Хранилища данных. Лекция 3. Методология построения

Антон Викторович Кудинов,
доцент кафедры ВТ

Содержание

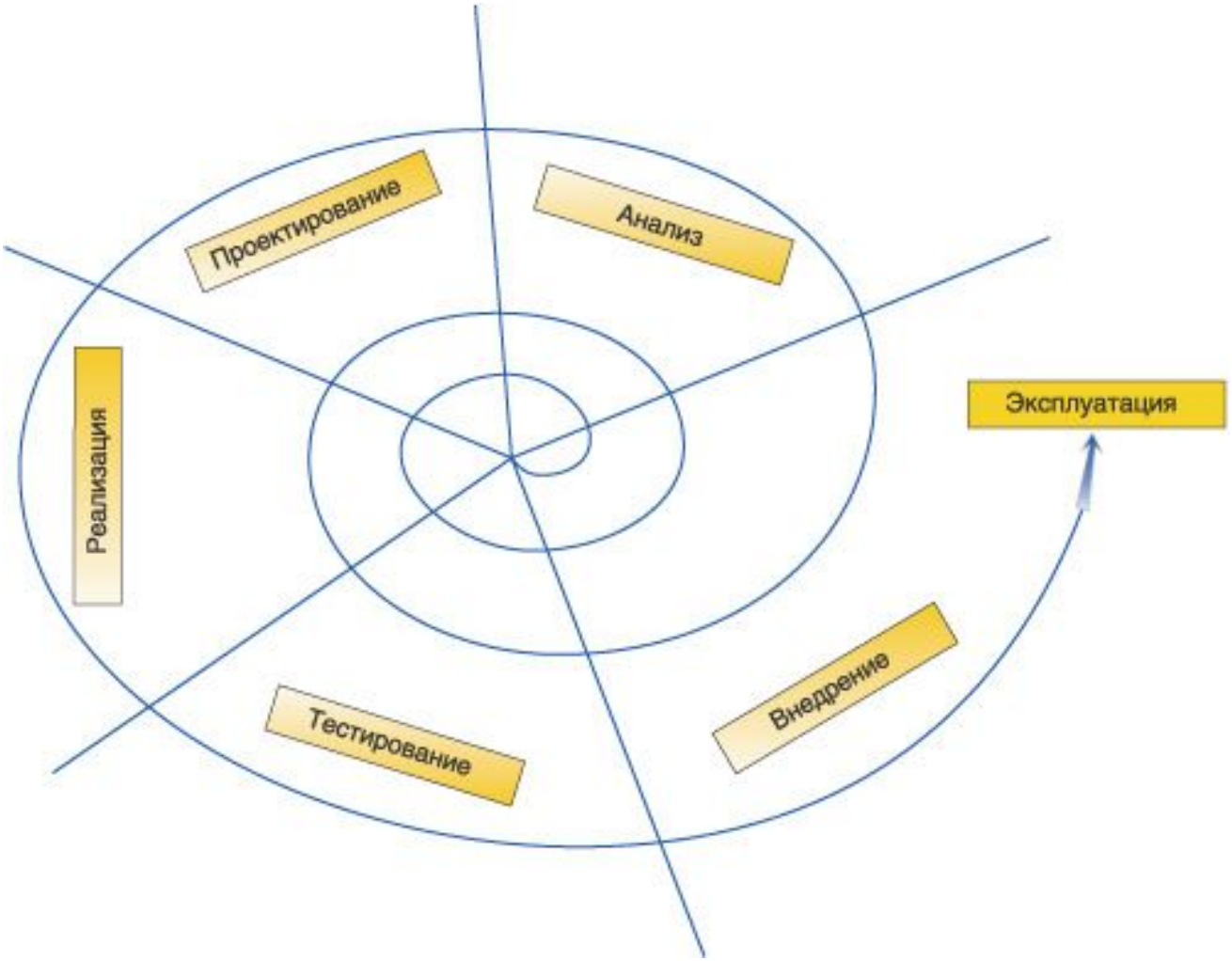
- Модели разработки
- Постановка задачи
- Проектирование
- Реализация
- Внедрение

Подходы к стратегии построения ХД

- построение сверху вниз,
- снизу вверх,
- динамическая интеграция данных и др.

- **Наиболее эффективный подход** - в процессе разработки и внедрения хранилища данных осуществляется его пошаговое наращивание на основе единой системы классификаторов и общей среды передачи и хранения данных – **спиральная модель процесса разработки**.

Спиральная модель процесса разработки



Стратегия пошагового наращивания позволяет

- по завершении каждого цикла ввести в кратчайшие сроки в промышленную эксплуатацию законченную систему, с определенной ограниченной функциональностью
- уменьшить потери при возможных проектных ошибках вследствие небольших масштабов каждого проектного цикла
- сократить время реализации каждой новой витрины за счет повышения опыта проектной группы, упрощения взаимодействия заказчика и исполнителя

Анализ (постановка задачи)

- Системно-аналитическое обследование
- Согласование технического задания

Системно-аналитическое обследование

- согласование и утверждения заказчиком плана и программы обследования
- проводятся интервью с основными участниками проекта со стороны компании-заказчика и лицами, ответственными за принятие управленческих реш
- уточняется организационная структура, фиксирую организационные и функциональные рамки проек
- выявляются и документируются особенности и недостатки существующих информационных решений
- формализуется схема бизнеса компании с учетом функциональных рамок
- производится сбор существующих отчетных материалов и прочих официальных документов

По итогам обследования

- уточняются стратегические и оперативные задачи управления компанией, решение которых должна обеспечивать СППР
- формализуются цели и задачи создания системы

- **Цель этапа анализа** – получение моделей данных и описание процедур принятия управленческих решений

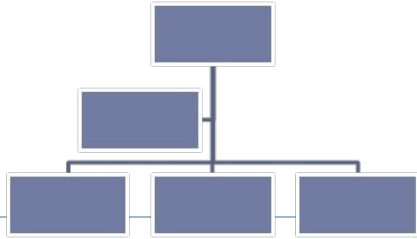
Техническое задание

- один из ключевых документов проекта, который определяет требования к созданию СППР и порядок этого создания
- Если время разработки > 12 месяцев нужно вводить очередность и сначала разрабатывать ТЗ первой очереди (на срок 3 месяца)
- Для большого проекта - основное ТЗ, частные ТЗ

Проектирование – разработка архитектуры

- Архитектура информационной системы рассматривается в четырех аспектах:
- **Логическая архитектура**. Представляет архитектуру системы с точки зрения пакетов базовых классов и их взаимосвязей. Определяются автоматизируемые процессы и функции, которые затем разделяются на задачи, подлежащие реализации на стадии разработки.
- **Архитектура процессов**. Определяет информационное обеспечение системы – состав и содержание процессов преобразования и передачи данных.
- **Компонентная архитектура**. Представляет архитектуру ПО системы, ее декомпозицию на подсистемы и компоненты.
- **Техническая архитектура**. Описывает физические узлы системы и связи между ними

Логическая архитектура



- **Объекты автоматизации** - технические процессы, связанные с информационным обеспечением управленческой и аналитической деятельности руководящего персонала и специалистов подразделений и высшего руководства компании
- **Цели системы**:
  - Интеграция ранее разъединенных детализированных данных
  - Разделение наборов данных, используемых для оперативной обработки и для решения задач поддержки принятия решений
  - Обеспечение всесторонней информационной поддержки максимальному кругу пользователей

Автоматизируемые процессы

- Сбор данных.
- Преобразование данных:
  - Очистка данных.
  - Согласование данных.
  - Унификация данных.
  - Агрегирование данных.
- Хранение данных:
  - Промежуточное хранение данных.
  - Накопление исторических данных.
- Предоставление данных потребителям.
- Сопровождение метаданных

Проектирование информационного обеспечения

Должны быть определены:

- **Измерения**, их иерархии и уровень детализации.
  - Измерениями хранилища часто служат организационная структура компании, справочник административно-территориального деления, план финансовых статей компании и пр.
- **Базовые показатели**, измерения, их индексирующие, и правила агрегирования каждого показателя по иерархиям. Правила агрегирования по иерархическому измерению зависят от показателя.
- **Производные показатели** и формулы их вычисления на основе базовых показателей.

Выбор подходящего источника данных

- Если имеется более одного источника, следует ли определить, какой из них лучше?
- Какие преобразования необходимо выполнить, чтобы приготовить источник к загрузке в хранилище?
- Согласуются ли структура источника и структура хранилища?
- Насколько согласуются данные источника с нормативно-справочной информацией?
- Что будет происходить, если источник имеет несколько месторасположений?
- Насколько аккуратны данные источника?
- Как источник обновляется?
- Каковы возраст и перспективность источника?
- Насколько полны данные?
- Что потребуется для интеграции данных источника в поток загрузки?
- Какова технология хранения данных в источнике?
- Насколько эффективно может осуществляться доступ к источнику?

Архитектурные решения

- **Определяются состав, содержание и источники** потоков данных, которые будут поступать из источников в хранилище.
- **Определяются преобразования**, которые должны быть выполнены над данными при загрузке, а также периодичность загрузки данных в хранилище.
- При необходимости **проектируются структуры** оперативного склада данных и транзитных файлов.
- **Выявляются данные**, которые отсутствуют в источниках информационного хранилища. Для таких данных, как правило, проектируются процедуры и регламенты ручного ввода

Общая структура репозитария хранилища

- **Персональная информация**. Чаще всего храниться в МБД.
- **Информация по бизнес-темам**. Обычно храниться в смешанных структурах: МБД и реляционных таблицах.
- **Детальные данные**. Обычно храниться в реляционных структурах.
- **Старые детальные данные** (ленты или библиотеки).

Компонентная архитектура

Два вида ПО:

- общее
- специальное

Общее ПО

- **ПО промежуточного слоя**, которое обеспечивает сетевой доступ к приложениям и БД. Сюда относятся сетевые и коммуникационные протоколы, драйверы, системы обмена сообщениями и пр.
- **ПО загрузки и предварительной обработки данных**. Набор средств для загрузки данных из OLTP-систем и внешних источников. Проектируется, как правило, в сочетании с дополнительной обработкой: проверкой данных на чистоту, консолидацией, форматированием, фильтрацией и пр.
- **Серверное ПО**. Представляет собой ядро всей системы. Оно включает в себя:
  - Серверы реляционных БД,
  - Серверы МБД,
  - Серверы приложений (поисковые, аналитической обработки, добычи знаний и др.).

Специальное ПО

- Подсистема загрузки данных
- Подсистема обработки запросов и представления данных
- Подсистема администрирования

- Проектируются модули, составляющие подсистему, и алгоритмы отдельных процедур, входящих в их состав

Техническая архитектура

- **Серверное ПО** работает под управлением серверов приложений и серверов БД на UNIX- или NT-платформах или мэйнфреймах.
- **Клиентское ПО** устанавливается на ПК конечных пользователей. «Тонкий» клиент (Web-броузер + JavaScript)
- Техническая архитектура во многом зависит от масштабов и требований, предъявляемых к ее производительности и надежности.
- Серверные компоненты (сегменты хранилища и витрины данных) могут

Реализация. Ее результаты

- Непосредственно сама **система** в виде общего и специального ПО, баз данных.
- **План внедрения** системы, который должен определять все работы по внедрению системы у заказчика, включая упаковку системы, доставку ее заказчику, инсталляцию системы на технических средствах заказчика, тестирование и доработку.
- **Набор тестов**, которые должны быть выполнены после установки системы у заказчика.
- **Пользовательская документацию** и учебные материалы для пользователей системы

Внедрение

- все по плану
  - монтаж и установка системы и отдельных ее компонентов у заказчика
  - первоначальная загрузка хранилища необходимыми данными
  - опытная эксплуатация системы
  - обучение пользователей и сотрудников службы технической поддержки
- Окончание этапа - переход к производственной эксплуатации хранилища

Спасибо за внимание!

KudinovAV@tpu.ru