

## Machine Learning 2018: Data virtualization: Using data virtualization for an integrated analytics platform- Manoj Mishra-Union Insurance

**Manoj Mishra**

*Union Insurance, UAE*

In order to possess the competitive advantage, organizations worldwide are driving the necessity for better analytics (historical, realtime, predictive and cognitive) of knowledge across various domains including customers, products, services and operations. thanks to this, the data available for such analytics is exploding in size, technology and complexity. for several years companies have invested in technologies like data warehouses, data marts, OLAP tools, Big Data/Hadoop systems and streaming real-time analytics platforms to require advantage of those opportunities. Total value proposition to the business is maximized only these are combined into an integrated analytics platform. However, traditional tools cannot integrate streaming data and data at rest especially when the data is spread on-premises, cloud, websites and documents everywhere. Data virtualization are often wont to provide cross platform logical views of knowledge and analytic insights across the enterprise to supply an integrated analytics platform. By utilizing native integration with in-memory data grids for processing , data virtualization can deliver a unified and centralized data services fabric with security and real-time integration across multiple traditional and large data sources, including Hadoop, NoSQL, cloud and software-as-a-service (SaaS). Hence data virtualization is becoming a requirement to deal with the unique challenges of knowledge explosion in today's changing business climate.

Data virtualization presents a contemporary approach to data integration. Unlike ETL solutions, which replicate data, data virtualization leaves the info in source systems, simply exposing an integrated view of all the info to data consumers. As business users drill down into reports, data virtualization fetches the info in real time from the

underlying source systems. Data virtualization proves that connecting to data is way superior to collecting it.

Data virtualization may be a unified, virtual data layer with which enterprise applications and users can access any enterprise information no matter its location, format, or protocol, using the methods that best suit their work needs like data discovery and search

Data virtualization is an approach to data management that permits an application to retrieve and manipulate data without requiring technical details about the info , like how it's formatted at source, or where it's physically located, and may provide one customer view (or single view of the other entity) of the general data

Unlike the normal extract, transform, load ("ETL") process, the info remains in situ , and real-time access is given to the source system for the info . This reduces the danger of knowledge errors, of the workload moving data around which will never be used, and it doesn't plan to impose one data model on the info (an example of heterogeneous data may be a federated database system). The technology also supports the writing of transaction data updates back to the source systems. To resolve differences in source and consumer formats and semantics, various abstraction and transformation techniques are used. this idea and software may be a subset of knowledge integration and is usually used within business intelligence, service-oriented architecture data services, cloud computing, enterprise search, and master data management.

Data virtualization can also be considered as an alternate to ETL and data warehousing. Data virtualization is inherently aimed toward producing

quick and timely insights from multiple sources without having to start a serious data project with extensive ETL and data storage. However, data virtualization could also be extended and adapted to serve data warehousing requirements also. this may require an understanding of the info storage and history requirements along side planning and style to include the proper sort of data virtualization, integration, and storage strategies, and infrastructure/performance optimizations (e.g., streaming, in-memory, hybrid storage).

**Biography :**

Manoj Mishra has completed his Bachelor of Engineering in Computer Science and a Certification in Data Science from Johns Hopkins University. He has more than two decade of experience spreading across multiple geographies (US, Europe, India and Middle East) working with organizations like Adobe Systems, Dell, Perot Systems, CEB-Gartner, Rolta and Tata Group. He is currently a Chief Manager of Business Intelligence and Data with Union Insurance and currently leading their data strategy and technology transformations through data analytics, research and various AI initiatives.

E-mail: [manumishra2001@gmail.com](mailto:manumishra2001@gmail.com)