Accelerating Business Value with a Multi-Model, Multi-Workload Data Platform
Overview

Leading enterprises are starting to weave advanced analytics into their core operational systems. The goal is to use the output from predictive models dynamically for decision-making. This requires access to multiple data stores of varying types and the use of many data processing techniques. Innovative applications simultaneously blend transaction processing and analytics to deliver business value while avoiding IT complexity, driving the need for a data platform based on a multi-model, multi-workload database. Through consolidation into a single data platform, businesses generate new value by delivering insight at transaction time while preventing yet another technology silo and minimizing data lag.

With the availability of new enterprise data sources to use as predictive measures, IT organizations are being asked to store an increasing volume of data in a variety of formats that may or may not be used in future projects. Different types of data have been traditionally managed using separate data stores. As a result, big data and analytics are commonly deployed separately from transaction processing systems, introducing time delays between when insights can be applied to affect business transactions. Hybrid applications combining transaction processing and analytics components with flexible access to multiple data models can help enterprises offer new value creation opportunities at an organizational scale—without the added burden on IT.

Multi-workload support

Operational processes, predictive analytics, natural language processing and data research have different workload signatures than transaction processing as well as unique system-level requirements. Many open source projects have been created to solve individual analytical use case needs. Despite the growth of programmatic access to data from the NoSQL open source movement, SQL remains central to the way organizations access and process data—both for reporting and transaction processing. For next-generation applications to be successful, it is critical that developers have access to both SQL query mechanisms as well as programmatic APIs to support different kinds of data models, different workloads and different skill sets.

Multi-analytical processes

There are multiple types of analytical processes including data research and discovery, predictive analytics and scoring, and key performance indicator reporting and dashboarding. Enterprises organize these teams differently depending on how integral analytics is to core business processes. More successful organizations tightly couple these teams, which in some cases can cause conflicts. Research teams want unstructured access to all data, while transaction processing teams require organized, predictable access to specific data. Therefore, organizations need a data platform with flexible access to data for analytics that can also support high-performance access and data ingestion to speed transaction processing.

Data interoperability

Analytical research processes need to interoperate with the data platform using analytics tools that can be applied from outside the data platform with cutting-edge data science algorithms. These tools don’t necessarily need to leverage SQL. They are more likely to take advantage of Spark or other kinds of machine learning tools. In order to be successful, a data platform needs to support the research side and enable high-performance applications of that research at the same time, in many cases third-party data sources and data provided via smart devices or the Internet of Things. This requires data interoperability as well as the ability to apply context. For example, combining data from medical devices with patient medical records requires healthcare data interoperability.

Making sense of unstructured text

Next-generation applications are benefiting from natural language processing techniques. A key requirement here is the ability to build ontologies or vocabularies linking words or phrases that mean the same thing and use that context in predictive analytics. For example, an apartment listing service in Germany implemented a data platform that allowed owners to enter structured information about the number of bedrooms and bathrooms, and unstructured text about the apartment. An analysis of key word searches by consumers revealed important insights for preferences that were only found in the unstructured text. The listing service is using natural language processing to make these apartments searchable in an
intelligent way.

Healthcare is another example. A considerable part of the data used in healthcare is only found in unstructured text. Doctors make notes that are transcribed into text. Better analysis requires translating these notes into structured data like diagnoses or symptoms. For instance, a large drug manufacturer was interested in identifying patients at higher risk for hepatitis C. Some of the best risk indicators included a history of incarceration and the number of tattoos. None of this data was included in a structured medical record, but it was commonly found in transcribed patient notes. Using natural language processing to analyze these notes, the drug manufacturer could identify patients with hepatitis C with 50% more accuracy than by looking at the structured data alone.

**Time lag**

Data lakes and data warehouses have been emerging as independent architectures for aggregating data. When log data and analytics are stored separately and analyzed independently from transactional data, time delays are inherent. Enterprises want to be able to affect what is going on in real time. Data lakes and data warehouses require an extra data integration and transformation step. This creates an additional data flow, impeding the ability to conduct inline analytics. Industries that rely on more real-time, inline analytics—such as global financial markets, emergency healthcare, severe weather alerting or utility providers—rely on the ability to apply real-time analytical models to make the best decisions or predictions.

**Centralize security, governance and scalability across data models**

Security, privacy and data governance are more complex across disparate systems. The challenge with maintaining multiple data systems lies in the need for separate management tools and processes for each data store. A key benefit of a centralized data platform is it provides a single, aggregated point of access with a single security model across all of the data stores. A data platform allows the enterprise to extend this beyond relational databases to other data models as well.

Many enterprises have spent years trying to architect the infrastructure to enable transaction processing and operational workloads to scale. Others have invested effort and money into scaling research-oriented workloads. The question enterprises face today is how to architect systems to deal with hybrid workloads that combine elements of both.

The operations side has been focused on processing a large number of transactions in parallel, with reliability and consistency. The analytics side has historically been focused on a batch-oriented approach for handling lots of data simultaneously on behalf of a few data scientists, which does not involve a lot of parallel activity. The requirements for these two approaches scale differently in terms of memory, compute and networking hardware. A data platform can simplify scaling applications across a large number of servers automatically with the right characteristics required for the job.

**Streamline the path to value creation**

Enterprises that want to create the next generation of applications need to consider the most efficient way to business value. Business value comes from developing applications rather than infrastructure. A unified data platform simplifies the ability to leverage multiple data models and processing techniques so that enterprises can focus on business value. An example is InterSystems' technology portfolio, which is based on its multi-model, multi-workload database Caché, and comprises an associated ecosystem of tools for integration, analytics and natural language processing.

The requirements for building world-class applications have changed—and will continue to change. If developers are not building applications that leverage big data and harness analytics, they will not be successful. Enterprises need to consider whether they can get what they need from a comprehensive platform or from deploying multiple separate pieces. In the long run, many enterprises will find that an integrated platform that manages the back-end complexity can save time and money, and increase the pace of developing this emerging style of hybrid operational/analytics applications.

Big data and analytics techniques need to be integrated into consolidated enterprise architectures to simplify the development of new applications. This helps enterprises generate new value faster and less expensively.
Enterprises need to consider the best approach for quickly implementing hybrid workloads across data models and disparate data sources. Weaving together multiple data storage and processing platforms can slow application development and create scalability challenges. It also complicates governance, risk management and compliance.

Combining data into offline data warehouses or data lakes can help, but this adds complexity and lag. An integrated data platform like InterSystems can reduce complexity so that the enterprise can focus on what brings value to more users in the organization. The InterSystems data platform has delivered on this promise to thousands of customers for more than 30 years.

For more information on how InterSystems can help your enterprise generate value faster and with less overhead, go to www.intersystems.com.

About InterSystems
InterSystems is the information engine that powers some of the world’s most important applications. In healthcare, finance, government and other sectors where lives and livelihoods are at stake, InterSystems has been a strategic technology provider since 1978. InterSystems is a privately held company headquartered in Cambridge, Massachusetts (USA), with offices worldwide, and its software products are used daily by millions of people in more than 100 countries. For more information, visit www.InterSystems.com.