Production Analytic Platform—A Shrinking Decision Cycle

DECEMBER 2017

A ThoughtPoint by
Dr. Barry Devlin, 9sight Consulting
barry@9sight.com

A blend of predictive analytics with production values and goals is a mandatory foundation for the tightly-coupled, closed-loop decision-making cycles characteristic of modern digital businesses. As time-to-decision decreases, reliability, maintainability, and other qualities of the Production Analytic Platform become increasingly important.

From discovery to action—and back

The MEDA model defines a closed-loop cycle of decision making and action taking—function that has been implemented in a siloed manner in today's analytic world. The Production Analytic Platform breaks down these silos and improves analytics across the whole business.

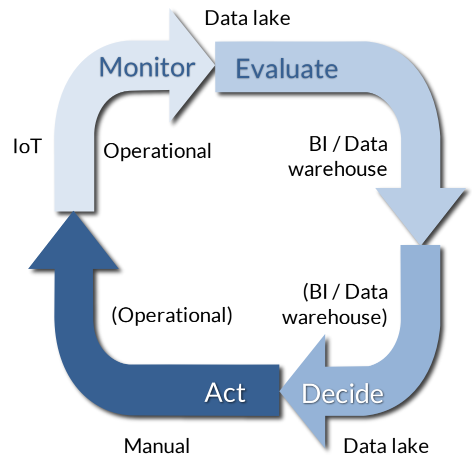
It's many years, perhaps decades, since it has been acceptable to suggest that a decision maker should go get a cup of coffee while the required data is being crunched. Of course, some decisions do demand considerable time for thought and some data is so big that crunching it is far from instantaneous. However, modern business must operate in cross-functional, tight and closed loops, avoiding any potential delay that introduces the possibility of being outmaneuvered by a competitor or ditched by a customer.

Nonetheless, much design thinking still starts from the simplicity of the silo: Optimize for performance of a single, well-scoped function—an app—with maximum control of the needed data and minimized external dependencies. This approach remains much favored by developers because it increases their chances of successful project delivery. However, the longer-term success of the business process is endangered by *ad hoc* data hand-offs, mismatched function, and time-devouring breaks in continuity.

A closed-loop, sense-and-respond approach is required, such as the MEDA model I have long promoted[1]. The acronym stands for:

- *Monitor* what is happening both within and outside the enterprise
- *Evaluate* implications and consequences, and possible actions

- *Decide* among possible/recommended courses of action
- *Act* to change behaviors, and/or processes and link back to monitoring



In traditional data warehousing, BI tools support this cycle—mostly the evaluate phase—but much work and coordination is manual. Analytic environments, as typically built today, show more automation but, with the IoT and data lake, lead to a more technically fragmented environment. The figure to the right shows how multiple systems support MEDA.

In the past, monitoring occurred only in operational systems. For production systems, this remains the case. As events in social media and IoT are of increasing interest, they are monitored in the data lake. Depending on the source, evaluation occurs in both the data warehouse and lake. Decision is increasingly automated for faster turn-around, and is also split between warehouse and lake. Action occurs in operational systems where possible or may be manually initiated.

As discussed in the first ThoughtPoint of this series[2], the current approach leads to complex data flows between the three environments—operational systems, data warehouse and data lake. The Production Analytic Platform reduces this data flow complexity by enabling all four phases of the MEDA model within an integrated and performant environment. And, as described in the second part of the series[3], it also addresses the temporal requirements of handling times series and bitemporal data.

With decision making and action taking in MEDA under ever tighter time constraints, two further characteristics of the Production Analytic Platform emerge. First is the principle of bringing the analytic function to where the data resides rather than moving the data to a tool with the right function. The Production Analytic Platform must incorporate a complete set of analytic function, all the way from simple aggregation to advanced analytics, from mathematical & statistical functions to machine learning. Second, the decision cycle must support a wide variety of users with a range of skills and tool preferences. Business analysts review data and make decisions, likely with SQL tools. Data scientists explore data with R/Python, for example. Developers operationalize the results with a combination of tools. The Production Analytic Platform must offer all needed tools.

We now examine the data and analytic function provided by the Production Analytic Platform, based on the MEDA model above.

From data to action on the Production Analytic Platform

A Production Analytic Platform offers multiple data storage formats and a wide range of powerful analytic function in a consolidated relational environment with strong operational characteristics across the full MEDA cycle.

Externally sourced data from the IoT and information from social media arrives in the enterprise in a variety of mostly simple, text-based formats. Internally sourced production data is almost exclusively relational. The Production Analytic Platform must therefore efficiently store this wide array of formats and offer appropriate functionality to process and analyze them with ease. This begins with the Monitor phase of MEDA.

Monitoring the world—within and without

A high-speed train departs from Madrid, its final destination Copenhagen, with only two stops *en route* in Paris and Frankfurt[4]. Soon reaching its top speed of 250 km/hour, it races across the Spanish plains, spewing not smoke but data describing every aspect of its performance. This data—multiple streams of events and measures—arrives continuously at the data processing center of Trans-European Rail (TER) in Berlin, where the progress and performance of all its rolling stock is monitored in real-time.

As discussed previously, this is time series data. TER's specific focus is on the payloads of these streams, containing detailed records of axle rotation speeds, temperatures, vibration levels, and lubrication pressures for every wheel of the train, as well as a range of events on the journey, such as applying brakes, accelerating, and crossing railroad switches. With multiple sensor and edge processor manufacturers, such data arrives in many formats, from simple comma separated variables and key-value pairs to sophisticated JSON and AVRO structures, in text and binary modes. To effectively monitor performance, the Production Analytic Platform stores all these formats natively within the relational database with equal ease. Using SQL with appropriate extensions, business users and applications monitor the incoming data streams and flag any unexpected events or measures. And with more traditional operational data in the same environment, any impacts on passengers or schedules can be easily spotted.

Evaluation and decision—models, analysis, and answers

In an environment where externally sourced data—be it from social media or the IoT—is the basis for evaluation and decision, only statistics and model-based analytics can operate at the scale and speed required. In the case of TER, the data exhaust from their high-speed train offers the opportunity to observe in near real-time the performance of their equipment, to note emerging problems, and to predict—and avert—potential failures.

The starting point is to develop models that correlate observed abnormal measures, such as increasing temperatures or vibration levels in the absence of braking events, with known failures in axle assemblies and subsequent train breakdowns. Such models, usually produced by data scientists in high-performance analytic data lakes, combine data from multiple time series over many train journeys to predict time to failure in many, varied situations. And one of these situations is emerging on TER's premier Madrid-Copenhagen service that has crossed into France just after midnight. Hydraulic pressure has risen while temperature is beginning to climb on bogie C of carriage 12.

To understand the potential implications of these changes, TER must run an array of models on the near real-time train data that is arriving on its Production Analytic Platform in Berlin. These models originate in the data lake and are maintained and enhanced there. They are executed on the Production Analytic Platform using the incoming IoT data using a variety of analytic tools.

For example, the recently introduced Teradata Analytics Platform, based on a combination of Teradata and Aster technologies, offers R, SAS, Jupyter, and KNIME analysis environments, and will include the Spark and TensorFlow engines in the near future. More tech-savvy analysts can work in Python or SQL as they desire.

But what of the Madrid-Copenhagen train? The news is not good: there is a 52% probability of failure by the time the train reaches Paris, rising to 78% by the time it arrives in Frankfurt. The carriage will have to be pulled from service before it arrives at its destination. But where?

Action—where the wheels meet the rail

Taking action demands data from a much broader set of sources than simply the IoT data from the train. A range of traditional business data will play into determining the best action to take. Occupancy of the affected carriage and availability of equivalent seating elsewhere on the train is an important consideration. In the event of there being insufficient alternative seating, a replacement carriage will need to be attached to the train. The immediate question is where is the nearest one and what knock-on effects on other trains might ensue. Or perhaps a temporary repair might suffice, in which case the locations of components and skilled staff will be required.

A key strength of the Production Analytic Platform is that such traditional operational and informational data is part of the same environment. With such data readily at hand, analyses based on complete information are easily undertaken and can be run repeatedly as updated predictions on likely time to failure are updated from the analytic models of ongoing real-time data from the train.

With hybrid row/column storage, in-memory optimization and vectorization, and automatic multi-temperature data management, as well as an intelligent cost-based query optimizer, the Teradata Analytics Platform provides an ideal environment for such complex and time-sensitive workloads.

Conclusion

Today's decision-making cycle demands tight integration of very different types of data and function, from collection and analysis of external data, through predicting future states, to taking immediate action based on ongoing operations. A Production Analytic Platform is vital to meet these demands.

Modern business needs for timeliness and cross-organization coordination drive a shrinking decision-making cycle. Real-time data from the external world—social media and IoT—is the foundation for ongoing modelling and analysis of the ever-changing behaviors of people and machines in the physical world. Seamlessly combining such data with traditional operational and informational data from internal systems is vital to ensure a closed-loop MEDA-style decision cycle. A Production Analytic Platform offers an ideal environment to bring the varied data and processing needs together with the required reliability, scalability, maintainability, and performance.

Dr. Barry Devlin is among the foremost authorities on business insight and one of the founders of data warehousing, having published the first architectural paper on the topic in 1988. With over 30 years of IT experience, including 20 years with IBM as a Distinguished Engineer, he is a widely respected analyst, consultant, lecturer and author of the seminal book, "Data Warehouse—from Architecture to Implementation" and numerous White Papers. His book, ***"Business unIntelligence—Insight and Innovation Beyond Analytics and Big Data"*** *(http://bit.ly/BunI-TP2) was published in October 2013.*

Barry is founder and principal of 9sight Consulting. He specializes in the human, organizational and technological implications of deep business insight solutions combining all aspects of internally and externally sourced information, analytics, and artificial intelligence. A regular contributor to Twitter (@BarryDevlin), TDWI, BACollaborative, and more, Barry is based in Cape Town, South Africa and operates worldwide.

[1] Devlin, B., *"Business unIntelligence—Insight and Innovation beyond Analytics and Big Data"*, (2013), Technics Publications LLC, NJ, http://bit.ly/BunI-TP2

[2] Devlin, B., *"And now… a Production Analytic Platform"*, October 2017, http://bit.ly/2k3QmQx

[3] Devlin, B., *"Production Analytic Platform—It's a Matter of Time"*, November 2017, http://bit.ly/2ABTpng

[4] Sadly, in the real word, this journey would require several high-speed trains operated by different companies.