблиц - сохранённый в СУБД под определённым именем запрос к данным одной или нескольких таблиц. Пользователь использует имена этих представлений также как имя таблицы (с полями, указанных в запросе полей таблиц).

3 Состав и свойства таблиц системы хранения данных

3.1 Общие положения

В БДХ используются следующие группы таблиц:

- служебные таблицы
- таблицы хранения данных Редактора
- таблицы хранения данных Пользователей

Заполнение и модификация данных всех перечисленных групп таблиц осуществляется через приложение **Big_Base**, поставляемое вместе с БДХ.

Таблицы хранения данных Редактора и данных Пользователя используют инфраструктуру, в основе которой лежат классы-контейнеры языка C++ или C# [1]. К ним относятся

- класс значений параметров (пары «вид значения – само значение»); возможные виды значений вынесены в справочную таблицу. К видам значений, например, относятся: текущее, максимальное, минимальное, заданное значение, и т.п.;
- класс параметров (пары «вид параметра – массив атрибутов параметра»); класс параметра является контейнером для экземпляров классов значений параметра; к видам параметров (внесены в справочную таблицу), например, относятся как параметры, имеющие числовые значения - время, скорость, масса, цена, и т.п., так и параметры не имеющие числовых значений (качественные, экспертные шкалы значений) – например, название какого либо компонента, (виды значений названия – полной, краткое, аббревиатура и т.п.), физическая размерность значений параметра, «уровень удобства использования» и т.п.
- класс «интерфейсных элементов взаимодействия с пользователем» (пара «вид элемента – атрибуты элемента»); перечень видов интерфейсных элементов внесён в справочную таблицу
- класс сущностей или юнитов (пара «вид сущности – атрибуты сущности»); используемые при испытаниях виды сущностей внесены в справочную таблицу. Этот класс является контейнером для экземпляров параметров, служащих для описания свойств данного вида сущности, также он включает контейнер компонентов этой сущности (иерархически вложенные сущности), а также включает контейнер экземпляров интерфейсных элементов для отображения и модификации свойств и их значений. К сущностям, например относятся: объект испытаний, его составные части, документация по объекту и испытаниям объекта, программные продукты обработки данных испытаний и т.п.

3.2 Служебные таблицы

К служебным таблицам относятся

- REF_TASK
- TBL_MGR_DATA
- TBL_SESS
- TBL_SESS_LOG
- TBL_SESS_U
- TBL_SESS_V

Две первые таблицы это – соответственно таблица зарегистрированных видов задач пользователей БДХ и таблица данных Менеджера БДХ

Таблица TBL_SESS предназначена для хранения данных о сессии работы конкретного пользователя с данными БДХ, т.е. идентификационных данных пользователя, решаемая им задача, начало и интервал времени работы конкретного пользователя с БДХ. Сессия обеспечивает привязку всех действий по добавлению и изменению данных в БДХ к конкретному автору этих изменений и времени их проведения. Таблица TBL_LOG служит для хранения протоколов работы пользователя во время сессии. Таблица TBL_SESS_U служит для хранения данных сессии, а именно, имя пользователя, с какой машины вошёл, через какое приложение, и т.п. Таблица TBL_SESS_V служит для хранения настроек интерфейсных данных приложений клиента (количество, размещение и наполнение интерфейсных компонентов приложений)

3.3 Таблицы хранения данных Редактора

К таблицам этой группы относятся:

- TBL_IDSTART - последние значения счётчиков (идентификаторов), если они изменились.
- TBL_DSE_UNIT –таблица сущностей редактора (юнитов) - папки, виртуальные таблицы, параметры (столбцы виртуальной таблицы), значения параметров (точнее, свойства значений параметра)
- TBL_DSE_ULINK –таблица связей сущностей (юнитов)
- TBL_DSE_CTRL –данные о интерфейсных элементах для отображения и изменения значений для всех свойств юнита, его параметров и их значений
- TBL_DSE_PRM – параметры сущностей
- TBL_DSE_VAL – значения параметров сущностей

Эти таблицы служат для создания и редактирования шаблонов структуры виртуальных таблиц с данными пользователей, а также для организации взаимосвязей виртуальных таблиц с данными. Инструментом для выполнения работ по созданию и редактированию шаблонов структуры виртуальных таблиц является клиент и сервер **BigBase_Dse** приложения **Big_Base**.

Перечисленные выше реальные таблицы (просто таблицы) в БД являются «узкими» таблицами вида «код - значение», которые также включают несколько служебных полей.

Пользователям в приложении **Big_Base** предоставляются «виртуальные» таблицы, которые формируются запросами к этим реальным таблицам. Эти виртуальные таблицы могут быть произвольной ширины. Число и состав виртуальных таблиц может меняться Редактором. Наполнение виртуальных таблиц определяется правами пользователя на предоставляемые данные. Виртуальные таблицы могут Редактором группироваться под конкретного пользователя с помощью виртуальных папок и подпапок.

3.4 Таблицы хранения данных пользователей

К таблицам хранения данных пользователей относятся:

Таблицы описания структуры и свойств виртуальных.

- TBL_DER_UNIT - таблица сущностей (юнитов) регистратора и пользователей - папки, виртуальные таблицы
- TBL_DER_ULINK – таблица связей сущностей
- TBL_DER_PRM – таблица параметров сущностей
- TBL_DER_VAL – таблица значений параметров (атрибутов) сущностей
- TBL_DER_CTRL – таблица данных о контролах для отображения и редактирования значений параметров

Таблицы с данными виртуальных таблиц

- TBL_ID_NAME – таблица для хранения виртуальных таблиц вида «имя - значение»
- TBL_REC_VAL_S – строковые данные
- TBL_REC_VAL_F – числовые данные
- TBL_REC_VAL_B – бинарные данные
- TBL_REC_VAL_O – данные с изменяющимся ведущим параметром (обычно это время)

Все данные пользовательских таблиц, таблицы которых имеют от 2х до 4х столбцов, хранятся в едином формате таблицы TBL_ID_NAME (виртуальные таблицы группы "идентификатор - имя").

Все данные пользовательских виртуальных таблиц, таблицы которых имеют более 2х столбцов и могут быть в процессе "жизни" БДХ произвольно увеличены на любое другое количество столбцов - хранятся в едином формате таблиц TBL_REC_VAL_S, TBL_REC_VAL_F, TBL_REC_VAL_B и TBL_REC_VAL_O. На базе запросов к этим таблицам формируются «виртуальные» таблицы. Это виртуальные таблицы группы "идентификатор - объект". Виртуальные таблицы группируются Редактором с помощью включения их в виртуальные "папки", которые в свою очередь, могут входить в "папки" более высоких уровней. Структура связей виртуальных Папок и Таблиц формируется Редактором и может быть различной для разных групп Регистраторов и не привилегированных Пользователей. Регистраторы и пользователи различаются данными, к которым они могут обращаться.

Примечание. Компонентами всех видов таблиц являются

- юнит с шаблоном для формирования полей записи таблицы и
- юнит для хранения собственно всех записей таблицы

Таблицы имеют схожую структуру, с обязательными полями (два первых и два последних поля), например:

```
create table TBL_REC_VAL_F
(
tbl_kind VARCHAR2(16) not null,           // вид таблицы
rec_id   VARCHAR2(40) not null,          // идентификатор записи
prm_kind VARCHAR2(16) not null,         // вид параметра
val_kind VARCHAR2(8) not null, // вид значения параметра
val_data NUMBER,                        // экземпляр значения параметра
sess_w   VARCHAR2(32) not null,         // сессия записи
sess_c   VARCHAR2(36)                  // сессия потери актуальности записи (иначе-NULL)
)
```

Доступ к пользовательским таблицам и их данным для регистрации и модификации значений этих таблиц осуществляется клиентом и сервером **BigBase_Dor** приложения **BigBase**.

Необходимо иметь ввиду, что «большие» данные испытаний хранятся в двух форматах; менее 20 Мб – хранятся как бинарные данные в поле `val_data` таблицы `TBL_REC_VAL_O`, а свыше 20 Мб хранятся ссылки на файлы данных в защищённых директориях на сервере СХД.

3.5 Базовый состав системы виртуальных таблиц СХД

Система виртуальных объектов СХД включает

- Перечень папок и таблиц **Редактора** виртуальных таблиц СХД, их состава и связей;
- Таблицы **Регистратора** данных в виртуальных таблицах СХД;
- Папки и таблицы не привилегированных **Пользователей** СХД.

3.6 Перечень объектов Редактора виртуальных объектов СХД, их состава и связей

Базовый состав папок Редактора включает:

- Папку СХД (корень дерева)
- Подпапку Шаблонов компонентов СХД (корень 1-го уровня)
- Подпапку схемы Регистратора (корень 1-го уровня)
- Подпапки схем не привилегированных Пользователей СХД - по правам доступа Пользователей (корни 1-го уровня).

Папка Шаблонов компонентов БДХ - корневой элемент дерева шаблонов компонентов БДХ. Не удаляемый элемент.

Шаблон компонента – служит для клонирования (копирования структуры и свойств) компонента в схеме Регистратора. При клонировании структура шаблона копируется, также копируется большая часть свойств; свойство «вид компонента» - модифицируется, а имя и описание компонента – могут быть модифицированы.

Папка шаблонов содержит подпапки:

- Шаблоны папок (корневая папка)
- Шаблоны таблиц «идентификатор – значение»
- Шаблоны таблиц «идентификатор – объект».

В данном контексте "объект" это макет записи с произвольным количеством параметров (столбцов), но фиксированным для записей таблицы определённого типа. Эта структура является жёсткой и не подлежит изменению.

Состав и перечень шаблонов объектов, и состав столбцов шаблонов (с соответствующими редакторами значений столбцов) таблиц не является жёстким. Все виртуальные таблицы существуют в единственном экземпляре, Оригинал таблицы находится в схеме редактора, В схемах Пользователей – представлены клоны этих таблиц. Структура таблиц-клонов не редактируется напрямую! При изменении структуры шаблона структура всех клонов модифицируется. При удалении столбца в шаблоне таблицы этот столбец удаляется из всех клонов; удаляются также все данные из этого столбца клонов (опасная операция). При добавлении столбца в шаблон - этот столбец добавляется во все клоны. В существующих записях значение этого столбца принимается пустым (NULL) или «по принятому умолчанию».

Папка схемы Регистратора содержит клоны компонентов (папок и таблиц) БДХ. Клонов одного компонента шаблона может быть несколько, но они должны входить в разные «деревья» из компонентов.

Папки схем (не привилегированных) Пользователей БДХ – обычно содержат поддерева дерева схемы регистратора, но могут иметь и оригинальное по структуре дерево компонентов.

3.7 Предопределённые виртуальные папки

К предопределённым папкам относятся следующие папки:

- Справочная система
- Справочники общего характера
- Служебные справочники
- Специализированные справочники
- Объекты испытаний
- Средства испытаний (стенды)
- Документация по испытаниям
- Испытания (исходные данные и результаты)

- Вспомогательные данные

3.8 Предопределённые таблицы "Идентификатор-имя"

К предопределённым таблицам "Идентификатор-имя" относятся следующие таблицы:

- Виды сущностей
- Коды групп параметров
- Коды использования значений параметра (текст, медийный, исполняемый,...);
- Коды видов параметров
- Виды значений параметров
- Кодовые страницы
- Виды связей сущностей
- Да-Нет
- Да-Нет-«Не определено»
- Виды типов данных
- Виды редакторов значения свойства

Эта группа таблиц используется Редактором для выбора и задания свойств виртуальных объектов.

Ещё одна группа таблиц обеспечивает описание объектов, импортируемых из старых баз данных на СУБД Access, созданных для подготовки к проведению испытаний. В неё, в частности входят следующие таблицы

- Типы градуировки датчиков
- Коды групп размерностей
- Принадлежность предприятий (организаций)
- Типы элементов архива
- Виды испытываемых изделий промышленности
- Категории документации
- Виды испытаний изделий
- Виды обработки данных
- Перечень категорий стендов

3.9 Предопределённые таблицы "Идентификатор-объект"

- Размерности физических параметров
- Предприятия и организации
- Перечень объектов испытаний
- Перечень средств испытаний (стенды)
- Перечень видов датчиков
- Проектная и конструкторская документация объекта
- Документы по подготовке и порядку проведения испытаний
- Акты и протоколы испытаний
- Отчётные и аналитические материалы
- Испытания объектов
- Испытания. Перечень работ
- Перечень сведений о документах
- Перечень архивов со сведениями о документах
- Перечень элементов архива
- Перечень документов (файлов документов)
- Перечень Баз данных для подготовки к испытаниям на рабочих местах стендов
- Перечень измерительных комплексов (ИК)
- Виды блоков измерительных комплексов
- Испытания. Перечень блоков ИК
- Испытания. Перечень датчиков
- Испытания. Измеряемые параметры
- Испытания. Калибровка датчиков
- Испытания. Параметры устройств (блоков) ИК

3.10 Автогенерация схем Пользователей

Таблица "Предприятия и организации" (эта таблица содержит также подразделения организаций) является ключевой для генератора схем Пользователей и поддеревьев с папками и таблицами в схеме Регистратора. Для создания новой схемы Пользователя и поддерева с папками и таблицами в схеме Регистратора необходимо выполнить следующие операции:

- войти в Big_Base под Регистратором,
- открыть таблицу "Предприятия и организации", пометить запись с организацией в столбце "Создать схему",
- выйти из регистратора с сохранением данных.
- войти в Big_Base под Редактором, проконтролировать создание схемы и поддеревьев (будут созданы узлы в редакторе);
- сохранить данные редактирования - при сохранении данных редактора будут сгенерированы объекты в схеме Регистратора и схема Пользователя.
- далее необходимо повторно войти в редактор - будут сгенерированы необходимые ссылки в таблицах схем.
- проконтролировать наличие ссылок и
- сохранить данные редактирования.

После этих операций вновь созданные таблицы пригодны для ввода данных и их модификации в регистраторе и для предоставления данных в схеме Пользователя.

Заключение

Схема построения и развёртывания представленной промышленной базы данных была реализована с некоторым

и модификациями в двух организациях в вариантах использования СУБД Oracle и СУБД PostgreSQL.

Литература

1. *А.В. Антонов*, Унификация моделирования классов объектов Труды Международная научная конференция МОДЕЛИРОВАНИЕ-2010, (SIMULATION-2010) 12-14 мая 2010 г. Институт проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины, г. Киев, стр.117,...120