Introduction

What is SDN?

Software Defined Networking (SDN) is a very promising networking paradigm that aims to decouple the control logic from the forwarding system by promoting a logically centralized entity, also known as the network controller or network operating system. The separation between the control plane, the logic that enforces network policies and take forwarding decisions, and the data plane, the underlying switching infrastructure, introduces a flexible way of monitoring, managing and evolving a network. SDN essentially makes network architecture simpler and the control plane more adaptive to changes.

SDN is an emerging network architecture that is highly adaptable. It offers the potential for multi-domain, automated, guaranteed bandwidth service, while providing for the programmatically configured, agile infrastructure that today's big data necessitates. The SDN approach allows network administrators greater control of network services by decoupling the network control and forwarding functions. SDN architecture is directly programmable and centrally managed, and its programmability allows managers to configure, manage, secure, and optimize network resources very quickly. By abstracting the underlying infrastructure for applications and network services, SDN lets administrators dynamically adjust network-wide traffic flow to meet changing needs, while maintaining a global view of the network.

Software-Defined Networking (SDN) is an emerging architecture that is dynamic, manageable, cost-effective, and adaptable, making it ideal for the high-bandwidth, dynamic nature of today’s applications. This architecture decouples the network control and forwarding functions enabling the network control to become directly programmable and the underlying infrastructure to be abstracted for applications and network services.

The SDN Architecture is:

- **Directly programmable:**

Network control is directly programmable because it is decoupled from forwarding functions.

- **Agile**:

Abstracting control from forwarding lets administrators dynamically adjust network-wide traffic flow to meet changing needs.

- **Centrally Managed:**

Network intelligence is (logically) centralized in software-based SDN controllers that maintain a global view of the network, which appears to applications and policy engines as a single, logical switch.

- **Programmatically Configured:**

SDN lets network managers configure, manage, secure, and optimize network resources very quickly via dynamic, automated SDN programs, which they can write themselves because the programs do not depend on proprietary software.

- **Open standards-based and vendor-neutral:**

When implemented through open standards, SDN simplifies network design and operation because instructions are provided by SDN controllers instead of multiple, vendor-specific devices and protocols.

Advantages of SDN

- Centralized controllers have global network view, so the operators can program the whole network from a single point.
- Network operators can dynamically change the whole network traffic flow to meet their demands.
- SDN provides vendor-neutral control because controller software development does not depend on devices types.

What is Data Virtualization?

Data virtualization is any approach to data management that allows an application to retrieve and manipulate data without requiring technical details about the data, such as how it is formatted at

source, or where it is physically located and can provide a single customer view (or single view of any other entity) of the overall data.

Data virtualization may also be considered as an alternative to ETL (extract, transform, and load) and data warehousing. Data virtualization is inherently aimed at producing quick and timely insights from multiple sources without having to embark on a major data project with extensive ETL and data storage. However, data virtualization may be extended and adapted to serve data warehousing requirements also. This will require an understanding of the data storage and history requirements along with planning and design to incorporate the right type of data virtualization, integration, and storage strategies, and infrastructure/performance optimizations (e.g., streaming, in-memory, hybrid storage). The increase in data sources, especially integration with Big Data and Unstructured data made Data Virtualization platform important part of enterprise data access strategy. Data virtualization provides the following attributes for efficient data access across enterprise.

Data Virtualization software provides some or all of the following capabilities:

- Abstraction – Abstract the technical aspects of stored data, such as location, storage structure, API, access language, and storage technology.
- Virtualized Data Access – Connect to different data sources and make them accessible from a common logical data access point.
- Transformation – Transform, improve quality, reformat, aggregate etc. source data for consumer use.
- Data Federation – Combine result sets from across multiple source systems.
- Data Delivery – Publish result sets as views and/or data services executed by client application or users when requested.

With the above benefits of the Data Virtualization Platform in mind, it is evident that enterprises will find it more useful if Data Virtualization platforms are built with Data Mining Models and Algorithms, so that effective Data Mining can be performed on top of Data Virtualization platform.

As the important part of Data Mining is about identifying the correct data sources and associated events of interest, effective Data Mining can be built if disparate data sources are brought under

the scope of Data Virtualization Platform rather than putting the Data Mining inside a single database engine.

Data virtualization software may include functions for development, operation, and/or management. The following extended view of Data Virtualization Platform signifies how Data Mining can be part of Data Virtualization Platform.

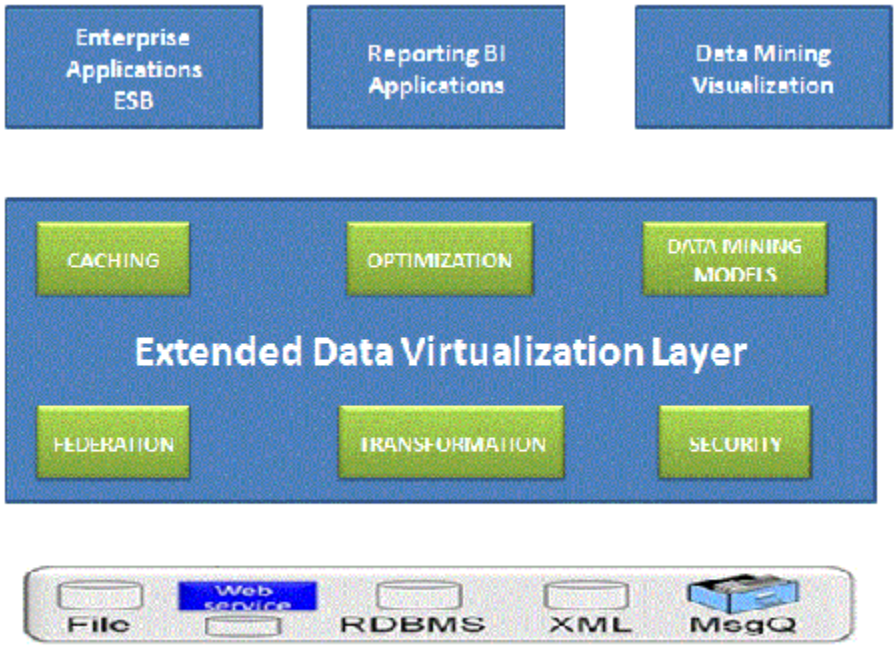


Figure 1: Data Virtualization in Data Mining

Benefits include:

- Reduce risk of data errors.
- Reduce systems workload through not moving data around
- Increase speed of access to data on a real-time basis
- Significantly reduce development and support time
- Increase governance and reduce risk through the use of policies.
- Reduce data storage required.

Drawbacks include:

- May impact Operational systems response time, particularly if under-scaled to cope with unanticipated user queries or not tuned early on.
- Does not impose a heterogeneous data model, meaning the user has to interpret the data, unless combined with Data Federation and business understanding of the data
- Requires a defined Governance approach to avoid budgeting issues with the shared services
- Not suitable for recording the historic snapshots of data. A data warehouse is better for this
- Change management "is a huge overhead, as any changes need to be accepted by all applications and users sharing the same virtualization kit".

Data Mining helps organizations to discover new insights from existing data, so that predictive techniques can be applied towards various business needs. The following are the typical characteristics of data mining.

- Extends Business Intelligence, beyond Query, Reporting and OLAP (Online Analytical Processing)
- Data Mining is cornerstone for assessing the customer risk, market segmentation and prediction
- Data Mining is about performing computationally complex analysis techniques on very large volumes of data
- It combines the analysis of historical data with modeling techniques towards future predictions, it turns Operations into performance

The following are the use cases that can benefit from the application of data mining:

- **Manufacturing / Product Development:** Understanding the defect and customer complaints into a model that can provide insight into customer satisfaction and help enterprises build better products
- **Consumer Payments:** Understand the payment patterns of consumers to predict market penetration analysis and discount guidelines.
- **Consumer Industry:** Customer segmentation to understand the customer base and help targeted advertisements and promotions.

- **Consumer Industry:** Campaign effectiveness can be gauged with customer segmentation coupled with predictive marketing models.
- **Retail Industry:** Supply chain efficiencies can be brought by mining the supply demand data

'In Database' Data Mining

Data Mining is typically a multi-step process.

1. Define the Business Issue to Be Addressed, e.g., Customer Attrition, Fraud Detection, Cross Selling.
2. Identify the Data Model / Define the Data / Source the Data. (Data Sources, Data Types, Data Usage etc.)
3. Choose the Mining Technique (Discovery Data Mining, Predictive Data Mining, Clustering, Link Analysis, Classification, Value Prediction)
4. Interpret the Results (Visualization Techniques)
5. Deploy the Results (CRM Systems.)

Initially Data Mining has been implemented with a combination of multiple tools and systems, which resulted in latency and a long cycle for realization of results.

Sensing this issue, major RDBMS vendors have implemented Data Mining as part of their core database offering. This offering has the following key features:

- Data Mining engine resides inside the traditional database environment facilitating easier licensing and packaging options
- Eliminates the data extraction and data movement and avoids costly ETL process
- Major Data Mining models are available as pre-built SQL functions which can be easily integrated into the existing database development process.

How Does Software-Defined Networking (SDN) Work?

Software-defined networking providers offer a wide selection of competing architectures, but at its most simple, the software-defined networking method centralizes control of the network by separating the control logic to off-device computer resources.

All software-defined network solutions have some version of an SDN Controller, as well as southbound APIs and northbound APIs:

- **Controllers:** The "brains" of the network, SDN Controllers offer a centralized view of the overall network, and enable network administrators to dictate to the underlying systems (like switches and routers) how the forwarding plane should handle network traffic.
- **Southbound APIs:** Software-defined networking uses southbound APIs to relay information to the switches and routers "below." OpenFlow, considered the first standard in SDN, was the original southbound API and remains as one of the most common protocols. Despite some considering OpenFlow and SDN to be one in the same, OpenFlow is merely one piece of the bigger landscape.
- **Northbound APIs:** Software-Defined Networking uses northbound APIs to communicates with the applications and business logic "above." These help network administrators to programmatically shape traffic and deploy services.

The Software-Defined Networking Solutions Framework

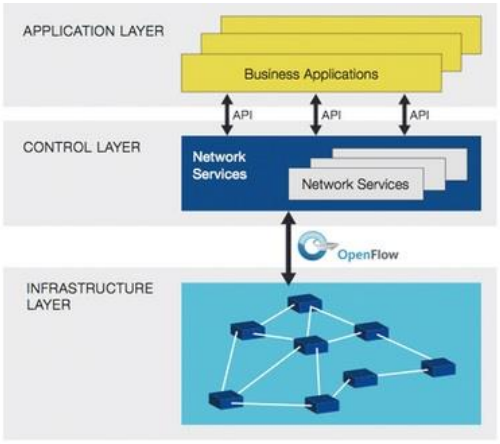
- 

Figure 2: Software-Defined Networking (SDN) Framework

Application areas of Software Defined Network in data mining

One of the application areas of SDN in data mining is for Data Center Networks. A data center network is an appropriate place to deploy SDN due to having full control by a provider of a data center over its entire set of devices. An SDN approach typically meets the requirement with a much better ability for flow control and virtual machine migration.

Application areas of data virtualization in data mining.

The increase in data sources, especially integration with Big Data and Unstructured data made Data Virtualization platform important part of enterprise data access strategy. Data virtualization provides the following attributes for efficient data access across enterprise.

- **Abstraction:** Provides location, API, language and storage technology independent access of data
- **Federation:** Converges data from multiple disparate data sources
- **Transformation:** Enriches the quality and quantity of data on a need basis
- **On-Demand Delivery:** Provides the consuming applications the required information on-demand

With the above benefits of the Data Virtualization Platform in mind, it is evident that enterprises will find it more useful if Data Virtualization platforms are built with Data Mining Models and Algorithms, so that effective Data Mining can be performed on top of Data Virtualization platform.

As the important part of Data Mining is about identifying the correct data sources and associated events of interest, effective Data Mining can be built if disparate data sources are brought under the scope of Data Virtualization Platform rather than putting the Data Mining inside a single database engine.

What can be done (our suggestion)?

Social media, mobile devices, and cloud computing are pushing traditional networks to their limits. Compute and storage have benefited from incredible innovations in virtualization and automation, but those benefits are constrained by limitations in the network. Administrators may spin up new

compute and storage instances in minutes, only to be held up for weeks by rigid and oftentimes manual network operations. Software-defined networking has the potential to revolutionize legacy data centers by providing a flexible way to control the network so it can function more like the virtualized versions of compute and storage today.

Conclusion on the future of SDN and Data Virtualization

Looking at all of the choices that companies have right now to source "programmable" network technologies, you could say that the future of software-defined networking is already here. One of the exciting things about SDN is its ability to simplify network design. If you order these things through a subscription, as a user, you may be able to see them as more transparent units – going back to that visual dashboard example, you can see them in a more visual way.

Data Virtualization is becoming part of the mainstream enterprise data access strategy, mainly because it abstracts the multiple data sources and avoids complex ETL processing and facilitates the single version of truth, data quality and zero latency enterprise. If value adds like a Data Mining engine can be built on top of the existing Data Virtualization platform, the enterprises will benefit further.