

ENHANCE YOUR IT STRATEGY

DIGITAL LEADERS

2015

bcs.org/digitalleaders

PLAN FOR THE
FUTURE





ANALYSIS AS A SERVICE

Analytics-as-a-service (AaaS) can provide on-demand access to an organisation's predictive and business intelligence operations. James P. Howard, II MBCS and Scott Beaumont introduce the concepts of AaaS and provide an architectural overview of this powerful technology.

Analytics-as-a-service deploys predictive models directly to enterprise as first-class services on a service-oriented architecture. Making these models first-class objects gives a number of advantages.

First, they are decoupled from their data sources and access data through abstraction layers. Second, they are decoupled from their callers and give the callers an abstract interface for analytic processing requests. Finally, the decoupling continues to receivers where the recommendations of the analytics service are presented through the abstraction layer.

These decouplings and abstraction layers are supported by asynchronous

communications through the use of message queues while the access abstraction layer is created by the enterprise service bus (ESB). Pushing abstraction enables the entire enterprise to move into cloud-based analytics, taking analytics and, importantly, decision support off of individual desktop computers.

Customer-facing systems with access to high-quality analytics at runtime can replace critical triage functions with decision support services through cloud-based solutions. These solutions will give faster response times with better accuracy when analytics are provided as a service. Further, by focusing on analytics-as-a-service and decoupling access paradigms, all parts of

the enterprise can access the system.

Architecture

The architecture of a data analytics platform is here broken down into four basic components: data storage, extraction – transformation – load (ETL) services, analytics services and real-time analytics services. Other components, such as ad hoc analytic tools, monitoring and reporting services, and end-user UI access should be included, but are left out here for brevity.

However, these additional components would depend upon the four core components described in the rest of this section.

The data store(s) used for holding the

to-be analysed data can range anywhere from a traditional relational database management system to a non-relational, distributed NoSQL database engine. More complex scenarios can even utilise a combination of storage engines. The type of storage engine(s) selected should depend upon the type and amount of data that needs to be analysed. Many enterprises are dealing with the current data explosion by moving to distributed NoSQL database systems and using the cloud to scale their analytics storage and processing.

The ETL services are responsible for fetching (and receiving) data, transforming it to a useful model, and ensuring that is loaded into storage. ETL services can be deployed to automatically fetch data at regular intervals, depending on computing costs, bandwidth, service level agreements and so on. This is the traditional approach to ETL.

However, in a service-oriented architecture, it is also possible to set up ETL services that receive data from publishers, especially internal business services that produce critical data during processing. This can provide enterprises with real-time analytical capabilities. This

communicates with the ETL service. An ESB or message queue can also provide reliability, ensuring that the ETL service eventually receives the message, even if it was unreachable during execution of the business process.

The analytics services are the consumer gateway to the data analytics platform. Some services might provide access to the underlying data stored within the platform. Other services are responsible for performing specific analysis and returning the results. Still other types of services might be responsible for scheduling and storing results for future reports. In an SOA, these services can be composed together or used separately by consumers to produce meaningful results.

A distinction needs to be made between long-running analytics services and those that provide real-time results. Currently, in-depth processing of huge data sets takes some amount of time, anywhere from seconds to minutes to hours, and maybe even days. It depends on the depth and amount of data to be processed, the algorithms used for processing, available bandwidth and so on. In many cases, this is acceptable.

However, there is a growing need for

Analytics-as-a-service can improve the implementation of enterprise-level reporting, decision support and intelligence.

is where data analytics is required to be a first-class citizen in the enterprise.

Business services need to be implemented knowing what, when, where and how: what data needs to be sent for analysis; at what points during processing it should be sent; where to send the data; whether to send it by ESB, queue, or direct service call; and how to format it for the service.

Some type of asynchronous approach, whether a full-fledged ESB or a light-weight message queue, is usually the best option. An asynchronous approach will not tie up the business process while it

in-process analytics; that is, analytics that can provide an almost immediate response during some business process. This might allow an enterprise to detect identity fraud or misuse of company resources while or even before it happens, and not after the fact.

Services that provide these capabilities are identified as real-time analytics services, but the classification of what qualifies as 'real-time' is really up to the enterprise.

A service that takes a second or two to provide a result would be acceptable for many enterprises, while other enterprises

can allow for more or require less time.

These 'real-time' services might meet their time constraints by using just a subset of available data or by pre-fetching and remodeling data for its specific uses. The point of the 'real-time' service is to provide quick value to some business process. In the previous example, a full analysis to detect identity fraud or misuse of resources should still be conducted regularly with all available data.

A key way to improve the quality of decision-support models is to close the feedback loop. Closing the loop captures the key features of the predicted case, along with its outcome, to revise the model.

This can be done in real-time or with a bulk update, but the important part is tracking a known outcome. Outcomes, both positive and negative, serve to strengthen the predictive quality of the model.

As the model is revised, its predictive power improves and increases the model's efficiency and accuracy. The enterprise's decision-support systems improve, providing better results to the business. As the predictions improve, so do the results and the circle is reinforcing. Further, while the accuracy is improving, so is the precision, and the narrower predictions support the enterprise's business decisions.

These sorts of self-updating models are used by technology companies to improve search results while retailers' recommended products use these sorts of models to connect shoppers to products they are likely to purchase. These models are revenue drivers when properly deployed.

Analytics-as-a-service can improve the implementation of enterprise-level reporting, decision support and intelligence.

There are a variety of tools available to analytics architects to improve deployment, increase capabilities or meet business constraints. Implementation is only limited by the requirements of the service and the ingenuity of its implementers.