WHITE PAPER

INTELLIGENT BUSINESS STRATEGIES

Accelerating Workload Migration to the Cloud Using Data Virtualisation

By Mike Ferguson
Intelligent Business Strategies
December 2020

Research sponsored by:
TIBCO™

Table of Contents

INTRODUCTION

THE MOMENTUM TO MOVE TO THE CLOUD IS NOW INCREASING

Over the last five years we have seen momentum to move to the cloud constantly increasing to the point where today it is unstoppable. The flood gates have opened. So much so that we are now at a point where 'cloud first' has become the deployment strategy in many enterprises. But is it more than that because we are now at a point where most companies have gone beyond a single cloud into a multi-cloud operational set-up. This can be seen from several surveys over the last few years such as the two shown in Figure 1 where it is evident that a hybrid multi-cloud set-up is now the norm.

A hybrid multi-cloud operating environment is now the norm in most enterprises

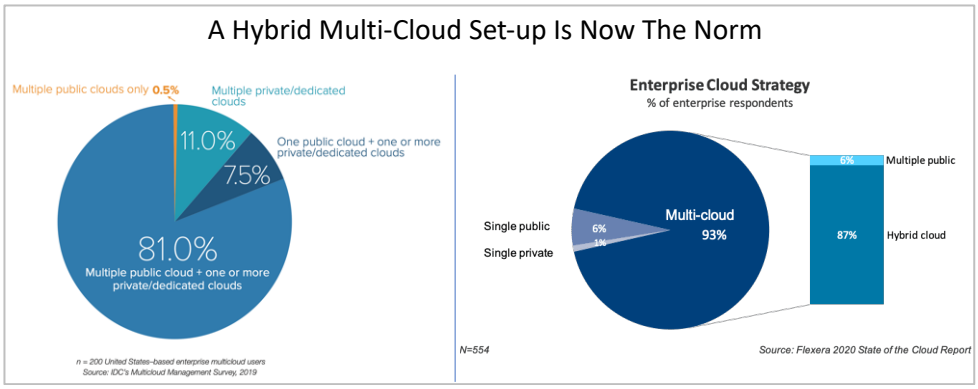


Figure 1

OPERATIONAL AND ANALYTICAL SYSTEMS ARE BEING MIGRATED

Both operational and analytical systems are being migrated to the cloud. In many cases the migration and uptake of operational transaction processing applications led the way with front and back office Software-as-a-Service (SaaS) applications now increasingly widely used as shown in Figure 2.

Many companies are now running SaaS transaction processing applications on one or more clouds

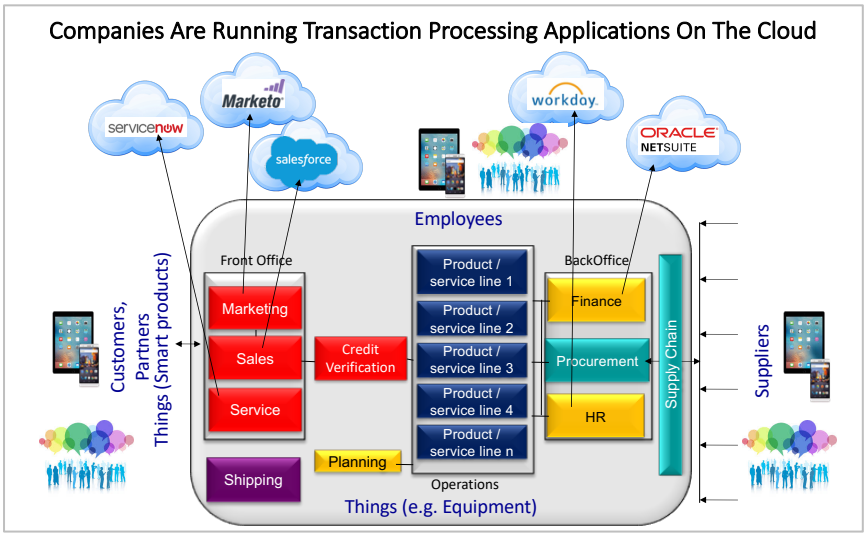


Figure 2

However, over the last two years the pendulum has swung towards migration of analytical systems. A recent survey[1] of five hundred enterprises clearly shows this shift in Figure 3. 51.8% of companies surveyed said they were in the process of migrating

[1] The 4th Industrial Revolution Survey, Intelligent Business Strategies, November 2019

The momentum has now shifted to migrating analytical workloads to the cloud

analytical workloads to the cloud with 42.4% citing data warehouse migration as already underway. In addition, when asked what percentage of analytical workloads they anticipated migrating through the end of 2022, 35.8% said that they expected to migrate as much as 60% of their analytical workload with a further 27.4% saying that up to 40% of their analytical workload would be migrated. That is a major push towards running data warehouses, data lakes, data science projects, graph analysis and streaming analytics all in a cloud computing environment.

42% of companies said they are migrating their data warehouses to the cloud

More migration of analytical workloads is planned

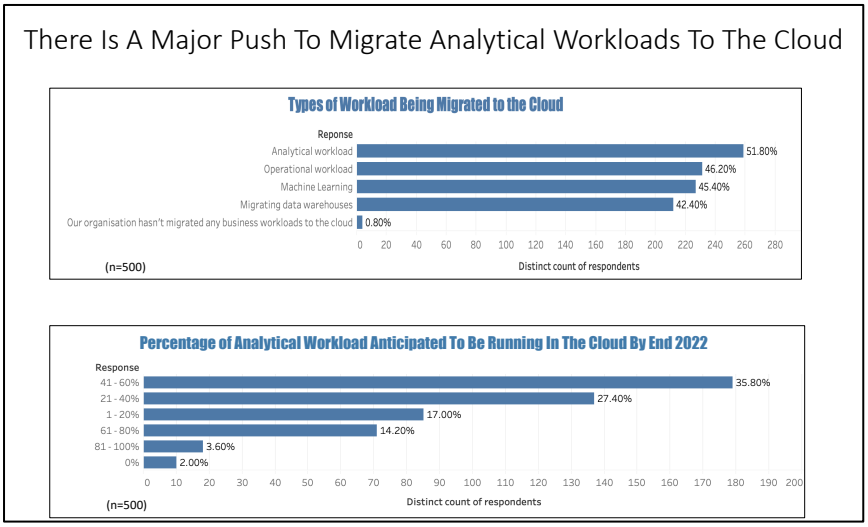


Figure 3

Given this kind of momentum, what then are the challenges facing companies in migrating to the cloud with respect to data? What kind of requirements need to be met to simplify access to data before, during and after migration? And how can data virtualisation address these challenges and requirements while also helping to accelerate workload migration to the cloud?

This paper tries to answer these questions and looks at how one vendor, TIBCO, is stepping up to tackle this with its TIBCO Data Virtualisation product.

DATA CHALLENGES FACING COMPANIES MIGRATING TO THE CLOUD

Companies face a number of challenges associated with data when migrating to the cloud

There are a number of data challenges that can occur when migrating systems to the cloud. These will differ system by system. Nevertheless, the following is a summary of typical data challenges associated with migrating operational transaction processing systems and analytical systems to the cloud.

DATA CHALLENGES ASSOCIATED WITH TRANSACTION SYSTEM MIGRATION

Operational transaction processing system migration is primarily a migration to a packaged SaaS application either from an existing on-premises version of the same packaged application or from one or more custom developed applications to a SaaS application. Transaction processing systems include applications for customer relationship management (CRM), sales force automation, order entry, customer service, marketing campaign management, inventory management, enterprise resource planning (ERP), procurement and supply chain management (SCM).

In practice, transaction processing applications rarely function as stand-alone systems. More often than not:

Data needs to flow between applications running on-premises and in the cloud as part of a business process execution

- Data has to flow between multiple transaction processing applications as part of executing a business process e.g. Order-to-cash, plan-to-manufacture, procure-to-pay

Real-time operational reporting using data in cloud and on-premises application databases is also needed

- Real-time operational reporting on live transaction data is needed to understand what is happening right now within some part of the business. This includes things like sales pipeline status, orders versus inventory right now and daily cashflow reporting. Operational reporting may also need near real-time data from one or multiple transaction processing systems to be integrated in an operational data store (ODS)

- Transaction systems are often data sources to analytical systems such as data warehouses

Therefore, when migrating transaction processing applications to the cloud the data challenges may go beyond the boundaries of the application and include:

Business process data flows still need to work even if one or more applications involved is migrated

- Maintaining inbound and outbound application-to-application integration data flows between the newly migrated cloud-based transaction processing application and other transaction processing applications that remain on-premises or that are also in the cloud or both

The length of time needed to migrate an application may be lengthened due to the need to re-develop operational reports

- Redevelopment or changes to operational reports when the data needed for reporting
  - Has been migrated to new data structures in a new SaaS application on the cloud
  - Is in a transaction processing application migrated to the cloud and in an on-premises application
  - Is in multiple cloud-based transaction processing applications

Data warehouse ETL jobs sourcing data from a migrated application may be impacted

- Extract, transform and load (ETL) jobs taking data from the transaction processing system data source that is being migrated to the cloud may break and need to be re-developed or changed

DATA CHALLENGES ASSOCIATED WITH ANALYTICAL SYSTEM MIGRATION

The approach taken to data warehouse migration will impact migration time and risk

With respect to analytical systems, we know from Figure 3 that 42.4% of companies surveyed are migrating data warehouses to the cloud. The data challenges when migrating a data warehouse to the cloud include:

- The length of time it takes to migrate the data warehouse and all of its dependent data marts to the cloud

- The impact on BI tool reports and dashboards and on analytical applications accessing the data warehouse / data marts being migrated can be significant. This includes the impact caused by:
  - Switching to a different database management system (DBMS) during data warehouse migration e.g.
    - The data types of columns in tables in the original data warehouse may not all be supported in a new DBMS
    - SQL incompatibilities between the original on-premises data warehouse DBMS and the new cloud-based DBMS
    - Privileges and user groups handling differences between old and new DBMSs

Changing the DBMS is likely to impact data warehouse migration time because of data type and SQL differences

Changes to data warehouse schema design and column names are very likely to slow the migration because it will impact on ETL jobs and reports

  - Changes to data warehouses and/or data mart schema design during data warehouse migration
  - Changes to column data names made during migration
  - One of multiple data stores needed to produce a report or dashboard being migrated when the others are not e.g. a data mart has been migrated to the cloud while another data mart also needed for a report remains on-premises

If data for a report is needed from multiple data marts and only one has been migrated then BI tools have to access data on both sides of the firewall

- Reports in a business intelligence (BI) tool with a semantic layer are being migrated to a different cloud based BI tool with or without a semantic layer. This is especially problematic when aggregate metric calculations and new derived columns are defined in semantic layers

Aggregate calculation formulae are not portable across BI tools

- ETL job redevelopment may be needed if:
  - Data warehouse staging tables in a migrated data warehouse are moved to a data lake as part of the migration
  - An application data source is migrated to the cloud irrespective of whether the data warehouse itself is also being migrated

ETL jobs can be impacted if the location of staging tables is changed during migration

The obvious question is can you avoid some or all of these problems during migration using data virtualisation software to shorten the time it takes to migrate? A good way to answer this is to define a number of requirements to simplify access to data that would make this possible.

REQUIREMENTS TO SIMPLIFY DATA ACCESS BEFORE, DURING AND AFTER MIGRATION TO THE CLOUD

The following is a list of requirements to simplify access to data that can help accelerate the migration of systems to the cloud.

Companies often want to simplify and modernise data warehouse achitecture during migration

It should be possible to:

- Simplify data warehouse architecture and access to shorten the time to migrate a data warehouse and data marts to the cloud while also improving agility and reduce total cost of ownership

They want the ability to change schema design and column names during migration without forcing re-development of existing reports and analytical applications

- Shield applications and BI tool reports / dashboards from any structural design changes made to a data warehouse or data mart schema during migration to the cloud.
- Shield applications and BI tool reports / dashboards from any changes made to a data warehouse or data mart schema data names during migration to the cloud.
- Test data mart migration through creation of virtual data mart before data, BI/tool report and applications access data marts are migrated to the cloud
- Reduce the time it takes to migrate BI reports and dashboards to the cloud as part of a data warehouse migration while also simplifying access to data

They want the ability to change the DBMS during migration without forcing re-development of existing reports and analytical applications

- Shield applications and user developed BI reports from any changes made to a data warehouse or data mart or differences introduced as a result of migrating to a different DBMS in the cloud. This includes:
  - Changes made as a result of differences in data types between the old DBMS and newly adopted cloud DBMS
  - Shielding applications and BI reports from SQL differences between the old DBMS and newly adopted cloud DBMS
  - Shielding applications and BI reports from privilege and user group handling differences between old and newly adopted cloud DBMS

It should be possible to migrate transaction processing systems to the cloud without forcing re-development of existing data warehouse ETL jobs using that system as a data source

- Hide the data structures of transaction system data sources from ETL tools so that ETL jobs are not impacted by changes to source system data structures caused by:
  - Routine maintenance
  - A release upgrade of the application system
  - Migration of a data source transaction system to the cloud

Companies want the ability to reduce the impact on data warehouse ETL jobs from moving staging areas to a cloud based data lake

- Reduce the impact of ETL jobs having to be re-developed as a result of data warehouse staging tables being moved to a data lake (e.g. cloud storage) during a data warehouse migration to the cloud

- Reduce the impact of ETL jobs populating an ODS from having to be re-developed as a result of migration of a source transaction system to the cloud

They also want the ability to access data on both sides of the firewall without forcing re-development of existing reports and analytical applications

- Support the ability to join data in a system migrated to the cloud with data in on-premises systems without forcing:
    - Re-development of BI reports and dashboards
    - Business users to have to know how to do this themselves in a BI tool
    - Changes to or redevelopment of applications accessing that data

It should be possible to join data across multiple clouds without forcing re-development of existing reports and analytical applications

- Support the ability to join data in a system migrated to the cloud with data in other cloud applications without forcing
    - Re-development of BI reports and dashboards
    - Business users to have to know how to do this themselves in a BI tool
    - Changes to or redevelopment of applications accessing that data

- Reduce the number of data stores to be migrated during a data warehouse migration without losing any data and without users knowing this has happened

- Easily migrate BI reports to another cloud based BI tool quickly without the need to redefine semantic layers

How Can Data Virtualisaton Accelerate Workload Migration To The Cloud?

Having understood some of the data challenges facing companies during migration to the cloud and some key requirements, this section of the paper looks at how data virtualisation software can be used to overcome these issues and help shorten the time to migrate.

Accelerating Operational System Migration To The Cloud Using Data Virtualisation

Data virtualisation de-couples applications from data in underlying systems

There are two key areas where data virtualisation can help accelerate operational transaction system migration to the cloud. These are by:

- Reducing the need to redevelop operational reports as part of a migration
- Avoiding the need to redevelop ETL jobs for downstream data warehouses as a result of the migration

It shields BI tools from needing to know where the data is in underlying systems

The former can be done by using data virtualisation software to create virtual views or information services (stored queries published as services) on top of an operational application database to shield BI tool users and reports from the data structures of the application that is being migrated. Alternatively, data virtualisation can be used to create a virtual operational data store (ODS) that integrates data from multiple operational systems. This is shown in Figure 4.

Therefore if an application is migrated, BI tools accessing the data within it for operational reporting are not impacted

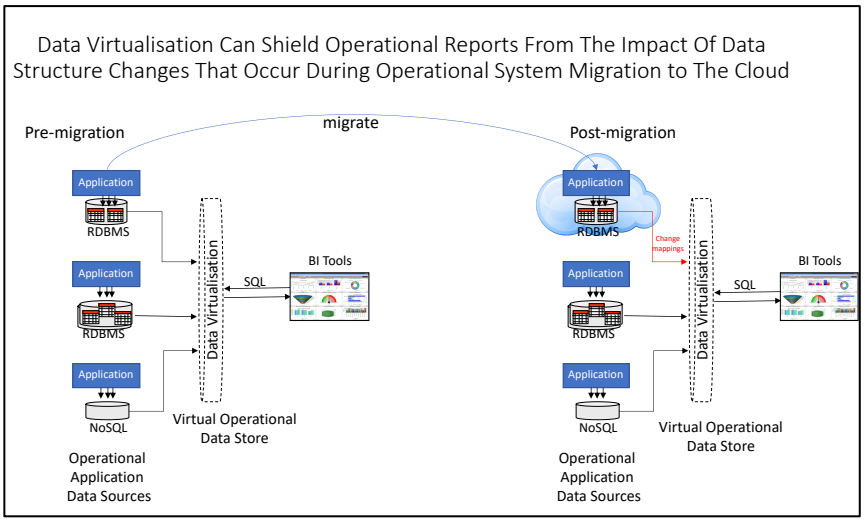


Figure 4

Therefore, even if a custom operational system is migrated to a new SaaS application, the data will need to be migrated, transformed and loaded into the new cloud application database. Once this is done, the data virtualisation server mappings from the new SaaS application into the virtual views being accessed by the reporting tools will need to be re-defined as shown. Other than that, there should be no impact on reports as long as the same data is available in the newly migrated system. This should hopefully shorten the time to migrate.

The second use of data virtualisation in accelerating migration of an operational system to the cloud is to use it to prevent ETL jobs, that take data from the operational transaction system being migrated to feed a downstream data warehouse, from

needing to be redeveloped as a result of the migration. This can be achieved by creating 'virtual data sources' so that ETL jobs extract data from virtual views of the source systems as shown in Figure 5.

Using data virtualisation to create virtual data sources shields ETL jobs from needing to know where source data is located

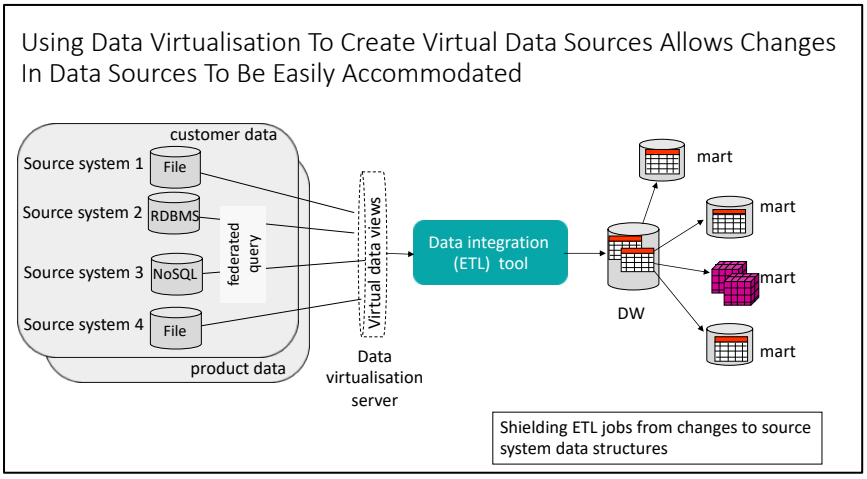


Figure 5

Introducing virtual data sources prior to migration of a source transaction system will avoid the need to have to re-develop ETL jobs after the system has been migrated

So, for example if Source system 2 is migrated to the cloud and the data structures in that system change as a result of the migration, all that would need to be changed are the mappings in the data virtualisation server virtual data source. The ETL jobs would not be impacted.

Introducing dimension and fact based virtual data sources can also improve clarity of ETL design

This approach also allows you to build target-oriented dimension and fact based ETL jobs such as the customer ETL job, the product ETL job, the orders ETL job etc with corresponding dimension and fact-based virtual data sources. It means that we highlight only the data needed in underlying physical sources for a specific dimension or transaction type based ETL job via the virtual source and then extract from there. We are not necessarily trying to join everything in the data virtualisation server, but we can at least 'ring fence' the data needed. Also, it keeps the design focused on the target rather than building ETL jobs for every source table.

Accelerating Analytical System Migration To The Cloud Using Data Virtualisation

Data virtualisation can be used to both de-risk and shorten the time it takes to migrate analytical systems to the cloud. There are several use cases with respect to analytical system migration.

Sliding data virtualisation between BI tools and underlying data warehouse and data mart databases before data warehouse migration will de-risk and accelerate the migration

De-Risking A Data Warehouse Migration

The first is to use data virtualisation to de-risk the migration of an analytical system such as a data warehouse for example. This can be done by introducing data virtualisation virtual views between BI tools accessing your data warehouse and data marts as shown in Figure 6.

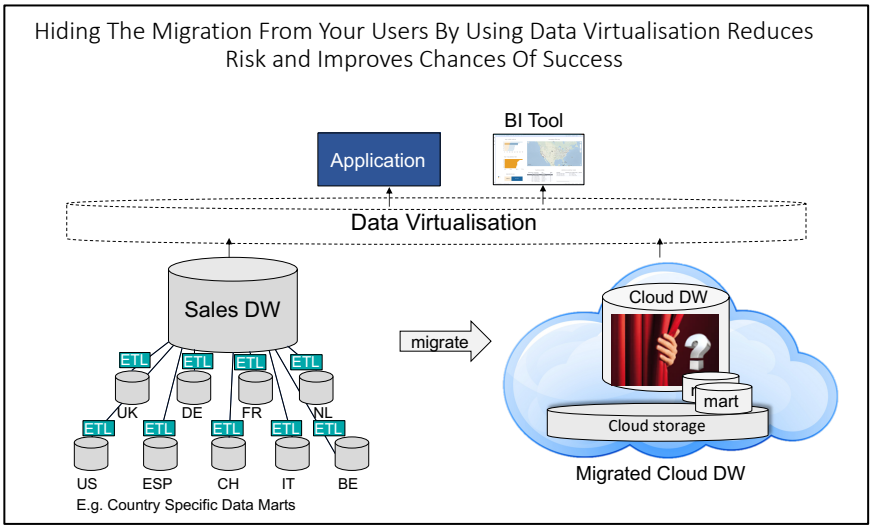


Figure 6

BI tools and analytic applications accessing the data in a data warehouse and data marts for reporting and analysis will not be impacted by the migration even if schema design changes and / or the DBMS is changed

Here, data virtualisation is used to break the dependency between business users utilising self-service BI tools and the physical schema of the underlying data warehouse and data marts that are being migrated. By introducing data virtualisation, any schema alternations made during data warehouse and data mart migration to the cloud (e.g. to optimise performance) can be hidden from business users because they only see and access virtual tables in the data virtualisation layer. If data types are changed or structural changes are needed to physical data warehouse tables, only the mappings between the underlying data warehouse or data marts and any virtual tables would need to be changed so that users remain unaware of those changes and unaware of the migration. This not only shields the impact of change and SQL differences from BI tool users and applications but also shortens migration time because it avoids the need to have to re-develop or change reports and dashboards as part of the migration.

Virtual Data Marts

Data virtualisation can be used to introduce virtual data marts prior to data warehouse migration

In addition, data warehouse architecture can be modernised and simplified to accelerate migration before migration takes place. This can be done by using data virtualisation software to replace physical data marts with virtual data marts before migration (see Figure 7). In this way, no data is lost but the number of data stores is reduced and ETL jobs taking data from data warehouses to dependent physical data marts are no longer needed. Therefore, data virtualisation can significantly reduce what has to be migrated while also introducing greater agility and lower cost of ownership at the same time.

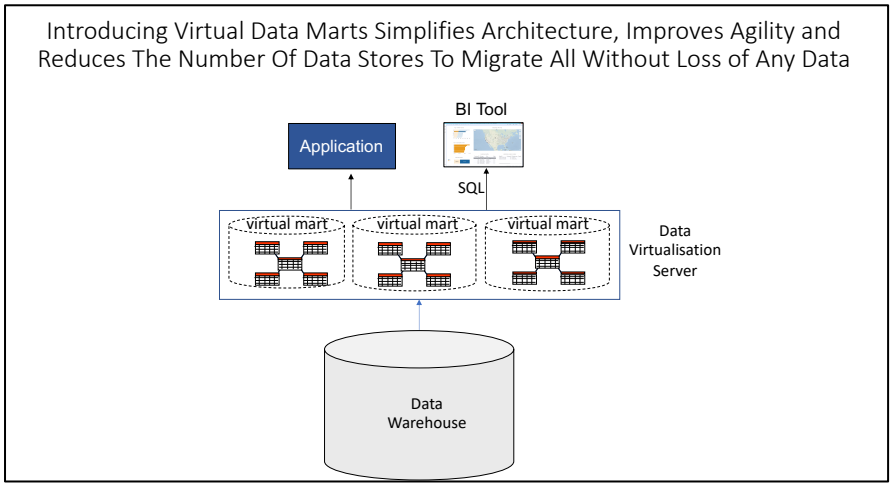


Figure 7

Replacing physical data marts with virtual data marts prior to data warehouse migration reduces risk because it reduces the number of data stores and ETL jobs to be migrated all without the loss of any data

Figure 8 shows how physical data marts can be converted to virtual ones.

The process of replacing physical data marts with virtual data marts is very straight forward and non-disruptive

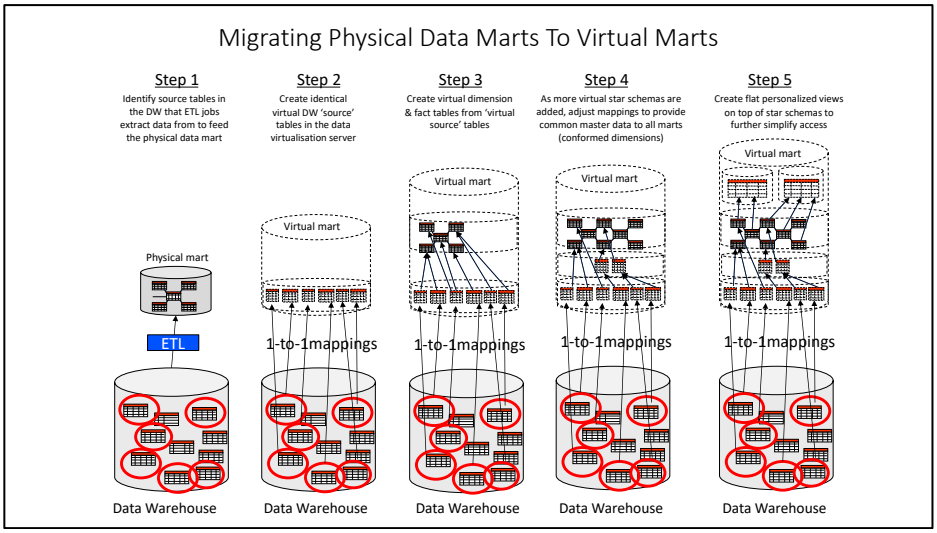


Figure 8

The first step is to identify the source tables in the data warehouse that are used to feed the physical data mart. Once this is done, you create an equivalent set of data warehouse virtual tables in a data virtualisation server. Then you create virtual star schema dimension and fact tables to represent your virtual data mart using columns in your virtual source tables. This is effectively layering virtual tables on top of virtual tables. You can then simplify access further by creating flat virtual table views on top of your virtual star schema data mart. This approach allows you to incrementally build up a set of virtual data marts with conformed (shared) dimensions across data marts. As each virtual data mart is created, you can test everything works and then switch off the old physical data mart. Doing this one data mart at a time eventually removes a number of physical data bases and ETL jobs leaving you with a much simpler migration, less risk, more agility and lower cost of ownership all without any data loss.

Joining Across Data Warehouses in A Hybrid Environment

Data virtualisation can be used to integrate data in multiple underlying data stores irrespective of whether they are in the cloud or on-premises or both

As data warehouses and data marts are migrated to the cloud there is often a need to join data in a newly migrated data warehouse to other analytical data stores that may still be running on-premises. For example, many companies have multiple data warehouses each covering different aspects of the business. Also, data warehouses often have multiple dependent data marts. Migration to the cloud could be done gradually data mart by data mart rather than attempting a complete migration of everything in one leap. For these reasons there may be a need to access data from multiple underlying data warehouses or data marts. For example, to produce management reports or corporate performance management (CPM) scorecards. If data from one data store is on premises and another has been migrated to the cloud, data virtualisation can be used to hide this complexity and present virtual views of integrated data that come from underlying data warehouses or data marts either side of the firewall as shown in Figure 9. This could be done across multiple clouds also.

This simplifies data access because BI tool users see the data already integrated as if it is in a single database

Introducing data virtualisation prior to migration avoids the need to redevelop reports and so accelerates the migration process

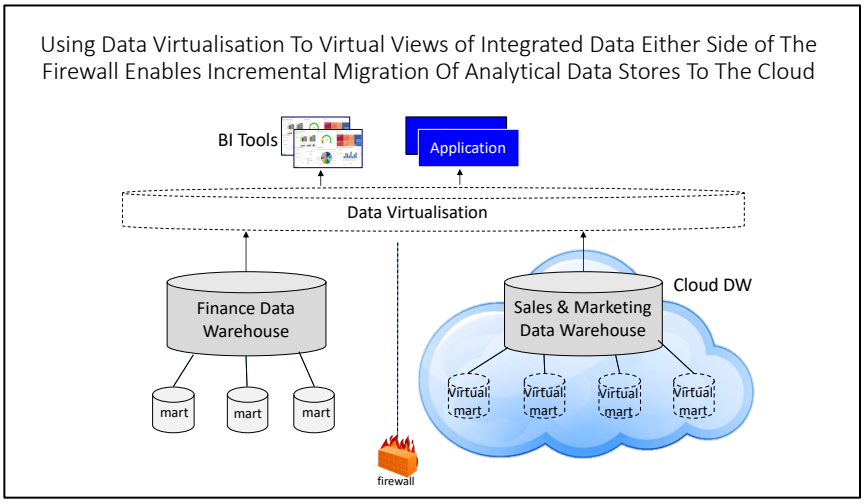


Figure 9

This is not just a use case for post migration. Introducing data virtualisation prior to migration to hide multiple data stores from the user allows migration to proceed without impacting on BI tool reports and applications that use data in the data warehouse. They would be unaware of the migration of one or more of the underlying data stores. Therefore, it accelerates migration by reducing the work that needs to be done to migrate to the cloud.

New Capabilities Enabled By Data Virtualisation Post Migration To The Cloud

Data virtualisation can also be used post migration to integrate data warehouse data with data and insights produced in other analytical data stores in the cloud or on-premises

Once an analytical system like a data warehouse has been migrated to the cloud, new capabilities become possible. For example, to bring together data from multiple cloud based analytical systems into a single integrated view. Simplifying access to data using data virtualisation can also speed up development in other analytical systems.

Logical Data Warehouse

The ability to integrate data warehouse data with big data at rest and in motion in a logical data warehouse produces richer, high value insights

It is often the case that companies want to integrate data warehouses with other analytical systems to bring together insights being produced from analysing data in each system. For example, companies may have a history of customer purchase activity in a migrated data warehouse but want to integrate this with customer related insights produced from analysing data in big data systems like Hadoop, cloud storage-based data lakes, graph databases and real-time streaming data. It could be that cloud storage and Apache Spark have been used by data scientists to analyse vast amounts of clickstream data to understand customer online behaviour. Data scientists may have also analysed customer sentiment posted on review websites and social networks about your products. That may have been done in Hadoop. Also, additional insights on customer relationships may be available from graph analysis done on a graph database. All of this represents customer related insights that could have been developed on different analytical systems on one or more clouds and in the data centre. However, what is missing is an integrated 360° view of customer insight. Data virtualisation can make this possible via a Logical Data Warehouse as shown in Figure 10.

Creating a logical data warehouse simplifies access to data across multiple analytical data stores by making it look like the data is all in one database

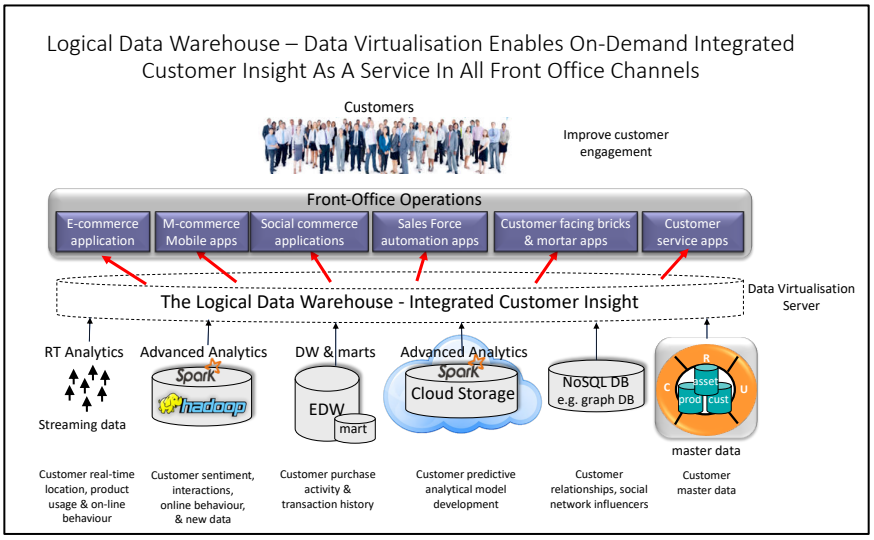


Figure 10

Virtual Provisioning of Trusted Data in A Data Marketplace

Data virtualisation can also be used to help modernise data architecture

In addition to a logical data warehouse many companies are also looking to modernise their data architecture to perhaps reduce data latency, change DBMSs, and to build trusted re-usable data assets. All of it can be achieved by decoupling BI tools and applications from physical data structures using data virtualisation. Also, by connecting to physical and virtual data sources across edge, multiple clouds and the data centre it becomes possible to build data analytics pipelines to produce these reusable data assets (See Figure 11).

It helps establish a logical data lake and virtually provision trusted data to multiple users and analytical systems to shorten time to value

Virtual data sources enable processing of data that is too big to move or that can't be moved for legal reasons

Provisioning data virtually avoids unnecessary data copying and better data governance

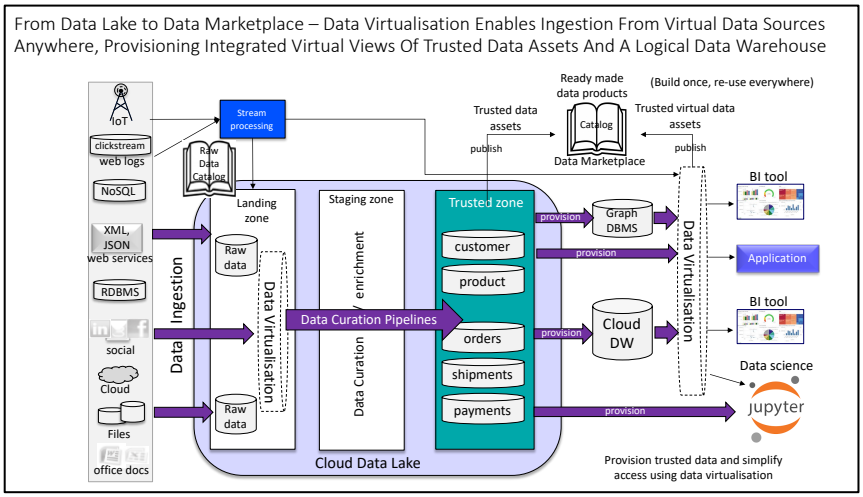


Figure 11

The idea here is not just to accelerate migration to the cloud but to accelerate complete digital transformation by building data assets once and publishing them in a data marketplace (a catalogue). This approach enables others to jump start their analytical project by shopping for ready-made trusted data assets irrespective of whether the data is needed for a data science project or in another analytical data store e.g. a graph database. It accelerates the whole data driven initiative and the whole digital transformation programme. Furthermore, data virtualisation enables the virtual provisioning of trusted data assets across the enterprise avoiding unnecessary copying of data and improving data governance. It also enables trusted virtual data assets to be created on top of trusted physical data assets and so helps incrementally build up more re-usable data. The result is faster and faster digital transformation through more reuse.

Data virtualisation accelerates digital transformation as well as migration to the cloud

ENABLING B2B INTEGRATION IN A HYBRID CLOUD COMPUTING ENVIRONMENT

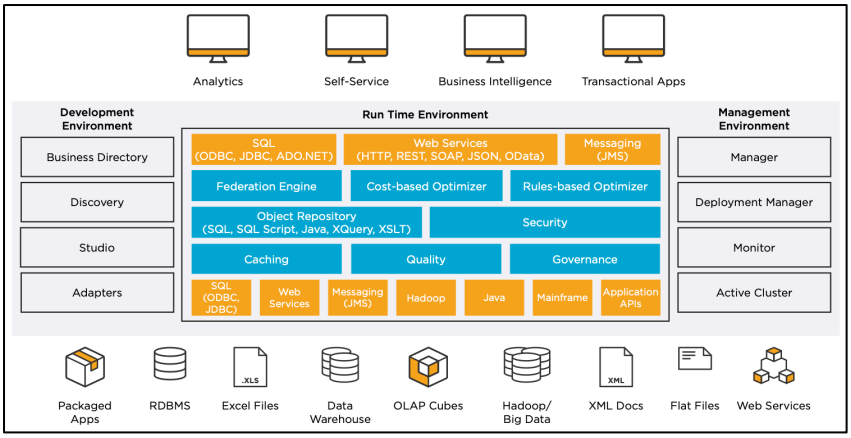
Data virtualisation in the cloud accelerates B2B digital transformation by enabling data to be integrated and shared across enterprises

Finally, data virtualisation can be used in the cloud to connect to data stores in multiple businesses to facilitate business-to-business data integration and data sharing. An obvious example of a beneficiary of this is all businesses participating in a supply chain or a partner network. It could also benefit a group of companies making it easy to share data across the group.

TIBCO DATA VIRTUALIZATION IN A HYBRID MULTI-CLOUD COMPUTING ENVIRONMENT

Now that we understand the data challenges companies are facing, the requirements that need to be met and how data virtualisation can accelerate workload migration to the cloud, we can look at how one data virtualisation vendor is stepping up to address these needs. That vendor is TIBCO. TIBCO is a provider of application integration, data integration, data management and analytics software products. Their data virtualisation product is TIBCO Data Virtualization (See Figure 12).

TIBCO Data Virtualization is run-time data integration software that includes a number of components



Source: TIBCO

Figure 12

It includes:

It can connect to a broad range of data sources both on premises and in one or more clouds

- Adapters to connect to a range of data sources including:
  - Relational DBMSs (on-premises or cloud)
  - NoSQL DBMSs (on-premises or cloud)
  - Hadoop (via Hive, Impala or HBase)
  - Files
  - Apache Spark
  - Web services (SOAP and REST)
  - Packaged applications e.g. SAP, Oracle EBS
  - SaaS Applications e.g. Salesforce.com, Oracle NetSuite
  - Social networks e.g. Facebook, Twitter, LinkedIn
- Caching to boost performance
- A cost-based optimiser
- Pushdown optimisation
- A massively parallel processing engine for scalable workloads
- A studio design tool
- Centralised metadata
- A self-service business directory (data catalogue)

Caching, massively parallel processing and cost-based pushdown optimisation allow it to support scalable query processing

TIBCO Data Virtualization can run in the cloud, on premises or both

The self-service business data catalog lets users search, categorise, and consume IT-curated data. TIBCO Data Virtualization can run on-premises, in a private cloud environment, or in a public cloud such as Amazon AWS, Google Cloud Platform, and Microsoft Azure. It is available on the AWS Marketplace and Azure Marketplace to simplify and accelerate deployment.

THE BENEFITS OF RUNNING TIBCO DATA VIRTUALIZATION IN A CLOUD COMPUTING ENVIRONMENT

There are a number of benefits to running TIBCO Data Virtualization in a cloud computing environment. These include the ability to:

Running TIBCO Data Virtualization in the cloud enables a shift towards operational expenditure and avoids the need for additional infratructure management

- Leverage cheaper cloud storage and better price/performance
- Get up and running more quickly
- Make use of a highly secure cloud computing environment
- Shift expenditure from a CAPEX to an OPEX pay-as-you-go model and to only pay for what is used
- Reduce infrastructure management and data centre costs
- Integrate data across edge, multiple clouds and the data centre quickly, at low cost with minimal disruption to existing systems
- Satisfy new requirements e.g. data architecture modernisation, migration of data sources to the cloud, IoT data sources, mobile access to BI reports using the TIBCO Data Virtualization software etc.
- Reduce risk of migrations to the cloud and changes to underlying data structures in cloud or on-premises data structures
- Improve collaboration, increase agility and reduce time to value
- To avoid upgrades of aging hardware
- To facilitate B2B data sharing with minimal disruption

ACCESSING DATA ACROSS SYSTEMS IN A MULTI-CLOUD COMPUTING ENVIRONMENT USING TIBCO DATA VIRTUALIZATION

Figure 13 shows how TIBCO Data Virtualization can be deployed in a cloud computing environment to access data across the edge, multiple clouds and the data centre.

TIBCO Data Virtualization can connect data data at the edge, in multiple clouds and in the data centre

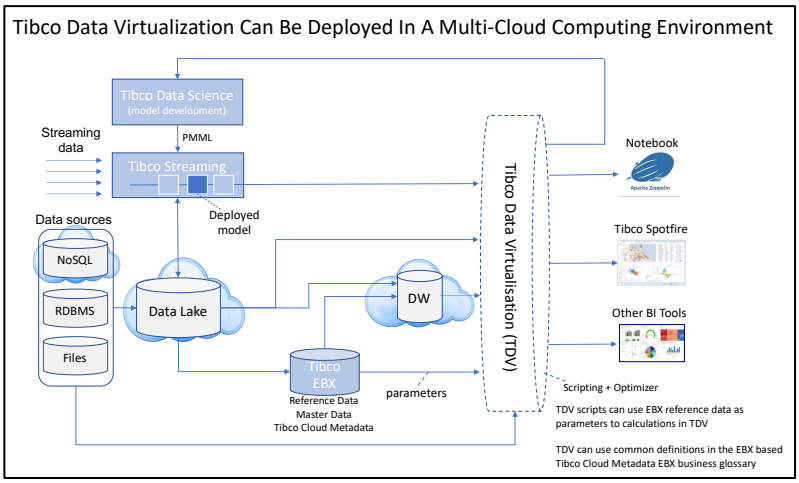


Figure 13

It simplifies data access for both business analysts using BI tools and data scientists

Migration of systems to the cloud is hidden from users which means reports and dashboards are not affected

TDV enables implementation of a logical data warehouse architecture

The figure shows that data scientists can use TIBCO Data Virtualization virtual views to access data needed to build a machine learning model using Apache Zeppelin notebooks and TIBCO Data Science. This data may be spread across multiple clouds and on-premises systems but that is hidden from the data scientist. Once the model is trained, tested and ready to be deployed, it can be deployed via PMML into TIBCO Streaming to analyse real-time data (e.g. IoT data) as it flows. This live streaming data can then be integrated with data at rest in a cloud data warehouse, master data in TIBCO EBX and live transaction data in cloud and on-premises operational system databases via 360° virtual views in TDV. Business analysts can then use TIBCO Spotfire

and other BI tools to access this data as if it was in a single database. Note that TDV can itself also run in the cloud.

THE POWER OF TIBCO DATA VIRTUALIZATION AND TIBCO CLOUD METADATA IN SIMPLIFYING DATA MANAGEMENT

Staying with Figure 13, there is integration between TDV and TIBCO EBX. TIBCO EBX offers a number of capabilities in helping to simplify data management. This includes:

- Multi-domain master data management (MDM)
- Enterprise reference data management (RDM)
- TIBCO Cloud Metadata[2]

TDV integrates with TIBCO Cloud Metadata running on TIBCO EBX

TIBCO EBX MDM is a multi-domain master data management system that manages the creation and maintenance of master data representing multiple master data entities such as product, customer, supplier, asset etc. It also synchronises this master data across both transaction and analytical systems running in the data centre and in one or more clouds. TIBCO EBX RDM offers a similar service for reference data such as code sets (e.g. product codes, currency codes, risk codes) to standardise codes across systems. TIBCO Cloud Metadata is centralised cloud-based metadata management system that can be used for data cataloguing and data governance. It utilises TIBCO EBX and runs on AWS or Azure. TIBCO Cloud Metadata offers a range of metadata services including:

- A data catalogue with auto data discovery, profiling and metadata extraction
- A metadata repository (store)
- Data governance
    - A business glossary
    - Governance classification, metadata lineage, user and role management, data access security
    - Data stewardship dashboards with data quality KPIs
- Search and collaboration including commenting and stewardship workflows

Common data names in the TIBCO Cloud Metadata business glossary can be used in TDV virtual views to hide the physical data names in underlying systems from the user

TDV integrates with TIBCO Cloud Metadata and can make use of common data names and definitions from the business glossary to ensure consistent data naming in all virtual views. This effectively means that TDV acts as a universal semantic layer for all BI tools to ensure common data names in all reports and dashboards accessing TDV. It also means that aggregate metric calculations can be removed from all BI tools and pushed down into TDV so that these types of calculations are done in one place and consistent across all tools. In addition, TIBCO EBX reference data can be used as parameters to calculations. There is also full lineage to understand where the data in TDV virtual views came from. Finally, TDV can integrate TIBCO EBX master data with any other data in TDV virtual views. This enables TDV to consistently provision trusted master data assets to any applications and tools. For example, master data in TIBCO EBX can be used to create common dimension data in all virtual data marts built using TDV.

This enables TDV to implement a unified semantic layer for all BI tools and so drive consistent data names and calculations across all reports and dashboards in all tools

[2] TIBCO Metadata Cloud is based on TIBCO EBX based

CONCLUSIONS

Migration of systems to the cloud is increasing

In this era of digital transformation, there is no question that the momentum behind moving to the cloud is unstoppable and if anything is increasing. Therefore, many companies are looking to migrate both operational and in particular, analytical systems to the cloud.

Companies want the ability to migrate more rapidly while also increasing agility and lowering total cost of ownership

However, companies need to overcome a number of challenges when migrating to the cloud that can slow down the migration process. These include the need to access data on-premises as well as on one or more clouds during and even after migration and the need to re-develop data integration and reports caused by changes to data structures that occur or when underlying technologies change during migration. These problems can occur when migrating both operational and analytical workloads.

Companies want to avoid re-development as much as possible during migration

Data virtualisation is a key technology that helps accelerate and de-risk migration to the cloud

The requirement is to be able to avoid some of these problems and accelerate migration to the cloud while also increasing agility and reducing total cost of ownership. What this paper shows is that data virtualisation is a key technology in helping to achieve this. Its core strength is to decouple applications and tools from data and provide a level of abstraction that has many benefits. This includes shielding applications, ETL jobs and user developed BI reports from any design and technology changes that occur during migration to the cloud. This both accelerates and de-risks the migration project.

It simplifies data access across a distributed data landscape and enables modernisation of data architecture

It can also be used to simplify and modernise data architecture before, during and after migration while reducing cost of ownership, increasing agility and reducing data latency. The use of data virtualisation to create virtual instead of physical data marts and to create a logical data warehouse are examples of this. Companies can also go further by using data virtualisation to connect to data in a distributed data landscape, build re-usable trusted data assets and use data virtualisation again to provision these data assets virtually to many different consumers and applications. This not only avoids the need for copying data, it lays the foundation for a data-driven enterprise by helping companies overcome the complexities of an ever-increasing distributed data landscape while simplifying access to data.

It lays the foundation for a data-driven enterprise and accelerates digital transformation

TIBCO Data Virtualization together with other TIBCO technologies goes a long way to helping companies make this happen as part of implementing a modern data strategy.

About Intelligent Business Strategies

Intelligent Business Strategies is an independent research, education, and consulting company whose goal is to help companies understand and exploit new developments in business intelligence, machine learning, advanced analytics, data management, big data, and enterprise business integration. Together, these technologies help an organization become an *intelligent business*.

Author

Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant, he specialises in business intelligence and enterprise business integration. With over 39 years of IT experience, Mike has consulted for dozens of companies on data strategy, data architecture, big data, machine learning, advanced analytics, data governance, master data management, and enterprise architecture. He has spoken at events all over the world and written numerous articles. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata on the Teradata DBMS and European Managing Director of Database Associates, an independent analyst organization. He teaches popular master classes in Data Warehouse Modernization, Designing, Managing and Operating a Multi-Purpose Data Lake, Machine Learning and Advanced Analytics, Big Data and Analytics Fundamentals, Enterprise Data Governance and Master Data Management, Modern Data Architecture and Real-Time Analytics and Operational BI.

Telephone: (+44)1625 520700
Internet URL: www.intelligentbusiness.biz
E-mail: info@intelligentbusiness.biz