WHITE PAPER

Automating Advanced AI at Scale

Using Teradata Vantage ModelOps to gain sustainable value from Advanced Analytics

By Dr. Will Fleury, Vantage ModelOps Product Owner, Teradata,
Pablo Escobar de la Oliva, Vantage ModelOps Product Manager, Teradata
and Dr. Chris Hillman, Director of Data Science EMEA, Teradata

06.22 / DATA ANALYTICS / WHITE PAPER

teradata.

Table of Contents

2 Executive Overview

4 Vantage ModelOps

5 Faster Deployment

6 Why deploy in Vantage?

8 Managing the whole lifecycle

9 Vantage ModelOps – streamlining deployment for faster time to value

Executive Overview

MLOps and ModelOps are emerging disciplines that seek to remove friction productionizing AI and improve time to value for enterprise analytic projects.

While MLOps is known to help data science processes with automation, ModelOps is focused on the model governance and integration, and both complement each other. ModelOps plays a key role in Teradata's Analytics 123 Strategy[1] by linking the building and training of models to their effective deployment and management in the business. As such it connects teams, tools and processes across lines of business, data science and IT departments. Modelled on the principles of DevOps which have eroded the barriers between software development teams and the deployment of software solutions in the business, ModelOps will become increasingly important as businesses look to deploy thousands and potentially millions of AI models to remain competitive in the digital economy.

The lack of clear methodologies and solutions to streamline the movement of models from exploration and development into scoring in production creates a great deal of friction in the process. Data Scientists, business and IT teams are all frustrated by the time and resources needed to get models delivering value to the business. Teradata developed ModelOps to help data scientists operationalize models at scale. While other solutions provide integrated data science environments for development and experimentation and include some basic model management capabilities, they seldom provide support in the critical phases moving from development to production. ModelOps specifically tackles challenges in the following areas:

- Model Deployment
- Model Lifecycle Management
- Data Drift and Model Monitoring
- Model Governance

1. https://www.teradata.com/Resources/White-Papers/Analytics-123-Enabling-Enterprise-AI-at-Scale

Challenges in each of these phases continue to hamper efforts to embed predictive analytics at scale in businesses across all sectors. Model deployment continues to be a challenge. Gartner suggests that up to 80 per cent of models never make it into production[2]. This wastes the efforts and insights of data scientists who have painstakingly built the model and crucially, failing to deliver business value. The gap between exploration and production is often articulated as the 'throw it over the wall' issue, signifying the lack of continuity between those that build models and the teams responsible for deploying them. Lack of process, understanding of what and how models work and protection of the integrity of live data conspire to prevent many models from ever delivering business value. It is important to remember that:

> **"Through 2022, only 20% of analytic insights will deliver business outcomes."[2]**

Manual processes slow the journey to production

Part of the issue lies in the heavily manual nature of moving from a trained model to a deployed model in production. Training is not the end of the process, nor is scoring live data the immediate next step. Evaluation and approval of new models are essential, but there are seldom well-formed processes or responsibilities for undertaking these steps. Data Scientists spend a great deal of time and effort getting a single use case to production as a one-off procedure. They essentially start again for the next model with little repeatability or opportunity to create and reuse best practice. Faced with these one-off, manual processes, it is unsurprising that it takes an average of five-months from development to deployment of new models at scale.[3]

Too many are flying blind

Once deployed, it can be a case of 'fire and forget.' With little ongoing monitoring of model performance, many organizations are effectively flying blind and risk being caught by changes in assumptions, data or situations. The recent pandemic, for example, immediately and abruptly impacted the vast majority of models across most industries as customer and business behaviors moved well outside of predicted parameters. As many businesses move towards creating and managing 'segments of one'[4] as a central facet of their digitalization and differentiation strategies, ongoing monitoring of potentially millions of micro models deployed across the enterprise will demand a more automated approach. Ensuring that performance, reliability and robustness of every model remains within expected parameters will become vital.

Governance increasingly important

AI model governance is how an organization controls access, implements policy, and tracks activity for models. As more decisions are made based on the scoring of data with increasingly sophisticated models, both the business and regulators will need to know how models were trained, and on what data, in order to audit and understand how those decisions were reached. For example, the rights of individuals guaranteed by GDPR[5] in the EU make full auditability and model lineage essential in any ML systems that rely on personal or customer data. Currently too much governance relies upon the notebooks, pipelines and personal experience of the data scientists who created each model. Not only can these be hard to track-down, but they could well leave the organization as individuals move on. The detailed and precise documentation of each step necessary for clear audit trails often simply does not exist.

---

2. https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/
3. https://analyticsfrontiers.uncc.edu/sites/analyticsfrontiers.uncc.edu/files/media/avi_misra_ai_and_ml%20%282%29.pdf
4. https://www.bcg.com/en-es/publications/1989/strategy-segment-of-one-marketing
5. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf

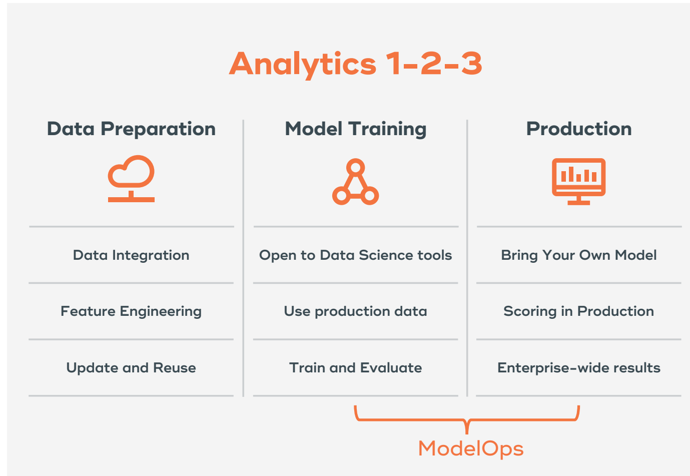
Automation needed

Collectively these issues contribute not only to the 80%[2] failure rate for deployment of predictive analytics models but to the 25% of time data scientists lose[6] in the manual back and forth interactions with IT and DevOps teams just trying to get models deployed. These figures are averages, and while companies will do better, many will do less well. Our experience suggests that in the real-world these figures may significantly underestimate the issues. As the demand for predictive AI models that drive value in data-driven organizations is only going to increase it is clear that new approaches are necessary. ModelOps offers a new approach to managing the interactions needed to facilitate, monitor and control moving analytics from the lab to the business. It provides a methodology and a toolkit to assist with orchestration, automation and governance of the critical steps between training a model and seeing it used to add value at scale in the business.

> **“Only a small fraction of real-world ML systems is composed of the ML code. The required surrounding infrastructure is vast and complex.”[7]**

Vantage ModelOps: A methodology and a toolkit for Vantage

Vantage ModelOps is extending model management capabilities in Vantage with a methodology and a toolkit. It bridges stages two and three of the Analytics 123 Strategy, helping to move models from experimentation to production. Many of the barriers to making this transition are created by a perspective that over-focuses on the machine learning code that sits in the middle of a wider process. Although essential, the code in the model is a small fraction of a wide and complex ecosystem. All the elements must work together to successfully operationalize predictive analytics. Teradata as an Enterprise Feature Store (EFS)[8] brings together many of the ‘upstream’ elements that form the inputs to model building and testing. ModelOps does the same for many of the downstream elements that ensure a model is deployed effectively at scale. As such, it is in essence a joint venture between data science teams, software engineering and IT, as well as business operations.



Meeting the needs of data science, business and IT teams

Data scientists want to get more of their models into production and scoring live data faster. Business leaders want significantly shorter time to value from analytics projects, and IT teams want to simplify infrastructure whilst maximizing return on existing investments. Key questions include, how do we run this; where do we run this; and, how do I serve it and monitor it? All get different answers at different times and from different parts of the organization. Without a consistent, agreed methodology for moving trained models into production the journey becomes disconnected, slow and frustrating.

A complementary extension

The ModelOps methodology represents a standardized series of steps that are transparent to all parties, supported by a software development toolkit and a modern user interface that automates many of the key steps. It links data science practices of

6. https://info.algorithmia.com/hubfs/2019/Whitepapers/The-State-of-Enterprise-ML-2020/Algorithmia_2020_State_of_Enterprise_ML.pdf
7. https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf
8. https://www.teradata.com/Resources/White-Papers/Efficiency-Productivity-and-Speed-to-Deployment

ideation, discovery and exploration with software engineering practices including governance, security and maintenance to create scalable deployments designed for enterprise roll-out. It unifies the different approaches, tools and languages used across the business to provide a consistent, auditable way of deploying predictive AI at scale.

For data scientists it is a way to get fast scalable scoring for their models, using production data at scale to deliver business critical insights in real time. For IT, ModelOps helps maximize investment in Teradata by combining a centralized location to deploy and score AI models, with the visibility and control to govern exactly which models are doing what with which data. Orchestration, automation and governance are built-in to ModelOps and transparency starts with the registration of a new model for deployment.

Faster Deployment

Vantage ModelOps is a ready-to go solution for model deployment and management – either directly in database, or in specialist, discrete systems and infrastructures. The ModelOps simplifies the steps from model registration to ultimate deployment whether in Teradata Vantage databases or elsewhere.

Specific user personas can be established for data scientists, data science managers, deployment engineers and anyone that needs a clear and transparent overview of the status of the projects and models which they have permission to view. The tool automates the key steps assisting users as they move their models through a lifecycle from registration to retirement. The process is configurable, but key automated steps include:

Registration

- New models definitions in any language or framework are checked in through either an intuitive user interface or command line.
- Models can be added using templates established by the organization to codify organizational practices. They offer pre-configured structures that are simple to populate with required information aiding standardization and repeatability.
- Templates can be set-up for multiple languages and frameworks to support the way the organization or even individual departments or groups work.

Training

- Data scientists are prompted to select a dataset from the enterprise feature store via a drop-down menu for training.
- They can tune hyper-parameters for each training and have these all recorded and documented automatically along with details of the dataset that was used and the CPU and memory resources required.
- The tool automates the training of the model and creates a new version once it has completed.

Evaluation

- The next step is to evaluate the trained model performance by comparing the predictions on an evaluation dataset with the real ground truth values.
- Data scientists can choose the metrics they want to use in evaluating the model and select a number of key metrics for ease of comparison.
- The tool automatically records user-defined charts which can include ROC curves, confusion matrices and performance metrics such as accuracy, recall and any other measures defined by the data scientist. It is also possible to identify which features, from the Enterprise Feature Store were the most important for the model. Successive evaluations of the model are all recorded and can be visualized over time to understand model drift and provide explainability.

Compare and Approve

- Data scientists can make side-by-side comparisons of any versions of their model, arranging metrics and performance data in order to make decisions about which to select for approval.
- Using whichever metrics and charts the data scientist prefers the ModelOps supports detailed explainability for every model. Feature importance charts help to interpret why a model is performing better or worse along with potential biases.
- Data scientists can also automate the comparison and documentation of challenger models with the tool. New model versions can be evaluated and compared to existing models within the dashboard. Based on the full range of metrics and charts selected, decisions can be made and clearly explained as to whether to promote new champion models.
- Detailed reports that support approval processes, can be produced for internal audits and regulatory compliance purposes.

Deploy

- Once a model is approved it can be deployed through the tool in one of three ways.
- Models can be deployed behind an API for RESTful application, with a scheduler for batch processes, or directly in the Vantage database. Selection and single-click deployment see the model go into production.
- Models can be deployed in isolated containers for portability either on premises, at the edge or in the cloud.
- A single model can be deployed in multiple ways and different versions of the model can be deployed in parallel to support A/B testing and other forms of control group or randomized testing. The tool remains a single point of reference for governance and monitoring (see below).

As outlined in our Bring your own model (BYOM)[9] white paper data scientists are free to use their preferred languages and approaches for project creation, exploration, discovery and model building and training. The only difference is that data scientists, rather than registering and then training a model using the tool, simply import their pre-trained model and go straight to the Evaluation step. From here on the approval and deployment process is the same.

Why deploy in Vantage?

Both data science and IT teams want to see more efficient and effective use of predictive models. Data scientists want their models to be as performant as possible and used frequently to add real value to the business. IT teams want to support this without needing to resort to buying and supporting new systems. Most data science tools do not include databases, and those that do often cannot scale to score data at an enterprise level, remaining essentially departmental solutions. As a result, many data scientists have overlooked the advantages of deploying models directly in Vantage. IT teams have been cautious in allowing access to models they had no visibility or control over, and data science teams were unaware of the ease of deploying their models in this way.

9. https://www.teradata.com/Resources/White-Papers/Bring-Your-Own-Model

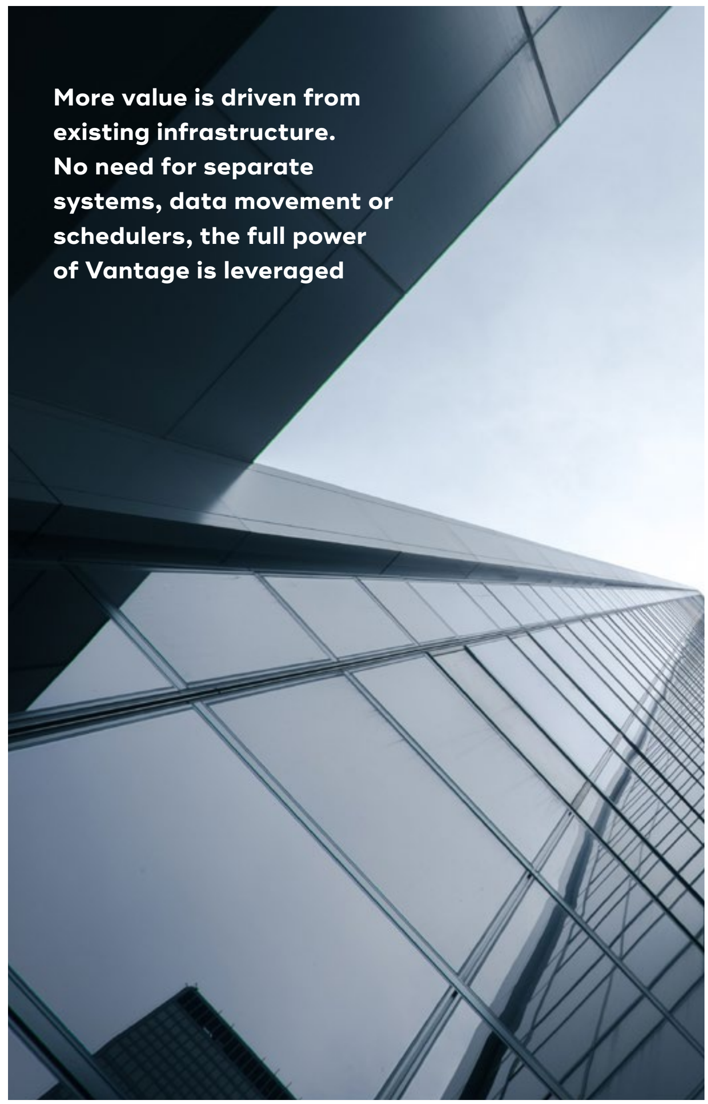
However, data scientists are increasingly considering scoring models in database. Our own BYOM approach offers data scientists the best of both worlds, the ability to build and train models using their favorite languages and libraries and then seamlessly import them to run in Vantage. As others have noted "The motivation is not that inference will perform better inside the database, but that the database is the best place to take advantage of enterprise features (transactions, security, auditing, HA, and so on)."[10]

The following describes a typical reporting process illustrating the significant advantages of deploying models directly in Vantage databases. A business lead requires a weekly forecasting report to inform purchasing decisions. The data science team has built a performant model to do this for her. The model is approved for use in production. The current workflow includes, packaging the model to run in a dedicated system; scheduling the job with data extracted from the data warehouse; executing the job and then copying the scored data back to the data warehouse from where it can be used to create a business intelligence report using Tableau or similar. It is a resource intensive operation that provides a snap-shot report that is out of date as soon as it is produced.

Using ModelOps to deploy the model directly to Vantage instead creates numerous advantages.

- Deployment is a one-step process. The model is simply published to the database, scoring live data that can feed directly into BI applications like Tableau.
- Reporting becomes real-time. Scoring live data in Vantage means that using the same model reports are always up to date. Static reports can be replaced with live dashboards.
- Any tool that can connect to the data warehouse can score data by deploying models in Vantage in this way.
- More value is driven from existing infrastructure. No need for separate systems, data moves or schedulers, the full power of Vantage is leveraged.
- Built for speed and scale, Vantage can score enterprise datasets. Full data sets of millions of rows with thousands of models in parallel can be used for insights.

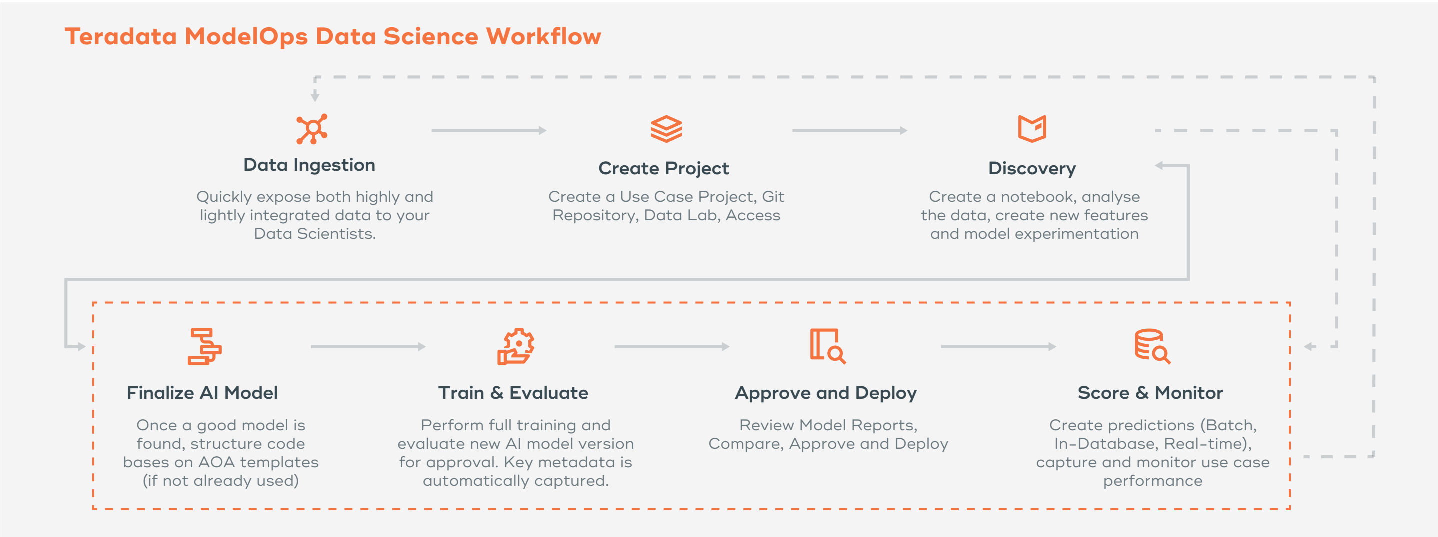
Most models can be run in Vantage with no additional work from either the data science team or the IT team. ModelOps provides a simple user interface that steps through the entire processes and publishes approved models to the database with one click. See the BYOM whitepaper for more details.



10. https://blog.acolyer.org/2020/02/21/extending-relational-query-processing/

Managing the Model lifecycle

The value of ModelOps does not end with the deployment of models and scoring against production data. Although streamlining the process of training, evaluating, approving and deploying models and making it repeatable is significant, it is also important to improve visibility and management of models once they are in use.



Model lifecycle management

As analytics become more important and as organizations seek to deploy more models more frequently, standardizing and automating the entire lifecycle will become essential. Managing a few hundred models at different stages is already hard, consistently transitioning thousands of models from testing through all the necessary stages to production will quickly become impossible. ModelOps presents one simple interface through which the whole process can be automated, governed and monitored. Different personas with different levels of access will see the information relevant and important to them through a consistent, clear interface.

Governance

Effective governance from end to end is already essential in many regulated industries and will quickly become necessary for all businesses. As more models are responsible for insights and decisions that impact all areas of the business it is essential that audits of their complete linage are easy to undertake. Explainability is fundamental to the widespread use of AI in all businesses. Regulations including GDPR but also encompassing specific regulations relevant to different industries, as well as the demand for transparent business operations in all sectors, mean that AI cannot operate as a black box. It must be possible to show exactly how each, and every decision was made and for those explanations to be easily auditable in many cases for years after the decision was made. This requires clear documentation not only of how the model was built, but what data was used to test and evaluate it.

Governance starts with the automated documentation and evaluation of every version of every model in the ModelOps methodology. The simple user interface shows when models were trained, using which data and by whom. Who evaluated, what were the results and who approved and deployed, exactly where and when. All this information is captured and stored automatically providing detailed audit chains for every model.

As models are used to score production data, further information on their performance and drift is added to the interface so that decisions can be made on their continued use or retirement. All are documented for governance purposes.

Monitoring

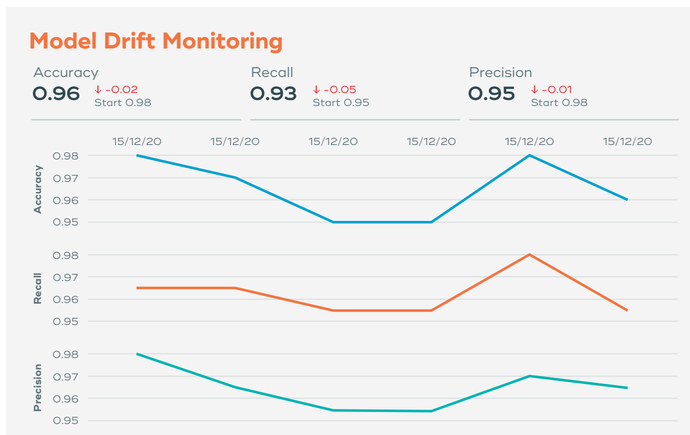
Monitoring itself is a crucial aspect of ModelOps, and one that is equally relevant to all the main audiences. It gives IT and data teams visibility over exactly what models are running in their systems, what they are doing and what data they are consuming. This provides several advantages. Not only can IT and data management teams act to stop or remove models that are performing in unexpected ways, but this visibility and control can give them confidence to allow more models to score in database. As noted above not only can this deliver better, real-time outputs for models, but can save money by reducing the need to provision additional resources, scheduling applications and data movements. In short, IT can maximise return on existing investments and reduce the complexity of systems to manage.

For data scientists and deployment engineers ongoing monitoring identifies when model prediction accuracy degrades (drift) and need retraining, increasing agility to changing environments. Historically this has been hard to do as data scientists had few windows through which to observe models scoring in production. The ModelOps user interface establishes a single view of all models at whatever stage in their lifecycle. It also presents opportunities to introduce competing challenger models that may deliver better results. Regular challenges and retirement or replacement of models ensures that the organizations is always deploying the best possible predictive models to support the business.

Model Catalogue

The monitoring and governance capabilities of ModelOps come together to support an enterprise-wide model catalogue. This is a single, unified and standardised library of all the models, at all stages of development across the business. It delivers significant advantages as organizations look to increase repeatability and speed of analytic projects.

- Democratization of AI – A single repository of models, managed by permission-based access, will help all parts of the business to find, utilise and build on analytic work done elsewhere. We estimate, from work with our customers that 50% of data in any project repeated in more than five other projects. So, it is clear that enhanced sharing and repurposing of analytic projects could dramatically increase efficiency. The Model Catalogue is an important step in increasing repeatability.
- Definition of a single methodology and infrastructure – the ModelOps methodology makes it easy to see, understand and replicate work for efficiency, consistency and auditability, irrespective of the tools and processes used to build and train models.
- Security – all of this is delivered via secure dashboards that only allow authorised access. For example, the marketing analytics team at a bank will not have access to models created by the fraud prevention team.



ModelOps – streamlining deployment for faster time to value

Significant global trends are redefining how organizations build, deploy and use analytics. The ongoing digitalization of all sectors means that data is not only an asset, but the heart of differentiation and success. Aligned with this, is the need to understand and effectively target segments of one and offer personalized responsive service. Data science teams are being driven to produce more highly performant predictive models at a faster rate.

Models that are tailored to 'segments of one' are increasingly core to the way organizations and their data scientists want to deploy AI. Training many

teradata.

small models on subgroups or partitions of data, rather than one huge model on all of your data, often results in better accuracy. It's possible and in many cases desirable to have an individual model for every customer, scoring just that customer's data. Managing models for segments of one naturally represents an order of magnitude increase in output. Customers are experimenting with predictive models for over 10,000 products, others are looking at potentially modelling every one of millions of customers. One customer in the US now maintains over 40,000 models whilst another is considering modelling each product in every store, globally over 2.5 million models! Clearly it is no longer tenable to have less than 20%[2] of models make it through to production or for each model to take 5 months to get there!

ModelOps provides a solution for all parties. Scheduling training and scoring of millions of models is easy in Vantage, taking advantage of Teradata's massive parallelism to train and score hundreds of models on millions of rows of data in seconds.

But ModelOps goes further. Quickly and easily deployed in the cloud, on premise or as a hybrid implementation, it establishes a common methodology that increases efficiency, repeatability and governance of the crucial steps between building a good model and having it deliver value to the business. Wide ranging integrations into existing ecosystems from containers for deployment to GIT for code repositories, RDBMS for metadata and more make it easy to apply ModelOps to specific customer requirements.

ModelOps gives data scientists and deployment engineers a clean and intuitive user interface (or a command line interface if they prefer) though which they can review and manage all their models. The lifecycle tool guides them through consistent steps that automate the key phases to get their models into production quickly and easily. It manages projects, datasets, and models in one place.

Security, plus an enterprise model catalogue, makes ModelOps entirely suitable for deployment across the whole business and not just one department. Consistency, repeatability, and transparency combine with scalability and performance to create an enterprise approach that complements the different data science tools and frameworks already in use across organizations. In addition to streamlining the deployment and management of models in production ModelOps can help data science and IT teams get the most out of Vantage. Leveraging its speed and scalability, models can become more performant. Scoring live data, they can transform reporting tools into live dashboards and deliver additional value to the business.

Ultimately ModelOps helps data scientists get more of their models into production, adding more value to the business faster.

About Teradata

Teradata leverages all the data, al the time, so you can analyze anything, deploy anywhere, and deliver analytics that matter. By providing answers to the complexity, cost and inadequacy of today's analytics, Teradata is transforming how businesses work and people live. Get the answer at Teradata.com.

About the Authors

Dr. Will Fleury is the Vantage ModelOps Product Owner

Pablo Escobar de la Oliva is the Vantage ModelOps Product Manager

Dr Chris Hillman is Director of Data Science EMEA