

Introduction to Data Virtualization: Technology and Use Cases

Course Outline

1. Introduction to Data Virtualization

- What is data virtualization?
- Use case of data virtualization: business intelligence, data science, democratizing of data, master data management, distributed data
- Differences between data abstraction, data federation, and data integration
- Open versus closed data virtualization servers
- Market overview: AtScale, Cirro Data Hub, Data Virtuality, Denodo Platform, FraXses, IBM Data Virtualization Manager for z/OS, RedHat JBoss Data Virtualization, Stone Bond Enterprise Enabler, and Tibco Data Virtualization

2. How Do Data Virtualization Servers Work?

- The key building block: the virtual table
- Integrating data sources via virtual tables
- Implementing transformation rules in virtual tables
- Stacking virtual tables
- Impact analysis and lineage
- Running transactions – updating data
- Securing access to data in virtual tables
- Importing non-relational data, such as XML and JSON documents, web services, NoSQL, and Hadoop data
- The importance of an integrated business glossary and centralization of metadata specifications

3. Performance Improving Features

- Caching of a virtual table for improving query performance, creating consistent report results, or minimizing interference on source systems
- Differences between full refreshing, incremental refreshing, live refreshing, online refreshing and offline refreshing
- Different query optimization techniques, including query substitution, pushdown, query expansion, ship joins, sort-merge Joins, statistical data and SQL override

4. Use Case 1: The Logical Data Warehouse Architecture

- The limitations of the classic data warehouse architecture
- On-demand versus scheduled integration and transformation
- Making a BI system more agile with data virtualization
- The advantages of virtual data marts
- Strategies for adopting data virtualization

- Application areas of data virtualization
- The need for powerful analytical database servers
- Migrating to a data virtualization-based BI system

5. Use Case 2: Data virtualization and Master Data Management

- How can data virtualization help with creating a 360° view of business objects
- Developing MDM with a data virtualization server – from a stored to a virtual solution
- On-demand data profiling and data cleansing

6. Use Case 3: From the Physical Data Lake to the Logical Data Lake

- Practical limitations of developing one physical data lake
- Shortening the data preparation phase of data science with data virtualization
- Sharing metadata specifications between data scientists
- Implementing analytical models inside a data virtualization server

7. Use Case 4: Democratizing Enterprise Data

- Increasing the business value of the data asset by making all the data available to a larger group of users within the organization
- The business value of consistent data integration
- Using lean data integration to make data available for analytics and reporting faster
- One consistent data view for the entire organization
- How the business glossary and search features help business users
- The coming of the data marketplace

8. Use Case 5: Dealing with Big Data

- Big data can be too big to move - data can't be transported to the place of integration
- Data virtualization pushes data processing to where the data is produced
- Hiding the physical location of the data
- With data virtualization, the network becomes the database

9. Closing Remarks

- The Future of Data Virtualization
- Data virtualization as driving force for data integration
- Potential new product features