WHITE PAPER

INTELLIGENT BUSINESS STRATEGIES

Containers and Kubernetes

- Business Value in An Analytical Environment

By Mike Ferguson
Intelligent Business Strategies
November 2019

Prepared for:

Qlik

Table of Contents

THE INCREASING DEMAND FOR BUSINESS INTELLIGENCE AND ANALYTICS

Data, BI and analytics are seen as critical to running a successful business

In most organisations today, demand to consume new data, BI and analytical services is now coming from everywhere in the enterprise with many more users looking to utilise insights to deliver business value. It's like data and analytics is now in the centre of the enterprise (see Figure 1) with demand coming from all departments and also from external users like customers.

Data, BI and analytics have moved to the centre of the enterprise with demand coming from every part of the business

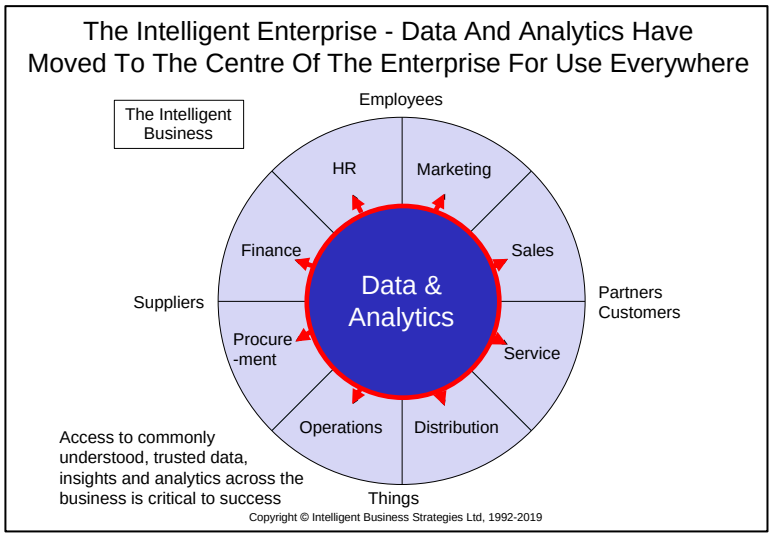

Figure 1

The data landscape is becoming increasingly complex with new data sources and data stores in a hybrid computing environment

This growth in demand is also causing a more complex data landscape to emerge as new data sources appear and new data is captured and persisted in different types of data store both on-premises and in the cloud. However, rather than having to prepare and analyse all this data, users now want 'ready-made' consumable data, BI, analytical services (e.g. recommendations) to shorten time to value.

The trend is towards more users, more data and ready made data and analytics to shorten time to value

Pressure is growing on BI and analytical platforms to be able to support these trends

The implications of all this are clear. We need simpler interfaces to broaden usage of BI and analytics so more users can deliver value. We also need APIs so on-premises and cloud-based applications can utilise BI and analytics. Both are driving up the number of invocations of BI queries, reports and dashboard services meaning BI platforms have to scale to cater for many more concurrent users. The same applies to on-demand invocations of machine learning models with multiple instances of these models likely to be needed to handle concurrent usage.

BI platforms need to be able to scale to handle more users and more data volumes

BI platforms need to run both on-premises and on multiple clouds with high availability

It doesn't stop there because analytical systems now run on premises and in the cloud. Therefore, BI and analytical servers need to run in both environments and offer high availability especially if BI and analytics are integrated with transaction processing systems that run 24 x 365. If this is the case, BI software platforms need to be easily upgradable and extensible without interrupting availability and internal and external users may need to connect via different identity providers.

Privacy is becoming more important

Finally, we want a continuous approach to developing queries, reports, dashboards to significantly reduce time to value.

NEW ANALYTICAL OPERATING REQUIREMENTS NOW NEEDED TO SUPPORT DEMAND AND DELIVER VALUE

New requirements have emerged to enable business users to deliver value

With demand for both data, insights and analytical services rapidly increasing, there are a number of new requirements that need to be supported now that most companies are operating in a hybrid computing environment to enable business users to deliver value.

It should be possible to:

Run on-premises and in multiple clouds

- Easily deploy, manage and operate data, BI and analytical platforms with minimal effort on-premises, in any cloud or both
- Declaratively configure and manage underlying infrastructure resources associated with a BI platform to

Dynamic scalability

  - Dynamically scale the platform to handle more concurrent users and more data as demand grows and more data is accessed
  - Scale specific BI platform microservices independently of others to cater for increases in demand for specific functionality e.g. security, data connectivity, analytical engines

Support continuous deployment using DevOps technologies

- Support continuous development / deployment of BI platform software using technologies such as Docker to place specific microservices in software containers for easy deployment of functionality on-premises or on the cloud. With this capability, new functionality can be provided in isolation at a much more granular level without the need to upgrade the entire platform

Leverage container management software to automatically deploy and scale your BI platform on-premises and in any cloud

- Exploit container management software (e.g. Kubernetes) to automatically:
  - Manage the deployment of BI platform software on any operating environment on premises or in any cloud
  - Scale a BI platform to run on a cluster of servers on any operating environment on premises or in any cloud
  - Load balance requests across containers to avoid bottlenecks
  - Provide high availability to avoid outage of BI platform software services running in any environment
  - Manage rolling software upgrades to BI platforms without disrupting usage
  - Manage the movement of analytical artifacts between development, test and production environments

Continuous development and deployment of reports, dashboards and models anywhere

- Support continuous development / deployment of data sets, queries, reports, dashboards, stories and machine learning models akin to a publish / subscribe style data and analytics hub and seamlessly port BI and analytical artefacts across cloud and on-premises environments

WHAT ARE CONTAINERS AND WHY DO THEY MATTER?

DevOps is key to meeting new operation requirements for BI software

These new operational requirements need the use of new DevOps[1] technologies that underpin BI servers to support a multi-cloud hybrid computing environment. One of the key technologies required is containerisation technology.

Docker packages up software as a container image that can run in any environment

A good example of this is Docker. Docker can package up an application such as a BI tool or a set of microservices representing specific functionality within it, along with all its dependencies and create a Docker container image. This contains everything needed to run the application or specific application functionality on any operating system including code, runtime, system libraries and any settings.

A Docker container is a running instance of a container image

A container image requires a Container Engine to run the application. In the case of Docker, this is the Docker Engine. When an application runs it becomes a docker container which is a running instance of a docker container image. Each container runs as an isolated process (see Figure 2) on a Docker Engine which itself runs on a range of operating systems (e.g. Windows and Linux) both on premises and on multiple clouds.

The Docker Container Engine enables containers to run regardles of underlying infrastructure and operating system

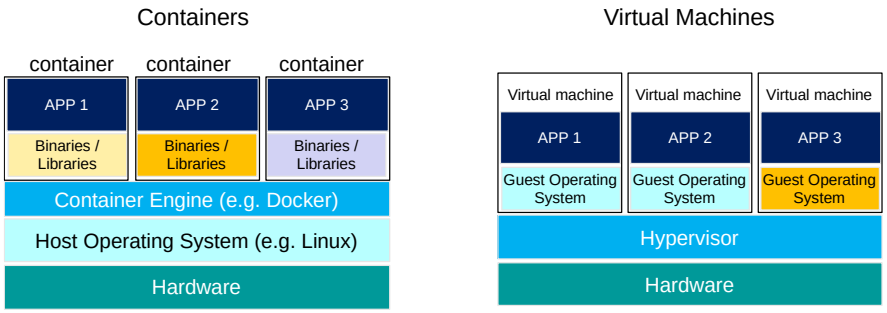

Figure 2

Running multiple containers on hardware infrastructure uses a lot less computing resources than virtual machines

Containers are also faster to spin up than virtual machines

The advantage of this is that containerised software runs the same *regardless* of the underlying infrastructure or operating system. Therefore containers matter because they make applications portable across on-premises and multiple cloud environments. They enable applications to run quickly and reliably in a number of computing environments.

[1] DevOps is a set of practices and technologies that automate the processes between software development and IT operations, so they can build, test, and release software faster and more reliably

WHAT IS KUBERNETES AND WHY DOES IT MATTER?

Kubernetes is open source container management software that manages containers in a cluster of servers on premises or in any cloud

Another key DevOps technology is Kubernetes. This is open source container management software for running, managing and orchestrating containers in a cluster of servers. Kubernetes can run on-premises, in the cloud or across multiple clouds. Many cloud providers offer it as a service in their cloud environments e.g. Amazon Elastic Kubernetes Service, Microsoft Azure Kubernetes Service or Google Kubernetes Engine.

You can define how you want your Kubernetes cluster to be configured and what types of objects you want to run in it

Kubernetes lets you define how you want a cluster to be configured and what you want in the cluster. It then ensures the system matches the configuration you have defined. Some of the main Kubernetes objects are

- PODs
- Services
- Replication Controllers
- Deployments

One or more software containers runs in a Kubernetes POD which can be restarted if it fails

PODs are the basic object in a Kubernetes cluster. They consist of 1 or more software containers. Normally only one container runs in a POD but it can run more. Each POD gets its own internal IP address if a POD fails, Kubernetes will automatically spin-up a new one to replace it. You specify what containers you want to run in each specific POD. Kubernetes then does the rest.

Requests can be load balanced across POD instances for scalability

Services act as internal load balancers between POD instances and enable web applications (e.g. a BI tool) to be accessed by external users.

PODs can be replicated for high availability

Replication Controllers (RCs) run multiple instances of a POD on a cluster and so provide both scalability and high availability.

Deployments enable automatic roll out of software updates without disruption

Deployments enable automatic roll out of software updates without bringing down existing PODs and the applications or application functionality that run in the containers associated with them.

Kubernetes provides automatic elastic scalability, load balancing, high availability and roll out of updates with no downtime

Kubernetes matters because of its ability to automatically:

- Create a cluster on premises or on any cloud to run and scale your containerized software on a cluster of servers to manage peak demands of high load
- Manage container execution on individual cluster nodes
- Scale containerised software by running containers in PODs, replicating PODs and load balancing requests across PODs using Kubernetes services to manage concurrent usage
- Recover from container failure by restarting them
- Manage rolling updates to software running in containers in any POD
- Maintain state if data is being stored in files or databases accessed by containerized software running in any PODs
- Create multiple cluster instances which are isolated from each other

WHAT BUSINESS VALUE DO CONTAINERS BRING?

REDUCE COST

Running containerised software on the cloud reduces costs and makes better utilisation of resources

A key advantage of containers is that they are lightweight which allows multiple containers to run on the same machine sharing the operating system with other containers. They therefore use less computing resources and spin up faster than virtual machines (see Figure 2) and so make better use of infrastructure. When deploying containerised software on the cloud, that translates to saving money. Also because containers run in isolation, failure of one is not going to impact on others running on the same machine.

INCREASE FLEXIBILITY

Different functional components of a microservices based BI platform can run in different containers

Another benefit is that different functionality in the form of microservices can be deployed in different containers allowing components of the same application or software product to be managed separately. In the case of a BI platform that means that new versions of functionality can be separately controlled and deployed rapidly and you can trust that a new build will run in production.
Also, functionality can be substituted by other containerised services (e.g. different security identity managers) and it becomes easy to extend the platform without impacting on any other microservices within it.

WHAT BUSINESS VALUE DOES KUBERNETES BRING TO AN ANALYTICAL ENVIRONMENT?

Having identified the above requirements and obtained a deeper understanding of DevOps technologies, the question is how does this apply to an analytical environment?

AUTOMATE DEPLOYMENT

Kubernetes enables your BI platform to be deployed on-premises and in any cloud

Kubernetes provides a number of benefits. For example, you can manage and automate the deployment of software (e.g. your BI platform or a containerised microservices based subset of its functionality) on any operating environment on premises or in any cloud. You can also dynamically configure the scalability of a BI platform and/or any individual containerised microservices within it to run on a cluster of servers on any operating environment on premises and in any cloud. This is important for scaling BI platforms to handle more concurrent users and analyse more data as described earlier. The same is true for any services developed on a BI Platform. For example if you create new queries, reports, dashboards, predictive models etc., and publish them as analytical services on your BI platform, then Kubernetes, by underpinning the running of your BI Platform, can effectively scale these services to support a growing number of user and application requests to consume the insights they produce.

BI platform and / or any of its individual containerised microservices can be dynamically scaled in any environment

Configure different instances of a BI Platform running in a Kubernetes cluster for development, test and production

In addition you could create multiple instances of a Kubernetes cluster (with different configurations) for development, test and production. You could also do this to isolate different types of analytical workloads or different types of users e.g. internal users and external users.

QUICKLY ADJUST TO MORE USERS AND MORE DEMAND

Automatically scale your BI platform using Kubernetes to handle more users, more data and to enable high availability for use by applications

Kubernetes can also replicate containers in a cluster and load balance requests across them (using PODs, services and RCs) to scale specific BI platform microservices functionality, avoid bottlenecks and provide high availability. Therefore individual containers can be scaled independently of others which may be needed as analytical workloads vary or as concurrent user numbers grow. User-defined configuration of replica PODs running containers allows you to enable and automatically manage high availability of your BI platform to avoid service outage in any environment. This is particularly important as BI and analytical services are made available to global and external users and integrated into ‘always on’ externally facing web applications.

Use Kubernetes to automatically roll out updates to your BI platform which keeping it available

Kubernetes deployments enable rolling upgrades to BI platform software to also be automated without disrupting usage. You can also automate the management and movement of queries, reports, dashboards and models between development, test and production environments.

LOWER OPERATIONAL COST

Kubernetes therefore offers many business benefits including lower operational cost (via dynamic adjustment of infrastructure), on-premises and multi-cloud portability, automated elastic scalability, high availability, workload management, and continuous integration / continuous delivery (CI/CD) of both BI software and analytical services developed on that software. It also supports automated rollout of software updates without disruption.

Manage the movement of queries, reports, and dashboards between development, test and production

Separately configure, manage and scale different aspects of BI platform functionality e.g. to handle more users

Add additional services to the cluster that integrate with BI

Run different production cluster for different analytical workloads

This plays well in the world of modern microservices based BI/ Analytics software as it allows specific BI/ analytical platform services to be isolated and separately managed in multiple different containers and container groups. For example for security or data connectivity for different groups of concurrent users can be handled separately from a BI platform analytical engine. Also multiple instances of an analytical engine could be created to scale the BI platform. This means you can separately configure resources by service type, replace service types with others if needs be (e.g. different identity providers for internal and external users) and even add new analytical services provided by other technologies e.g. to integrate BI with machine learning services for example.

Also analytical artefacts such as machine learning models can be developed and deployed as services in containers. These could then be managed, scaled separately based on concurrent usage and availability requirements and all upgraded automatically using Kubernetes. The same could be said for other types of analytical artefacts (reports, dashboards…) developed on that platform.

INCREASE BI USAGE

Develop queries, reports, dashboards and models on-premises and publish to the cloud for mass consumption

Centrally administer users and security entitlements and push to the cloud

Kubernetes also enables a develop on-premises, deploy on the cloud approach to flourish in an analytical environment. You could set up development, test and production clusters either side of the firewall and manage resources to them all separately. Similarly, security entitlements could be defined on premises and pushed to the cloud (centralised authorisation). And as more concurrent users emerge using simpler natural language user interfaces, mobile apps, and accessing BI indirectly via operational applications, you can manage and scale it all using Kubernetes. It makes the BI platform a highly available and scalable hub at the centre of the enterprise and encourages mass consumption of insights to enable more employees to contribute to delivering business value as depicted in Figure 1.

How Is Qlik Exploiting The Power of Containers and Kubernetes?

Qlik Sense Enterprise has a microservices based architecture

Given the increasing demand for data, BI and analytics, the impact this is having on BI/analytical platforms and the new analytical operating requirements it brings, we have seen the emergence of DevOps technologies like Docker containers and Kubernetes to help provide flexibility, scalability and high availability now needed. High availability is important to enable BI and analytical services to be available around the clock so that these insights can be utilised on a continuous basis throughout the enterprise and beyond. In this section of the paper we look at how one vendor, Qlik, is taking advantage of this within its analytical platform.

Different aspects of Qlik Sense Enterprise functionality can be created as different Docker containers and scaled independently

Qlik Sense Enterprise is Qlik's flagship cloud-native BI platform. It has a modern elastic microservices based architecture that allows it to group specific functionality in to separate docker containers and run them on-premises or on any cloud using Kubernetes. The architecture for this is shown in Figure 3.

Qlik Sense Enterprise can run in a multi-cloud environment

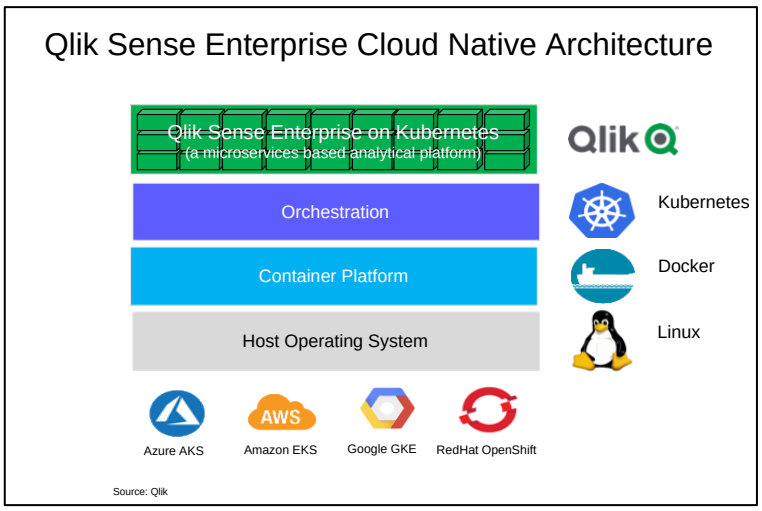

Figure 3

Qlik Sense Enterprise has different Docker containers for user and administative clients, the analytical engine, data connectivity, security and infrastructure

Within Qlik Sense Enterprise there are a number of components including:

- A browser based portal
- A browser based administrative client
- Client and content PODs
- Engine containers including the associative data indexing engine, in-memory and load balancing
- Data containers to manage connectivity to data sources, reloading and scheduling
- Security containers to configure and manage authentication and entitlements
- Infrastructure containers for network external load balancing, datastore (MongoDB) and in-memory caching (Redis)
- Storage of Qlik Sense applications and content in MongoDB and Kubernetes Data Volumes

You can see right away that it is not one monolithic software application. Far from it. Multiple containers enable different elements of Qlik Sense Enterprise functionality such as security, the analytical engine, data connectivity, and Qlik

Sense infrastructure to be managed in separate PODs and configured and scaled accordingly via Kubernetes to meet increasing demand of more users.

BUSINESS BENEFITS OF RUNNING QLIK IN A KUBERNETES ENVIRONMENT

Given this architecture there are several business benefits from running Qlik Sense Enterprise in a Kubernetes environment.

Kubernetes makes it possible to scale Qlik Sense Enterprise to handle more users and more applications requesting insights on-demand

Enable Increased Usage

The first is that Kubernetes really enables businesses to put data, BI and analytics at the centre of their company and open Qlik Sense Enterprise up to many more internal and external users taking advantage of simpler user interfaces. In addition it also allows a separate high availability Kubernetes cluster to be created to make use of BI and machine learning services (queries, reports, dashboards, etc.) by integrating Qlik Sense Enterprise with 'always on' transaction processing and mobile applications. This enables operational BI and machine learning models to be embedded right into core business processes. It also enables scalability to handle more data via Kubernetes automated scale up and scale out.

Each Qlik Sense Enterprise containers can be separately scaled on a Kubernetes cluster on-premises and on any cloud

Furthermore, Qlik Sense Enterprise's microservices architecture means that different types of microservices run in different containers in different Kubernetes PODs and can be separately scaled via Kubernetes POD replicas and services that load balance requests across those containers. It can also distribute these across more nodes in a Kubernetes cluster as and when they needed. This enables specific BI platform functionality to scale separately as well as the platform as a whole to scale.

Kubernetes allows different Qlik Sense Enterprise clusters to be configured for development, test and production all with support for automatic roll out of upgrades

In addition these Qlik Sense Enterprise containers can be brought together in separate Kubernetes deployments that are configured for development, test and one or more Qlik Sense Enterprise production clusters all with automated continuous rollout of software upgrades without disruption. All of this can be managed and adjusted declaratively via YAML. In addition, Kubernetes POD replication makes Qlik Sense Enterprise highly available (e.g. to serve up operational BI via APIs into customer facing applications on the web) and different production Kubernetes clusters could exist for different groups of users, different analytical workloads and/or different Qlik Sense applications.

Cost And Consumption Optimisation

Develop on Qlik Sense Enterprise on premises and publish to the cloud

Automatically scale Qlik Sense Enterprise to handle more data sources both on-premises and on multiple clouds

Qlik Sense Enterprise is platform agnostic, with a cloud native microservices architecture utilising Docker and Kubernetes so that you can deploy it on-premises / private cloud, in a single cloud like AWS, Azure or Google Cloud or across multiple clouds. It can also be run on Qlik Cloud Services fully managed as a service by Qlik. This flexibility allows you to create different Qlik Sense Enterprise deployments on premises or on different sizes of Kubernetes cluster for consumption in different environments. This helps optimise the cost of consumption and allows BI content authoring, security administration and data sources to be set up once on-premises and pushed to the cloud.

Meet The Demand For More Data

Also, Qlik Sense Enterprise can not only run on-premises and in a multi-cloud environment but it can bridge the corporate firewall and connect to both cloud-based (e.g. Amazon Redshift, Azure SQL Data Warehouse, Google Big Query, Snowflake) and on-premises data sources.

Support Both On-Premises and Cloud Environments

Easily move your Qlik Sense Enterprise environment to a different cloud with no rework

Running Qlik Sense Enterprise in the cloud and on-premises inside the firewall means you can connect to on-premises data sources and push Qlik Sense applications to the cloud in as many engines and clusters as you like. This enables a develop on premises and deploy in the cloud approach. It also makes migration to the cloud easier and allows one or more analytical engines and analytical clusters in cloud production environments. You can also support multiple clouds and move off one cloud onto another with no rework.

Multi-Level Security

By separating Qlik Sense Enterprise security microservices into its own container, Kubernetes allows you to support different identity providers for internal and exteral users

Qlik Sense Enterprise allows you to keep single user and user group developed content separate using spaces to maintain privacy before publishing to production spaces for easy consumption

When running in a multi-cloud environment we know already that providing access to BI by internal and external users on the cloud or on mobile apps elevates the importance of data privacy and access security. Qlik Sense Enterprise accommodates this by supporting identity provider and edge authentication containers to manage different identity providers authenticating via the OpenID standard. Identity providers allow for a common identity to be used in all Qlik Sense Enterprise on-premises and cloud deployments so that common security rules and policies can be defined once and used everywhere. It also supports user and user groups with personal and shared collaborative areas called *spaces* that allow for single user and collaborative development on Qlik Sense Enterprise. In this case the user and user group attributes are provided by the Identity Provider. Once Qlik Sense applications have been developed they can publish into managed production spaces (e.g. in the cloud) for consumption by authorised users either within or outside the enterprise.

CONCLUSIONS

A modern microservices architecture allows for functionality to be separately scaled to meet different needs such as more users and more data

A modern microservices architecture enables key components of platform functionality to be isolated, packaged up into containers and run on Kubernetes. This enables different aspects of the platform functionality to scale to handle more users, more data, more data sources, and integration of BI with more on-premises or cloud applications in a modern data driven enterprise. Kubernetes provides flexible deployment of Qlik Sense Enterprise on premises and on one or more clouds.

Kubernetes allows multiple BI to be deployed for different user communites and different workloads

All of it can be configured, managed and continuously updated in a multi-cloud environment

It also makes it possible to define and configure different independent clusters for development, test and different production workloads so that BI content, policies and rules can be centrally developed in a low-cost environment and pushed to workload and cost optimised production Kubernetes clusters running in a multi-cloud hybrid computing environment. All of this can be centrally managed and administered with a common approach to access security, user management, licensing and more. Furthermore, it can all be done with high availability where needed, automated rolling software upgrades with no disruption to service and with continuous integration and continuous delivery (CI/CD).

The value of Kubernetes underpinning a BI Platform like Qlik Sense Enterprise is that can fulfil the role of a BI and analytics hub at the centre of the enterprise offering up insights on demand to all that need it while also running at lower total cost of ownership either on-premises, on the cloud or both. It enables specific functionality to scale independently of other services and to easily scale to meet increasing demand for more users and process more data. As a result, it can make data, BI and analytics easy to access from anywhere to maximise contribution to business value.

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 and advanced analytics, data management, big data 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 analyst and consultant he specialises in data management and analytics. With over 38 years of IT experience, Mike has consulted for dozens of companies on data strategy, big data, machine learning and advanced analytics, data governance, master data management, data warehousing and enterprise architecture. He has spoken at events all over the world, written numerous articles and is Chairman of Big Data LDN, one of the fastest growing and largest data and analytics conferences in Europe. 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 Data Warehouse Modernisation, 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, Real-Time Analytics and Operational BI.

8 Paddock Chase, Poynton
Cheshire, SK12 1XR
England
Telephone: (+44)1625 520700
Internet URL: www.intelligentbusiness.biz
E-mail: info@intelligentbusiness.biz