



Data Migration and Access in a Cloud Computing Environment

By Mike Ferguson
Intelligent Business Strategies
March 2014

Prepared for:



Table of Contents

| | |
|--|----|
| Introduction | 3 |
| The Growing Popularity of SaaS Applications | 3 |
| Changing Landscapes - The New Hybrid Computing Environment | 4 |
| Data Requirements In A Hybrid Computing Environment | 6 |
| On-Premises to Cloud | 7 |
| Cloud to On-Premises | 7 |
| Hybrid | 8 |
| Managing Cloud Data Connectivity And Access Using Progress DataDirect Cloud | 9 |
| What is Progress DataDirect Cloud? | 9 |
| Simplification of Cloud Data Access | 11 |
| Cloud Data Integration And Migration..... | 11 |
| Progress DataDirect Cloud Support For Data Security..... | 11 |
| Progress EasyI | 12 |
| Conclusion | 13 |

INTRODUCTION

The compelling attraction of cloud based computing is growing

For many organisations, the use of cloud computing has got to the point where it is becoming a very attractive deployment option. There are a number of reasons for this including:

- Faster time to value
- Lower entry and maintenance costs via subscription based services to immediately start using a service, reduced operations cost and less software license cost
- Sustainability resulting from better hardware utilization and so lower energy costs
- Centralised and reduced system administration
- Flexible and scalable processing power on-demand
- Rapid configuration deployment and extension
- Reduced risk
- Faster change management and release management

Faster time to value, reduced cost and scalability on-demand are all driving cloud adoption

Cloud expenditure is expected to surge in 2014

We are also at the point where the cloud computing market has reached a level of maturity where confidence is growing. This is reflected in predictions by one analyst firm¹ that cloud spending, including cloud services and the technology to enable these services, will surge by 25% in 2014, reaching over \$100 Billion. The \$100 Billion figure includes software, services and cloud infrastructure. It is also expected that there will be a dramatic increase in the number of datacenters as cloud players race to achieve global scale.

THE GROWING POPULARITY OF SAAS APPLICATIONS

SaaS transaction processing applications are becoming increasingly popular

One of the biggest attractions over the last few years has not just been infrastructure as a service but the emergence of popular customizable software-as-a-service (SaaS) transaction processing applications such as Salesforce.com, Microsoft Dynamics, Workday, NetSuite and Eloqua.

SaaS BI is also growing

In addition, Business Intelligence SaaS providers such as Birst, GoodData, Indicee, and PivotLink have all seen gradual growth with many mainstream BI vendors also now entering the market to boost.

All of this growth means that migration to the cloud from on-premises systems is starting to gather pace with demand to connect to more and more cloud based systems to migrate and acquire data on the increase. In this white paper we look at the challenge that companies are facing when migrating to the cloud and how one vendor, Progress Software is stepping up to the challenge of simplifying access to data to make migration and cloud data access much easier to deal with.

¹ IDC Predictions 2014

CHANGING LANDSCAPES - THE NEW HYBRID COMPUTING ENVIRONMENT

The adoption of SaaS based applications has meant that many companies now have a hybrid computing set-up

In the last few years, the growth of SaaS based applications on the public cloud has meant that many companies have already taken the decision to move some of their applications to the cloud to support specific business functions. The result is that they are immediately pushed into a situation where they are operating and managing a hybrid-computing environment with some of their corporate systems deployed on-premises and some of them deployed on the cloud. That means their data is both inside and outside the firewall.

There has been an increase in data migration to the cloud

The move to a hybrid-computing mode of operating has implications for data in terms of access, migration, integration and governance. From a business operations perspective many companies making the move to deploy a SaaS applications have to consider data migration as an initial exercise but once data migration is over, they are then in a position where business process execution may have to 'straddle' the corporate firewall with process execution starting in a SaaS based application in the cloud and finishing on-premises, starting on-premises and finishing in the cloud or bouncing back and forth between on-premises and SaaS based applications in a more complex end-to-end process. This is shown in Figure 1.

Business process execution now straddles the corporate firewall

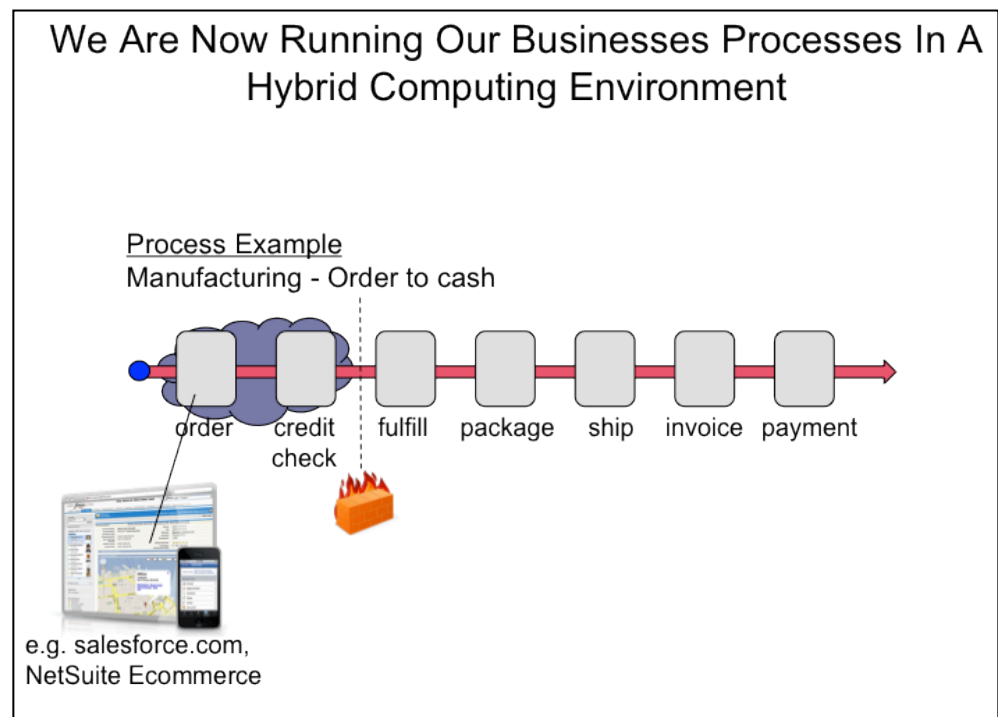


Figure 1

Similarly analytical systems can also be deployed in a hybrid-computing environment as shown in Figure 2 where dependent data marts can be created outside the corporate firewall. Some

organisations may go further by creating enterprise data warehouses on the cloud. This is particularly true for small and medium size enterprises where much of their core transaction data may already be stored on SaaS transaction processing applications.

SaaS BI adoption has resulted in a hybrid analytical environment

Data integration now needs to have access to on-premises and cloud based sources and targets

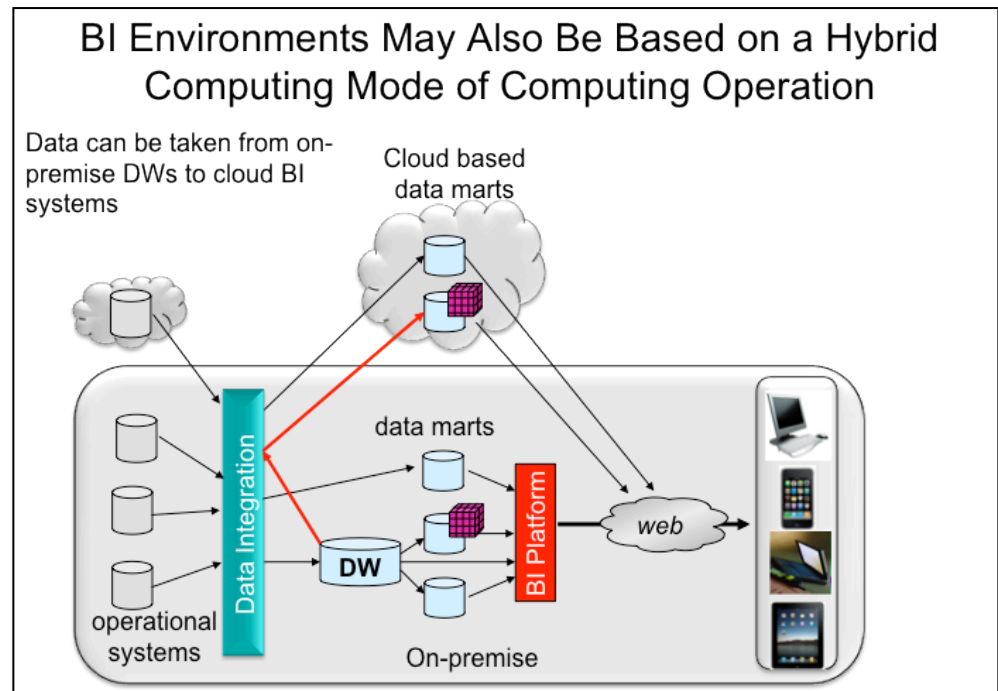


Figure 2

In addition the emergence of big data has meant that more and more big data is being collected on the cloud for analytical processing. We are also seeing an increase in archiving of data to cloud based data stores as well as external parties making value add data sources available in public cloud environments.

The net result of this is that the hybrid computing environment that has evolved in many organisations may include systems deployed across multiple clouds as well system deployed on-premises.

Given this complexity, what then are the requirements for accessing and managing data in a hybrid-computing environment? Let's take a look.

DATA REQUIREMENTS IN A HYBRID COMPUTING ENVIRONMENT

Data needs to flow in both directions in a hybrid computing environment

In any hybrid-computing environment there is a need to access and integrate data from one or more on-premises systems and provision that data to SaaS based transaction-processing applications and SaaS based BI systems. Equally there is a need to access and integrate data from one or more SaaS based transaction processing applications and other cloud-based data stores and provision that data to on-premises transactional systems to continue process execution or to analytical systems to produce new insights that can lead to better decisions. This need for two-way access is shown in Figure 3. While this figure looks relatively simple, the number of potential data sources of value in the cloud could grow rapidly and so just accessing them can be a challenge. Also the complexity of data migration to get data into new SaaS applications will increase as SaaS applications are more widely deployed. Similarly the complexity of accessing data for reporting and analysis will increase as different SaaS applications on many different clouds need to be accessed for operational reporting or for integration with on-premises data for analysis by users of self-service BI tools. It means the tools have to have connectors to more and more SaaS systems unless there are another options available to access data across the hybrid environment.

Simplifying access to data is becoming especially important in the cloud where the number of sources is growing exponentially

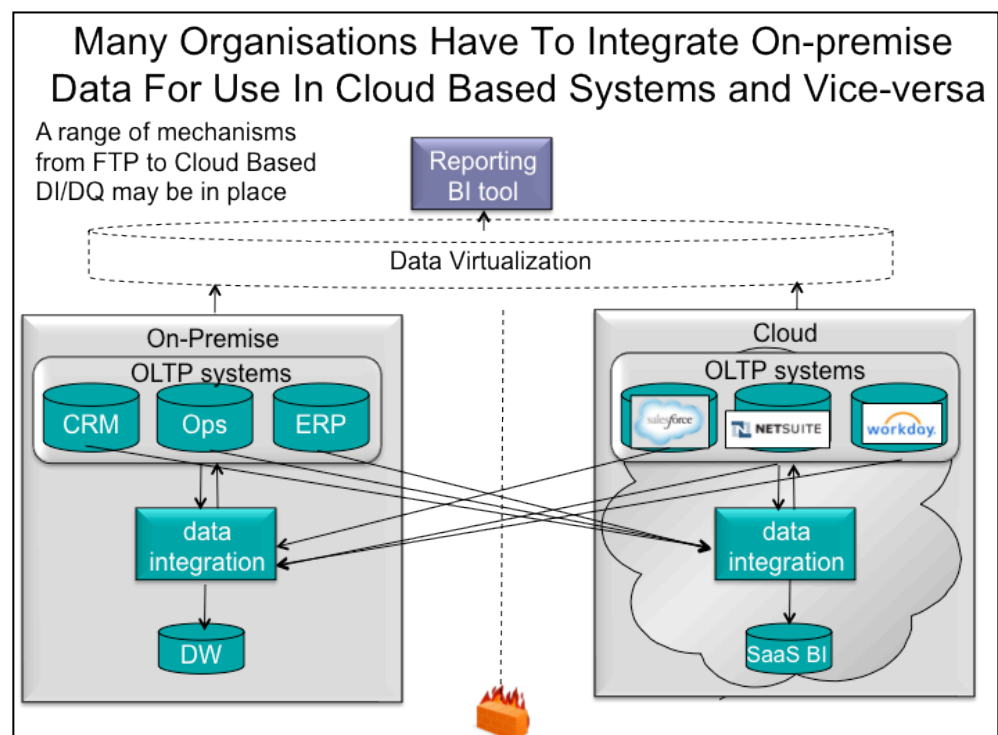


Figure 3

Given that this is the case, the following is a list of requirements for accessing and managing data in a hybrid-computing environment:

ON-PREMISES TO CLOUD

Need to migrate data to SaaS applications

It should be possible to:

- Access and integrate data from multiple on-premises data sources for migration to newly deployed SaaS-based transaction processing systems such as Salesforce.com, Workday, NetSuite, Microsoft Dynamics etc.
- Access and integrate on-premises data for use in SaaS transaction processing systems during real-time business process execution in a hybrid environment
- Access and integrate data from multiple on-premises data sources for use in

Access and integrate data required by SaaS BI

- SaaS BI systems (including cloud data warehouses and data marts)
- Cloud-based Master data management systems
- Cloud-based big data exploratory analysis

Allow cloud operational and historical reporting on on-premises data

- Access data from on-premises transaction processing systems using cloud-based BI tools for real-time operational reporting
- Access data from on-premises BI systems using secure cloud-based BI tools for historical reporting and analysis
- Synchronise on-premises master data across multiple SaaS applications

CLOUD TO ON-PREMISES

Processes starting in the cloud should be able to flow data into the enterprise

It should be possible to:

- Access and integrate cloud data for use in on-premises enterprise transaction processing systems during real-time business process execution in a hybrid environment

Operational reporting is needed on SaaS transaction processing systems

- Connect to popular SaaS transaction processing applications such as Salesforce.com, Workday, NetSuite, Microsoft Dynamics etc, from on-premises based self-service BI tools for real-time operational reporting

SaaS application and cloud storage data is needed inside the enterprise for analysis and for master data management

- Connect to popular cloud based storage (e.g. Amazon S3) and SaaS transaction processing applications (such as Salesforce.com, Workday, NetSuite, Microsoft Dynamics etc,) from cloud or on-premises data integration software for near real-time extraction of data from these applications for on-premises
 - Data warehousing
 - Master data management
 - Data virtualization
 - Big data exploratory analysis
- Synchronise cloud-based master data across multiple on-premises applications

Simplified access to a growing number of cloud sources is needed

- Connect to a general-purpose cloud connector rather than have to build or buy connectors for every cloud data source. This general-purpose cloud connector in turn should connect or be capable of being extended to connect to any cloud based data source required

HYBRID

On-demand data integration and data consolidation of cloud and on-premises data is needed

It should be possible:

- For data virtualization tools to federate access to multiple cloud and on-premises data sources to create on-demand information services for access by SaaS and on-premise applications, business processes, BI tools and portals
- To access and integrate data from multiple cloud and on-premises applications for cloud-based or on-premises data warehouses and data marts to simplify access to data

MANAGING CLOUD DATA CONNECTIVITY AND ACCESS USING PROGRESS DATADIRECT CLOUD

Progress has three products that improve access to data

Having defined the requirements for access to data in an hybrid cloud-computing environment in part one of this paper, we are now in a position to see how one vendor, Progress Software is stepping up to meet these requirements to help customers be successful with cloud computing and in managing data in a hybrid environment of both on-premises and SaaS-based applications.

Founded in 1981, Progress Software has offices in over twenty countries and more than 140000 customers. It offers three products to help organisations get access to data. These are:

- Progress DataDirect
- Progress DataDirect Cloud
- Progress EasyI

Progress DataDirect Cloud and EasyI are focussed on simplifying access to cloud data

The focus of this paper is Progress DataDirect Cloud in the main but also Progress EasyI.

WHAT IS PROGRESS DATADIRECT CLOUD?

Progress DataDirect Cloud allows organizations to connect to cloud data regardless of source using a single ODBC or JDBC driver. This includes connectivity to ODBC or JDBC compatible Software-as-a-Service (SaaS) CRM, ERP and marketing automation applications including:

- SaaS CRM applications
 - Microsoft Dynamics CRM
 - Oracle RightNow (beta)
 - Salesforce.com
 - ServiceMax (beta)
 - SugarCRM (in development)
 - Veeva CRM (beta)
- SaaS ERP applications
 - Financial force (includes bulk loading)
 - Workday (in development)
- SaaS Marketing Automation applications
 - HubSpot
 - Marketo
 - Oracle Eloqua

Progress DataDirect Cloud provides connectivity to many SaaS applications from a single driver

It allows BI tools to easily access SaaS application data to produce operational reports

This simplifies access to many cloud based data sources

Access to cloud based data warehouses and data marts is also in development to facilitate historical reporting of cloud data

Planned rollout of access to NoSQL databases and Hadoop will allow NoSQL operational reporting and analysis of big data respectively

- SaaS applications developed on Progress Rollbase

Connectivity to SaaS applications via a single ODBC or JDBC driver means that business users using self-service BI tools can access any of these SaaS applications to produce real-time operational reports and dashboards. It also means that users of self-service BI tools like Tableau, Qlikview, Tibco Spotfire and others could potentially connect to multiple SaaS applications and use their data blending capabilities to mash the data from the applications they require to produce the necessary reports. There is no need to export data to build reports given that BI tools can simply download and configure the Progress DataDirect Cloud driver, test the connection and then access SaaS application data directly. It is even possible for Microsoft Excel users to access cloud data via Progress DataDirect Cloud. This is especially important with respect to cloud based ERP and HR applications.

Connectors to cloud based deployments of relational DBMSs are also in development². This includes support for: Amazon Redshift data warehouse, and cloud deployments of IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL and SAP HANA. Once available, users of self-service BI tools like Tableau, Qlikview, Tibco Spotfire and others can potentially connect to data warehouses and data marts, blend cloud based historical data with on-premises enterprise data or with personal data to produce reports and dashboards containing actionable insights. These can then be published to information consumers to act on.

With respect to access to cloud deployments of NoSQL data stores, Progress DataDirect Cloud currently offers access to Database.com and Google BigTable. Access to MongoDB is in development. Connectivity to these systems would allow for operational reporting of NoSQL transactional data.

Finally, connectivity to cloud deployments of Hadoop³ is in development to facilitate access to big data. This includes connectivity to the following Hadoop distributions:

- Apache Hadoop via Hive
- Cloudera CDH via Impala
- PivotalHD via HawQ
- Hadapt

² Note that RDBMS access is already available in Progress DataDirect and that cloud deployments of these RDBMSs is what is being added

³ Note that access to some on-premise Hadoop distributions are available in Progress DataDirect and that cloud deployments of these distributions is what is being added

In addition to the above sources there is also a Progress DataDirect Cloud on-premise Connector available on Windows desktops and Windows Servers that support access to on-premise Progress OpenEdge and Progress Rollbase data stores. Support for other platforms are in development.

SIMPLIFICATION OF CLOUD DATA ACCESS

Simplified access to a growing number of cloud based data sources is becoming increasingly important

The attraction of Progress DataDirect Cloud is that it simplifies access to cloud data. SaaS application data, cloud deployed relational and NoSQL databases as well as big data in cloud-based Hadoop distributions can all be accessed by using a SQL-based single ODBC or JDBC driver.

Use of proprietary APIs of each and every SaaS application can be avoided

Access can be from popular BI tools and also from applications. The latter is possible because Progress DataDirect Cloud connectivity can be embedded in applications. This means that applications needing to connect to multiple cloud based applications and databases do not have to use proprietary APIs of each and every SaaS application (e.g. Salesforce.com SOQL), nor are they impacted when SaaS application vendors change any APIs. They simply use a SQL interface to Progress DataDirect Cloud to read and write data.

CLOUD DATA INTEGRATION AND MIGRATION

Progress DataDirect Cloud also makes it easier for data integration tools to connect to multiple cloud data sources and targets

It is not just BI tools and applications that can make use of Progress DataDirect Cloud. Data integration software can also make use of it to simplify access to cloud data sources and targets. This applies to both data consolidation (also known as extract, transform and load – ETL) where data is integrated to populate a target system and also data virtualisation software that can integrate data on the fly in response to on-demand requests. This is important because data integration tools can be used to:

- Migrate data to SaaS applications in the cloud
- Provide data from on-premises systems to SaaS BI systems, or to synchronize master data across SaaS applications etc.
- Pull data from cloud to on-premises systems e.g. to provide cloud data to on-premises BI systems, or to capture changes to master data made in SaaS applications to bring into an on-premises MDM system

Accessing cloud data from data integration tools is extremely important in a hybrid-computing environment

Simplifying data integration tool access to cloud data using Progress DataDirect Cloud is extraordinarily valuable in a hybrid-computing environment where complexity can spiral out of control.

PROGRESS DATADIRECT CLOUD SUPPORT FOR DATA SECURITY

Progress DataDirect Cloud ensures that data connectivity is kept secure when accessing cloud data

In addition to cloud connectivity, Progress DataDirect Cloud provides a single API with consistent security and error handling. In terms of security, Progress DataDirect Cloud uses encryption. All communication between the DataDirect Cloud Driver (ODBC or

JDBC) and the DataDirect Cloud connectivity service, including user IDs and passwords, is encrypted using Secure Sockets Layer (SSL). Also authentication to the cloud provider and the data store or SaaS application is handled as part of the login to each data store. This is an important point given the concerns about data security when connecting to cloud data. Protecting data privacy is a matter for the respective SaaS systems and cloud databases themselves.

PROGRESS EASYL

EasyI allows business users to integrate on-premise and cloud data quickly to produce cloud based data marts

Another product from Progress Software that is important in a hybrid cloud-computing environment cloud is Progress EasyI. This new product allows users to create cloud based data marts from multiple cloud and on-premise data sources. The cloud-based data marts can then be accessed via self-service BI tools using Progress DataDirect Cloud. Progress EasyI is aimed at business analysts to help them quickly create data marts by integrating data rather than exporting data to spreadsheets and trying to integrate data manually by crude mechanisms such as Excel cut and paste. Creating dependent data marts in the cloud that source data from existing on-premise data warehouses is also possible and best practice. By combining trusted data from existing data warehouses and new cloud data, Progress EasyI allows higher value cloud-based data marts to be created and is intended to help users produce reports on newly integrated data quickly through the use of business templates. This helps speed up time to value.

CONCLUSION

Cloud adoption is expected to grow rapidly

The attraction of a maturing low cost cloud computing environment and the large number of SaaS based applications available, has led to a growing number of companies deploying and migrating to SaaS transaction processing applications. That means a spike in data migration to get up and running on SaaS applications is inevitable. However, once established, many organisations soon realise that they now need to manage access to data in a hybrid-computing environment in order to establish process integration and end-to-end business process execution across on-premises and SaaS based applications.

Two-way data flows between on-premises and cloud-based systems is needed in a hybrid environment

Data integration of cloud and on-premise data is inevitable and will grow in importance

In addition, data needs to be integrated and provisioned to SaaS BI systems from within the enterprise and from cloud-based data sources to on-premises data warehouses and data marts. Also, if a lot of core data held in SaaS applications, many organisations naturally want to integrate that data with on-premises data to quickly create departmental analytical data marts on the cloud for the purposes of reporting and analysis. These data marts may be short lived or become dependent data marts to established on-premises data warehouses. Either way, a growth in cloud based analytical data stores is inevitable.

The need for cloud-based analytical data stores will also increase

Also if your e-commerce system is in the cloud, then it is likely that your web log clickstream data is also there. This added to the fact that sensor data is being collected on the cloud means that the need to access and analyse big data in a cloud computing environment is also growing.

Demand for access to cloud based external data is growing rapidly as it potentially offers significant business value

Finally, once established in the cloud, many organisations soon become aware of the explosion of cloud-based external data sources, often made available through data marketplaces, that could be of significant value to their business.

Maintaining ease of access to data in a hybrid environment will be critical to success

It is this rapid evolution that quickly leads companies to realise that ease of access to data in a hybrid-computing environment is extremely important to success and as the thirst for external cloud data grows, it should be possible to add new sources without increasing complexity. It is clear that Progress Software has already recognised this need and with Progress DataDirect Cloud and Progress EasyI together with their on-premises Progress DataDirect connectivity, they are laying the foundations to allow organisations to get to the data they need.



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 analyst and consultant he specialises in business intelligence and enterprise business integration. With over 32 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 analyst organisation. He teaches popular master classes in Big Data Analytics, New Technologies for Business Intelligence and Data Warehousing, Enterprise Data Governance, Master Data Management, and Enterprise Business Integration.

INTELLIGENT
BUSINESS
STRATEGIES



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

Data Migration and Access in a Cloud Computing Environment

Copyright © 2014 by Intelligent Business Strategies
All rights reserved