

End-to-End GPU-Accelerated Machine Learning Pipeline with Kinetica and RAPIDS

Enterprises looking to incorporate machine learning into their active analytical applications need to enable data scientists to explore the entire data corpus interactively to develop better insights and improve model accuracy. Organizations also need to accelerate model training to shorten model development cycles, and integrate models into their active analytical applications, while maintaining corporate governance.

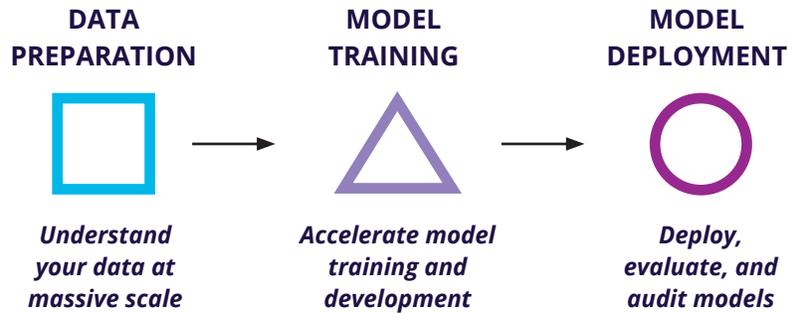
Kinetica's integration with RAPIDS, the popular GPU-accelerated open source library for data science supported by NVIDIA, dramatically accelerates the machine learning pipeline. This integration can help organizations dealing with huge training data sets and/or large volumes of inbound production data for inference or those that simply wish to run experiments faster and at a lower cost thanks to the

massive parallel processing power of the GPU. Together, Kinetica and RAPIDS make it easier for every company or government agency to maximize the benefits of machine learning.

The Kinetica Active Analytics Platform delivers at-scale data exploration and model deployment which, when combined with RAPIDS for accelerated model training, enables end-to-end NVIDIA GPU-accelerated data science and operational deployment. Kinetica and NVIDIA customers can run GPU-accelerated machine learning models on data as it streams into the Kinetica platform, as well as develop and inference on models of their own. Business decision-makers can simultaneously implement models and track and evaluate outcomes, with features available in Kinetica to support governance, transparency, and repeatability.

Enterprises looking to incorporate machine learning into their analytical applications need to:

- Enable data scientists to explore the entire data corpus (not a snapshot) interactively, to develop better insights and improve model accuracy
- Accelerate model training, to shorten model development cycles
- Integrate models into their active analytical applications while maintaining corporate governance



Kinetica and RAPIDS dramatically accelerate the entire machine learning pipeline. Key capabilities of the integrated solution include:

- **Managed Jupyter Notebooks:** Located within the Kinetica Active Analytics Workbench, they are pre-integrated with RAPIDS, offering a multi-user, GPU-accelerated data science environment.
- **Interactive Data Exploration:** Kinetica's GPU-accelerated OLAP capabilities, which include tiered storage, enable data scientists to explore the entire data corpus interactively. In cases where location analysis is required, data scientists can visualize data geospatially right within the Jupyter notebook.
- **Large-Scale Feature Transformations:** Data scientists can use the power of the Kinetica Active Analytics platform to perform large-scale feature transformations at scale, be it normalizations, filters, joins, or complex geospatial functions.
- **GPU-Accelerated Training:** Once feature transformations are complete, data can be seamlessly transferred to RAPIDS for training via a Python API call, right on the GPU.
- **Model Deployment and Inference:** Kinetica automates model deployment and GPU-accelerated inference – in continuous, on-demand, or batch modes. No need to worry about deployment, network configuration, or scaling. Once deployed,

Kinetica automatically orchestrates the full analytical pipeline – from ingest to database to model and back to database and downstream applications.

- **Audit:** Kinetica tracks the full data lineage, including raw data, feature transformations, and model output for all of the model deployments. Easy-to-use search tool provides an instant ability to do a full model audit or find a “needle in a haystack” for a specific inference.

These new machine learning capabilities can enable organizations to develop more accurate models faster. Kinetica's integration with RAPIDS empowers data scientists and brings them to the center of the modern enterprise experience where they're able to access operational data and drive intelligence throughout their organization.

With Kinetica and RAPIDS, chief data officers have the tools to impose data-driven decision-making across the organization, in an accessible and fully integrated manner, from data scientist to developer to business intelligence analyst. Data scientists can dramatically streamline the development and improve the accuracy of their models, by GPU-accelerating their entire pipeline. Data engineers and developers can seamlessly integrate the models into their analytical applications, while ensuring corporate governance. Developers deploying the models no longer need to rewrite them to put them into production, with limited transparency into the process throughout the organization.