Simplifying Data Management with a Modern Cloud-based Data Warehouse

Technical White Paper

The life sciences industry has an unusually large number of third-party data vendors, faces strict regulatory and privacy requirements, and must manage continuously-changing data sources. That's why companies have traditionally taken a conservative approach to data management. Medical literature doubles every three years, which complicates pharma's physician education efforts. But as the life sciences innovation model evolves, a new approach to data warehousing is necessary.

Veeva Systems helps life sciences organizations navigate the complex landscape from research and development through commercialization. Veeva offers a multi-tenant commercial data warehouse to help companies like Antares Pharma and Karyopharm, drive innovation, build a foundation for machine learning, and deliver new drugs to market. Veeva Nitro standardizes the commercial life sciences data model by using adaptive Intelligent Sync™ connectors and pre-built integrations with industry-relevant data sources.

A productized data warehouse approach pulls many disparate industry data sources together in one place for reporting and insights. Veeva Nitro simplifies how customers gather data in one source and push to other applications, such as the CRM system, content analytics, and a variety of Amazon services. By making it available on the cloud, it also offers significant scalability, speed, and ease of access to data. Industry-standard data connectors go beyond typical cloud data warehouse platforms by eliminating the need for 'mechanics' to maintain source data integrations, freeing those resources to focus on more strategic IT priorities.

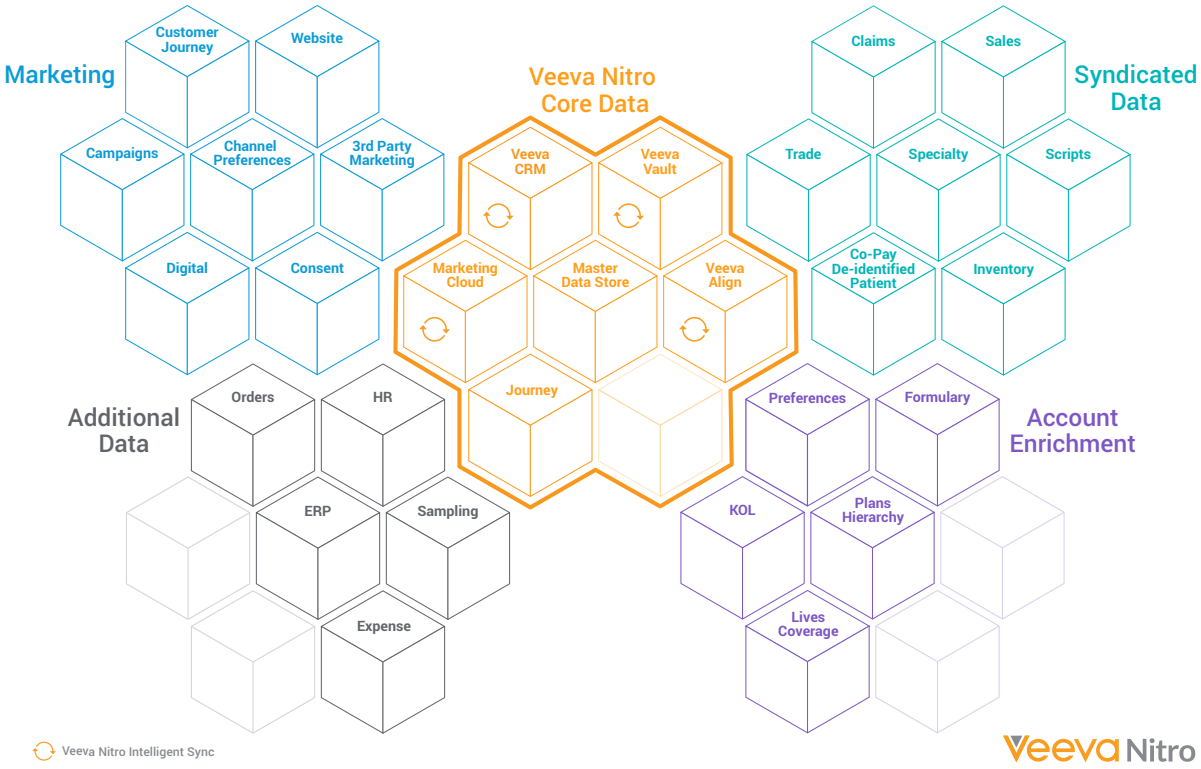

Figure 1: Veeva Nitro leverages a continuously growing number of industry-standard data connectors. Adaptive Intelligent Sync™ integration means changes are automatically updated and data remains current.

Competitive differentiation: delivering real-time data from the data warehouse directly to sales representatives

In the life sciences space, agility and speed provide field representatives a significant advantage. Readily accessible and actionable data is critical to success.

Because our standardized data model keeps data current, life sciences field teams and their managers can get immediate and actionable insights about their accounts and activities, including tracking key metrics against their goals and business objectives, updated in real-time directly from the data warehouse.

Figure 2: Example dashboard available to field reps on mobile device. Near real-time data is accessible directly from the data warehouse.

Using offline data, field reps can work in protected areas, such as hospitals, where no Wi-Fi access is available, and still get complete access to the most current insights on their sales activities and performance.

For most back-office analytics users, their primary Veeva Nitro interaction is through their own visualization and reporting tools of choice. These are configured to their Redshift instance to provide clean, fast performance access to their entire commercial warehouse content.

Architectural Principles of Veeva Nitro

Veeva focuses on delivering a centrally managed, out-of-the-box data warehouse model. This solution combines AWS technology with an advanced data warehousing methodology to deliver a turnkey solution for high-speed integration, rapid implementation, fast reporting, and easy-to-use analytics. This centralized and managed application environment is specifically configured to control source-ETL-warehouse data flows.

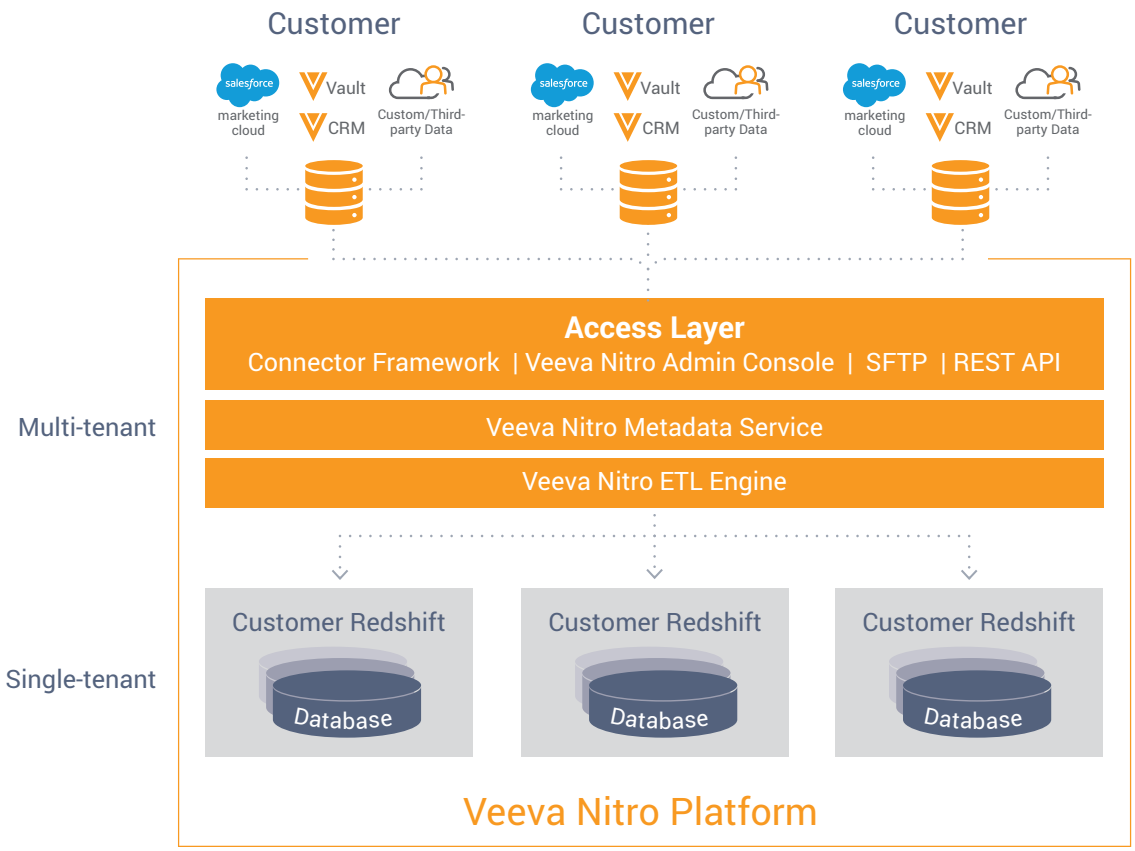

Figure 3: Logical architecture and data flows within the Veeva Nitro platform

Leveraging Amazon Redshift, a data platform optimized for warehousing

Each Veeva Nitro deployment includes a single-tenant Amazon Redshift cluster to support the data warehouse. Redshift is an ideal platform for data warehousing because of its ability to scale based on the needs of each life sciences customer, readily growing to support more data, faster queries, and more users. Redshift's massively parallel processing (MPP) architecture provides greater performance for large scale queries and bulk imports than traditional RDBMS systems, and supports many business intelligence platforms.

In Veeva Nitro's pre-packaged solution, each Nitro-to-Redshift cluster is pre-configured and ready for use. Access control, security, and workload management provide companies isolated access to just their data - encrypted and secured for both data-at-rest and data-in-transit. As the application data sources expand, more reports and dashboards are developed, and more users leverage the system, each isolated Redshift cluster can grow with it and maintain the required level of performance for each client.

An extensible approach to multi-tenant data warehousing

While each Veeva Nitro customer has their own Redshift cluster, the platform provides a centralized management environment for all commercial data warehousing customers. It uses multi-tenant techniques to deliver access to large, scalable clusters of ETL servers, connectors, and security services that isolate each of their applications into segregated, single-company sections. The environment can be easily extended and tailored to meet the needs of each customer's specific commercial life sciences practice. Companies benefit from Veeva Nitro's centralized management and continuous enhancements.

To support the unique commercial life sciences analytic needs of each company, the Veeva Nitro platform delivers pre-packaged application connectors, tailored specifically for the analytic specifications of common life sciences application data sources. Out of the box, the IntelligentSync™ connectors interact directly with applications' APIs to understand their configuration, identify key objects and attributes, perform optimal data ETL extracts, and monitor ongoing changes of back end systems. They are easily extensible to incorporate the specialized requirements for each company, integrating with new life sciences data sources in days or weeks, instead of months or longer and accelerate time to value for new customers. And as the company's data needs grow, their Veeva Nitro platform quickly scales with modular expansion of new connectors and source systems, supporting dozens of common industry data sources including but not limited to Veeva CRM, Veeva Vault, Veeva Align, Salesforce CRM, and Salesforce Marketing Cloud.

In commercial life sciences, data frequently comes from data subscriptions, commonly distributed in the form of downloadable CSV data feeds with specialized content. The platform also supports specialized connectors for custom content from many life sciences data providers, while also enabling free form customization for any type of client specific sources.

Moving from Extract-Transform-Load to Extract-Load-Transform

Traditional data warehousing ETL projects are composed of two labor-intensive integration phases. The initial phase translates business requirements into specific data source objects to extract and consultants build specific ETL maps to integrate that data. In the second phase, as continuous changes occur on the backend system, person-hours are needed to extend the data model, grab new objects and attributes, and monitor ongoing semantic changes of the source application.

IntelligentSync™ connectors are designed to simplify both of these phases and ensure that data remains current:

1. Utilizing intelligent integration techniques to plug into the configuration of the source system applications, they interrogate the configurations, objects, attributes, types, and hierarchies to automatically build a data warehouse model that conforms to the source system.
2. As business requirements drive changes to the source applications, Veeva Nitro automatically "shapeshifts" the data model to keep up with the new model without the need for any IT developer intervention.

Veeva Nitro ingests data via "Extract-Load-Transform" (ELT), a fast integration model supported on MPP-style data warehousing platforms like Amazon Redshift. Redshift extracts data from the source system, such as Amazon S3, and quickly bulk loads it into a staging area. SQL queries transform the

data into a consumable format. Individual million row tables are processed in seconds and entire complex applications can be loaded, fully transformed and starred in tens of minutes in most cases, ready for analysis without stressing operational load windows.

The above “ELT” model contrasts sharply with traditional “Extract-Transform-Load” (ETL) models often leveraged in other third-party ETL tools and implementations. As ETL requires a separate upfront step to transform and stage the data prior to load, it can add additional complexity and overhead for customers. Veeva Nitro “ELT” model ensures existing data is always ready and accessible. New data tables are swapped in transparently and quickly without disconnecting the reporting tools sitting on top of it. In this scenario, users experience no downtime, even during “ELT” load windows.

Best Practice Data Warehousing Modeling - Built for Extensibility

Veeva Nitro’s analytic models are built from the ground up using advanced data warehouse modeling techniques, intended to support the unique complexities of commercial life sciences applications. Its commercial life sciences operational data model support data sources from across the industry and is designed to be:

- extensible to new applications,
- durable to evolve over time, and
- easy to use for analytics and business intelligence.

Veeva Nitro utilizes an operational data store (ODS) layer that is normalized, extensible, and time-series aware. It maintains a long-running history of applications’ activities and continuously captures state changes and transitions of objects over a 10-year period. For the duration of that history, dimension objects are recorded with full historical effective dating to help keep track of ongoing changes in any large commercial environment. The ODS layer is also built on the RedShift platform to allow for detailed queries and advanced analyses.

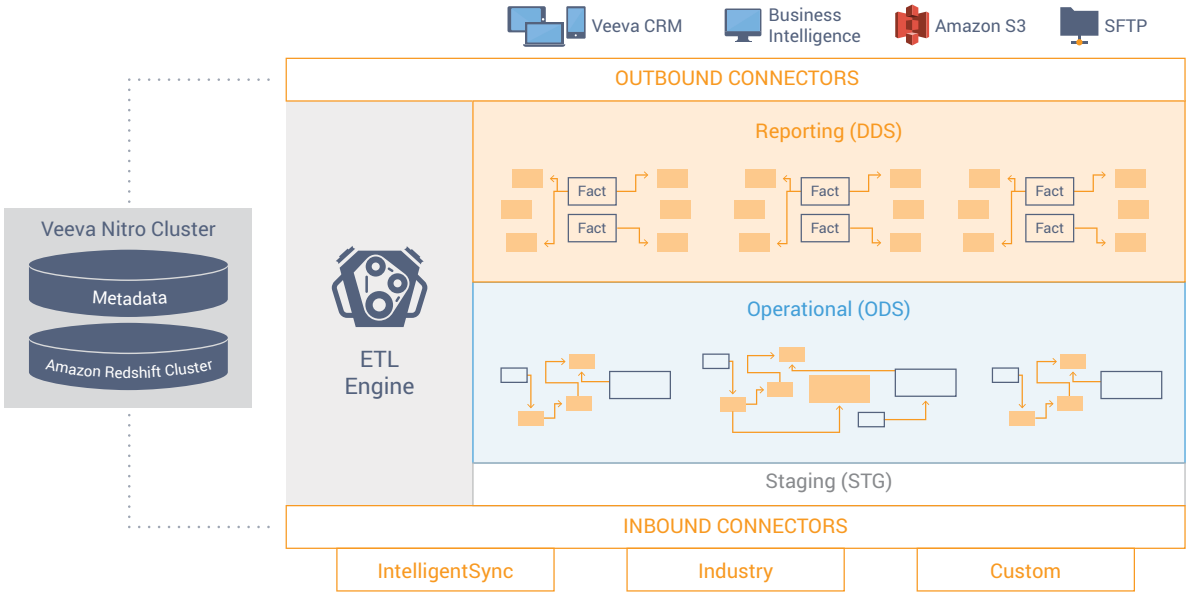

Figure 4

In the traditional data warehousing approach, such broad and expansive data stores with hundreds of objects can be challenging to decipher with business intelligence tools, requiring a full understanding of the relational semantics of various data elements that span across multiple applications. This also makes it difficult to build extensible reports and dashboards for answering common business questions without pulling in IT experts fully versed in the data.

To simplify data analysis, Veeva Nitro uses easy, business-targeted, and fit-for-purpose star schema models that are designed for seamless integration with modern reporting, discovery, and visualization tools. This approach targets specific business metrics (e.g. corporate, regional, district, territory, account) and is both time-series aware as well as filterable for specific aspects of the company. These star schemas abstract the complex semantic relationships of the underlying applications so that business analysts can create ad hoc analysis, dashboards, or reports in a matter of minutes. And since star schemas are summarizations, reports and dashboards get fast response times with a minimum of data transfer, even for data sources relying on millions/billions of transaction history rows.

Additionally, the detailed and summarized data structures within Veeva Nitro enable powerful data science initiatives, including machine learning, model training, and advanced analytic discovery.

Focus on Bringing Therapies to Patients

Veeva Nitro's productized data warehouse, built on AWS, significantly simplifies the life sciences industry's data management challenges by offering:

- a multi-tenant cloud data warehouse with standard industry data connectors,
- an adaptive architecture that keeps data current, and
- its own roadmap with regularly scheduled product enhancements.

Comments Sal Paolozza, director of sales operations at Antares Pharma, "Without any handoffs, we received data and we were able to get it to our reps with no manipulation from me. All reporting for the launch came out of Nitro from day one, which was phenomenal."

"Veeva Nitro offered a much better alternative. We were able to, in pretty short order, deliver useful information to the field. In traditional ways, there's no way we would have been to make those deployments as fast as we have with Nitro," adds Jason Magyar, senior director of enterprise applications at Karyopharm.

To learn more about how Antares Pharma uses Veeva Nitro, watch the complete interview.

About Veeva Systems
Veeva is the global leader in cloud software for the life sciences industry. Committed to innovation, product excellence, and customer success, Veeva serves more than 1,100 customers, ranging from the world's largest pharmaceutical companies to emerging biotechs. As a Public Benefit Corporation, Veeva is committed to balancing the interests of all stakeholders, including customers, employees, shareholders, and the industries it serves. For more information, visit **veeva.com**.

Veeva Systems
Global Headquarters
Pleasanton, California, USA
4280 Hacienda Drive
Pleasanton, California 94588
+1 925 452 6500 | veeva.com/contact | veeva.com

Learn more at veeva.com | 925-452-6500 | veeva.com/nitro