



Общая информация о системе

DWH система представляет собой специально организованную базу данных, для которой обеспечивается сбор информации, ее хранение и быстрый доступ к предметно-ориентированным данным при помощи широкого спектра BI инструментов. При помещении данных в хранилище так же происходит их трансформация, агрегация и устранение противоречий, что позволяет использовать DWH различным пользователям в качестве единого источника достоверной информации.



Компоненты системы и инструменты их реализации:



Описание компонент

1.Источники данных. Источники могут быть представлены файлами, сервисами, разнородными СУБД.

2.Загрузка, трансформация. ETL (extract, transform, load) процессы выполняют загрузку первичных данных, очистку (выявление противоречий), трансформацию в структуры хранилища. ETL был построен на собственной разработке BarsETL в идеологии продукта Oracle Data Integrator (ODI) выполнение пакетов и интерфейсов. Возможна настройка данного процесса на промышленный ETL продукт Oracle Data Integrator. Работа с данным продуктом даст возможность в дальнейшем абстрагироваться от понятия pl/sql и настраивать процессы загрузки, трансформации данных непосредственно в графическом интерфейсе.

3.Загрузка, трансформация. ETL (extract, transform, load) процессы выполняют загрузку первичных данных, очистку (выявление противоречий), трансформацию в структуры хранилища. ETL был построен на собственной разработке BarsETL в идеологии продукта Oracle Data Integrator (ODI) – выполнение пакетов и интерфейсов. Возможна настройка данного процесса на промышленный ETL продукт Oracle Data Integrator. Работа с данным продуктом даст возможность в дальнейшем абстрагироваться от понятия pl/sql и настраивать процессы загрузки, трансформации данных непосредственно в графическом интерфейсе.

4.Хранилище данных. Была выбрана СУБД Oracle EE.

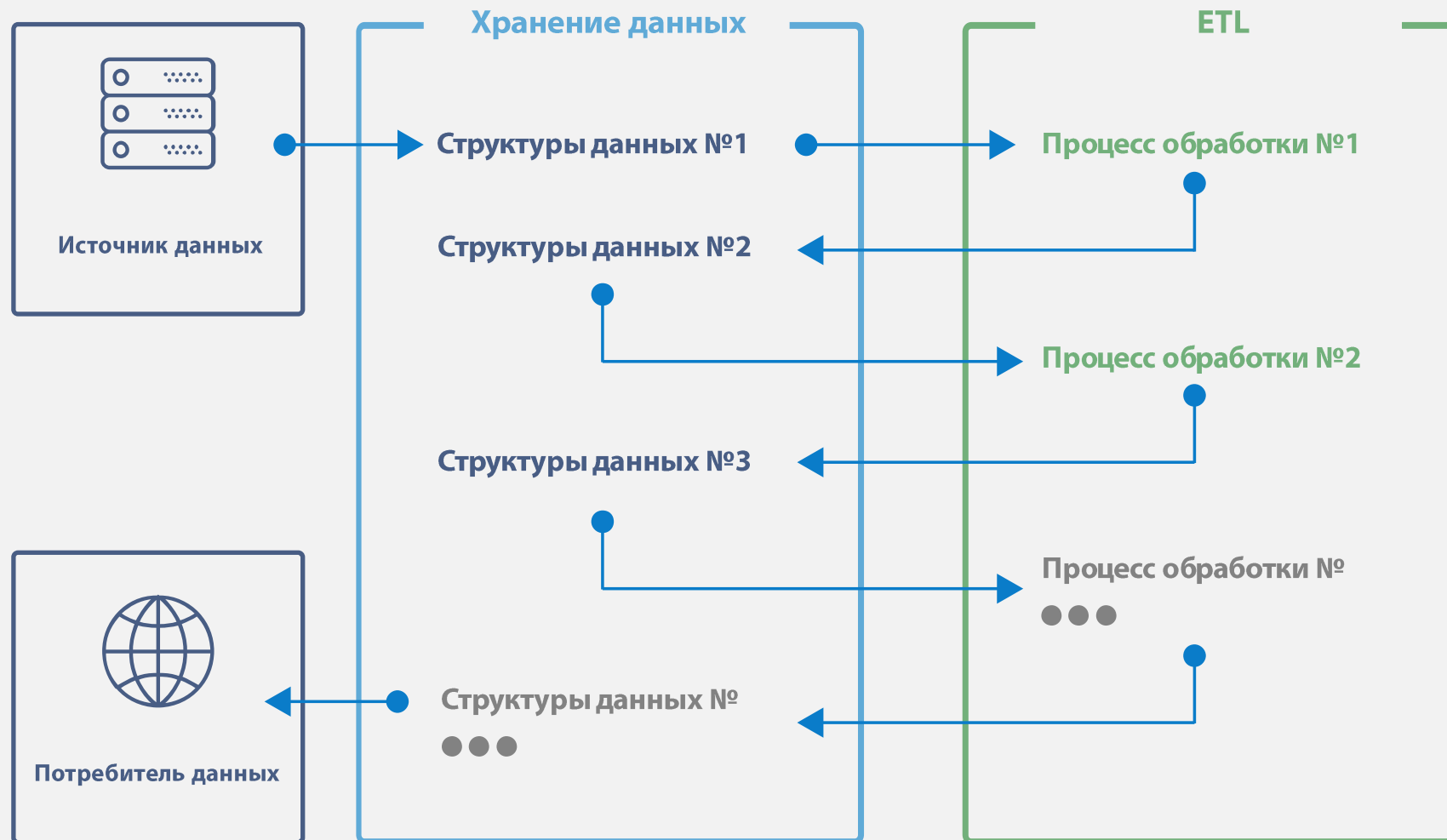
Схема данных. Схема подразумевает типовую банковскую модель данных, которая на физическом уровне превращается непосредственно в структуры хранения (таблицы) в хранилище данных. Схема всегда дорабатывается под специфику заказчика.

5.Анализ данных. Анализ данных поддерживается двумя направлениями:

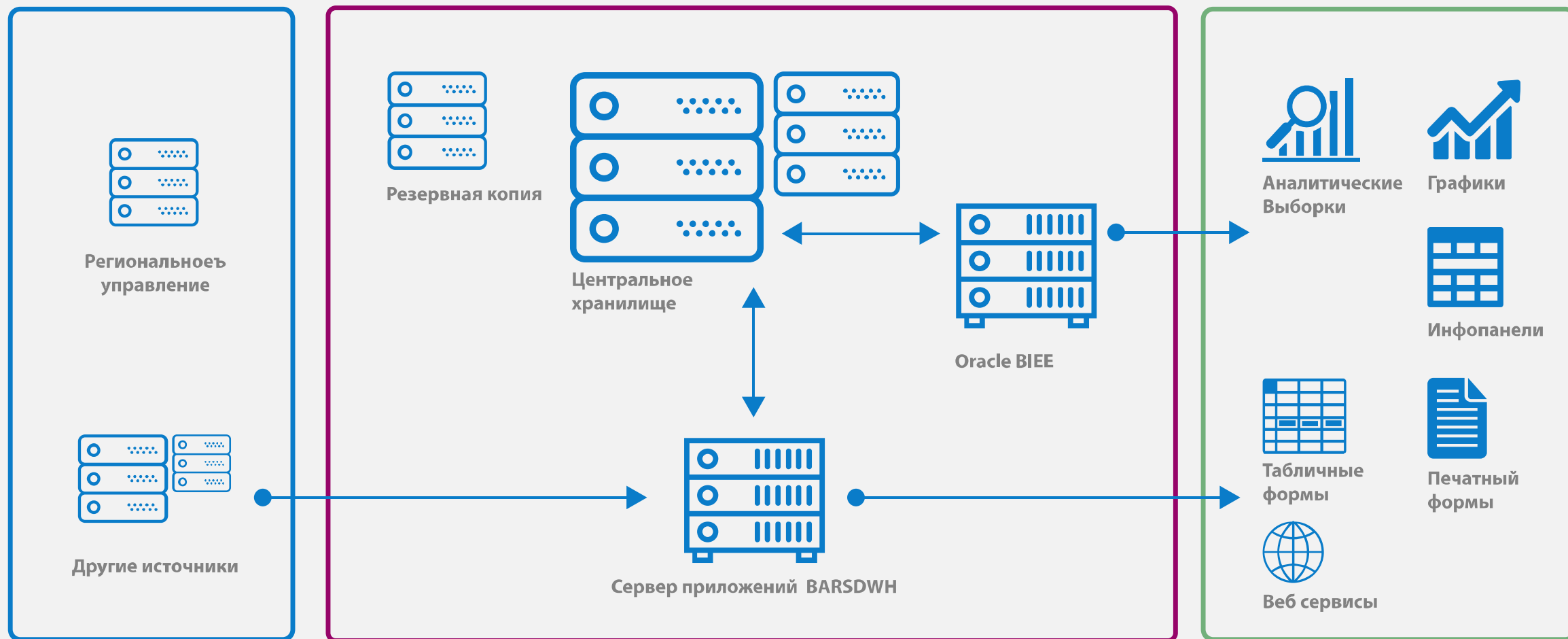
BarsDWH – собственная веб разработка для формирования отчетных и табличных форм, а так же использованием промышленного BI продукта Oracle BIEE, который позволит выполнить более гибкий и глубокий анализ данных на основании динамических строителей табличных форм в виде ‘проваливающихся’ деревьев от агрегата к значению, динамические представления графиков, построение информационных панелей.

6.Координатор процесса. Для организации работы всех компонент в непрерывную цепочку, разработано веб приложение BarsDWH. При помощи данного приложения оператор системы производит мониторинг работы и принимает оперативные решения по устранению проблем.

Логическая схема



Физическая схема



Дополнительные требования к системе

DWH ориентирована не только на процесс 'загрузка – трансформация - анализ, а так же: отказоустойчивость, гибкость в настройках, прозрачность в масштабировании – ключевые пункты в построении системы.

Предусматриваются такие требования к работе хранилища данных:

1. **Загрузка информации от источников как он-лайн режиме так и при помощи файлового обмена**
2. **Возможность повторной перегрузки данных, в случае ошибочных расчетов или сбоев в системе хранилища**
3. **Возможность автоматического и ручного управления процессом загрузки и трансформации**
4. **Бесперебойная загрузка данных, не зависящая от содержимого загрузок**
5. **Параллельная загрузка из нескольких источников данных**
6. **Поддержка полной и частичной загрузки**
7. **Настройка дополнительных критериев проверки качества данных**