

ПОДХОДЫ К ПОСТРОЕНИЮ ХРАНИЛИЩ ДАННЫХ

Рассматриваются два основных подхода к построению хранилища данных. Предлагается комбинированный метод построения хранилища данных для совмещения преимуществ основных подходов.

ВВЕДЕНИЕ

На сегодняшний день существует два основных подхода к построению Хранилища данных. Это хранилище данных с архитектурой шины Ральфа Кимбола и с корпоративной информационной системой Билла Инмона.

I. ПОДХОД БИЛЛА ИНМОНА

Рассматриваемая структура хранилища данных представлена на рисунке 1.

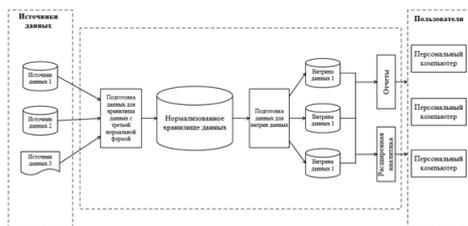


Рис. 1 – Структурная схема хранилища данных с корпоративной информационной системой

Данные согласованно извлекаются из источников в централизованное хранилище с третьей нормальной формой, откуда поступают в витрины данных для дальнейшего анализа. Такая структура предоставляет возможности для манипулирования атомарными данными, но не обеспечивает высокой производительности.

II. ПОДХОД РАЛЬФА КИМБОЛЛА

Структура рассматриваемого хранилища данных представлена на рисунке 2.

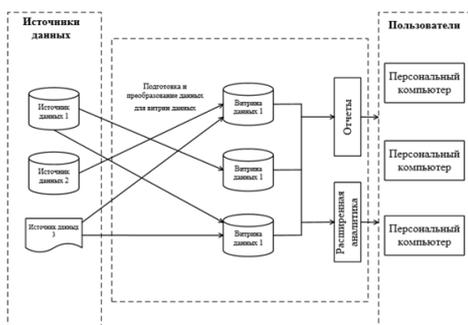


Рис. 2 – Структурная схема хранилища данных с архитектурой шины

Геврасёва Ирина Петровна, магистрант кафедры информационных технологий автоматизированных систем БГУИР, irina_gevraseva@mail.ru.

Научный руководитель: Навроцкий Анатолий Александрович, заведующий кафедрой информационных технологий автоматизированных систем БГУИР, кандидат физико-математических наук, доцент, navrotsky@bsuir.by.

Данные из источников поступают непосредственно в пространственно структурированные витрины данных, позволяет добиться высокой производительности без дополнительного перепрограммирования. С другой стороны, в этой схеме нет необходимой гибкости структуры.

III. КОМБИНИРОВАННЫЙ МЕТОД ДЛЯ ПОСТРОЕНИЯ ХРАНИЛИЩЕ ДАННЫХ

Рассмотрим комбинированное хранилище, структура которого представлена на рисунке 3.

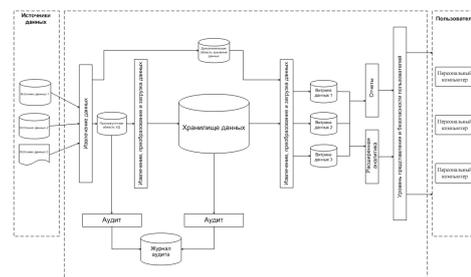


Рис. 3 – Структурная схема комбинированного хранилища данных

Данные извлекаются из источников данных в стейджи данных. Затем данные преобразуются и поступают в пространственно структурированное хранилище, откуда поступают в витрины данных для дальнейшего анализа. Для уменьшения задержки попадания данных к конечному пользователю предлагается использовать дополнительную область хранения данных, заполняемую чаще, чем хранилище данных.

ВЫВОДЫ

Предложенный подход обеспечивает высокую производительность и гибкость работы с атомарными данными. Минусом данной системы является ее сложность и ресурсоемкость.

Список литературы

1. Kimball, R. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition / R. Kimball, M. Ross // Wiley Publishing, Inc. – 2013. – С. 564.
2. Inmon, W. H. Building the Data Warehouse, Fourth Edition / W. H. Inmon // Wiley Publishing, Inc. – 2005. – С. 576.