

ОРГАНИЗАЦИЯ СИСТЕМ ХРАНЕНИЯ ДАННЫХ В СЕЛЬСКОХОЗЯЙСТВЕННОЙ ОРГАНИЗАЦИИ

Чернышева К.В., Афанасьева С.И., Королькова А.П., Горошкина У.Е.

Чернышева Кира Владимировна, Афанасьева Светлана Ильинична

Российский экономический университет имени Г.В. Плеханова,

Россия, 115054, г. Москва, Стремянный переулок, д. 36

Королькова Антонина Павловна

Российский научно-исследовательский институт информации и технико-экономических исследований по инженерно-техническому обеспечению агропромышленного комплекса",

Россия, 141261, Московская область, Пушкинский район, р.п. Правдинский, ул. Лесная, д. 60

Горошкина Ульяна Евгеньевна

Национальный исследовательский университет "Высшая школа экономики"

Россия, 109028, г. Москва, Покровский бульвар, д. 11

E-mail: CHernysheva.KV@rea.ru, Afanaseva.SI@rea.ru, 52_kap@mail.ru, u.goroshkina@yandex.ru

В статье изложены результаты авторских исследований организации систем хранения данных в сельскохозяйственных организациях на разных уровнях управления, от оперативного до стратегического. Предложены методологические подходы к выбору систем хранения данных в зависимости от класса решаемых задач, вида обрабатываемой информации. Изложены рекомендации по проектированию информационных хранилищ для крупных сельскохозяйственных организаций и агрохолдингов, сельскохозяйственных организаций, не имеющих собственных перерабатывающих мощностей, крестьянских и фермерских хозяйств в соответствии с объемом поступающих новых данных, текучестью источников данных и изменчивостью их метаданных.

Ключевые слова: цифровая трансформация, хранилища и витрины данных, модели данных, Data Vault, якорная модель, «Снежинка», «Звезда»

ORGANIZATION OF DATA STORAGE SYSTEMS IN AN AGRICULTURAL ORGANIZATION

Chernysheva K.V., Afanasyeva S.I., Korolkova A.P., Goroshkina U.E.

Chernysheva Kira Vladimirovna, Afanasyeva Svetlana Ilyinichna

Plekhanov Russian University of Economics,

Russia, 115054, Moscow, Stremyanny Lane, 36

Korolkova Antonina Pavlovna

Russian Research Institute of Information and Technical and Economic Research on Engineering and Technical Support of the Agro-Industrial Complex,

Russia, 141261, Moscow, Moscow Region, Pushkinsky District, Pravdinsky Settlement, Lesnaya Street, 60

Goroshkina Ulyana Evgenievna

National Research University Higher School of Economics

Russia, 109028, Moscow, Pokrovsky Boulevard, 11

E-mail: CHernysheva.KV@rea.ru, Afanaseva.SI@rea.ru, 52_kap@mail.ru, u.goroshkina@yandex.ru

The article presents the results of the author's research on the organization of data storage systems in agricultural organizations at different levels of management, from operational to strategic. Methodological approaches to the choice of data storage systems are proposed, depending on the class of tasks being solved and the type of information being processed. Recommendations on the design of information repositories for large agricultural organizations and agricultural holdings, agricultural organizations that do not

have their own processing facilities, peasant and farm farms are presented in accordance with the volume of incoming new data, the fluidity of data sources and the variability of their metadata.

Keywords: digital transformation, data warehouses and storefronts, data models, Data Vault, anchor model, "Snowflake", "Star"

Для цитирования:

Чернышева К.В., Афанасьева С.И., Королькова А.П., Горошкина У.Е. Организация систем хранения данных в сельскохозяйственной организации. *Современные наукоёмкие технологии. Региональное приложение*. 2026. № 1. С. 89-99. DOI: 10.6060/snt.20268501.00010

For citation:

Chernysheva K.V., Afanasyeva S.I., Korolkova A.P., Goroshkina U.E. Organization of data storage systems in an agricultural organization. *Modern high technology. Regional application*. 2026. N 1. P. 89-99. DOI: 10.6060/snt.20268501.00010

ПОСТАНОВКА ПРОБЛЕМЫ

Современное управление в соответствии с новым национальным проектом «Экономика данных и цифровая трансформация государства» должно стать управлением на основе данных, формирующихся в цифровых системах организации [1]. Причем цифровизация касается всех процессов преобразования информации, включая сбор, передачу, обработку, хранение, распространение и предоставление.

Так, сбор данных выполняется, в том числе и с применением высокочувствительных датчиков, их передача в режиме реального времени, хранение с использованием отечественных облачных платформ, центров обработки данных, а обработка и анализ данных с использованием технологий искусственного интеллекта на базе отечественных платформ и сервисов.

В рамках стратегического направления в области цифровой трансформации отраслей агропромышленного и рыбохозяйственного комплексов РФ на период до 2030 года реализуется проект «Цифровая трансформация агропромышленного и рыбохозяйственного комплексов Российской Федерации» [2]. Отмечается, что в настоящее время отсутствует единый подход к сбору, верификации и хранению данных, позволяющих организовать принятие качественных управленческих решений в отрасли. Результатом указанного проекта должно явиться единообразие подходов к взаимодействию с участниками информационного обмена, формирование единой цифровой платформы с унифицированным подходом к сбору, верификации и хранению данных, позволяющих организовать принятие качественных управленческих решений для органов управления отраслью, рост производительности хозяйствующих субъектов, в том числе из-за наличия программ поддержки, доступности к данным для принятия управленческих решений [2].

Предусматривается создание отраслевого центра обработки данных (ЦОД) на базе создаваемого Ситуационного цифрового центра Министерства сельского хозяйства РФ, связанного сверхвысокоскоростными каналами передачи данных с сельхозтоваропроизводителями.

Исходя из вышесказанного, совершенствование организации систем хранения данных разных видов, объемов, поступающих из различных источников для решения разных классов экономических задач на всех уровнях управления организацией, становится особенно актуальным. Рассматривается использование транзакционных и аналитических информационных систем, реляционных и нереляционных баз данных (в т.ч. графовых и документальных), а также хранилищ данных, построенных на моделях «Звезда», «Снежинка», Data Vault, якорной модели, выполненных на базе отечественных программных решений.

Цель исследования – разработка методологического подхода к выбору систем хранения данных в зависимости от уровня управления, класса решаемых задач, вида обрабатываемой информации; формирование рекомендаций по проектированию витрин и хранилищ данных для сельскохозяйственных организаций различной специализации, размера, в зависимости от объема поступающих новых данных, текучести источников данных и изменчивости их метаданных.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Исследования выполнялись на основе данных Росстата, нормативно-правовых документов, справочных материалов, личных наблюдений авторов с применением основных положений системного подхода. Проведен анализ подходов в организации систем хранения данных (реляционных и нереляционных (NoSQL) баз данных, хранилищ данных), систем обработки данных (транзакционных и аналитических) на разных уровнях управления сельскохозяйственной организации,

обоснован выбор критериев для использования систем обработки и систем хранения. Рассмотрены методологические подходы и разработаны рекомендации информационно-логического проектирования реляционных витрин и хранилищ данных по разным моделям («Звезда», «Снежинка», Data Vault, якорная модель), физического проектирования на основе отечественных систем управления базами данных.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ И ОБСУЖДЕНИЕ

Современное сельское хозяйство является динамично развивающейся отраслью, активно использующей современные цифровые технологии. Так, по данным Федеральной службы государственной статистики, число организаций, использовавших цифровые технологии, увеличилось в 2023 году по сравнению с 2022 годом на 0,7% и составило 5738 единиц [3]. Почти на 30% возросло количество хозяйств, использующих различные

цифровые платформы (916 единиц в 2023 году и 709 в 2022 году), цифровые двойники (96 и 75 единиц соответственно). Многие организации используют электронный обмен данными между своими и внешними информационными системами по форматам обмена для отправки или получения данных в органы государственной власти, местного самоуправления, геоинформационные системы, технологии радиочастотной идентификации объектов (RFID) для сбора данных [3].

В настоящее время пользователи должны использовать отечественные программные продукты, в т.ч. российское программное обеспечение для разработки и проектирования баз данных информационных систем.

В табл. 1 приведено количество сельскохозяйственных организаций, использующих отечественное программное обеспечение для ЭВМ и баз данных (по классам программ) [3].

Таблица 1

Число организаций, использовавших российское программное обеспечение для ЭВМ и баз данных (по классам программ)
Table 1. Number of organizations using Russian computer and database software (by program class)

Показатели	2022 г.	2023 г.	2023 к 2022, %
Число организаций, использовавших цифровые технологии (зарубежного и российского производства), всего	5 696	5 738	100,74
Число организаций, использовавших специальные программные средства российского производства, всего	4402	4244	96,41
Средства управления базами данных	355	360	101,41
Средства хранения данных	207	223	107,73
Предметно-ориентированные информационные базы данных (EDW - Enterprise Data Warehouse)	2	4	200,00

Источник: Составлено авторами

Наблюдается некоторый рост количества организаций, использующих средства хранения и управления базами данных, но само число таких организаций по сравнению с общим количеством сельскохозяйственных организаций, пользующихся цифровыми технологиями, остается очень низким. Отсутствие единой системы хранения данных, долгий процесс получения оперативной информации указываются как одна из проблем использования систем аналитики данных в сельском хозяйстве [4-6]. Системы сбора информации представлены транзакционными системами типа учетных бухгалтерских, логистических систем, систем управления взаимоотношениями с клиентами, систем управления складом, персоналом, корпоративных информационных систем и пр., а также

системами управления технологическими и производственными процессами типа SCADA, MES-системами и др. Перечисленные транзакционные системы функционируют на оперативном уровне управления и накапливают данные транзакций с переменной информацией в реляционных базах данных, работающих с хорошо структурированными данными. Однако, в эпоху Big Data возрастает объем неструктурированных или слабоструктурированных данных, поступающих в потоковом режиме с использованием технологий радиочастотной идентификации объектов (RFID), датчиков промышленного интернета и пр. Это вызывает необходимость использования нереляционных (NoSQL) баз данных, в т.ч. графовых и документоориентированных.

В графовых базах данных информационные объекты представляются в виде узлов, а ребра отражают связи между ними. Каждый узел характеризуется определенным набором свойств, и ребра имеют свойства, характеризующие связи между узлами [7]. Графовые базы данных эффективно работают с взаимосвязанными данными разных типов, что определяет их использование для моделирования сложных отношений.

Специфика использования графовых баз данных в сельском хозяйстве заключается в том, что они могут быть использованы для описания сложных взаимосвязей как непосредственно процесса производства продукции, связанным с биологическим объектом (растением, животным), при отражении геномов, молекулярных взаимодействий; так и в логистике поставок, маршрутов реализации скоропортящейся сельскохозяйственной продукции; маркетинговых исследованиях данных социальных сетей по выявлению предпочтений пользователей и пр. Кроме того, важна последующая интеграция графовых баз данных с графовой аналитикой, машинным обучением, базами знаний для выявления закономерностей и принятия решений.

Документные (документальные, документоориентированные) базы данных, работающие с неструктурированными данными, хранящие информацию в виде документов форматов, аналогичных JSON (например, XML, JSON, BSON), имеющие иерархическую структуру по коллекциям, в управлении агропромышленным комплексом практически не используются. На уровне управления отрасли используется база данных полнотекстовых информационных ресурсов по наилучшим доступным технологиям в сфере сельского хозяйства (ДБД НДТ в АПК), база данных «Инженерно-техническое обеспечение АПК», реализованные в Федеральном государственном бюджетном научном учреждении «Российский научно-исследовательский институт информации и технико-экономических исследований по инженерно-техническому обеспечению агропромышленного комплекса» [8]. Данные базы реализованы на основе системы автоматизации библиотек ИРБИС 64, позволяют обеспечить структурированное представление данных и анализ полнотекстовых информационных ресурсов. Указывается, что данные базы являются документальными, но, на наш взгляд, это фактографические базы данных, работающие со структурированной информацией. Также используется фактографическая база данных «Машины и оборудование для сельскохозяйственного производства», информацию

по названию машины и оборудования, о его марке и модификации, конструктивным особенностям, технической характеристике, адресам изготовителей и разработчиков, источнике информации [9]. Однако на уровне сельскохозяйственной организации и фактографические, и документные базы данных в настоящее время не используются. На наш взгляд, возможно использование документных баз данных непосредственно в хозяйствах, особенно в крупных агрохолдингах и крупных фермерских хозяйствах. Это могут быть специализированные коллекции документов по отдельным производственным бизнес-процессам, а также локальным документам организации.

Кроме того, использование документных баз данных необходимо для проведения аналитики большого объема фрагментов текстовых данных. Таким образом, на оперативном уровне управления сельскохозяйственной организации рекомендуется использование реляционных систем хранения хорошо структурированных данных для использования в транзакционных информационных системах, графовых баз данных для описания сложных взаимосвязей производственных процессов, документных баз данных для хранения неструктурированной и слабоструктурированной информации. Применение перечисленных систем хранения зависит от размера и специализации хозяйства. Крупным агрохолдингам, имеющим распределенную территориальную инфраструктуру, собственные перерабатывающие мощности, осуществляющих сложный производственный процесс, работающим с разными видами информации, рекомендуется использование как реляционных, так и графовых, и документных баз данных.

Предметно-ориентированные информационные базы данных (EDW - Enterprise Data Warehouse), хранилища данных, рекомендуется использовать на тактическом и стратегическом уровнях управления организацией, а также на уровне муниципальных образований управления отраслью (район, область).

В едином информационном пространстве организации источниками данных для хранилища сельскохозяйственной организации являются информационные базы транзакционных систем, офисные документы, электронные архивы, расположенные как внутри организации, так и во внешнем окружении, а также данные, поступающие непосредственно с RFID-меток, элементов IoT в процессе управления технологическим процессом производства. Архитектура данных хранилища выбирается и проектируется с учетом общекорпоративной модели данных (Enterprise Data Model,

EDM), включающей концептуальные, логические и физические модели сущностей и связей для предметных областей организации, описание их взаимосвязей (диаграммы потоков данных), а также методику возможного расширения и адаптации, готовые модели анализа бизнеса, типовые алгоритмы трансформации данных в типовые структуры модели [10]. Как правило, централизованное информационное хранилище строится в виде размерностной модели типа «Звезда» или «Снежинка», с пакетной загрузкой структурированных данных баз данных транзакционных систем, файлов различных форматов, данных Интернет-источников и пр. [11].

Фактами (мерами) в таких хранилищах могут быть стоимостные и натуральные показатели, такие, как себестоимость и выручка, объемы производства и реализации продукции и пр.

Измерениями являются содержательные (кто, что), временные (когда), пространственные (где), причинные (почему, как) и прочие характеристики объекта. Централизованное хранилище может содержать витрины по подразделениям организации [12].

Приведем описание предметной области производства молока с помощью реляционной модели. Описание структуры сущностей реляционной модели витрины данных представлено в табл. 2.

Таблица 2

Пример описания элементов реляционной модели
Table 2. An example of a description of relational model elements

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 с указанием типа и характеристик.
Feed	Feed_ID (PK), Feed_Type, Supplier, Nutritional_Info_ID (FK)	Данные о кормах, 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: 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: произведенное молоко, дата продажи, покупатель и цена.
Buyers	Buyer_ID (PK), Name, Contact_Info	Данные о покупателях молока, включая контактную информацию.

Источник: Составлено авторами

ER-модель базы данных, описывающая бизнес-процесс производства молока, представлена на рис. 1.

Такой подход, характеризующийся простотой и устойчивостью модели хранения данных, их избыточностью, целесообразно использовать для небольших и средних сельскохозяйственных

организаций, фермерских хозяйств, функционирующих по линейной или линейно-функциональной структуре управления, в которых незначителен объем поступающих новых данных, количество источников данных, их структурный состав остаются неизменными в течение длительного периода времени.

В качестве технологий хранения в централизованных хранилищах схем «Звезда» и «Снежинка» возможно использование Single-Node Processing баз данных типа Oracle, SQL Server и др. Однако, недостатком модели, ограничивающей ее использование в эпоху Big Data, является

ограниченное горизонтальное масштабирование, заключающееся в сложности добавления новых объектов при изменениях в общекорпоративной модели данных, а также работа только со структурированными данными [13].

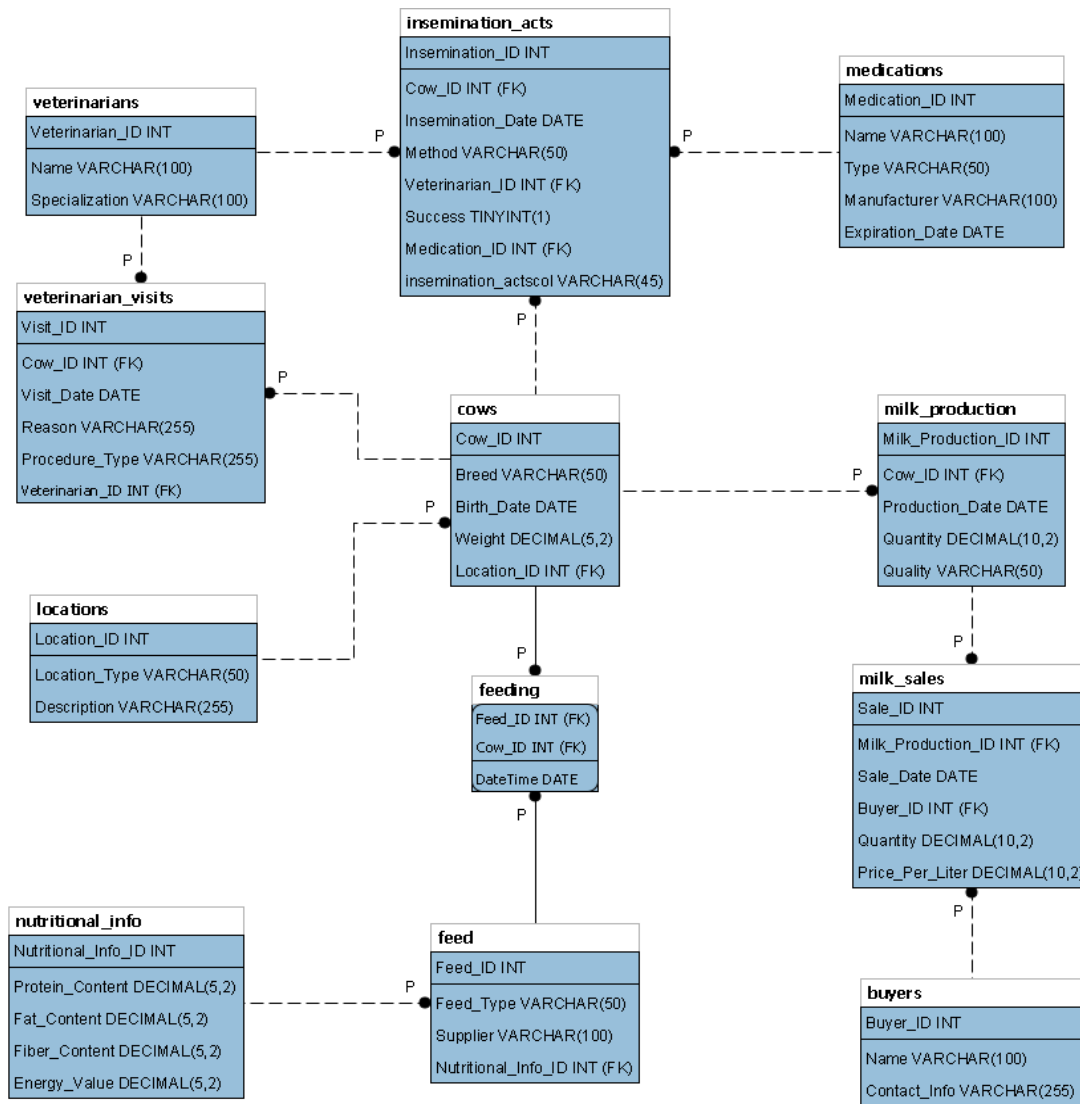


Рис. 1. Реляционная модель данных
Fig. 1. Relational data model

Расширением реляционного подхода к моделям данных информационных хранилищ являются модели Data Vault 2.0 и якорная модель (Anchor Modeling) [13].

В модели Data Vault 2.0 главным элементом является хаб (Hub), в котором хранится основная информации о сущностях хранилища данных. Определение Хабов происходит на основании бизнес-сущностей. Так, для проектирования витрины данных бизнес-процесса производства молока в нотации Data Vault 2.0 возможно выде-

ление следующих хабов: биологические объекты, корма, медикаменты, готовая продукция (молоко). Дополнительными хабами в проектируемой модели данных могут быть хабы о персонале, технологическом оборудовании, затратах на продукцию и др. Хабы соединяются между собой связью «многие-ко-многим» через таблицы-ссылки (Link) по суррогатным ключам. Ссылки определяются через выявление возможных отношений между бизнес-объектами [14].

Моделирование контекста каждой бизнес-сущности и отношений между ними предопределяют таблицы Спутники (Satellite). Предлагается отдельно выделять Сателлиты с условно-постоянной информацией, не меняющейся длительное время (порода, дата рождения животного и др.), и Сателлиты с переменной информацией, содержащейся в первичных документах (например, ведо-

мость учета расхода кормов, отражающая переменную информацию по расходу кормов).

В табл. 3 приведено описание структуры Хаба биологического объекта, связанных с ним Сателлитов, описывающих детальную информацию о биологическом объекте, а также Связи с Хабом корма.

Таблица 3

Пример описание элементов модели Data Vault 2.0
Table 3. Example description of Data Vault 2.0 model elements

Название элемента	Атрибуты	Тип данных	Связи	Описание
Hub_Animal	Animal_ID	INT	Sat_Animal_Details Sat_Animal_Insemination_Act Sat_Animal_Milk_Production Sat_Animal_Veterinarian_Visits Sat_Animal_Location Link_Animal_Feed Link_Animal_Medication Link_Animal_Milk	Хранит уникальные идентификаторы коров
	Load_Date	DATE		Дата загрузки
	Record_Source	VARCHAR(255)		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 сателлит
	Record_Source	VARCHAR(255)		Источник данных
Sat_Animal_Insemination_Act	Animal_ID	INT	Hub_Animal	Уникальный идентификатор коровы
	Insemination_Date	DATE		Дата осеменения
	Insemination_Method	VARCHAR(100)		Метод осеменения
	Veterinarian_ID	INT		Идентификатор ветеринара, который проводил осеменение
	Success	BOOLEAN		Успех осеменения (да/нет)
	Load_Date	DATE		Дата загрузки записи в сателлит

Источник: Составлено авторами

Hub_Animal: Хаб для уникальных идентификаторов животного, связанный с сателлитами, которые содержат данные о породе, местополо-

жении, 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 животном, такие как порода, дата рождения, вес и пол;

Sat_Cows_Insemination_Act: Сателлит, содержащий данные о проведенных 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.

На рис. 2 приведена модель витрины данных Data Vault 2.0.

Спецификой модели является возможность параллельной загрузки данных в таблицы, поскольку они не зависят друг от друга, что важно при появлении систем массивно-параллельной обработки (MPP) типа Teradata, Vertica, Greenplum и др. [15, 16]. Кроме того, еще одной особенностью Data Vault 2.0 от других подходов к проектированию хранилищ данных является использование ETL (Extract, Load, Transform) технологии, когда данные загружаются из источников без преобразований, трансформация данных может выполнена позже.

Модель данных Data Vault 2.0 применима для крупных организаций, работающих со структурированной и слабоструктурированной информацией, с учетом характеристик источников. К характеристикам источников данных можно отнести их объем, изменчивость количества и структуры (добавления или удаления полей) [17].

В зависимости от названных характеристик источников компания «Qlever Solutions» предлагает алгоритм выбора модели хранилища. Если объем новых данных в организации незначителен (менее 200 Гбайт в месяц), их структура не меняется в течение длительного времени, текучесть источников также незначительна, то рекомендуется модель «Снежинка».

Модель Data Vault следует использовать при увеличении объема новых данных до 500 Гбайт в месяц и обновлении структуры не более одного раза в полугодие.

Если же объем новых данных в организации превышает 500 Гбайт в месяц и структура источников меняется чаще одного раза в полугодие, то рекомендуется якорная модель данных информационного хранилища Anchor Modeling.

Данную методику можно рекомендовать и для организаций агропромышленного комплекса. Так, например, внедряемая в настоящее время в компании «Тюменские молочные фермы» система управления стадом КРС «Молоко 2.0» функционирует на основе колончатой системы управления базами данных ClickHouse с открытым кодом, позволяющей выполнять аналитические запросы в режиме реального времени на структурированных больших данных [18].

Применение модели Data Vault 2.0 является перспективным направлением совершенствования организации информационных хранилищ данной информационно-аналитической системы.

Также выбор модели данных информационного хранилища определяется целями решаемых аналитических задач. Это оперативный анализ с построением OLAP-кубов по агрегированным данным хранилищ моделей «Звезда» и «Снежинка» в разрезе разных измерений для фермерских хозяйств, небольших сельскохозяйственных организаций.

Для крупных агрохолдингов, помимо оперативного анализа необходимо использование предсказательного и описательного моделирования Data Mining на основе данных хранилищ моделей как «Снежинка», так и Data Vault 2.0. Это прогнозирование натуральных и стоимостных показателей по моделям линейной регрессии и ARIMAX, построение классификационных деревьев решений, использование логистической регрессии для задачи классификации, кластеризация контрагентов организации, товаров и продукции методами k-means, DBSCAN, статистический анализ исторических данных организации, сегментация контрагентов с использованием ABC анализа и пр. [12]. Существенно расширяет возможности анализа структурированных, слабо структурированных и неструктурированных данных применение озер данных (Data Lakes) с модификациями [19, 20]. Поступление данных в названные организационные структуры хранения может осуществляться в потоковом режиме (Data streaming) без предварительной предобработки и трансформации.

Это актуально для данных, поступающих с датчиков температуры и влажности, систем радиочастотной идентификации животных (RFID) и др. Причем в традиционных реляционных хранилищах, используемых в организационных архитектурах Data Lakehouse наравне с озерами данных, возможно использование различных моделей данных («Звезда», «Снежинка», Data Vault).

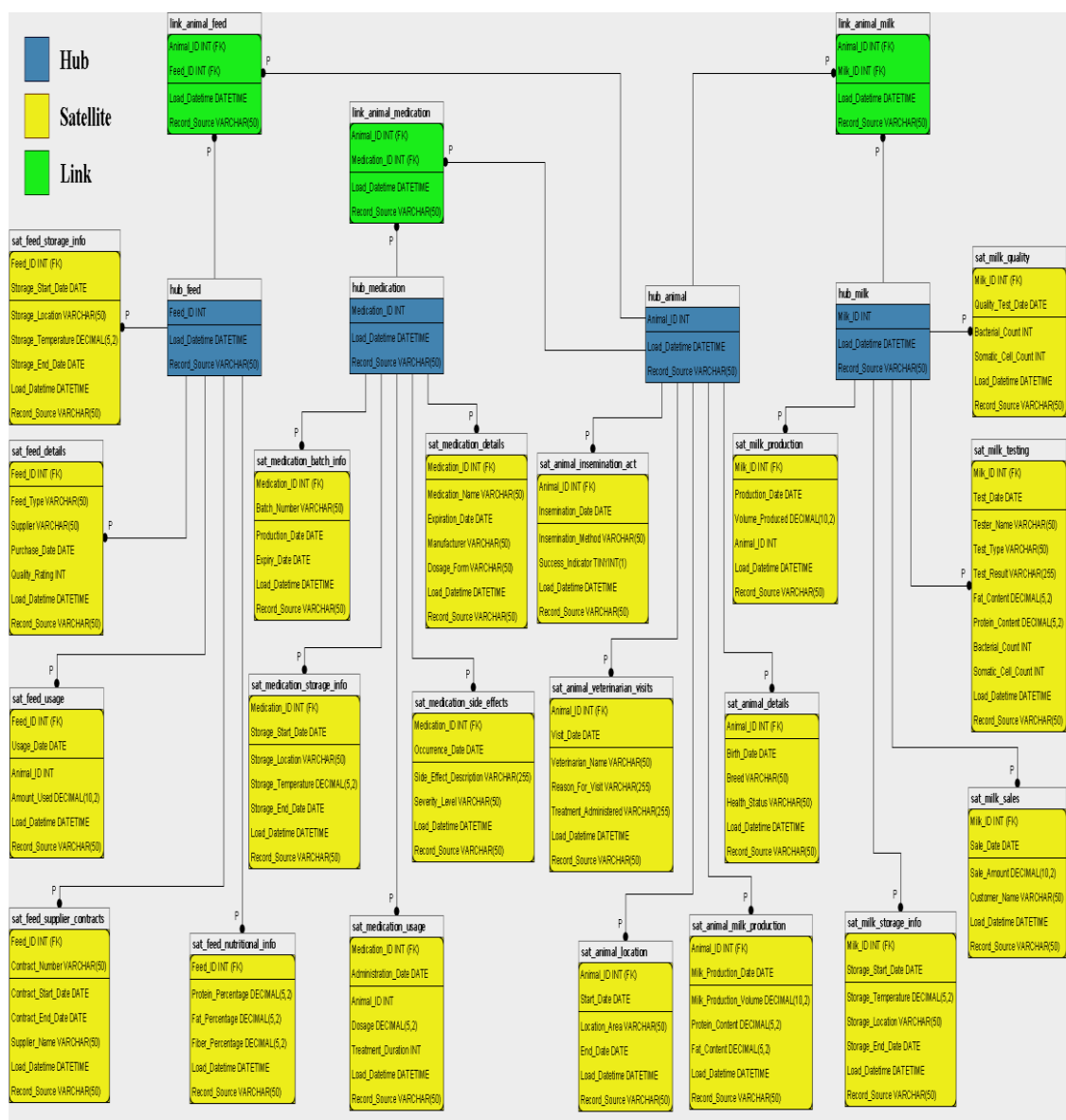


Рис. 2. Модель данных Data Vault 2.0
Fig. 2. Data Vault 2.0 data model

При анализе данных Data Lakehouse могут использоваться модели предсказательной аналитики с использованием методов машинного обучения с подкреплением (reinforcement learning), а также активного обучения (active learning). Также область применимости озер данных расширяется на проведение анализа Text Mining (глубинное обучение нейросети для анализа отзывов клиентов, фильтрация спама и др.), Process Mining (анализ бизнес-процессов организации).

Озера данных рекомендуется использовать также для управления отраслью на уровне различных муниципальных образований (район, область и пр.) для хранения и обработки данных, поступающих из информационных хранилищ сельскохозяйственных организаций.

Новая распределенная децентрализованная организационная архитектура хранения данных Data Mesh, в которой домены различных подразделений, филиалов и пр. организации создают и предоставляют другим доменам и структурам свои Data-продукты (тематические витрины, готовые наборы данных, аналитические отчеты, модели машинного обучения и пр.), рекомендуется к использованию в крупных, территориально распределенных агрохолдингах, имеющих собственные перерабатывающие мощности, предприятия оптовой и розничной торговли, логистические мощности. Подобную архитектуру можно использовать также и в управлении отраслью на уровне муниципального образования, где необходим как оперативный анализ данных, так и модели Data,

Text, Process Mining [21]. Отечественные программные решения для организации систем хранения данных целесообразно рассматривать с учетом так называемых «Кругов Громова» по ВІ-системам, ETL&Connectors- системам, системам управления базами данных (СУБД), каталогам данных по российским провайдерам облачных услуг [22, 23].

ВЫВОДЫ

1. На оперативном уровне управления небольшим сельскохозяйственным организациям, крестьянским и фермерским хозяйствам, на наш взгляд, рекомендуется использование реляционных систем хранения хорошо структурированных данных для использования в транзакционных информационных системах. Крупным агрохолдингам, работающим с разными видами информации (структурированной, слабоструктурированной и неструктурированной), необходимо также применение нереляционных баз данных, а именно графовых баз данных для описания сложных взаимосвязей производственных процессов, документных баз данных для хранения неструктурированной и слабоструктурированной информации.

2. На тактическом и стратегическом уровне управления рекомендуется проектирование информационных хранилищ. Модель данных хранилища определяется как размером организации, так и характеристикам источников данных (объем, изменчивость количества и структуры). Небольшим и средним сельскохозяйственным организациям возможно использование централизованных хранилищ схем «Звезда» и «Снежинка», источниками данных для которых являются в том числе и

данные транзакционных систем. Загрузка структурированных данных в информационные хранилища производится в пакетном режиме с использованием технологий подготовки и преобразования данных ETL (Extract, Transform, Load). Крупным организациям, объем данных, количество и структура источников данных для хранилища у которых меняется достаточно часто (около одного раза в полугодие или чаще) рекомендуются модели данных Data Vault 2.0 или якорная модель соответственно с потоковой загрузкой данных и ELT (Extract, Load, Transform) технологиями.

3. Выбор организационной архитектуры системы хранения (традиционное централизованное реляционное хранилище Data Warehouse, озеро данных Data Lake, Data Lakehouse, распределенная архитектура Data Mesh и пр.) определяется также размерами, специализацией организации, видами обрабатываемой информации, а также целями анализа. Централизованные реляционные хранилища с витринами данных по подразделениям, филиалам организации должны обеспечивать проведение оперативного анализа данных, предсказательное и описательное моделирование Data Mining. Область применимости озер данных с различными источниками, в т.ч. документными и графовыми базами данных, расширяется на проведение анализа Text Mining, Process Mining, графовой аналитики.

Авторы заявляют об отсутствии конфликта интересов, требующего раскрытия в данной статье.

The authors declare the absence a conflict of interest warranting disclosure in this article.

ЛИТЕРАТУРА

1. Ведомственная программа цифровой трансформации Минэкономразвития России на 2023 год и плановый период 2024-2025 годов (с учетом изменений от 16 мая 2023 года). <https://www.economy.gov.ru/material/dokumenty>.
2. Стратегическое направление в области цифровой трансформации отраслей агропромышленного и рыбохозяйственного комплексов Российской Федерации на период до 2030 года. <http://actual.pravo.gov.ru/content/content.html>.
3. Официальный сайт Росстата. <https://rosstat.gov.ru/statistics>.
4. Эффективные отечественные практики на базе технологий искусственного интеллекта в сельском хозяйстве. Аналитический отчет. АНО «Цифровая экономика», 2023. <https://files.data-economy.ru>.
5. Малегин А.А. Параметрическая оценка и мониторинг устойчивости регионального аграрного производства: растениеводство. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2023. № 02(56). С. 62-70. DOI: 10.6060/ivecofin.2023562.644.

REFERENCES

1. Departmental program for digital transformation of the Ministry of Economic Development of the Russian Federation for 2023 and the planning period 2024-2025 (including amendments of May 16, 2023). <https://www.economy.gov.ru/material/dokumenty>.
2. Strategic direction in the field of digital transformation of the agro-industrial and fisheries sectors of the Russian Federation for the period up to 2030. <http://actual.pravo.gov.ru/content/content.html>.
3. Official website of Rosstat. <https://rosstat.gov.ru>.
4. Effective domestic practices based on artificial intelligence technologies in agriculture. Analytical report. ANO "Digital Economy", 2023. <https://files.data-economy.ru>.
5. Malygin A.A. Parametric assessment and monitoring of regional agricultural production sustainability: group production. *Ivecofin*. 2023. N 02(56). P. 62-70. DOI: 10.6060/ivecofin.2023562.644 (in Russian)
6. Yurk N.A., Diner Yu.A. Innovation in the agro-industrial complex as a key factor in the sustainable region development. *Ivecofin*. 2025. N 04(66). P. 69-76. DOI: 10.6060/ivecofin.2025664.746 (in Russian)

6. **Юрк Н.А., Динер Ю.А.** Инновации в агропромышленном комплексе как ключевой фактор устойчивого развития региона. *Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2025. № 04(66). С. 69-76. DOI: 10.6060/ivcofin.2025664.746
7. **Фаткуллин Р.В., Кислицын Е.В.** Графовые базы данных: основные подходы к проектированию. Информационные технологии и когнитивная электросвязь: Сборник научных трудов IX Всероссийской научно-практической конференции. 2023. С. 180-184.
8. **Чернышева К.В., Королькова А.П., Афанасьева С.И., Карпузова Н.В.** Использование хранилищ данных в экономике и менеджменте АПК. *Техника и оборудование для села*. 2022. № 4 (298). С. 44-48.
9. **Середенко Н.Н., Чернышева К.В., Афанасьева С.И.** Информационно-аналитические системы: учебное пособие. М.: ООО «Издательство «КноРус», 2025. 208 с.
10. **Чернышева К.В., Королькова А.П., Афанасьева С.И.** Цифровая трансформация агропромышленного комплекса на основе систем аналитики данных. *Техника и оборудование для села*. 2024. № 1 (319). С. 44-48.
11. Управление данными в госсекторе. Навигатор для начинающих / под ред. О. М. Гиацинтова, В. А. Сазонова, М. С. Шклярук. М.: РАНХиГС. 2022. 198 с.
12. **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.
13. **Вичугова А.** Проектирование хранилища данных с методологией Data Vault в архитектуре Lakehouse.
14. Проектирование DWH с помощью Data Vault Хабр <https://habr.com/ru/companies/otus/articles/816219>.
15. Data Vault, Anchor Modeling. Какая методология проектирования DWH подойдет для вашего бизнеса? <https://www.qlever.ru/press/dwh-methodology>.
16. Молоко 2.0 Плинор. <https://plinator.ru/wp-content/uploads/2023/07/moloko-2.0>.
17. **Gorelik A.** The Enterprise Big Data Lake Delivering the Promise of Big Data and Data Science. O'Reilly Media, Inc., 2019.
18. Первое в России внедрение «умного озера данных» – комплексной платформы обработки, хранения и анализа данных с интегрированным Data Governance. <https://globalcio.ru/projects/3040>.
19. **Горожанкин В.** Data Mesh вместо информационных «бункеров»: как управлять данными с пользой для бизнеса. <https://rb.ru/opinion/data-mesh>.
20. Круги Громова - исследование русских ИТ-вендоров и российского ПО. <https://russianbi.ru>
21. **Дейт К. Д.** Введение в системы баз данных. 8-е издание. Пер. с англ. М.: Издательский дом «Вильямс», 2005. 1328 с.: ил.
22. **Лакшманан Валиappa, Тайджани Джордан.** Google BigQuery: Все о хранилищах данных, аналитике и машинном обучении. Пер. с англ. СПб: Питер, 2021. 496 с.: ил. (Серия «Бестселлеры O'Reilly»). ISBN 978-5-4461-1707-9
23. **Бобков С.П., Астраханцев Р.Г., Павлова Е.А.** Исследование структуры потоков в технологических аппаратах с использованием дискретных динамических моделей. *Современные наукоёмкие технологии. Региональное приложение*. 2024. № 1(77). С. 95-101. DOI 10.6060/snt.20247701.00013.
7. **Fatkullin R.V., Kislitsyn E.V.** Graph databases: main approaches to design / Information technologies and cognitive telecommunications: Collection of scientific papers of the IX All-Russian scientific and practical conference. 2023. P. 180-184.
8. **Chernysheva K.V., Korolkova A.P., Afanasyeva S.I., Karpuzova N.V.** Use of data warehouses in the economics and management of the agro-industrial complex. *Machinery and equipment for the village*. 2022. N 4 (298). P. 44-48.
9. **Seredenko N.N., Chernysheva K.V., Afanasyeva S.I.** Information and analytical systems: a textbook. M.: ООО "Izdatelstvo" Knorus ". 2025. 208 p.
10. **Chernysheva K.V., Korolkova A.P., Afanasyeva S.I.** Digital transformation of the agro-industrial complex based on data analytics systems. *Machinery and equipment for the village*. 2024. N 1 (319). P. 44-48.
11. Data management in the public sector. Navigator for beginners / edited by O. M. Giatsintov, V. A. Sazonov, M. S. Shklyaruk. M.: RANEPa, 2022. 198 p.
12. **Chernysheva K.V.** Information Support for Management of the Agricultural Sector: Dairy Farming: Dis. Cand. of Economics and Management of the National Economy: 08.00.05. Moscow. 2007. 218 p.
13. **Vichugova A.** Designing a Data Warehouse with the Data Vault Methodology in the Lakehouse Architecture. <https://bigdataschool.ru/blog/data-vault-for-lakehouse.html>, (accessed on 20.01.2025).
14. Designing a DWH with Data Vault.Habr: <https://habr.com/ru/companies/otus/articles/816219>.
15. Snezhinka Data Vault, Anchor Modeling. Which DWH Design Methodology is Right for Your Business? <https://www.qlever.ru/press/dwh-methodology>.
16. MILK 2.0 Plinor. <https://plinator.ru/wp-content/uploads/2023/07/moloko-2.0>.
17. **Gorelik A.** The Enterprise Big Data Lake Delivering the Promise of Big Data and Data Science. O'Reilly Media, Inc., 2019.
18. Russia's First Implementation of a "Smart Data Lake" – a Comprehensive Platform for Processing, Storing, and Analyzing Data with Integrated Data Governance. <https://globalcio.ru/projects/3040>.
19. **Gorozhankin V.** Data Mesh Instead of Information "Silos": How to Manage Data for Business Benefits. <https://rb.ru/opinion/data-mesh> accessed January 26.
20. Gromovs Circles - a study of Russian IT vendors and Russian software. <https://russianbi.ru>.
21. **Date K.J.** Introduction to Database Systems. 8th edition: Translated from English. Moscow: Williams Publishing House, 2005. 1328 p.: ill.
22. **Lakshmanan Valiappa, Taijani Jordan.** Google BigQuery: All About Data Warehousing, Analytics, and Machine Learning. Translated from English. St. Petersburg: Piter, 2021. 496 p.: ill. (O'Reilly Bestsellers Series). ISBN 978-5-4461-1707-9.
23. **Bobkov S.P., Astrakhancev R.G., Pavlova E.A.** Study of the structure of flows in technological devices using discrete dynamic models. *Modern high technology. Regional application*. 2024. N 1(77). P. 95-101. DOI 10.6060/snt.20247701.00013.

Поступила в редакцию (Received): 26.05.2025
Принята к опубликованию (Accepted): 02.12.2026