WHITE PAPER

Data Warehouse Automation and Real-time Data - Reducing Time to Value in a Distributed Analytical Environment

By Mike Ferguson
Intelligent Business Strategies
December 2016

Prepared for: ATTUNITY

Table of Contents

TRADITIONAL ANALYTICAL ENVIRONMENTS

For over 20 years the traditional data warehouse has been the approach taken for integrating and analysing data to produce business intelligence

In the last 20+ years we have been analysing structured data using a traditional data warehouse pattern. This involves extracting data from multiple operational systems, cleaning and integrating it using data integration tools and populating a data warehouse (DW). Subsets of that data are then moved into data marts for analysis and reporting using business intelligence (BI) tools to produce insights. This classic data warehouse architecture is shown in Figure 1.

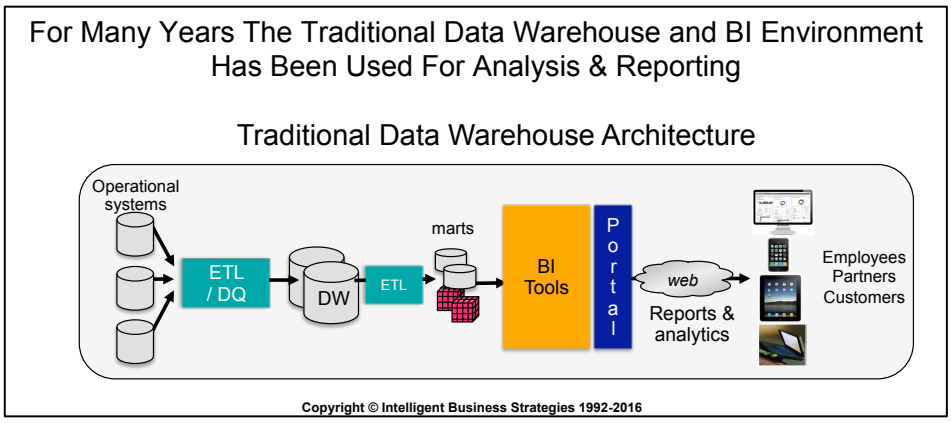

Figure 1

Traditional data warehouses have become 'production' systems that are heavily controlled and slow to change

Today, these systems are 'nailed down'. They are heavily controlled *production* systems. That means they are often subject to formal IT change management procedures. Therefore, if changes are requested to a data model(e.g. to accommodate a new data source), it can be a slow process of dependency checking to understand what might be impacted by extending the capabilities of the system.

New techniques and technologies have made some inroads into reducing data warehouse development times over recent years

Over the last 5+ years, this type of process has frustrated business users who want to move forward in an agile way to deliver new competitive insights. Yes, we have had some advances to speed up delivery. For example, Data Vault as a modelling technique has made it much quicker to change data warehouse data models. Many companies now use this approach. In addition, data virtualisation has been adopted to build virtual data marts instead of physical ones and self-service BI is now the norm for business users building their own reports and dashboards as opposed to IT developing everything. Nevertheless, the speed at which disruption is occurring from new competitors continues to quicken. Therefore, business continues to demand greater agility, shorter delivery times, more data and more insight to remain competitive, which in turn is driving demand for big data and analytics.

However business is demanding more agility and more insight to remain competitive

Data warehouse automation may hold the key to governed productivity gains as complexity grows

This paper looks at this problem and how the adoption of automation tools in traditional data warehouse environments, can help companies break free from slow waterfall development approaches to get ahead of business requirements rather than always be behind them. We also explore how companies can remain productive across traditional and lower cost, scalable big data analytical platforms while also dealing with the increasing complexity of a more distributed data landscape.

THE THIRST FOR NEW DATA, AGILITY AND NEW ANALYTICAL WORKLOADS

Most data warehouses and data marts are still restricted to analysing master data and transaction activity to understand business performance

Figure 1 shows the classic analytical data warehouse set-up that many companies have in place today. In this environment, data warehouse data is typically taken from online transaction processing (OLTP) system data sources such as Customer Relationship Management (CRM) systems, Enterprise Resource Planning (ERP) systems, and Supply Chain Management (SCM) systems. While the number of data warehouses and data sources feeding data warehouses has increased over the years and some OLTP systems are now wholly in the cloud (e.g. Salesforce, Workday etc.), the vast majority of traditional analytical systems are still restricted to analysing structured master data[1] and transaction data[2] to understand business performance and customer behaviour in terms of transactional activity.

More data is needed to gain a deeper understanding about customers

However, if there is no transactional activity, then there is very little to help distinguish one customer from the next and that presents a problem in a world where customer retention is now just as important as growing market share. Therefore, it's not surprising that the thirst for more data and new insights about customers has become a top priority. New data sources from inside and outside the enterprise that are in demand include:

This includes data on online behaviour, opinions as well as machine data, sensor data and external data from government and vertical industry data sources

- Online behaviour data such as clickstream (every click a visitor makes when navigating your website) from desktop and/or mobile devices
- Machine data from log files beyond just web logs that provide clickstream e.g. IVR system logs, application server logs, database logs, network router logs etc.
- Opinion data from blogs, social networks, review websites and news feeds
- Sensor data from new instrumentation in business operations e.g. on production lines and supply chains
- Sensor data from smart products (Internet of Things (IoT)) that record usage and / or location (e.g. GPS sensor in mobile phones)
- Other open and vertical industry specific information becoming available on the internet e.g. Open Government data

New data sources differ from transaction data because they may vary in data type, be large in volume and be generated at very high rates

The difference between this new data and data from traditional data sources is that many of these data sources are huge in data volume and generated at very high rates. Also, although many new sources are still structured in nature (e.g. web logs, sensor data), some are semi-structured (e.g. JSON, XML) and some are unstructured (e.g. Twitter text).

[1] E.g. Customer, Product, Employee, Asset, Supplier, Material, Site

[2] E.g. Orders, Shipments, Payments, Adjustments

BEYOND THE DATA WAREHOUSE - NEW PLATFORMS IN AN EXTENDED ANALYTICAL ECOSYSTEM

New big data sources and new types of analytical workloads have led to the adoption of new data stores and analytical platforms beyond the data warehouse

The emergence of these new big data sources has resulted in a corresponding emergence of new analytical workloads not well suited to traditional data warehouses. These new workloads have caused many companies to extend their analytical environments beyond the data warehouse to include new kinds of analytical platforms (see Figure 2).

While the data warehouse is still needed, new data characteristics (variety, velocity and volume) and new types of analysis to analyse this data, has brought new platforms into the enterprise. These include cloud storage (not shown in Figure 2) such as Amazon S3, OpenStack Swift or Microsoft Azure storage. It also includes NoSQL data stores like Cassandra, HBase, MongoDB and Basho Riak. Then there is Hadoop with its Distributed File System (HDFS) and analytical relational DBMSs. Furthermore, some companies are also adding streaming analytical platforms to analyse streaming high velocity data before it is stored anywhere.

Multiple platforms now exist in the enterprise to support different analytical workloads

As a result data integration and data movement has increased rapidly across data stores in this new analytical ecosystem

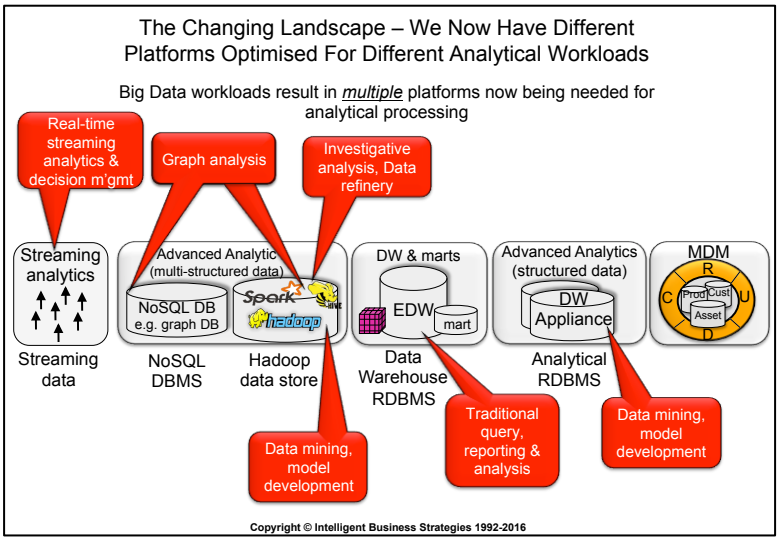

Figure 2

Therefore the analytical set-up in many companies today, has grown far beyond the data warehouse into more of an '*analytical ecosystem*' consisting of multiple data stores and platforms. Data is now in multiple platforms both on-premises and in the cloud. This has resulted not only in a rapid increase in the amount of data ingestion, but also in the amount of data movement across the modern analytical ecosystem. Figure 3 shows some of the popular data integration paths that have emerged. Again, cloud storage is missing but could also have been incorporated into the diagram.

Complexity has increased significantly in analytical environments

The point however, is that complexity has increased significantly. More data, more data stores, more data types, more data sources, hybrid computing across on-premises and cloud based systems, relational DBMSs with SQL as a standard data access API, NoSQL data stores with proprietary access APIs. Data is in multiple places, not just one. We are drowning in data.

Despite the increase in the number of data sources and the increased complexity of the analytical ecosystem, business is demanding greater agility

Yet business is demanding event more aglity

and responsiveness in order to gain the timely analytic insights necessary to compete in a world where disruption is coming from all directions.

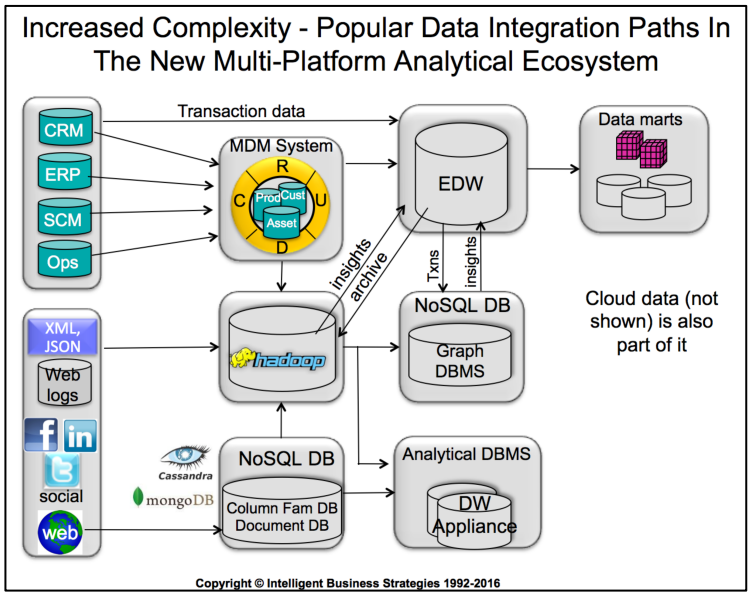

Figure 3

Structured, semi-structured and unstructured data now need to be integrated

Also, the number of data flows have increased dramatically between sources and systems and between systems

NEW CHALLENGES – A MULTI-PLATFORM ANALYTICAL ECOSYSTEM AND A DISTRIBUTED DATA LAKE

Opposing forces – business wants more agility and to produce insight faster while complexity increases around them

The dilemma is that the business needs to increase agility, reduce time to value and simplify in an environment where the data and analytical landscape is getting increasingly more complex. However, greater complexity requires more resources. Add to that the fact that we are dealing with potentially tens or even hundreds of new data sources and you can quickly see why business now sees IT as a bottleneck. The business viewpoint is that it takes too long to implement changes in 'production' data warehouses. So, what are IT's chances of being able to deliver on time with all these new data sources and analytical platforms? Business users want to help.

Business now sees IT as a bottleneck and wants more self-service capability

IT recognises that their productivity needs to improve and are worried about governance in a self-service environment

Equally from IT's perspective they too recognise the problem. They need to be able to introduce agility into production data warehouses so that they can rapidly make changes in a governed manner in order to free up resources to work with data scientists and business analysts on new data sources and analytical systems. Some IT industry analysts are suggesting that a two-speed 'bi-modal' analytical environment is needed consisting of slow-to-change, carefully controlled production data warehouses alongside agile investigative big data environments, I don't agree. I think agility is needed *both* in traditional data warehouses *and* in new analytical environments. Without this, IT won't be able to keep pace with business need, leaving business self-service data cleansing and data integration[3] to potentially degenerate into an unmanaged 'wild west' environment. It is precisely this situation that companies need to avoid.

Agility is now needed both in traditional data warehouse and in new big data analytical environments

[3] Self-service data cleansing and data integration is also known as Data Preparation or Data Wrangling.

REQUIREMENTS – THE NEED FOR GOVERNANCE, AGILITY AND PRODUCTIVITY IN A MORE COMPLEX ANALYTICAL ENVIRONMENT

Given this situation, what are the requirements that need to be met to help provide governance whilst also providing agility and productivity in analytical environments? Productivity improvements are needed both in the development and change of traditional data warehouses as well as in new big data analytical systems.

Let's consider the complexity of the emerging analytical environment (see Figure 4) to understand what is needed.

Companies now have to manage a distributed data landscape consisting of multiple data stores, both on-premises and on the cloud, as well as multiple analytical systems

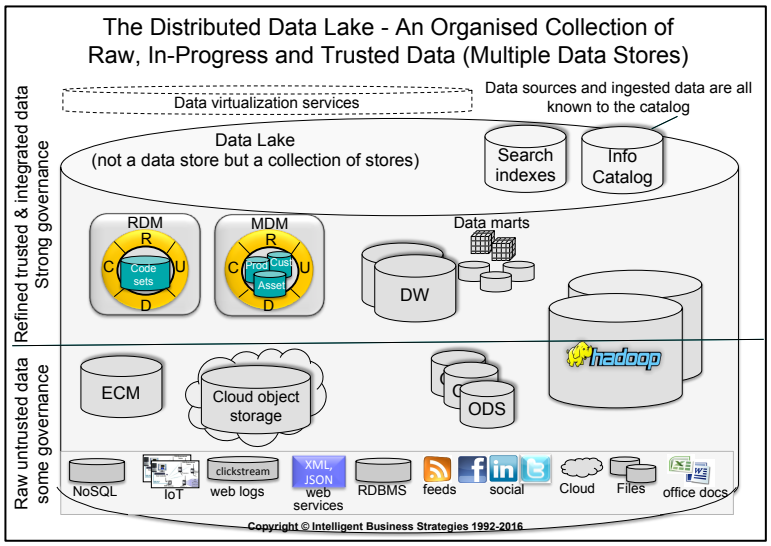

Figure 4

A new set of requirements need to be met to remain agile and productive in this more complex set-up

Governance across multiple types of analytical system is also needed

Looking at Figure 4, many companies are now dealing with a distributed data landscape consisting of multiple data stores including multiple operational data stores (ODS), multiple data warehouses and data marts, master data management systems, reference data management systems, one or more big data Hadoop systems, cloud storage, NoSQL data stores and an explosion of different internal and external data sources. This is much more complex than the set-up shown in Figure 1 and therefore warrants a new set of requirements if companies are to remain agile and productive while still governing multiple types of analytical system as shown in Figure 2. Key requirements that need to be supported in this more complex set-up are as follows:

It should be possible to:

We need to automate the development and maintenance of data warehouses to easily accommodate change and free up IT resources

- Automate the building and maintenance of data warehouses and data marts to accelerate development and change management while also maintaining governance over traditional analytical environments and reducing risk. In this way, IT resources can be freed up to work alongside business on new data and new analytical workloads running on new analytical platforms

Collaborative and guided development is needed across the analytical ecosystem

- Enable web based projects with collaborative design and guided development of jobs needed to build and maintain different types of analytical systems both on-premises and in the cloud. These include traditional data warehouses, data marts, virtual views on a data virtualisation server and big data sandboxes used by data scientists. Collaborative development enables virtual communities of IT and business users to work together to create systems that deliver competitive insights and value

We need to capture best practices and enable governance by creating templates

- Leverage pre-built customisable templates to accelerate the design, creation and maintenance of analytical systems whether they be on-premises or in the cloud
- Create repeatable models and custom templates to capture best practices, build-in governance, facilitate re-use and accelerate development

Automation of data profiling and mapping can help accelerate development

- Import data warehouse and data mart data models into automation software to accelerate development
- Automate the mapping of data sources to data models in traditional data warehouse environments to accelerate development
- Automate the profiling of data sources to accelerate detection of anomalies in data sources so that corrective action can be taken to improve data quality

Automation should also make it possible to generate data cleansing and integration jobs

- Use the metadata created during collaborative and guided design and development to automatically generate data processing[4] jobs to clean, transform and integrate data to build the aforementioned systems. This is known as model-driven development. These jobs should automatically exploit the scalability of underlying platforms (e.g. a Hadoop, Apache Spark or a massively parallel relational DBMS) to maximise performance. This means that software needs to be 'platform aware' to enable the same data processing job to run in different underlying environments.
- Distribute the execution of a job if needs be to optimise performance such that parts of a job execute on different underlying platforms

Sharing metadata between business and IT enables collaborative development and reuse

- Share metadata between business and IT users working together in a collaborative development environment to maximise the ability to reuse rather than re-invent

Continuous change data capture (CDC) enables low latency data streaming to replace batch ETL processing

- Replicate data changes in any data source into a common messaging platform (e.g. Apache Kafka) to move from batch data extract toward continuous streaming of data as it changes. This reduces data latency enabling analysis to happen closer to real-time and actions to be taken more rapidly to avoid risks and exploit opportunities
- Maintain a history of data in a data warehouse including dimension data

[4] Cleansing, transformation, integration

It should also be possible to automate the building of MDM systems and offloading of data warehouse ETL processing

- Monitor execution and provide metadata lineage to facilitate confidence, trust and governance in data
- Automate the building and maintenance of centralised master data management (MDM) systems to speed up the creation and synchronisation of consistent master data across the enterprise
- Automate the offloading of data warehouse ELT processing to Hadoop to reduce the total cost of ownership of staging and processing data
- Have the software automatically select the best underlying technology to ingest data for subsequent processing in building of data warehouses and big data sandboxes

Governance rules also need to be applied at runtime

- Apply governance rules on any platform at runtime to guarantee that data is correctly governed no matter where it is in the distributed data landscape or where it goes in the distributed data landscape

STAYING IN CONTROL – LEVERAGING AUTOMATION AND CENTRALISED METADATA TO BUILD ANALYTICAL SYSTEMS AND SANDBOXES

Governance, agility and acceleration of data warehouse development is made possible using automation software

Looking at these requirements, it's clear that if companies want to free up IT resources from traditional data warehouse (DW) development to work on innovative new analytical workloads on big data platforms then governance, agility and acceleration of DW development need to come from somewhere. That somewhere is likely to be data warehouse automation (DWA) software, which is now expanding beyond data warehousing into big data environments. Given that we are some twenty-five years into DW development, it's not surprising that this kind of software is now available to capture best practices, guide development, apply governance and introduce agility into traditional DW environments. The beauty about automation software is that it:

Automation software can also support multiple data warehouse development projects simultaneously

- Provides a common, governed approach to design, development, maintenance and testing
- Manages incremental change to data sources, data models and ETL jobs
- Provides centralised, documented metadata
- Can manage the development of multiple projects at the same time

All of these are needed to improve productivity without loss of governance.

Figure 5 shows how far we could potentially go in expanding the use of automation software to introduce agility, governance and accelerated development across today's more complex analytical and data landscape.

Automate the building of data warehouses, data marts, MDM systems, virtual views, plus automate data prepartion for analysis in big data sandboxes.

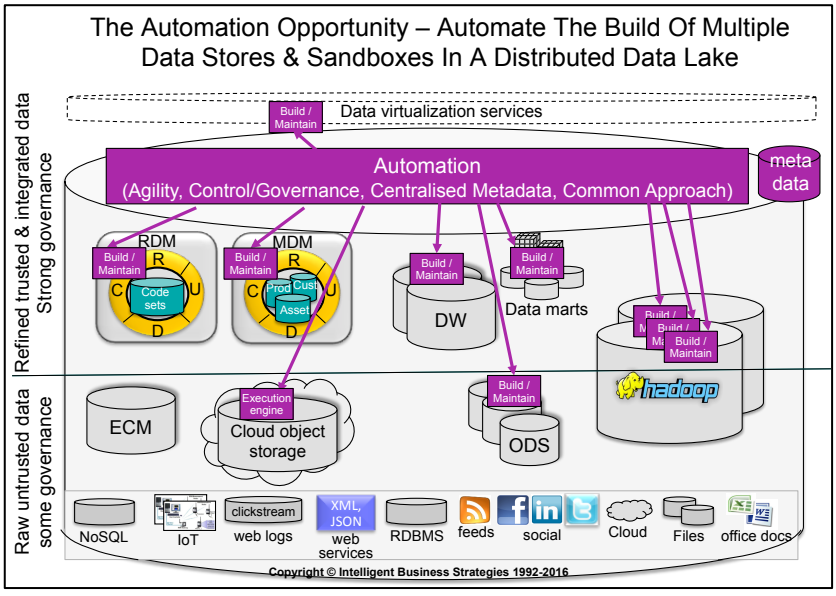

Figure 5

It is potentially possible to extend the use of automation software beyond data warehouses and data marts to stretch across the entire analytical ecosystem and beyond

It's potentially possible to grow the use of automation software to rapidly:

- Build data warehouses and data marts
- Build Master Data Management (MDM) systems
- Create and maintain virtual views of data across multiple data stores in a data virtualisation server
- Offload DW ELT processing to Hadoop
- Prepare data in big data sandboxes for analysis on Hadoop or Spark (See Figure 6)
- Ingest data from operational systems and data sources using scalable change data capture and any utilities deemed best suited to the job
- Ingest data by interfacing with continuously streaming data platforms

This could be done all off common metadata both on premises and in the cloud

Automation can control data processing and movement across systems

For example, generated scalable ETL jobs could potentially be run on Hadoop or Spark

Newly produced trusted data and insights could then be automatically moved into DWs and MDM systems to add to what is already known

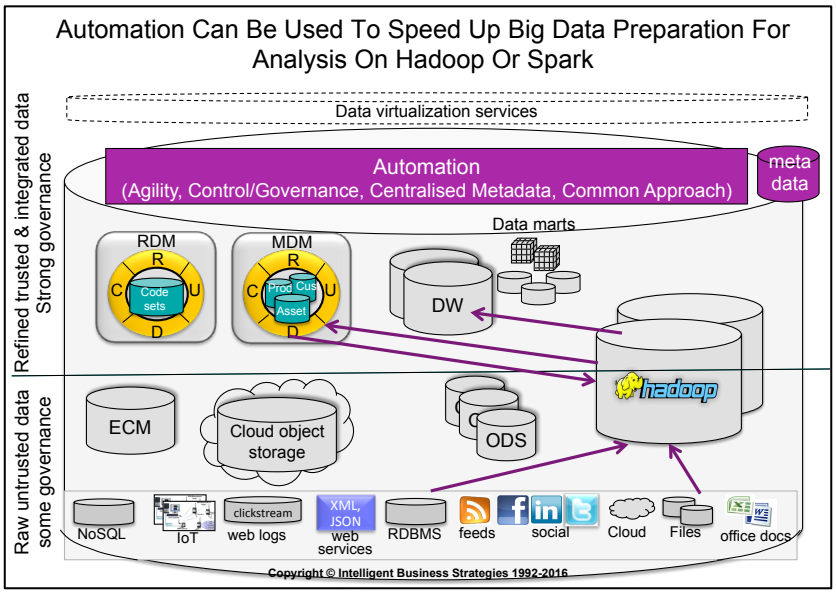

Figure 6

A common metadata repository makes it possible to share reusable metadata across multiple projects where automation is being used

The potential is there for data warehouse automation software to be used as a common approach to accelerating the production of valuable data in different data stores and analytical systems across a distributed data landscape. Ultimately, it could become software that acts as a 'production line' to control multiple development projects to rapidly build, maintain and govern core data and analytical systems across the enterprise. The reason that is possible is because common reusable metadata is available to multiple projects in a shared metadata repository.

Realistically, innovation may first need to take place using manual development but once new data of value has been identified, automation software could be used to automate the process of producing new high value data and insights for provision into MDM systems, data warehouses or big data sandboxes. This would allow valuable business and IT resources to move on to new exploratory analytical workloads in response to new business requirements while automation takes over the responsibility for governing, maintaining what has already been created.

ENABLING AGILITY AND REDUCING TIME TO VALUE USING ATTUNITY COMPOSE

Having understood the transition from traditional data warehouses to the analytical ecosystem, the increasing complexity of the data landscape and the valuable role that automation software could have in improving agility while retaining control in this environment, this section looks at how one vendor's agile data warehouse automation software can potentially help companies achieve this. That vendor is Attunity.

Attunity is a data management vendor offering data management and automation software

Attunity was founded in 1988. They currently operate globally and have more than 2000 customers including many of the Fortune 100. Attunity has a number of products in the area of data integration and management. The two we will discuss in this paper are Attunity Replicate and Attunity Compose.

ATTUNITY REPLICATE

Attunity Replicate captures changes from source system data stores and automates their transfer to target environments to keep systems in sync and for analysis

Attunity Replicate is web-based software that can replicate, synchronise, and distribute data across multiple different data stores. Users can replicate data in bulk, or real-time with Change Data Capture (CDC) technology. Replication can be applied to each of multiple sources to bring together data into a single data store for subsequent transformation, integration and analysis. Attunity Replicate automates the transfer of data changes into various target environments for data synchronisation and analysis. This includes all major relational databases, data warehouse analytical relational databases, Hadoop, cloud and popular NoSQL data stores. It also integrates with the Apache Kafka messaging system, which is rapidly becoming a popular mechanism for ingesting streaming data from hundreds of applications into Hadoop and for feeding streaming data into real-time analytic applications that analyse the data before it is stored.

It reads from source system log files to avoid impacting on source system transaction performance

Attunity Replicate CDC identifies real-time changes by scanning source-system log files and therefore doesn't interfere with source system operational performance. Sources supported include:

Most widely used data sources and targets are supported including RDBMSs, cloud applications and Hadoop

- Legacy NoSQL databases and file systems (e.g. IBM IMS and VSAM)
- Popular relational DBMSs and data warehouses – both on-premises and in the cloud
- Cloud based applications such as Salesforce
- Several Hadoop distributions, e.g. Cloudera, Hortonworks and MapR

It uses compression, in-memory processing and parallel data transfer to boost performance

In addition, most of the data sources supported can also be targets and replication can be managed to and from the cloud and between clouds. To accelerate performance, Attunity Replicate also makes use of data compression, in-memory processing, parallel data transfer and configurable batch sizes for change data capture processing. Also data can be filtered from data sources and fed into multiple targets simultaneously thereby making it possible to land the changes in a data warehouse or in Hadoop (see later) while at the same time, feeding the data into real-time analytical applications to enable companies to act on business conditions signalled by real-time events. Simultaneous replication in batch and in real-time means that Attunity

Replicate supports the increasingly popular Lambda architecture that is being adopted by many enterprise organisations

ATTUNITY COMPOSE

Attunity Compose is a data warehouse automation product that runs on-premises and in the cloud

The second product is one of Attunity's latest offerings - Attunity Compose.

Attunity Compose provides a metadata-based platform for data warehouse automation. It is built for data architects rather than developers, that automates the manual, repetitive aspects of data warehouse design, development, testing, deployment, operations, impact analysis, and change management. Its capabilities span development processes, requirements gathering, deployment and maintenance.

It can be used to build DWs and data marts and integrates with many data warehouse platforms such as Teradata and Oracle Exadata

Attunity Compose provides a guided step-by-step approach to designing, building and maintaining data warehouses and data marts and can run on-premises and/or in the cloud. It automatically generates the data warehouse structures, ETL jobs and documentation needed to implement projects while tracking data lineage and ensuring integrity. Attunity Compose automates routine IT tasks associated with designing, creating, populating, and managing data warehouses and data marts (DMs). It also accelerates tasks that cannot be completely automated. This allows IT teams to rapidly improve productivity. Once projects are approved, it's possible to deliver completed data warehouses, data marts, and BI environments in much less time than if data warehouse development was entirely manual. Enterprises gain productivity efficiencies, a more agile development approach, best practice automation, reduced risk, and can also lower the cost of change.

In addition, Attunity Replicate integrated with Attunity Compose to ingest data in bulk or real-time into data warehouses that Compose builds.

Data architects can apply best practices by using proven methodologies and design patterns such as Inmon, Kimball or Data Vault. Attunity Compose supports:

It supports data modelling, automated source data profiling and automatic generation of ETL jobs

Real-time source change data capture is also supported

- The creation, import[5] or reverse engineering of data models
- Derived data attributes in data models
- An business glossary for all the attributes in the data warehouse
- Source data profiling
- Automated mapping of source system databases to the data model
- An expression builder to define transformation rules
- A workflow designer and scheduler
- Automated ETL generation and support for staging tables, slowly changing dimensions, change tables and lookups
- Real-time source change data capture (via Attunity Replicate) and data integration
- Metadata lineage and impact analysis
- Generation of data marts (including support for aggregate fact data)

[5] E.g. from CA ErWin

Metadata lineage provides governance by allowing users to track data back to where it came from

- Automated generation of project documentation
- Physical management and monitoring of data warehouse operations
- Managed migration between development, test and production systems making it easy and less risky to promote to production

Auditing is also included

Also changes to data sources, models, mappings and business rules can be managed and audited to govern a traditional data warehouse environment as depicted in Figure 1. Modern analytical relational databases are supported as target data warehouse data stores.

Development and maintenance can be automated both on-premises and in the cloud

Attunity Compose therefore speeds up traditional data warehouse development and maintenance enabling skilled IT professionals to be freed up to work alongside business colleagues on new big data analytical projects and platforms.

Leveraging Attunity In New Big Data And Traditional Analytical Environments

Attunity Replicate and Compose also integrate with big data environments

However, much like the requirements we discussed earlier in the paper, Attunity is taking its replication and automation software beyond traditional data warehousing by integrating it with Big Data platforms and technologies as well as traditional data warehouses and data marts. They are doing this by integrating Attunity Replicate with Apache Kafka and Hadoop

Attunity Replicate Integration with Apache Kafka and Hadoop

Attunity Replicate integration with Apache Kafka and Hadoop allows changes to be capturred and streamed at scale into big data analytical applications and data stores

The integration of Attunity Replicate with Apache Kafka and Hadoop is getting a lot of attention. The reason is because it makes it possible to turn changes to legacy and modern OLTP system data stores into low latency data streams that can be analysed as soon as the changes occur.

Attunity Replicate change data capture allows streaming data changes to be processed by Hortoworks DataFlow before they are landed in Hadoop

To make this possible Attunity Replicate treats Kafka just like another endpoint. Attunity Replicate change data capture can then stream changes from any supported source as they happen into Kafka where they can be picked up by data-in-motion processing technologies such as Hortonworks DataFlow (HDF). From here, a product like HDF could, for example, then filter and pre-process the streaming data changes, integrate this data with other data streams from Hadoop, other application data stores or IoT devices and then do any or all of the following:

- Land the streaming data changes in Hadoop or other NoSQL data stores for further analysis.
- Convert the data into another file format like optimised record columnar (ORC) file format before landing it in Hadoop so that the data is immediately ready for SQL query processing via BI tools interfacing to Hadoop HDFS data via Hive
- Pass the changes into Apache SOLR to index the data for use in real-time search

Attunity Replicate changes can be made available for SQL access on Hadoop and also indexed to search near real-time data

This kind of combined real-time processing capability makes it possible to act on live streaming data as it changes even in legacy systems without interfering with transaction path lengths.

Attunity Compose enables collaborative, accelerated development of data warehouses and data marts

Besides automation that quickens development, test and deployment, Attunity Compose's graphical interface hides the complexity of creating and populating data warehouses and data marts. Also because little of no actual coding is required, it makes it easier for the business and IT implementation teams to communicate.

CONCLUSIONS

Companies need to find a way to manage complexity while increasing agility and productivity without loss of governance

Manual building and maintenance of data warehouses is too slow a process

Data complexity is also slowing us down andself-service development across the extended analytical environment could become chaotic if not well managed

Now is a good time to introduce automation software to accelerate and govern traditional DW and data mart development

Replication can be used to lower the latency of data to enable analysis and decision making closer to real-time

The use of automation software can grow beyond the data warehouse to automate the build of other systems and services

Attunity already automates DW development and is expanding into big data environments

Looking at how the data landscape and analytical ecosystems are evolving, there is no question that companies are going to have to find a way to conquer complexity, increase agility and fuel productivity in a governed way if they are to keep delivering value in a timely fashion

Manual maintenance to try to control changes to traditional data warehouses is too slow a process. It also locks in valuable IT resources to mundane maintenance rather than innovation. Equally, the growing complexity of data also slows us down. New analytical environments, data science and the demand for self-service data preparation in big data environments may provide agility, but development could also degenerate into chaos if it is not well managed.

The objective, therefore, is to fuel innovation in big data environments while using replication and automation to:

- Reduce the latency of data being made available for analysis to enable business to be more responsive by taking actions in near real-time
- Govern and accelerate the build of traditional data warehouses and data marts while facilitating agile change management
- Accelerate data preparation in big data environments to reduce the time it takes to produce new insights
- Formalise and automate completed analytical projects on big data environments so that data scientists can move on to new projects while automation takes over the role of delivering value from new data
- Govern the flow of data between ‘data science lab’ big data platforms and traditional data warehouses and accelerate our ability to enrich what we already know
- Go beyond data warehousing and big data to automate the build of MDM systems and virtual data views to speed up delivery of trusted information services

Looking at this list, it is clear that Attunity has already tackled data warehouse automation and is now extending the use of their replication and automation software to introduce low latency data, agility and productivity into the larger analytical ecosystem. They are starting to accelerate development in big data environments while also driving governed integration between analytical data stores. There is no doubt that this is needed.

But we are not done yet. The automation opportunity depicted in Figure 5 is much bigger than this. If companies want to continue to reduce time to value across analytical systems in a governed manner while data complexity grows around them then they would surely need to take a closer look at what Attunity has to offer.

About Intelligent Business Strategies

Intelligent Business Strategies is a research and consulting company whose goal is to help companies understand and exploit new developments in business intelligence, analytical processing, data management and enterprise business integration. Together, these technologies help an organisation become an *intelligent business*.

Author

Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an independent IT industry analyst and consultant he specialises in Big Data, BI/Analytics, Data Management and enterprise business integration. With over 35 years of IT experience, Mike has consulted for dozens of companies on BI/Analytics, big data, data governance, master data management and enterprise architecture. He has spoken at events all over the world and written numerous articles and blogs providing insights on the industry. 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 IT industry analyst organisation. He teaches popular master classes in Big Data Analytics, New Technologies for Business Intelligence and Data Warehousing, Data Virtualisation Enterprise Data Governance, Master Data Management, and Enterprise Business Integration.

Water Lane, Wilmslow
Cheshire, SK9 5BG
England
Telephone: (+44)1625 520700
Internet URL: www.intelligentbusiness.biz
E-mail: info@intelligentbusiness.biz

Automation – Reducing Time To Value In A Distributed Analytical Environment