

BANKS SHOULD PURSUE EFFICIENT AND SCALABLE “MODEL FACTORY” MANAGEMENT

BY STEVE TURNER, WES WEST, AND KAUSHIK DEKA

Although regulator-driven at the outset, quantitative thinking is now firmly in the DNA of many banks. As a result, many are struggling to manage hundreds of models that require an increasing number of resources.

Novantas believes that such model management can become more efficient

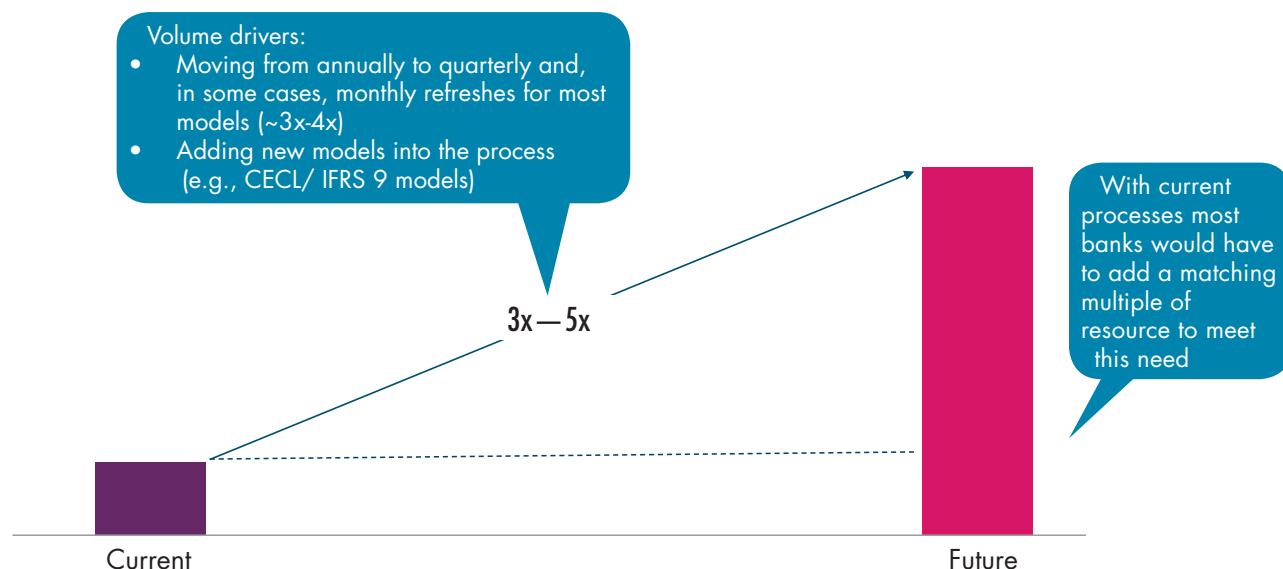
by using sophisticated technology, including machine learning, to automate the process.

A growing number of banks are already using a “factory” approach for managing their models, but the process is still daunting and demands a significant amount of technical expertise. New approaches address this growing need

for oversight, inventory management, and maintenance (see Figure 1).

The number of models in use varies dramatically according to the specific institution. A bank with 150 models can have more than 600 model monitoring runs a year.

A traditional reaction would be to pile on more bodies — modelers, data

FIGURE 1: ESTIMATED INCREASES IN VOLUME OF MODEL REFRESHES PER YEAR

Source: Novantas Analysis

managers, validators — to take on the tasks of data review and validation, model refreshes and reruns, and test results reviews. These demands would require three to five times the staff used today. Most banks don't want to grow their teams by this magnitude and, even if they did, the ability to find and keep a qualified team is nearly impossible.

Instead, they can ease the burden by leveraging capabilities that are available to the industry and, in some cases, using artificial intelligence to speed the process.

TURNING DOWN THE HEAT

As a first step, some banks have created a more structured model management process we describe as a “model factory.” In a model factory, bankers focus on efficiency and effectiveness by looking at each stage of the planned process: determine what needs to be done, figure out ways to do it with the least effort, set rules on data source usage, and establish common development standards.

This aligns processes to reduce a complex set of production tasks into what we visualize as a “model factory”

that has distinct and rigorously-defined stages: “architect”, “data scientist”, “data engineers”, “modelers”, and “model validators.” This “factory” environment structures the approach, eliminates unnecessary redundancies, and decreases the amount of rework. Many banks are benefiting from this streamlined and consistently-conducted process, but the number of monitoring events keeps increasing. **Most teams will be swamped in the next 24 months without a radical re-evaluation of their workflow.**

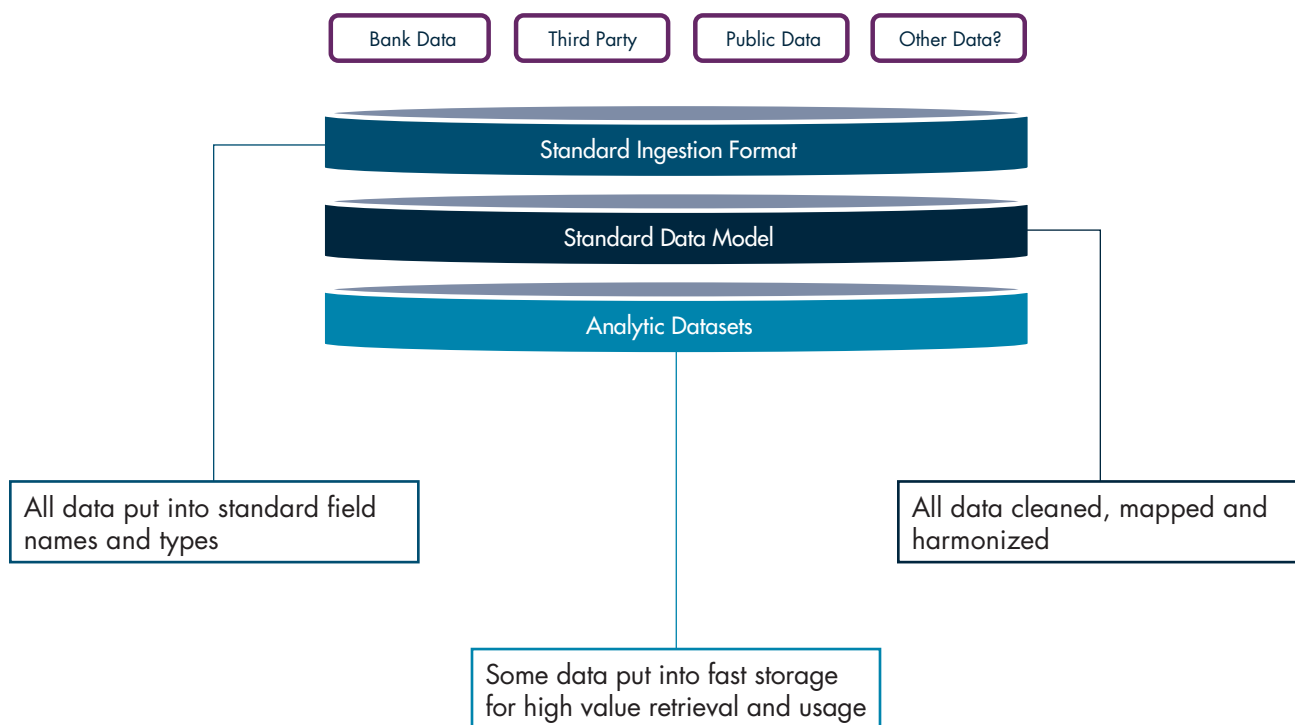
TAKING THE POT OFF THE STOVE

Banks have done everything “humanly possible” to improve upon this situation. Now banks need to take advantage of technology to improve this process in three dimensions.

Data. It is surprising that the biggest challenge banks continue to face is getting and maintaining data in usable condition for modeling. There are no guardrails to direct modelers into where and what data can be used allowing for variations in sourcing that then requires separate validation processes.

And once in production, source data is often maintained in multiple data store — each with separate rules, refresh routines, and validation processes. Even banks that are rolling out an enterprise data lake as a single source of truth are struggling to operationalize their data assets and build scalable processes. Seldom is there complete registry and lineage to sources making validation a cumbersome process. Further, the data used to build models are often not the same data used in production.

To solve this conundrum, advanced banks are investing in high-performance models that position the data to be centrally managed and curated and usable for modeling and production with automated checks and balances in place (see Figure 2). All data is cleaned, mapped and harmonized on first time ingestion. Subsequent refreshes can then be fully automated with rules in place for verifying sourcing and validating each element. **As a result, the dream of validating a model without the lengthy and low-value debate on data viability is quickly becoming a reality.**

FIGURE 2: HIGH PERFORMANCE DATA MANAGEMENT

Source: Novantas Technology Schematic

Modeling. Establishing rules up front for model families, acceptance criteria, validation tests, and driver preferences significantly decrease the range of decisions required to develop a model. Models can then be developed using machine-learning processes within imposed constraints for many CCAR and ALM-type models, resulting in the best 100 or so models being generated at a push of a button. They can then be served up for higher order review by an experienced modeler.

Most banks adhere to model development paths where modelers play decisive roles in identifying preferred models. Supervised machine learning, which cuts down the number of viable models, fits into this path. A separate type of machine learning occurs in unconstrained modeling processes, e.g., fraud detection, where the driver variables are unknown, changing over time, or both. **Both approaches — supervised and unsupervised machine learning — can be dropped into structure**

processes for data management and model monitoring.

Monitoring. This is the primary source of the massive increases of model management. It also is the place where all the data, model family, and processing rules come together to create significant improvements in efficiency and effectiveness. Machine learning can be harnessed to create automated efficiencies for integrating newly-created models into the monitoring process. When a bank combines automated data ingestion, standardized development, and optimized reporting within an integrated operational framework that supports multiple constituencies, efficiency results are exponentially greater than the sum of the parts.

THE BIG COOLDOWN

In the current environment, the model management process is a set of tedious and repetitive rote tasks. High-value resources are often needed to ensure effective task completion, but will tire

quickly, and move on. Alternatively, using lower quality resources is highly procedural dependent and midstream problems will often go unnoticed.

In the new world, the repetitive tasks are automated, new tasks are machine learned, and reporting occurs without human intervention. High value resources are only called upon when models break, and new model solutions are needed. The result is higher quality outcomes, lower cost to maintain, and happier model management teams. ■



Steve Turner
Managing Director, New York
sturner@novantas.com



Jonathan 'Wes' West
Managing Director, New York
jwest@novantas.com



Kaushik Deka
Director, New York
kdeka@novantas.com