

## 10 лучших решений для работы с графическими базами данных

### Описание

Графовые базы данных хранят плотные данные с высокой степенью связанности и эффективно обрабатывают запросы. Но знаете ли вы, когда следует использовать ту или иную графовую базу данных? Читайте, чтобы узнать больше. “Данные – это новая нефть”. Рост любой организации зависит от того, насколько эффективно она хранит и использует данные. Ежедневно генерируется 2,5 квинтиллиона байт данных. Поэтому нам нужны отказоустойчивые системы и хранилища, где данные можно эффективно хранить и управлять ими. Изначально использовались реляционные базы данных. Но со временем количество и тип данных стремительно менялись. Возникла потребность в хранении видео, аудио, изображений и т. д. Это послужило толчком к развитию SQL, NoSQL баз данных, Hadoop, графовых баз данных и т. д. Каждая из них имеет свои сценарии использования и работает с различными форматами данных. Графовые базы данных были разработаны для упрощения операций с данными и их эффективного хранения.

### Графовые базы данных

Граф – это структура данных, представленная в виде узлов и ребер. База данных – это набор таблиц, в которых хранятся данные и связи между ними. Графовая база данных – это база данных, в которой данные хранятся в виде узлов, а отношения, существующие внутри данных, – в виде ребер. Графовые базы данных помогают обрабатывать запросы в реальном времени и эффективно управлять отношениями “многие-ко-многим” между сущностями. К популярным моделям графовых данных относятся графы свойств и RDF-графы. Аналитика и запросы в основном

выполняются с использованием графов свойств. Интеграция данных осуществляется с помощью RDF-графов. Разница между графами свойств и графами RDF заключается в том, что графы RDF представлены в виде трипов, то есть субъекта, предиката и объекта. Графовые базы данных хранят данные в виде узлов, а отношения между данными – в виде ребер между узлами. Ребра в графе могут быть направленными (однонаправленными) или ненаправленными (двунаправленными). Обработка запросов осуществляется путем обхода графа. Для эффективного ответа на запросы используются алгоритмы обхода графа, которые помогают найти путь от одной вершины к другой, расстояние между вершинами, обнаружить закономерности, петли внутри графа, возможность образования кластеров и т. д.

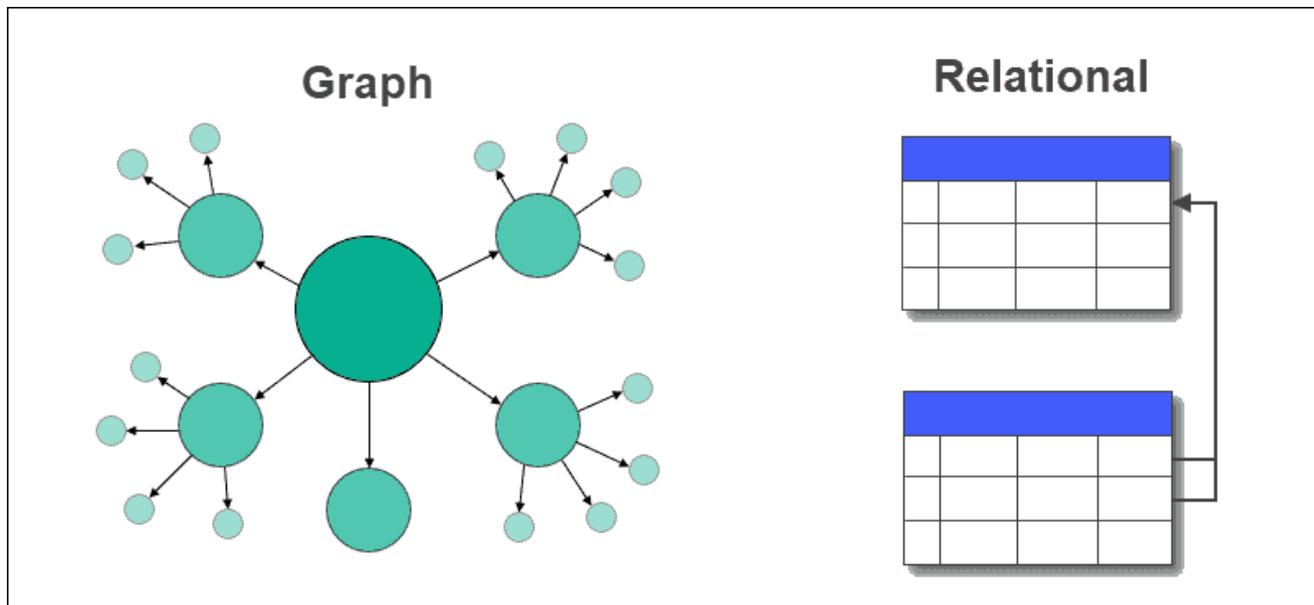
## **Области применения баз данных на основе графов**

Графовые базы данных используются для обнаружения мошенничества. Узлами/сущностями могут быть имена, адреса, даты рождения и т. д. людей, а также некоторые мошеннические IP-адреса, номера устройств и т. д. Когда мошеннический узел взаимодействует с немошенническим узлом, между ними образуются связи, которые помечаются как подозрительные. Сайты социальных сетей используют базы данных графов, чтобы показывать рекомендации людей, с которыми мы могли бы общаться, и контент, который мы хотим просмотреть. Это происходит с помощью обхода графов в базе данных. Картирование сети и управление инфраструктурой, элементы конфигурации и т. д. также эффективно хранятся и управляются с помощью графовых баз данных.

## **Графовые базы данных в сравнении с реляционными базами данных**

В графовой базе данных таблицы со строками и столбцами заменяются узлами и ребрами. В графовой базе данных связи между данными хранятся на ребрах. В реляционной базе данных хранятся связи между таблицами, использующими внешние ключи, и другими таблицами. Извлечение данных или запрос просты и не требуют сложных соединений в базе данных графов, но это не так в реляционных базах данных. Реляционные базы данных лучше всего подходят для использования в случаях, связанных с транзакциями, в то время как графовые базы данных подходят для приложений с большим количеством взаимосвязей и данных. Графовые базы данных поддерживают структурированные, полуструктурированные

и неструктурированные данные, в то время как реляционные базы данных должны иметь фиксированную схему. Графовые базы данных удовлетворяют динамическим требованиям, в то время как реляционные базы данных обычно используются для решения известных и статических задач.



Граф и реляционные базы данных

Давайте рассмотрим лучшие решения для работы с базами данных графов.

## Cayley

Cayley – это база данных графов с открытым исходным кодом, разработанная в Apache 2.0. Она была создана с использованием Go и работает со связанными данными. Cayley – это база данных, использованная при создании Freebase и графа знаний Google. Она поддерживает множество языков запросов, таких как MQL и Javascript, с графовым объектом на основе Gremlin. Он прост в использовании, быстр и имеет модульную конструкцию. Он может интегрироваться и взаимодействовать с различными внутренними хранилищами, такими как LevelDB, MongoDB и Bolt. Он поддерживает различные API сторонних разработчиков, написанные на разных языках, таких как Java, .NET, Rust, Haskell, Ruby, PHP, Javascript и Clojure. Его можно развертывать в Docker и Kubernetes. Ключевыми областями, в которых используется Cayley, являются информационные технологии, компьютерное программное обеспечение и финансовые услуги.

## Amazon Neptune

Amazon Neptune известен своей исключительной эффективностью при работе с высокосвязанными наборами данных. Он надежен, безопасен, полностью управляем и поддерживает открытые графические API. Он может хранить миллиарды взаимосвязей и выполнять запросы к данным с чрезвычайно низкой задержкой в несколько миллисекунд. Графовая модель данных Neptune состоит из 4 позиций, а именно: субъект (S), предикат (P), объект (O) и граф (G). Каждая из этих позиций используется для хранения позиции исходного узла, целевого узла, отношений между ними и их свойств. Также используется кэш, который ускоряет выполнение запросов на чтение. Данные хранятся в виде кластеров БД. Каждый кластер состоит из основного экземпляра БД и копий экземпляров БД для чтения. Neptune отличается высокой степенью безопасности, поскольку использует аутентификацию IAM, сертификацию SSL и мониторинг журналов. Кроме того, в Amazon Neptune легко переносить данные из других источников. Кроме того, система обеспечивает отказоустойчивость за счет создания реплик и периодического резервного копирования. Среди компаний, использующих Neptune, – Herren, Onedot, Juncture и Hi Platform.

## Neo4j

Neo4j – это масштабируемая, безопасная, надежная база данных графов, создаваемая по требованию. Neo4j была создана на Java с использованием языка запросов Cypher. Она использует протокол Bolt, и все транзакции происходят через конечную точку HTTP. По сравнению с другими реляционными базами данных она гораздо быстрее отвечает на запросы. В ней нет накладных расходов на сложные соединения, а ее оптимизация хорошо работает при большом объеме набора данных с высокой степенью связанности. Она предлагает преимущества хранения графов наряду с ACID-свойствами реляционной базы данных. Neo4j поддерживает различные языки, такие как Java, .NET, Node.js, Ruby, Python и т. д., с помощью драйверов. Она также используется для работы с графовыми данными, аналитики и машинного обучения. Neo4j Aura DB – это отказоустойчивая и полностью управляемая облачная база данных графов. Neo4j используют такие компании, как Microsoft, Cisco, Adobe, eBay, IBM, Samsung и др.

## ArangoDB

ArangoDB – это многомодельная база данных с открытым исходным кодом. Многомодельный подход позволяет пользователям запрашивать данные на любом языке запросов по своему выбору. Узлы и ребра ArangoDB представляют собой документы в формате JSON. Каждый документ имеет уникальный идентификатор. Отношения между двумя узлами обозначаются в виде ребер, и их уникальные идентификаторы хранятся. Хорошая производительность объясняется наличием хэш-индекса.

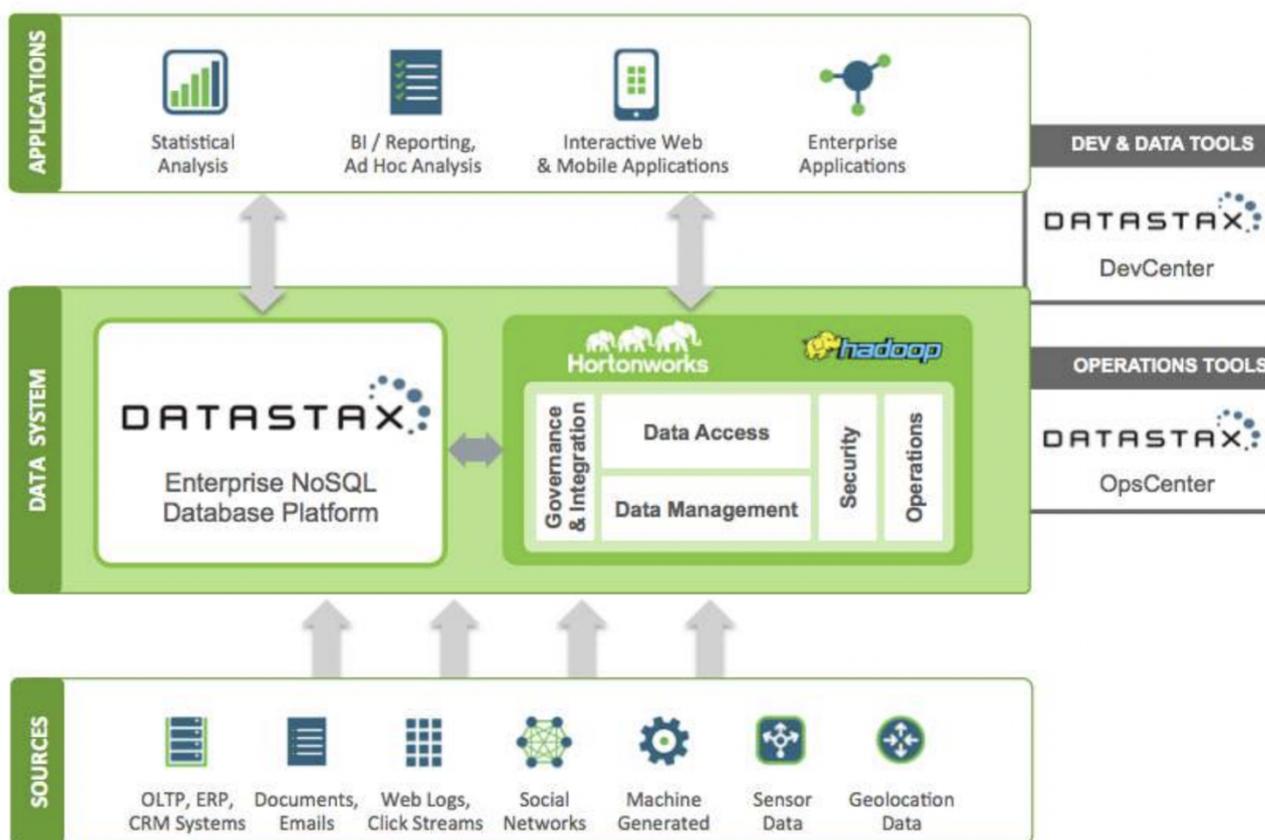


Улучшены обходы, соединения и поиск в базах данных. Она помогает в проектировании, масштабировании и адаптации к различным архитектурам. Он играет важную роль в сложных задачах науки о данных, таких как извлечение признаков и расширенный поиск. ArangoDB может работать в облачной среде и совместима с Mac Os, Linux и Windows. LDAP-аутентификация, маскировка данных и

алгоритмы шифрования обеспечивают безопасность базы данных. Она используется в системах управления рисками, IAM, обнаружения мошенничества, сетевой инфраструктуре, рекомендательных системах и т. д. Accenture, Cisco, Dish и VMware – вот некоторые организации, использующие ArangoDB.

## DataStax

DataStax – это облачная база данных NoSQL как услуга, построенная на базе Apache Cassandra. Она обладает высокой масштабируемостью и использует облачную нативную архитектуру. Она надежна и безопасна. Каждый документ, хранящийся в DataStax, имеет индекс, который помогает легко искать и быстро извлекать данные. Над проиндексированными данными создаются шарды. Различные источники данных могут быть использованы для создания приложений с помощью инструментов Datastax Enterprise, Kafka и Docker.



Данные, собранные из источников, отправляются в экосистему Hadoop и DataStax. Hadoop управляет безопасностью, операциями, доступом к данным и управлением,

взаимодействуя с DataStax. Данные уточняются с помощью инструментов разработки и эксплуатации Datastax. Проанализированная информация затем используется для статистического анализа, корпоративных приложений, составления отчетов и т. д. Поскольку решение является облачным, клиенты платят за то, что используют, и цены вполне приемлемы. Verizon, CapitalOne, TMobile и Overstock – вот некоторые компании, использующие DataStax.

## **Orient DB**

OrientDB – это графовая база данных, которая эффективно управляет данными и помогает создавать визуальные представления для их демонстрации. Это многомодельная графовая база данных, созданная на языке Java. Она хранит данные в виде пар ключ-значение, документов, объектных моделей и т. д. Она состоит из трех важных компонентов: редактора графов, студии запросов и консоли командной строки. Для визуализации и взаимодействия с данными используется редактор графов. Интерфейс запросов Studio используется для выполнения запросов и немедленного предоставления результатов в графическом и табличном формате. Консоль командной строки используется для запросов к данным из OrientDB. OrientDB имеет распределенную архитектуру с несколькими серверами, которые могут выполнять операции чтения и записи. Серверы-реплики используются для выполнения операций чтения и запросов. Поддерживается индексирование, а также ACID-совместимость. Среди компаний, использующих OrientDB, – Comcast Corporation и Blackfriars Group.

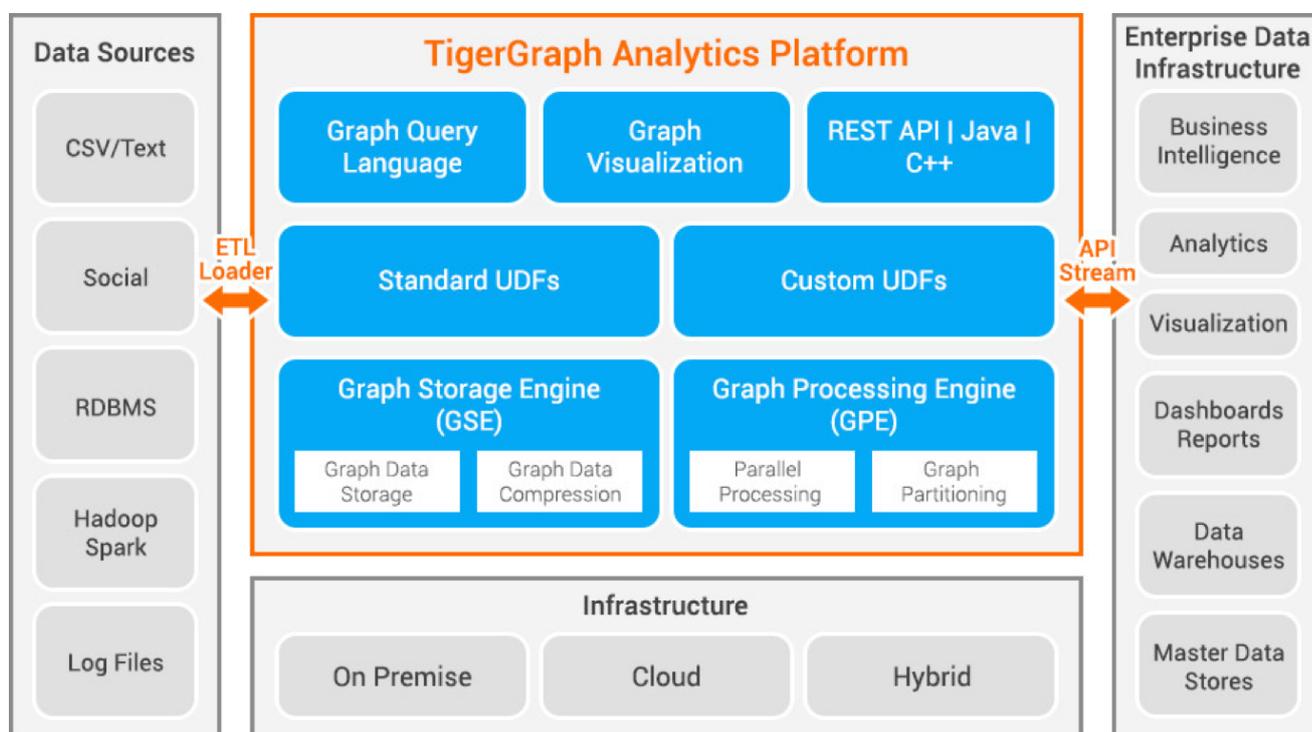
## **Dgraph**

Dgraph – это облачная база данных графов, поддерживающая GraphQL. Она была создана с использованием Go. Она минимизирует количество сетевых вызовов и снижает задержки за счет максимальной одновременной обработки запросов. Бесшовная интеграция Dgraph с GraphQL помогает легко разрабатывать бэкэнд-приложения на GraphQL. Мутация GraphQL передается через функцию Lambda, которая взаимодействует с базой данных и конвейером данных. Это упрощает обработку запросов. Она горизонтально масштабируема, то есть количество ресурсов увеличивается по мере роста запросов и данных. Он предоставляет различные функции, такие как авторизация на основе JWT, визуализатор данных, облачная аутентификация, резервное копирование данных и т. д. Среди

организаций, использующих Dgraph, – Intuit, intel и Factset.

## Tigergraph

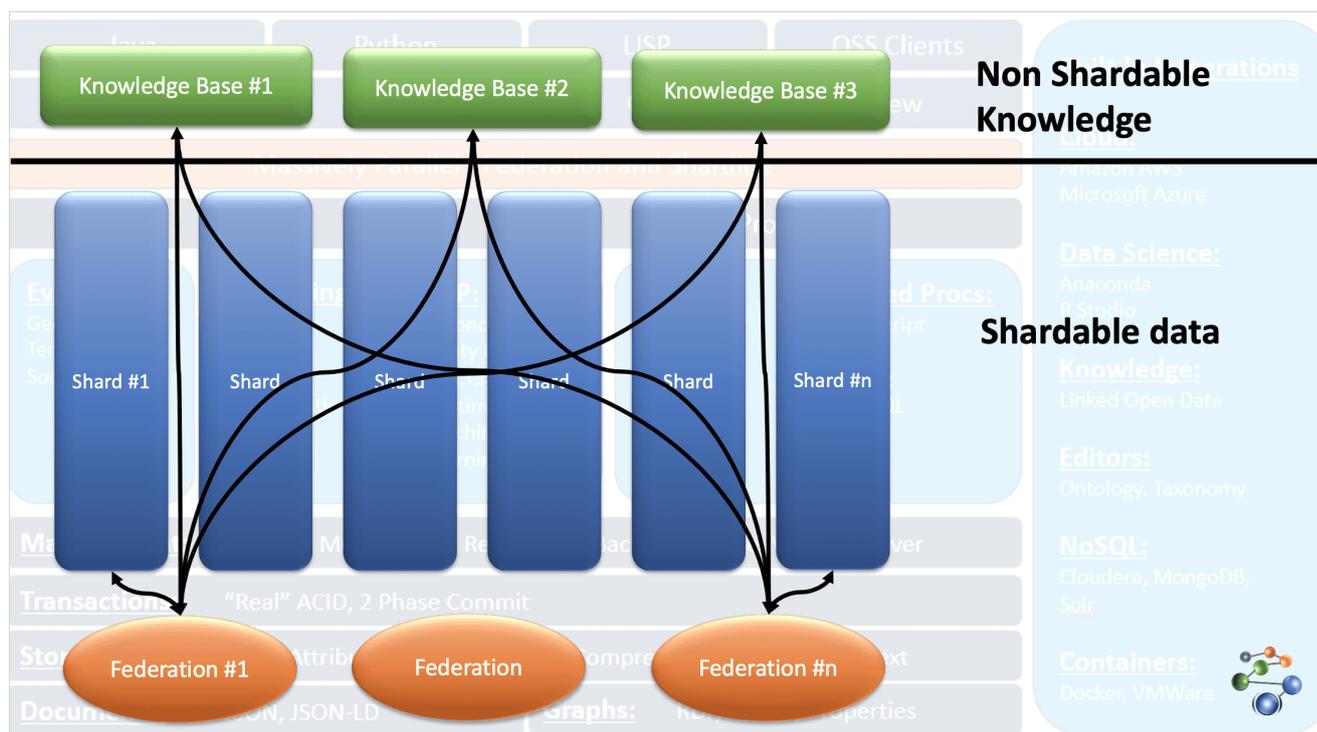
Tigergraph – это база данных графов свойств, разработанная на C++. Она обладает высокой масштабируемостью и позволяет выполнять расширенную аналитику на высокосвязанных данных. Для хранения данных используется собственная структура графов, а для их обработки – движок обработки графов. База данных хранится на диске и в памяти, а также использует кэш процессора для быстрого поиска. Для параллельной обработки данных используется функция Map Reduce.



Он чрезвычайно быстр и масштабируем. Он выполняет параллельные вычисления и обеспечивает обновления в режиме реального времени. Она использует методы сжатия данных и сжимает их в 10 раз. Разделение данных по серверам происходит автоматически, что избавляет пользователя от необходимости тратить время и силы на разделение данных вручную. Она используется для выявления мошенничества в домашних хозяйствах, управления цепочками поставок и улучшения медицинского обслуживания. Среди организаций, использующих Tigergraph, – JPMorgan Chase, Intuit и United Health Group.

## AllegroGraph

AllegroGraph использует технологию графов знаний “сущность-событие” для проведения аналитики и принятия решений на основе высокосвязанных, сложных и плотных данных. Данные хранятся в формате JSON и JSON-LD в узлах графа. Используется архитектура протокола REST. Она также позволяет работать с очень большими массивами данных, разделяя их по определенным критериям и распределяя по нескольким хранилищам баз знаний.



Это возможно благодаря функции FedShard базы данных AllegroGraph. Выполнение запросов происходит путем объединения федераций с хранилищами баз знаний. Поддерживаются типы схем XML и используются тройные индексы. Хранит геопространственные данные, такие как широты и долготы, и временные данные, такие как дата, временная метка и т. д. Она совместима с Windows, Mac и Linux. Используется для обнаружения мошенничества, в здравоохранении, идентификации субъектов, прогнозирования рисков и т. д.

## Stardog

Stardog – это графовая база данных, которая выполняет виртуализацию графовых

данных и связывает данные из хранилищ и озер данных без физического копирования данных в новое место хранения. Stardog построена на основе открытых стандартов RDF. Он поддерживает структурированные, полуструктурированные и неструктурированные данные. Подобная материализация, реализованная в Stardog, обеспечивает гибкость. Это единственная графовая база данных, которая сочетает в себе графы знаний и виртуализацию.



Stardog использует механизм выводов, основанный на искусственном интеллекте, для эффективной обработки и предоставления результатов запросов. Это ACID-совместимая база данных графов. Поддерживается одновременное чтение и запись. Благодаря современной архитектуре она легко обрабатывает сложные запросы. Она используется для управления ИТ-активами, управления данными и аналитики и обеспечивает высокую доступность. Среди компаний, использующих Stardog, – Cisco, eBay, NASA и Finra.

## **Заключительные слова**

Графовые базы данных позволяют легко запрашивать отношения “многие-ко-многим” и эффективно хранить данные. Они масштабируемы, безопасны и могут быть интегрированы со многими сторонними инструментами, API и языками. В последние годы они были интегрированы с облаком и обеспечивают наилучшую производительность. Они упрощают сложные соединения в простые запросы, что облегчает задачу разработчикам. Задачи, требующие больших объемов данных, такие как IoT и Big Data, также являются графовыми базами данных. Они будут продолжать развиваться и в будущем, несомненно, распространятся на другие сферы применения.

## **Дата Создания**

22.05.2024