

How Cisco Transformed Data Management with Virtualization

Data virtualization reduces IT costs and improves information accuracy and timeliness for business decision-making.

EXECUTIVE SUMMARY	
CHALLENGE	<ul style="list-style-type: none"> • Data trapped in multiple sources and silos • Ever-increasing costs for data storage and integration • Diverse access patterns for data consumption, from SQL to web services • Long development cycles for making data available to business applications and users
SOLUTION	<ul style="list-style-type: none"> • Cisco Data Virtualization avoids expensive and time-consuming data replication by enabling federated queries across multiple data technologies • Cisco Unified Computing System and Cisco Application Centric Infrastructure for virtualization platform
RESULTS	<ul style="list-style-type: none"> • Reduced IT costs for data integration and storage • Faster application development • Unified cloud and on-premises data • Improved data access and visibility to business applications and users
LESSONS LEARNED	<ul style="list-style-type: none"> • Reuse data resources across application projects • Design for scalability and new data sources • Establish application best practices • Update the data service frequently
NEXT STEPS	<ul style="list-style-type: none"> • Leverage a custom SDK for access to more NoSQL data sources and cloud-application adapters

Background

Cisco IT has 400+ databases and about 3000 applications in its environment. With more than 50 petabytes of data stored today and growing data volumes, we face a constant demand to deliver high-quality information views to application developers and business users.

Challenge

Within Cisco, business users are trying to identify where all of their data can be found and how to get more value from that data by using it more strategically for decision-making. User concerns include:

- Increasing demand for faster access to reliable, current data that is delivered securely and consistently across the customer interaction lifecycle.
- Increased challenges for protecting customer data if it is accessed from different sources of records without consistent access controls, as well as monitoring and auditing capabilities.
- Avoiding confusion when multiple “versions” of what should be the same data are presented to users when that data is requested at different times or from different applications.
- Maintaining a central point for measures that control data access, such as strong passwords, encryption, data segmentation, and filtering.

A big problem in solving these challenges is that our data is scattered across databases, applications, and cloud services.

Additionally, data is often stored in functional and technology silos, with diverse access methods such as SQL and web services. Furthermore, users want applications that can easily access unstructured data (e.g., documents or video) and integrate those elements with structured data files and reports.

Traditional methods for replicating or integrating data into repositories do not support the speed or scalability to meet these needs nor the exponential growth of both data and new applications. Moreover, replication only adds to the ever-increasing costs of data storage and integration, and can reduce the coherency of data over time.

Another challenge is the time and resources required to finalize requirements around the processing, analytics, and consumption patterns for specific data sets. The inability to quickly iterate the requirements definition process can lead to project failure, cost overruns, and frustration for everyone. One reason for this challenge is that it's usually necessary to work with the data in advance to identify the nuggets of information that are relevant to applications and users. Another reason is that the access requirements often evolve or change once business users see the processed data.

PRODUCT LIST
Cloud and Systems Management Services
<ul style="list-style-type: none">• Cisco Information Server and components:<ul style="list-style-type: none">◦ Monitor◦ Deployment Manager◦ Business Directory• Cisco Application Centric Infrastructure (ACI)
Security
<ul style="list-style-type: none">• Cisco Enterprise Policy Manager
Servers—Unified Computing
<ul style="list-style-type: none">• Cisco UCS B-Series blade servers

“If the data definition process from initial requirements to a prototype is too long or expensive, the project has a very high chance of failure or not being completed,” says Piyush Bhargava, distinguished engineer, Cisco IT. “Often so much investment goes into this data preparation phase that the business objectives of data consumption are left unrealized.”

The requirements definition process takes so much time because data must be extracted from multiple applications, each with its own data model, into a common data model in the reporting application. This is a cumbersome manual effort that requires significant expertise in data modeling, data integration, and the associated business process.

Given all of these challenges, it became clear that data virtualization, as part of a holistic data architecture, would deliver the capabilities we needed for data management and application development.

Solution

Cisco® Data Virtualization is a data integration software suite that creates a logical abstraction for internal and external data spread across multiple, disparate sources and technologies. Anchored by the flagship product, the Cisco Information Server (CIS), Cisco Data Virtualization technology makes it easy for applications to access all data, no matter where it's managed, and query it across the network as if the data were in a single place.

Data virtualization includes functionality to connect applications or web services to data sources, execute queries to retrieve requested data, combine or federate data from those sources, and deliver the result to the consuming application in the desired format. Data sources include enterprise data warehouses, operational data stores, transactional sources, file systems, and big data (Figure 1).

Figure 1. Cisco Data Virtualization High-Level Architecture



Virtualization offers several distinct advantages over traditional data access and integration methods:

- Data isn't moved or replicated so it preserves a high degree of consistency.
- Logical data modeling can help create an abstraction that combines data from multiple sources into a single model.
- The data supply chain shrinks dramatically for application developers. Their focus changes from moving data between disparate data models and technologies to logically integrating data in the application.
- All users see current information because the consuming applications always retrieve the latest data from a single source.

Cisco Data Virtualization technology offers additional advantages for our application development, including:

- A clear understanding of data access patterns helps developers model data views on large datasets across multiple sources.
- Application developers can work with iterative data integration and consumption patterns for rapid prototyping and development of data integration requirements.
- Access is simpler for business applications that integrate or analyze small data volumes.
- Cloud data integration capabilities enable rapid integration of enterprise data with a cloud data API, eliminating the need for separate, manual integration of cloud data.
- Lower-cost data platforms due to improvements in performance and capabilities for data retrieval and integration offered by virtualization.
- Providing application access to information via data virtualization can help isolate users from changes in the underlying data technologies. This design also means we avoid the expense and complexity of the associated change management.

Cisco IT is using several elements of the Cisco Data Virtualization Suite. The primary element is the Cisco Information Server, a complete data virtualization platform that supports consistent API access to data from multiple and diverse sources. Cisco Information Server accesses, integrates, transforms, and delivers data to consuming applications such as business intelligence, analytic tools, and information dashboards as well as transactional web and mobile applications.

Cisco Information Server supports a multitenant model for application teams. This model provides a dedicated workspace for developing the interaction of application web services with the Cisco Information Server. To avoid duplication, we publish commonly used enterprise information views, which the application teams can combine with function-specific datasets to build relevant views for the associated business group.

Solution Deployment

We implemented the Cisco Information Server software on two virtual clusters of Cisco Unified Computing System™ (Cisco UCS®) B-Series blade servers, running with a load-balanced design across data centers in Active-Active mode. We recently migrated to the Cisco Application Centric Infrastructure (ACI), which improves the agility and manageability of the server infrastructure while minimizing its overall cost. Additionally, we can provision capacity seamlessly and securely without manually configuring every infrastructure component. The move to ACI gives us a scalable and secure, multitenant infrastructure with complete visibility into application performance across the data virtualization clusters.

We use a mission-critical production cluster to serve the most important business applications and a standard cluster for non-mission-critical applications. This design allows us to manage the clusters efficiently and support the agile development needs of application teams.

We use the Cisco Information Server Monitor tool for real-time performance monitoring of the Cisco Data Virtualization Suite environment and the overall health of the cluster. Performance information also helps the data architects create better models for the application queries and views. The tool's alert process monitors system resources such as cluster health and CPU as well as memory and storage usage by each application. In order to maintain the availability and stability of the virtualization service, application owners receive alerts when their application queries create abnormal usage of system resources.

The Cisco Information Server Deployment Manager helps us quickly and easily migrate resources, cache settings, server configurations, security profiles, and other information from one Cisco Information Server instance to another across lifecycle environments. Application teams create and manage the deployment plan for the instances.

The Cisco Information Server Business Directory provides a self-service directory of virtualized business data without requiring the higher level of technical expertise needed to view the data using Cisco Information Server Studio. Users can search, browse, and collaborate on all available data and categorize large, diverse datasets. Business analysts can use their preferred analytic or business intelligence tools to identify the data in Business Directory that will best support their business decisions and actions.

Application developers can use the Cisco Enterprise Policy Manager as an option for defining secure, differentiated access to corporate data.

Use Cases and Examples

Cisco Data Virtualization technology is helping us address multiple use cases for data access and integration by applications.

Data Federation

Collecting and integrating data from multiple sources is the task of federation. This use case enables the Cisco Supply Chain Risk Management group to combine data from two different sources (Oracle and Teradata), then apply business logic to the combined dataset for analytics and reporting.

Results:

- Information that helps the group proactively highlight and mitigate supply chain risks from the perspective of component, supplier, manufacturing, resiliency, product, and revenue.
- Applications access data through a modern, standards-based framework that connects legacy data sources.
- Readiness for the Internet of Everything (IoE) because the risk management team can track components and products through the entire supply chain lifecycle and can respond in real-time to any issues.
- 10 percent reduction in costs of the operational infrastructure for this application.
- 50 percent improvement in report-generation performance and user experience.

Big Data Integration

This use case addresses the need to control the storage and integration costs of ever-growing data. For example, a logically expanded data warehouse leverages data virtualization capabilities for federating multiple data sources.

Traditionally, data warehouses have been designed to process and store large volumes of enterprise data for mission-critical and historical analytics reporting by integrating data from disparate sources in a central repository. To reduce the usage of disk space, CPU, and I/O resources in our Teradata enterprise data warehouse platform, we offloaded infrequently used data into a Hadoop repository. With an extended data warehouse, we are able to make both current and infrequently used reference data readily accessible for business analytics.

Results:

- Online reports are delivered with no delays. Previously, business users had to wait up to five days for database administrators to retrieve the data from the archive and run the reports manually. The performance improvements for report delivery and reductions in manual data retrieval are estimated to produce cost savings of more than USD \$100,000 per year.
- Greater flexibility for ad hoc data requests because it is easy to add new data sources to the environment.

Data Abstraction

Enterprise views provide key business information based on datasets available in platforms such as Hadoop, Teradata, and SAP HANA. Application teams that support data needs across business functions are able to connect to these published views and combine them with function-specific data sets, allowing application users to explore the data and generate business insights.

Example: Cisco Services AutoQuote. The Cisco AutoQuote software helps Cisco Services account managers and channel partner sales representatives easily identify and act on sales opportunities for service contract renewals and upgrades. The software retrieves data about the customer, contract, and renewal pricing to automatically generate price quotes that are emailed to salesperson.

Results:

- Cisco AutoQuote simplifies the renewal process for customers while increasing renewal rates for service contracts.
- Data on the top 20 opportunities for service contract cross-sell and upsell are available in multiple applications used by sales representatives.
- Sales people spend less time on high-volume but low-revenue renewals and can focus on selling higher-value service contracts.
- Automatically generated quotes provide consistency by using application rules for defining, recording, and crediting sales opportunities, quotes, and bookings.
- Developers can thread the same data through different applications and platforms.
- Decoupling of the presentation layer hides from user applications the complexity of the underlying source of records (SoR) and single source of truth (SSoT) data models.
- Provides a model of high-level API reuse and data set integration across IT delivery teams.

Cloud Data Integration

In order to get a holistic view of the business and empower decision makers with actionable insights, there is always a need to combine existing data with new data, whether those datasets are in on-premises systems or in the cloud. This hybrid access must be simple for application developers and deliver acceptable performance for application users.

Example: Salesforce.com integration. Cisco uses Salesforce.com to store data about customer support cases and return authorizations in the cloud, with data uploaded using an API.

Results:

- Cisco customer sales teams receive push notifications whenever a customer creates a high-severity support case, helping to deliver fast response to customer issues.
- Reduction in case-open time because customers, account teams, and partners receive updates when a Cisco TAC Engineer has been assigned to the support case, when the customer raises the severity level, or when another change has been made to the case status.

Data Prototyping

Developers often want to examine and extract data from multiple sources, combine it with data stored on a user's desktop computer, then import the resulting dataset into an analytics tool or a spreadsheet for the user to review or share. With traditional access methods, data from the various sources would need to be moved to a separate file for examination, then transferred to the application. With data virtualization, developers can use the Cisco Information Server to review the data, then receive ready-to-use queries for integrating that data in their applications.

Results:

- Faster to generate reports because developers don't need to move data in order to identify and extract it.
- Offers an easier method for application developers to work with business users on identifying the types and sources of data they need.
- Four times faster to develop new data warehouse projects by identifying the right data for a Cisco Information Server view instead of directly building an extract, transform, and load (ETL) script.

Results

As of late 2015, the Cisco Data Virtualization implementation is complete and serving many applications. Virtualizing data helps our efforts toward business objectives that include improving business performance, increasing agility, reducing costs, decreasing risk, and ensuring compliance. Other early results from this implementation show multiple benefits for Cisco IT, application developers, and information users.

Lower costs and faster application development. Application development is proving to be significantly faster, with data virtualization enabling time savings of up to 40 percent because less custom code is required. The Cisco Data Virtualization API allows developers to reuse code components among applications and to rapidly model and prototype new application functionality, capabilities that also reduce development effort by 20 percent, leading to lower application costs.

"Data virtualization changes the way data applications are built," says Bhargava. "Instead of moving data between databases or writing complex data access code, the data stored in various repositories across the enterprise can be made available via a SQL or REST API with one click, and without writing any code."

Design once and consume everywhere. By always retrieving data directly from the source, virtualization eliminates inconsistent views of data and avoids the costs of manual retrieval and replication efforts. It also allows reuse of the same data source while supporting the preferred input format of the consuming application.

Secure access to all data. Manual data preparation and distribution leads to major concerns around securing highly confidential data. Cisco Information Server uses unified security methods and controls to consistently apply security rules across all data sources and consuming applications and users.

Ability to unify cloud and enterprise data. The traditional approach to integrating cloud data with on-premises data is not standardized, so it can require a lot of custom coding. Cisco Data Virtualization offers a wide range of data transformation utilities, as well as access to cloud-based data sources and consumers via industry-standard APIs that simplify and speed new development. Support for multiple security models ensures proper data authentication, authorization, and encryption of web services, both on-premises and in the cloud.

Lessons Learned

Cisco customers can learn from the insights we have gained in implementing Cisco Data Virtualization technology and using virtualized data in business applications.

Reuse data resources across application projects. Teams that develop applications across business functions can share common objects in the multi-tenant Cisco Information Server environment.

Design for scalability and new data sources. Prepare to expand the data service for more use cases and data sources over time. Scalability allows for data analysis from new and varied data sources, leading to improved data insights for business users.

Establish application best practices. Guidelines for application developers will help them better use the data service in order to optimize data retrieval and application performance. Cisco IT created an online support community where application developers and business users can find documentation and other resources for using data virtualization effectively.

Update the data service frequently. Regular releases of product service packs with new features will help drive rapid adoption of the virtualized data service.

Next Steps

Cisco IT will continue to support more business use cases by leveraging a custom software development kit (SDK) to support access to more NoSQL data sources and adapters for cloud applications.

For More Information

Products and services: [Cisco Data Virtualization](#), [Cisco UCS servers](#), [Cisco Enterprise Policy Manager](#), and [Cisco Application Centric Infrastructure](#)

Learn more about concepts and business benefits in this [data virtualization white paper](#).

To read additional Cisco IT case studies on a variety of business solutions, visit Cisco on Cisco: Inside Cisco IT <http://www.cisco.com/go/ciscoit>

Note

This publication describes how Cisco has benefited from the deployment of its own products. Many factors may have contributed to the results and benefits described; Cisco does not guarantee comparable results elsewhere.

CISCO PROVIDES THIS PUBLICATION AS IS WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Some jurisdictions do not allow disclaimer of express or implied warranties; therefore, this disclaimer may not apply to you.



Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)