

**Цифровое сельское хозяйство****ИСПОЛЬЗОВАНИЕ МОДЕЛИ ДАННЫХ DATA VAULT 2.0 В  
ИНФОРМАЦИОННОМ ХРАНИЛИЩЕ АНАЛИТИЧЕСКОГО КОМПЛЕКСА  
СЕЛЬСКОХОЗЯЙСТВЕННОЙ ОРГАНИЗАЦИИ**

Статья рекомендована к публикации главным редактором Т. В. Ершовой 02.12.2024.

**Чернышева Кира Владимировна**

*Кандидат экономических наук  
Российский экономический университет имени Г. В. Плеханова, доцент  
Москва, Российская Федерация  
chernysheva.kv@rea.ru*

**Горошкина Ульяна Евгеньевна**

*Российский экономический университет имени Г. В. Плеханова, студент  
Москва, Российская Федерация  
u.goroshkina@gmail.com*

**Аннотация**

*В статье изложены результаты авторских исследований существующих подходов к организации информационных хранилищ аналитических систем, используемых моделях данных. Рассмотрена модель данных Data Vault 2.0, возможности ее применения в аналитическом комплексе сельскохозяйственной организации, в частности, в зоотехнической службе. Рассмотрены сущности (хабы), характеризующие бизнес-процесс производства молока: биологический объект, корма, ветеринарные мероприятия, готовая продукция, их связи и спутники (сателлиты).*

**Ключевые слова**

*цифровая экономика; экономика данных; информационно-аналитические системы; информационные хранилища; модели данных; Data Vault 2.0; Хаб; сущность; связь*

**Введение**

В условиях цифровой экономики и перехода к экономике данных возрастает значимость данных для принятия эффективных управленческих решений. В программном документе «Стратегия развития информационного общества в РФ на 2017 – 2030 гг» сказано, «Цифровая экономика – это хозяйственная деятельность, в которой ключевым фактором производства являются данные в цифровом виде, обработка больших объёмов и использование результатов анализа которых по сравнению с традиционными формами хозяйствования позволяют существенно повысить эффективность различных видов производства, технологий, оборудования, хранения, продажи, доставки товаров и услуг» [1].

Целью нового Национального проекта «Экономика данных» является внедрение управления на основе данных. Основными направлениями проекта являются сбор данных с применением высокочувствительных датчиков, их передача в режиме реального времени, хранение с использованием отечественных облачных платформ, центров обработки данных, а также обработка и анализ данных, в т. ч. и с использованием технологий искусственного интеллекта, на базе отечественных платформ и сервисов [2].

В связи с этим актуальным становится проектирование хранилищ данных, как элемента информационно-аналитической системы организации.

Цель исследования – разработка методик проектирования хранилища данных зоотехнической службы сельскохозяйственной организации, специализирующейся на производстве молока, на основе модели данных Data Vault 2.0. Выбор предметной области обусловлен значимостью отраслей

---

© Чернышева К. В., Горошкина У. Е., 2025

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «С указанием авторства – С сохранением условий» версии 4.0 Международная», размещенной по адресу: <https://creativecommons.org/licenses/by-sa/4.0/legalcode.ru>

[https://doi.org/10.52605/16059921\\_2025\\_03\\_178](https://doi.org/10.52605/16059921_2025_03_178)

агропромышленного комплекса, молочного скотоводства, в частности, для обеспечения продовольственной безопасности страны. Кроме того, цифровизация отрасли, поступление информации из систем радиочастотной идентификации животных (RFID), прочих датчиков изменяет требования к объему и организации информационных хранилищ.

Исследования выполнялись на основе нормативно-правовых документов, справочных материалов с применением основных положений системного подхода. Проведен анализ существующих подходов к проектированию информационных хранилищ аналитических систем. Обоснованы методологические подходы к выбору и проектированию архитектуры информационных хранилищ аналитических систем.

## 1. Проблемы хранения данных и современные ИТ-архитектуры

В настоящее время проблема хранения и анализа накопленных данных есть в каждой организации, как и в госсекторе, так и в бизнесе. Так, по данным главы Минцифры Максута Шадаева, 85% компаний хотят наладить работу с данными в своих организациях, включая их оперативный анализ для принятия управленческих решений, но 91% организаций считают эту работу сложной [3]. Необходимо построение целостной системы управления данными, разработку ИТ-архитектура организации и архитектуры данных.

Современная ИТ - архитектура организации представляет собой совокупность различных интегрированных информационных систем, реализующих необходимый функционал по различным уровням управления, направлениям деятельности организации. Это транзакционные бухгалтерские системы, системы управления взаимоотношениями с клиентами, персоналом, логистические, корпоративные информационные системы и пр. Часто эти программные продукты бывают реализованы на разных платформах, построены на разных даталогических моделях, что затрудняет формирование единого информационного пространства организации. Разрозненность, недостоверность, недостаточность информации для принятия решения при высокой потребности в единой информационной картине, общей согласованной модели данных являются причинами появления информационно-аналитических систем в ИТ-архитектуре организации.

## 2. Транзакционные и аналитические системы

Транзакционные и аналитические системы направлены на решение разных типов управленческих задач по реализации основных функций управления, таких как среднесрочное и долгосрочное планирование и прогнозирование, учет, анализ и контроль.

Транзакционные информационные системы обеспечивают учет и контроль в организации. Возможно проведение и традиционного анализа, включающего расчет основных экономических и финансовых показателей деятельности фирмы, анализ их изменения во времени, сравнение плановых и фактических показателей и пр. В подобных системах реализованы OLTP (Online Transaction Processing)-технологии. Регистрация транзакций насыщает информационную базу данными, которые в дальнейшем могут использоваться в информационно-аналитических системах.

Планирование, прогнозирование, анализ с использованием статистических, экономико-математических методов, описательных и предсказательных моделей машинного обучения, методов и моделей искусственного интеллекта выполняется в информационно-аналитических системах. Данные программные продукты предназначены для решения широкого спектра бизнес-задач, требующих обработки больших объемов данных, реализации сложной логики и применения методов машинного обучения. Эти системы, наряду с OLTP-технологиями, включают OLAP-технологии оперативного анализа данных (Online Analytical Processing), KDD (Knowledge Discovery in Databases) - технологии, Data Mining - технологии интеллектуального анализа данных и машинного обучения.

## 3. Сбор, обработка и хранение данных в информационно-аналитических системах

Информационно-аналитические системы обеспечивают сбор данных из различных источников, в т. ч. из транзакционных систем и Интернет-источников, с использованием механизмов автоматического объединения/слияния данных; их преобразование, включая очистку и обогащение путем возможного расчета дополнительных показателей (мер); а также загрузку в информационное хранилище (Extract, Transform, Load, ETL--процессы). В настоящее время существуют варианты потоковой загрузки данных в архитектурные модификации хранилищ (озера, платформы данных), а также в реляционные хранилища

модели Data Vault. Кроме того, в аналитическом комплексе организации присутствуют инструменты анализа данных, включающие методы машинного обучения (рис. 1).

Источники данных	Обработка данных	Хранилище данных	Анализ данных	Визуализация данных
Отдельные файлы	Пакетная (ETL, ELT)	Хранилище данных (DWH)	Отчеты	BI дашборды
Датчики интернета вещей		Озеро данных (Data Lake)	Произвольные запросы	
Интернет-источники	Платформа данных		Оперативный многомерный анализ	
Базы данных транзакционных систем		Извлечение данных		

Рисунок 1. Архитектура аналитического комплекса организации

#### 4. Архитектура хранилища данных

Хранилище данных информационно-аналитических систем (Data Warehouse, DWH) строится на основе общекорпоративной модели данных (Enterprise Data Model, EDM), включающей модели сущностей и связей для предметных областей организации (концептуальные, логические и физические), описание их взаимосвязей (диаграммы потоков данных) и другие модели. Кроме того, в EDM содержится методика возможного расширения и адаптации, готовые модели анализа бизнеса с учетом ключевых бизнес-индикаторов, а также типовые алгоритмы трансформации данных в типовые структуры модели [4].

Именно на основе общекорпоративной модели данных выполняется проектирование информационных хранилищ. Из EDM производится отбор сущностей, их атрибутов и взаимосвязей для хранилища, выполняется расчет промежуточных полей, определяется степень детализации данных в хранилище и др. Также с учетом имеющейся общекорпоративной модели данных организации выбирается архитектура данных проектируемого хранилища.

#### 5. Классические модели данных для хранилищ

Традиционно Централизованное информационное хранилище в информационно-аналитической системе строится на основе реляционной модели данных 3НФ (третьей нормальной формы), размерной модели типа «Звезда» или «Снежинка», с пакетной загрузкой структурированных данных баз данных транзакционных систем, файлов различных форматов, данных Интернет-источников и пр.

Недостатками перечисленных моделей являются ограниченная горизонтальная масштабируемость, необходимость переработки структуры хранилища при изменениях в общекорпоративной модели данных, высокие затраты на создание, низкая производительность при выполнении аналитических запросов, сложность подготовки отчетов, а также работа только со структурированными данными [6, 7, 8]. Это является существенным минусом в эпоху Big Data, большой аналитики, когда возникает потребность в потоковой передаче данных, обработке структурированных, слабоструктурированных и неструктурированных данных. Но простота модели, избыточность данных, ее высокая устойчивость делает ее применимой и в настоящее время.

Данную модель данных можно рекомендовать для информационных хранилищ организаций, имеющих линейную или линейно-функциональную структуру управления, а также для небольших муниципальных образований.

В качестве технологий хранения в централизованных информационных хранилищах схем «Звезда» и «Снежинка» используются Single-Node Processing базы данных, такие как Oracle, SQL Server и др. [6].

#### 6. Модель данных Data Vault 2.0

Современная концепция моделей данных информационных хранилищ Data Vault 2.0 представляет методологию проектирования гибких и масштабируемых хранилищ данных. Главным элементом данной модели является хаб (Hub), в котором хранится основная информация о сущностях хранилища данных. Хаб в терминах Data Vault — таблица фактов. Хаб всегда включает следующие поля: уникальный

суррогатный ключ, ключ сущности (бизнес-ключ) – ИНН, табельный номер и пр., ссылка на источник записи, время добавления записи. Помимо перечисленных полей, возможно добавление новых.

Хабы соединяются между собой связью «многие-ко-многим» через таблицы-ссылки (Link) по суррогатным ключам.

Все изменяемые атрибуты сущностей хранятся в отдельных таблицах Спутниках (Satellite). Как правило, количество и содержание полей таблиц Спутников определяется из логической связности и частоты изменений данных. Так, в одном Спутнике можно хранить условно - постоянную информацию, в другом – переменную информацию.

У каждого Хаба и у каждой Связи может быть неопределенное количество вспомогательных таблиц Спутников с различными наборами полей.

К модели данных можно добавлять новые таблицы-хабы (Hub) и/или загрузка данных в таблицы может происходить параллельно, поскольку они не зависят друг от друга, что важно при появлении систем класса MPP – массивно-параллельной обработки типа Teradata, Vertica, Greenplum и др. [9].

Еще одной особенностью Data Vault 2.0 от других подходов к проектированию хранилищ данных является отсутствие необходимости преобразований и дополнений при переносе данных из источников. Трансформация данных может быть выполнена позже, т.е. возможно использование использоваться ETL-технология, в которой данные сначала загружаются из источников, а затем выполняется их преобразование.

Кроме того, в Data Vault возможна реализация point-in-time (PIT) таблиц. PIT-таблица создаётся для конкретного хаба или линка и связанных с ними сателлитов. Она предоставляет снимок данных в конкретный момент времени, консолидируя загруженные даты из таблиц сателлитов [10].

## 7. Пример проектирования витрины данных

### 7.1 Постановка задачи

Нами рассмотрен пример проектирования витрины данных информационного хранилища в нотации Data Vault 2.0 для сельскохозяйственной организации, специализирующейся на производстве молока. Бизнес-процесс производства молока связан с жизнедеятельностью животного основного стада, включающей в себя кормление, содержание и разведение.

Формирование единого информационного пространства сельскохозяйственной организации подразумевает использование информационно-аналитических комплексов, работающих с хранилищами данных, интегрирующих информацию их различных источников, в т.ч. из MES (manufacturing execution system) систем управления производственным процессом, из систем радиочастотной идентификации объектов (RFID), датчиков температуры и влажности и пр., обрабатывающих ее для принятия обоснованного управленческого решения. Возможно использование различных моделей данных для проектирования информационных хранилищ. Это реляционные модели, размерные модели типа «Звезда» и «Снежинка» [11]. Однако, увеличение объема поступающей информации, частота изменения источников данных и их структуры вызывают необходимость применения модели Data Vault 2.0.

Продемонстрируем примеры построения витрины данных информационно хранилища для бизнес-процесса производства молока на основе модели Data Vault 2.0 и реляционной модели.

Бизнес-процесс производства молока включает несколько этапов, в том числе уход за животными, кормление, ветеринарный осмотр, процесс дойки, хранения молока, а также его дальнейшую реализацию. Все эти этапы взаимосвязаны и требуют организации для обеспечения высокого качества продукции и устойчивого производства.

При инфологическом проектировании на первом этапе определяются основные бизнес-сущности и их использование в предметной области. Возможно выделение следующих объектов в проектируемой витрине данных [12]:

- биологические объекты (содержит информацию о состоянии здоровья животных, параметрах их продуктивности, времени и методах кормления, осеменения и ветеринарных процедурах);
- корма (данные о типах и количестве кормов, затраты на кормление биологических объектов, информация о закупках и качестве кормовых запасов);
- медикаменты (содержит сведения о лекарствах, включая тип, количество, назначение и период применения);

- готовая продукция - молоко (информации о процессе производства молока, объёмах, качестве, условиях хранения и продажах).

## 7.2 Пример проектирования витрины данных в нотации Data Vault 2.0

Витрина данных бизнес-процесса производства молока в нотации Data Vault 2.0 содержит следующие элементы: Хабы, Ссылки и Сателлиты. Определение Хабов происходит на основании бизнес-сущностей. Нами выделены следующие хабы: биологические объекты, корма, медикаменты, готовая продукция (молоко) [13]. Возможно добавление дополнительных хабов, линков и саттелитов в проектируемую модель данных. Это могут быть хабы о персонале, используемом технологическом оборудовании, затратах на продукцию и др. На втором этапе проектирования определяются Ссылки через выявление возможных отношений между бизнес-объектами. Моделирование контекста каждой бизнес-сущности и отношений между ними предопределяют Сателлиты. Предлагается отдельно выделять Сателлиты с условно-постоянной информацией, не меняющейся длительное время (порода, дата рождения коровы и др.), и Сателлиты с переменной информацией, содержащейся в первичных документах (например, ведомость учета расхода кормов, отражающая переменную информацию по расходу кормов).

Рассмотрим описание структуры Хаба биологического объекта, связанных с ним Сателлитов, описывающих детальную информацию о биологическом объекте, а также Связи с Хабом корма (таб. 1):

- Hub\_Animal: Хаб для уникальных идентификаторов коров, связанный с сателлитами, которые содержат данные о породе, местоположении, осеменении, надое молока и ветеринарных осмотрах. Связан с другими объектами через линки, например, с кормлением и медикаментами;
- Link\_Animal\_Feed: Связь между хабами коров (Hub\_Animal) и кормов (Hub\_Feed), описывающая, какой корм используется для каждой коровы. Содержит идентификаторы коровы и корма для отслеживания рациона;
- Sat\_Cows\_Details: Сателлит, хранящий детализированную информацию о коровах, такие как порода, дата рождения, вес и пол;
- Sat\_Cows\_Insemination\_Act: Сателлит, содержащий данные о проведённых актах осеменения коров, включая дату, метод осеменения и информацию о ветеринаре, который проводил процедуру. Также содержит отметку об успехе осеменения, что помогает отслеживать репродуктивную историю животных.

Таблица 1 – Пример описание элементов модели Data Vault 2.0

Название элемента	Атрибуты	Тип данных	Связи	Описание
Hub_Animal	Animal_ID	INT	Sat_Animal_Details Sat_Animal_Insemination_Act	Хранит уникальные идентификаторы коров
	Load_Date	DATE	Sat_Animal_Milk_Production	Дата загрузки
	Record_Source	VARCHAR(255)	Sat_Animal_Veterinarian_Visits Sat_Animal_Location  Link_Animal_Feed Link_Animal_Medication Link_Animal_Milk	Источник данных
Link_Animal_Feed	Link_ID	INT	Hub_Animal Hub_Feed	Содержит информацию о связи коров и используемых кормов
	Animal_ID	INT		Уникальный идентификатор коровы
	Feed_ID	INT		Уникальный идентификатор корма
	Load_Date	DATE		Дата загрузки

	Record_Source	VARCHAR(255)		Источник данных
Sat_Animal_Details	Animal_ID	INT	Hub_Animal	Уникальный идентификатор коровы
	Breed	VARCHAR(100)		Порода коровы
	Birth_Date	DATE		Дата рождения
	Weight	DECIMAL(5,2)		Вес коровы (в кг)
	Load_Date	DATE		Дата загрузки записи в спутник
	Record_Source	VARCHAR(255)		Источник данных
Sat_Animal_Insemination_Act	Animal_ID	INT		Hub_Animal
	Insemination_Date	DATE	Дата осеменения	
	Insemination_Method	VARCHAR(100)	Метод осеменения (естественный, искусственный)	
	Veterinarian_ID	INT	Идентификатор ветеринара, который проводил осеменение	
	Success	BOOLEAN	Успех осеменения (да/нет)	
	Load_Date	DATE	Дата загрузки записи в спутник	

На последнем этапе происходит проектирование модели витрины данных (рис. 2).

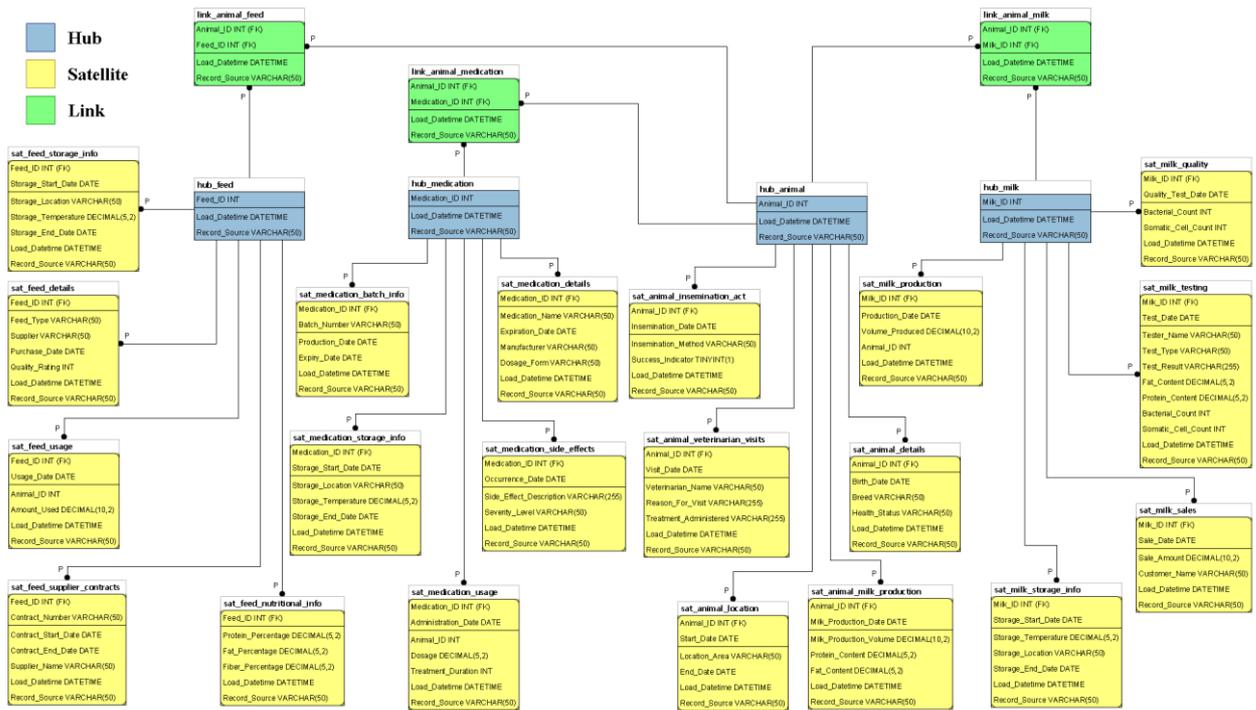


Рисунок 2. Модель данных Data Vault 2.0

Авторами выполнено информационно-логическое проектирование рассматриваемой витрины данных. Построение физической модели данных в СУБД Greenplum, наполнение витрины в наше исследование не входило и не выполнялось.

### 7.3 Применение информационных систем в сельском хозяйстве

В сельском хозяйстве, в животноводстве, в частности, используются различные информационные системы, предназначенные для автоматизации бизнес-процессов [14]. Это отраслевые и специализированные решения на платформе «1С: Предприятие», такие как «1С:Предприятие 8. ERP Агропромышленный комплекс», «1С: Цифровое животноводство. Оперативный учет и управление производством. КРС», «1С: Селекция в животноводстве КРС» и др.[15], федеральные государственные системы в области ветеринарии «Цербер», ВетИС и пр. [16, 17]. Эти системы широко применяются в отрасли, однако в основном они сосредоточены на обеспечение оперативного сбора информации, в т.ч. и с использованием RFID-датчиков, учете, контроле выполнения зоотехнических и ветеринарных мероприятий, хранении нормативно-справочной информации и работают с жестко фиксированными схемами данных, что ограничивает их гибкость и аналитические возможности.

Отдельного упоминания заслуживает Федеральная государственная информационная система ВетИС, разработанная Россельхознадзором, которая предназначена для обеспечения учета, контроля и управления в сфере ветеринарии на территории Российской Федерации [17]. Особенностью данной системы является ориентированность не только на сбор и хранения данных, а также на анализ информации и формирование отчетности. Такой подход отражает актуальную тенденцию применения методов анализа данных в сельском хозяйстве, включая использование кластеризации, других технологий искусственного интеллекта. Например, в материалах компании «СБЕР Про Цифровое развитие» представлены результаты внедрения «умных технологий» для оптимизации процессов в животноводстве, что позволяет предприятиям повысить эффективность. Согласно оценкам АНО «Цифровая экономика», применение ИИ способно увеличить объемы производства продукции растениеводства на 5%, а животноводства – на 3% [18].

Кроме того, особо значимым проектом в настоящее время является система управления стадом КРС «Молоко 2.0», курируемым Министерством сельского хозяйства РФ и внедряемая в компании «Тюменские молочные фермы»[19]. Данный программный продукт разрабатывается фирмой «Плино» при поддержке фонда «Сколково» в рамках национальной программы «Цифровая экономика РФ». Информационно-аналитическая система «Молоко 2.0» интегрируется с транзакционными информационными системами типа 1С:ERP, государственными ветеринарными информационными системами, программными продуктами, обеспечивающими работу доильного оборудования для

осуществления сбора информации, выполняет учет, анализ производства молока, в т.ч. и предиктивную аналитику [20]. Программный комплекс функционирует на основе колончатой системы управления базами данных ClickHouse с открытым кодом, позволяющей выполнять аналитические запросы в режиме реального времени на структурированных больших данных. На наш взгляд применение модели Data Vault 2.0 является перспективным направлением совершенствования организации информационных хранилищ данной информационно-аналитической системы.

В рамках настоящего исследования продемонстрированы преимущества применения модели Data Vault 2.0, позволяющей создавать эффективные витрины данных для анализа и прогнозирования. Эта модель интегрируется с инструментами искусственного интеллекта и машинного обучения, расширяя возможности существующих отечественных решений. Например, анализ больших данных с использованием методов кластеризации или прогнозирования заболеваний может не только повысить продуктивность животноводческих хозяйств, но и снизить риски благодаря раннему выявлению проблем и оптимизации управленческих решений.

Для работы с хранилищами данных, построенных по модели Data Vault, подходит массово-параллельная СУБД на основе PostgreSQL Greenplum [21]. Greenplum разработана на основе архитектуры MPP (Massively Parallel Processing), что позволяет распределять нагрузку между множеством серверов и выполнять параллельную обработку данных, а значит, данная СУБД хорошо подходит для реализации Data Vault. Для этого необходимо создать основные элементы хранилища данных (Хабы, Ссылки и Сателлиты), загрузку данных в Greenplum можно осуществить через различные ETL-инструменты, такие как Apache NiFi, Talend, или собственные инструменты Greenplum, например, gpload. Реализация point-in-time (PIT) таблиц в рассматриваемой СУБД делает Greenplum отличным средством для развертывания витрин данных [22].

Стоимость внедрения информационной системы включает расходы на приобретение самой системы и системы управления базами данных (СУБД). Например, цена решения «1С: Цифровое животноводство. Оперативный учет и управление производством. КРС» составляет 121 600 рублей, что является разовым платежом. Дополнительно стоимость лицензий зависит от размера стада: для фермы с 10 000 голов она составит до 515 300 рублей [15].

Для хранения и обработки данных в рамках традиционных подходов могут использоваться СУБД, такие как Oracle Database, стоимость которой составляет около 110 000 рублей в месяц для среднего предприятия. Однако для систем, использующих модель Data Vault 2.0, предпочтение может быть отдано СУБД Greenplum, которая стоит около 105 000 рублей в месяц [23, 24]. Greenplum предоставляет встроенную поддержку параллельной обработки больших объемов данных, что делает ее оптимальным выбором для построения хранилищ данных на основе Data Vault.

Таким образом, при использовании Data Vault 2.0 в сочетании с Greenplum затраты на поддержание системы остаются на уровне традиционных реляционных решений, но при этом достигается высокая гибкость и масштабируемость архитектуры. Это позволяет снизить общие издержки за счет оптимизированного хранения и обработки данных, а также уменьшить стоимость внедрения новых аналитических витрин и интеграции с современными инструментами анализа, такими как искусственный интеллект и машинное обучение.

#### 7.4 Пример проектирования витрины данных по реляционной модели

Для сравнения, приведем описание рассматриваемой предметной области с помощью реляционной модели. Описание структуры сущностей реляционной модели представлено в таб. 2.

Таблица 2. Описание структуры реляционной модели

Таблица	Атрибуты	Описание
Cows	Cow_ID (PK), Breed, Birth_Date, Weight, Location_ID (FK)	Информация о коровах, включая идентификатор, породу, дату рождения, вес и местоположение.
Locations	Location_ID (PK), Location_Type, Description	Описание местоположений (фермы, пастбища и т. д.) с указанием типа и характеристик.
Feed	Feed_ID (PK), Feed_Type, Supplier, Nutritional_Info_ID (FK)	Данные о кормах для скота, включая тип корма, поставщика и информацию о питательных свойствах.

Nutritional_Info	Nutritional_Info_ID (PK), Protein_Content, Fat_Content, Fiber_Content, Energy_Value	Содержит информацию о питательных свойствах кормов: белок, жир, клетчатка и энергетическая ценность.
Medications	Medication_ID (PK), Name, Type, Manufacturer, Expiration_Date	Сведения о медикаментах, включая название, тип, производителя и срок годности.
Milk_Production	Milk_Production_ID (PK), Cow_ID (FK), Production_Date, Quantity, Quality	Данные о производстве молока: корова, дата, количество и качество (жир, белок, бактерии).
Veterinarian_Visits	Visit_ID (PK), Cow_ID (FK), Visit_Date, Reason, Procedure	Информация о визитах ветеринара, включая причину визита и назначенные процедуры.
Insemination_Acts	Insemination_ID (PK), Cow_ID (FK), Insemination_Date, Method, Veterinarian_ID (FK), Success	Данные об актах осеменения: корова, дата, метод, ветеринар и результат.
Veterinarians	Veterinarian_ID (PK), Name, Specialization	Сведения о ветеринарах, включая их специализацию.
Feeding	Cow_ID (PK), Feed_ID (PK), DateTime	История кормления скота: корова, корм и дата кормления.
Milk_Sales	Sale_ID (PK), Milk_Production_ID (FK), Sale_Date, Buyer_ID (FK), Price_Per_Liter	Информация о продажах молока: произведённое молоко, дата продажи, покупатель и цена.
Buyers	Buyer_ID (PK), Name, Contact_Info	Данные о покупателях молока, включая контактную информацию.

ER-модель базы данных, описывающая бизнес-процесс производства молока, представлена на рис.

3.

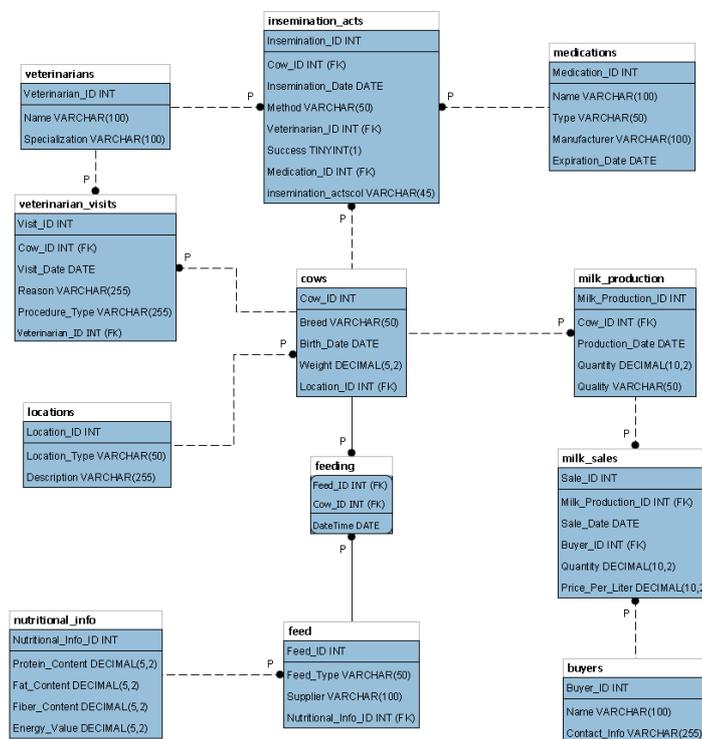


Рисунок 3. Реляционная модель данных

Данную модель данных можно рекомендовать для небольших сельскохозяйственных организаций, фермерских хозяйств, в которых незначителен объем поступающих новых данных, количество источников данных, их структурный состав остаются неизменными в течение длительного периода времени [25].

## 7.5 Сравнение реляционной модели и Data Vault 2.0

Выбор между реляционной моделью и Data Vault зависит от конкретных требований к системе. Реляционная модель лучше подходит для управления транзакциями и операционных систем, небольших хранилищ данных, тогда как Data Vault 2.0 – для аналитических хранилищ с высокими требованиями к историзации и масштабируемости.

В таб. 3 приведен сравнительный анализ классической реляционной модели данных и модели Data Vault 2.0 для информационных хранилищ аналитических комплексов организации.

Таблица 3. Сравнительный анализ моделей

Критерий	Реляционная модель	Data Vault 2.0
Простота проектирования	Простой и понятный подход	Сложное проектирование, требует опыта
Гибкость к изменениям	Ограниченная гибкость	Высокая гибкость, легко добавлять новые сущности
Историзация данных	Требует дополнительных решений	Поддерживается изначально через сателлиты
Масштабируемость	Проблемы с масштабируемостью	Хорошо масштабируется для больших объемов данных
Управление зависимостями	Требует внешних ключей и сложных зависимостей	Управление через хабы, ссылки и сателлиты
Оптимальность транзакционных операций	Хорошая производительность для транзакций	Неоптимальные запросы для оперативных операций
Поддержка целостности данных	Строгая поддержка через ограничения	Меньше ограничений, фокус на гибкости

## Заключение

Таким образом, применение модели данных Data Vault 2.0 для организации информационного хранилища в аналитическом комплексе сельскохозяйственной организации предоставляет ряд преимуществ, таких как гибкость, масштабируемость и возможность интеграции данных из различных источников, в т. ч. поступающих в потоковом режиме. Использование подхода Data Vault 2.0 позволяет эффективно организовать хранение данных, обеспечивая целостность и структурированность данных благодаря разделению на хабы, ссылки и спутники. Возможно применение модели данных Data Vault 2.0 не только в отдельных информационных хранилищах, но и в Data Warehouse как элементе организационной архитектуры Data LakeHouse аналитического комплекса организации.

## Благодарности

Работа выполнена в рамках гранта РЭУ имени Г. В. Плеханова «Совершенствование методов моделирования и организации систем управления информацией для реализации требований экономики данных».

## Литература

1. Указ Президента Российской Федерации от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы» [Электронный ресурс]. URL: <http://static.kremlin.ru/media/acts/files/0001201705100002.pdf> (дата обращения: 12.10.2024).
2. Национальный проект «Экономика данных» [Электронный ресурс]. URL: <http://government.ru/rugovclassifier/909/events/> (дата обращения: 12.10.2024).
3. Глава Минцифры Максют Шадаев на CNews FORUM – о развитии космических технологий, сроках импортозамещения и судьбе российских ИТ-решений за рубежом [Электронный ресурс]. URL: [https://biz.cnews.ru/articles/2023-11-09\\_glava\\_mintsifry\\_maksut\\_shadaev\\_na\\_cnews](https://biz.cnews.ru/articles/2023-11-09_glava_mintsifry_maksut_shadaev_na_cnews) (дата обращения: 12.10.2024).

4. Руководство данными. Конспект на русском языке [Электронный ресурс]. URL: <https://dataliteracy.ru/dmbok-notes-3> (дата обращения: 12.10.2024).
5. Arenadata: «Технологический цикл решений для обработки данных» [Электронный ресурс]. URL: [https://www.youtube.com/watch?v=i\\_3\\_DdB4pk4&list=PLThlRr26UjnPchbO0fhIJasSe5QTICgYH](https://www.youtube.com/watch?v=i_3_DdB4pk4&list=PLThlRr26UjnPchbO0fhIJasSe5QTICgYH) (дата обращения: 12.10.2024).
6. Пять подходов к построению современной платформы данных [Электронный ресурс]. URL: <https://arenadata.tech/about/blog/5-ways-to-dwh/> (дата обращения: 12.10.2024).
7. Горожанкин В. Data Mesh вместо информационных «бункеров»: как управлять данными с пользой для бизнеса [Электронный ресурс]. URL: <https://rb.ru/opinion/data-mesh/> (дата обращения: 12.10.2024).
8. Зенкова М. Хранилища данных. Обзор технологий и подходов к проектированию [Электронный ресурс]. URL: <https://systems.education/dwh-technologies-and-design> (дата обращения: 12.10.2024).
9. Вичугова А. Проектирование хранилища данных с методологией Data Vault в архитектуре Lakehouse [Электронный ресурс]. URL: <https://bigdataschool.ru/blog/data-vault-for-lakehouse.html> (дата обращения: 12.10.2024).
10. Проектирование DWH с помощью Data Vault // Хабр [Электронный ресурс]. URL: <https://habr.com/ru/companies/otus/articles/816219/> (дата обращения: 12.10.2024).
11. Чернышева К.В., Карпузова Н.В., Афанасьева С.И., Королькова А.П. Использование хранилищ данных в экономике и менеджменте АПК. //Техника и оборудование для села. 2022. № 4 (298). С. 44-48.
12. Как строить Data Vault. Построение модели Data Vault: пошаговая инструкция и рекомендации [Электронный ресурс]. URL: <https://telegra.ph/Kak-stroit-Data-Vault-Postroenie-modeli-Data-Vault-roshagovaya-instrukciya-i-rekomendacii-06-19> (дата обращения: 12.10.2024).
13. Чернышева К. В. Информационное обеспечение управления отраслью АПК: молочное скотоводство: дис. канд. Экономика и управление народным хозяйством наук: 08.00.05. М., 2007. 218 с.
14. Малетин В. Применение автоматизированных систем в отрасли животноводства // Экономика и социум. 2019. №3(58). С. 280-283.
15. 1С:Цифровое животноводство. Оперативный учет и управление производством. КРС // Отраслевые и специализированные решения 1С:Предприятие URL: <https://solutions.1c.ru/catalog/mes-krs/buy> (дата обращения: 26.11.2024).
16. Цербер Реестр поднадзорных объектов // Цербер URL: <https://cerberus.vetrf.ru/cerberus/> (дата обращения: 26.11.2024).
17. ВетИС Федеральная государственная информационная система в области ветеринарии // Россельхознадзор Федеральная служба по ветеринарному и фитосанитарному надзору URL: <https://vetrf.ru/> (дата обращения: 26.11.2024).
18. Железный нянь. Как ИИ помогает животноводам // Сбер Про Цифровое развитие URL: <https://sber.pro/digital/publication/zhelezni-nyan-kak-ii-pomogaet-zhivotnovodam/> (дата обращения: 26.11.2024).
19. ОЗП «МОЛОКО 2.0: СИСТЕМА УПРАВЛЕНИЯ СТАДОМ КРС» // ЦЕНТР ЦИФРОВОЙ ТРАНСФОРМАЦИИ В СФЕРЕ АПК URL: <https://cctmcx.ru/o-tsentre/novosti/3214/> (дата обращения: 27.11.2024).
20. МОЛОКО 2.0 // Плинор URL: <https://plinor.ru/wp-content/uploads/2023/07/moloko-2.0-sistema-upravleniya-stadom-krs.pdf> (дата обращения: 27.11.2024).
21. Greenplum DB // Хабр [Электронный ресурс]. URL: <https://habr.com/ru/companies/tbank/articles/267733/> (дата обращения: 12.10.2024).
22. DataVault на Greenplum с помощью DBT // Хабр [Электронный ресурс]. URL: <https://habr.com/ru/articles/671836/> (дата обращения: 12.10.2024).
23. Oracle Technology Global Price List // Oracle URL: <https://www.oracle.com/in/a/ocom/docs/corporate/pricing/technology-price-list-070617.pdf> (дата обращения: 26.11.2024).
24. Managed Service for Greenplum® pricing policy // Yandex Cloud URL: [https://yandex.cloud/en/docs/managed-greenplum/pricing/?utm\\_referrer=https%3A%2F%2Fwww.google.com%2F](https://yandex.cloud/en/docs/managed-greenplum/pricing/?utm_referrer=https%3A%2F%2Fwww.google.com%2F) (дата обращения: 26.11.2024).
25. Qlever [Электронный ресурс]. URL: <https://www.qlever.ru> (дата обращения: 12.10.2024).

# USING THE DATA VAULT 2.0 DATA MODEL IN THE INFORMATION STORAGE OF THE ANALYTICAL COMPLEX OF AGRICULTURAL ORGANIZATION

**Chernysheva, Kira Vladimirovna**

*Candidate of economic sciences*

*Plekhanov Russian State University of Economics, associate professor*

*Moscow, Russian Federation*

*chernysheva.kv@rea.ru*

**Goroshkina, Ulyana Evgenievna**

*Plekhanov Russian State University of Economics, student*

*Moscow, Russian Federation*

*u.goroshkina@gmail.com*

## Abstract

*The article presents the results of the author's research of existing approaches to the organization of information repositories of analytical systems used by data models. The Data Vault 2.0 data model and the possibilities of its application in the analytical complex of an agricultural organization in the zootechnical service, are considered. The essences (hubs) characterizing the business process of milk production are considered: biological object, feed, veterinary measures, finished products, their connections and satellites.*

## Keywords

*digital economy; data economy; information and analytical systems; information storages; data models; Data Vault 2.0; hub; essence; communication*

## References

1. Ukaz Prezidenta Rossiiskoi Federatsii ot 09.05.2017 no 203 «O Strategii razvitiia informatsionnogo obshchestva v Rossiiskoi Federatsii na 2017–2030 gody». URL: <http://static.kremlin.ru/media/acts/files/0001201705100002.pdf> (accessed on 12.10.2024).
2. Natsional'nyi proekt «Ekonomika dannykh». URL: <http://government.ru/rugovclassifier/909/events/> (accessed on 12.10.2024).
3. Glava Mintsifry Maksut SHadaev na CNews FORUM – o razvitii kosmicheskikh tekhnologii, srokhakh importozameshcheniia i sud'be rossiiskikh IT-reshenii za rubezhom. URL: [https://biz.cnews.ru/articles/2023-11-09\\_glava\\_mintsifry\\_maksut\\_shadaev\\_na\\_cnews](https://biz.cnews.ru/articles/2023-11-09_glava_mintsifry_maksut_shadaev_na_cnews) (accessed on 12.10.2024).
4. Rukovodstvo dannymi. Konspekt na rusском iazyke. URL: <https://dataliteracy.ru/dmbok-notes-3> (accessed on 12.10.2024).
5. Arenadata: «Tekhnologicheskii tsikl reshenii dlia obrabotki dannykh». URL: [https://www.youtube.com/watch?v=i\\_3\\_DdB4pk4&list=PLThlRr26UjnPchbO0fhIJasSe5QTICgYH](https://www.youtube.com/watch?v=i_3_DdB4pk4&list=PLThlRr26UjnPchbO0fhIJasSe5QTICgYH) (accessed on 12.10.2024).
6. Piat' podkhodov k postroeniiu sovremennoi platformy dannykh. URL: <https://arenadata.tech/about/blog/5-ways-to-dwh/> (accessed on 12.10.2024).
7. Gorozhankin V. Data Mesh vmesto informatsionnykh «bunkerov»: kak upravliat' dannymi s pol'zoi dlia biznesa. URL: <https://rb.ru/opinion/data-mesh/> (accessed on 12.10.2024).
8. Zenkova M. Khranilishcha dannykh. Obzor tekhnologii i podkhodov k proektirovaniu. URL: <https://systems.education/dwh-technologies-and-design> (accessed on 12.10.2024).
9. Vichugova A. Proektirovanie khranilishcha dannykh s metodologiei Data Vault v arkhitekture Lakehouse. URL: <https://bigdataschool.ru/blog/data-vault-for-lakehouse.html> (accessed on 12.10.2024).
10. Proektirovanie DWH s pomoshch'iu Data Vault // KHabr. URL: <https://habr.com/ru/companies/otus/articles/816219/> (accessed on 12.10.2024).
11. Chernysheva K.V., Karpuzova N.V., Afanas'eva S.I., Korol'kova A.P. Ispol'zovanie khranilishch dannykh v ekonomike i menedzhmente APK. // Tekhnika i oborudovanie dlia sela. 2022. No 4 (298). S. 44-48.

12. Kak stroit' Data Vault. Postroenie modeli Data Vault: poshagovaia instruktsiia i rekomendatsii. URL: <https://tegra.ph/Kak-stroit-Data-Vault-Postroenie-modeli-Data-Vault-poshagovaya-instrukciya-i-rekomendacii-06-19> (accessed on 12.10.2024).
13. Chernysheva K.V. Informatsionnoe obespechenie upravleniia otrasl'iu APK: molochnoe skotovodstvo: dis. kand. Ekonomika i upravlenie narodnym khoziaistvom nauk: 08.00.05. M., 2007. 218 p.
14. Maletin V. Primenenie avtomatizirovannykh sistem v otrasli zhivotnovodstva // Ekonomika i sotsium. 2019. No 3(58). S. 280-283.
15. 1S: Tsifrovoe zhivotnovodstvo. Operativnyi uchet i upravlenie proizvodstvom. KRS // Otrasleyve i spetsializirovannye resheniia 1S: Predpriatie URL: <https://solutions.1c.ru/catalog/mes-krs/buy> (accessed on 26.11.2024).
16. Tserber Reestr podnadzornykh obektov // Tserber URL: <https://cerberus.vetrf.ru/cerberus/> (accessed on 26.11.2024).
17. VetIS Federal'naia gosudarstvennaia informatsionnaia sistema v oblasti veterinarii // Rossel'khoznadzor Federal'naia sluzhba po veterinarnomu i fitosanitarnomu nadzoru URL: <https://vetrf.ru/> (accessed on 26.11.2024).
18. Zheleznyi nian'. Kak II pomogaet zhivotnovodam // Sber Pro TSifrovoe razvitie URL: <https://sber.pro/digital/publication/zhelezni-nyan-kak-ii-pomogaet-zhivotnovodam/> (accessed on 26.11.2024).
19. OZP «Moloko 2.0: sistema upravleniia stadom KRS» // tsentr tsifrovoi transformatsii v sfere APK. URL: <https://cctmcx.ru/o-tsentre/novosti/3214/> (accessed on 27.11.2024).
20. MOLOKO 2.0 // Plinor URL: <https://plinor.ru/wp-content/uploads/2023/07/moloko-2.0-sistema-upravleniya-stadom-krs.pdf> (accessed on 27.11.2024).
21. Greenplum DB // KHabr. URL: <https://habr.com/ru/companies/tbank/articles/267733/> (accessed on 12.10.2024).
22. DataVault na Greenplum s pomoshch'iu DBT // KHabr. URL: <https://habr.com/ru/articles/671836/> (accessed on 12.10.2024).
23. Oracle Technology Global Price List // Oracle URL: <https://www.oracle.com/in/a/ocom/docs/corporate/pricing/technology-price-list-070617.pdf> (accessed on 26.11.2024).
24. Managed Service for Greenplum® pricing policy // Yandex Cloud URL: [https://yandex.cloud/en/docs/managed-greenplum/pricing/?utm\\_referrer=https%3A%2F%2Fwww.google.com%2F](https://yandex.cloud/en/docs/managed-greenplum/pricing/?utm_referrer=https%3A%2F%2Fwww.google.com%2F) (accessed on 26.11.2024).
25. Qlever. URL: <https://www.qlever.ru> (accessed on 12.10.2024).