

# ПРОГРАММНОЕ СРЕДСТВО ВИЗУАЛИЗАЦИИ СТАТИСТИЧЕСКИХ И ОПЕРАТИВНЫХ ДАННЫХ ПРЕДПРИЯТИЯ ПО РЕАЛИЗАЦИИ ВЕЛОСИПЕДОВ И ВЕЛОЭКИПИРОВКИ

*Буйновская С.В., студент*

*Белорусский государственный университет информатики и радиоэлектроники,  
Институт информационных технологий,  
г. Минск, Республика Беларусь*

*Шелягович А. С. – ст. препод. каф. ИСиТ*

В представленной работе описан результат разработки программного средства визуализации статистических и оперативных данных предприятия по реализации велосипедов и велоэкипировки, приведено описание основных возможностей программного средства и обоснована актуальность разработки.

Визуализация данных является ключевым инструментом в современной аналитике, поскольку она помогает аналитикам превратить сырые данные предприятия в понятную информацию, а также увидеть скрытые взаимосвязи и тенденции, которые трудно обнаружить, просто просматривая данные в таблицах баз данных, хранилищ данных или файлах.

В превращении неструктурированных фрагментов данных во что-то полезное помогает визуализация данных, а именно создание BI-системы.

BI-система (Business Intelligence) – набор инструментов и технологий для сбора, анализа, визуализации и обработки данных о состоянии бизнеса. Проще говоря, это целый класс приложений, которые преобразуют информацию из разных источников в удобную и понятную аналитику с визуализацией данных [1].

Программное средство представляет собой BI-систему, архитектура программного средства представлена на рисунке 1.

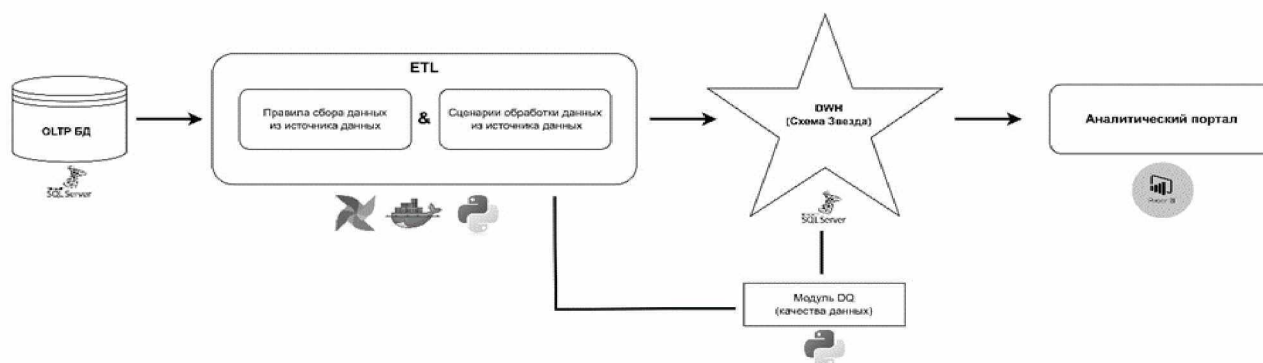


Рисунок 1 – Архитектура программного средства

Основными элементами архитектуры программного средства являются:

- источник данных в виде OLTP базы данных, развернутой на MS SQL Server;
- ETL-процесс инкрементной загрузки данных в хранилище данных, реализованный с использованием Apache Airflow, Docker и Python;
- само хранилище данных, построенное по схеме звезда;
- модуль проверки качества данных, реализованный при помощи Python;
- аналитический портал на основе средства визуализации данных Microsoft Power BI.

Разработанное хранилище данных было построено, используя метод размерного моделирования, поскольку типичная структура хранилища данных существенно отличается от структуры обычной реляционной СУБД.

Размерное моделирование – это метод структуры данных, оптимизированный для хранения данных в хранилище данных, цель его использования – оптимизировать базу данных для более быстрого поиска данных. Данная концепция была разработана Ральфом Кимбаллом и состоит из таблиц фактов и измерений.

Этапы размерного моделирования:

- определить бизнес-процесс;
- определить зерно (уровень детализации);
- определить измерения;
- определить факты;
- построить звезду (схема данных) [2].

Концептуальная модель разработанного хранилища данных представлена на рисунке 2.

Программное средство выполняет следующие функции:

- осуществляет первоначальную загрузку данных в хранилище данных;
- ежедневно осуществляет инкрементную загрузку данных в хранилище данных при помощи платформы Apache Airflow;
- проверяет качество данных в хранилище данных по различным стандартным атрибутам качества данных (полнота, согласованность, целостность данных, отсутствие дубликатов в данных) после инкрементной загрузки данных;
- уведомляет пользователей программного средства в корпоративном мессенджере Slack в случаях, когда качество данных ухудшилось после инкрементной загрузки данных;
- отображает изменения данных в хранилище данных на разработанных отчетах в режиме реального времени.

Программное средство формирует 3 интерактивных отчета: отчет, отражающий общие показатели эффективности предприятия за год; отчет, позволяющий увидеть покупали ли определенный товар на определенной территории в выбранный период времени; отчет, позволяющий узнать сколько продуктов за месяц купили покупатели на определенной территории, а также сколько продаж совершил выбранный продавец за выбранный период времени.

Перечисленные выше отчеты являются интерактивными, т.е. у пользователя есть возможность самостоятельно устанавливать перечень отображаемых показателей, сортировать данные и выстраивать фильтры, это позволит провести более качественный анализ данных и не упустить важные аспекты.

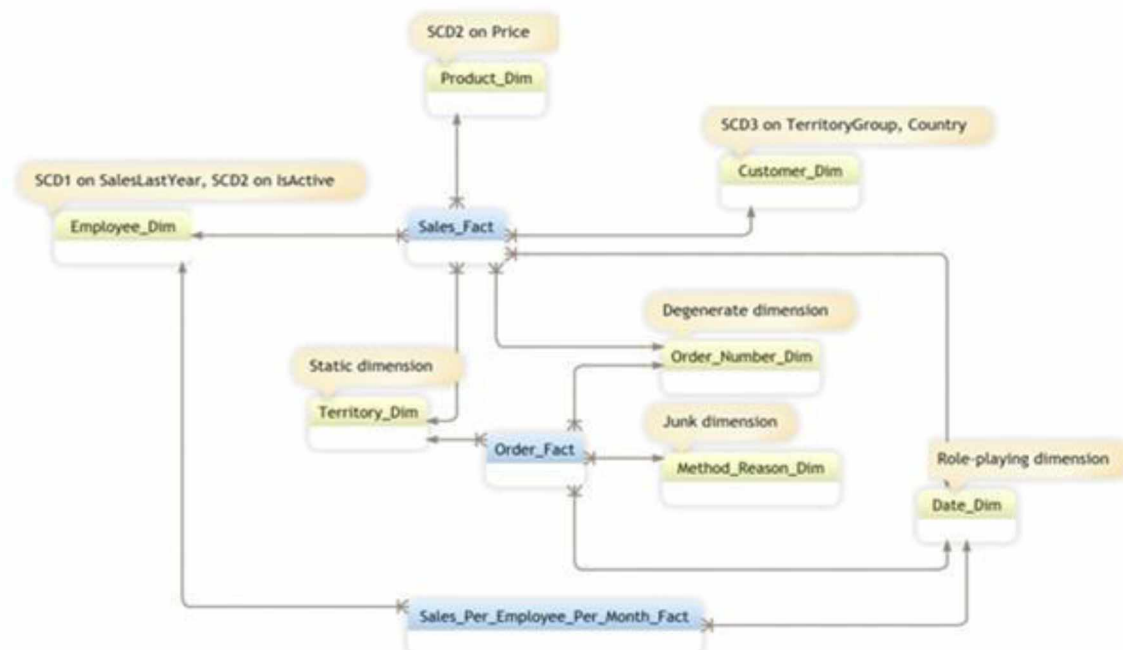


Рисунок 2 – Концептуальная модель хранилища данных

Разработка является востребованной, так как с помощью визуализации данных можно облегчить процесс принятия решений, ведь вместо того, чтобы пытаться понять сложные таблицы и числа, можно просто посмотреть на график или диаграмму и увидеть, что происходит, что является значительной экономией времени аналитиков данных, а также делает данные понятными и доступными для всех, даже тех, кто не является экспертом в области данных. Все это упрощает и ускоряет принятие решений на основе данных, что является ключевым аспектом в бизнес-среде.

**Список использованных источников:**

1. Business Intelligence (BI) [Электронный ресурс]. – Режим доступа : <https://www.unisender.com/ru/glossary/business-intelligence>.
2. Введение в OLAP [Электронный ресурс]. – Режим доступа : [http://www.olap.ru/basic/olap\\_intro2.asp](http://www.olap.ru/basic/olap_intro2.asp).