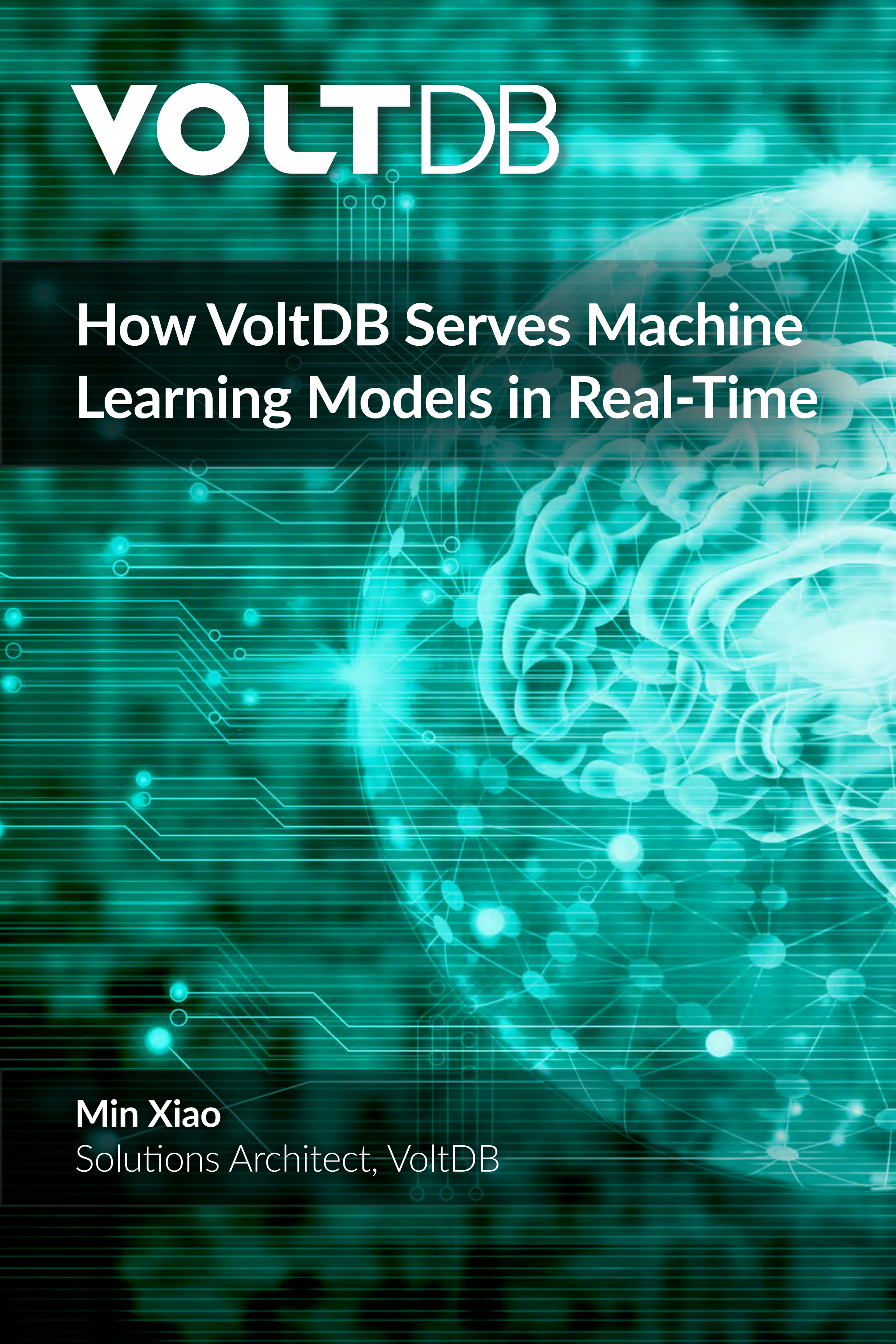
VOLTDB



How VoltDB Serves Machine Learning Models in Real-Time

Min Xiao
Solutions Architect, VoltDB

Table of Contents

Introduction

In my previous blog, we discussed how to apply your big data analytics to real-time applications. The idea is that, if you have built analytics on your data, the next step is to use the analytics directly in the applications to automate your business workflow and remove the manual components. In this blog, we will use VoltDB's application in machine learning as another example to show how VoltDB's real-time processing capabilities can help take your business to new heights.

The Deployment of Machine Learning Models

As we are all aware, machine learning has transformed many companies in how they leverage data as a part of their daily business activities. Historical business data is collected to train AI algorithms to better understand and develop business rules representative of the data. When new data comes in, they pass through the machine learning models and results are computed based on the models. Essentially, the process is learning and applying business logic to drive business activity in an automated fashion. It is a significant step closer to business automation, vs. the traditional approach, e.g. analyzing the BI reports and then asking human to take actions.

An Example of using Machine Learning

Many insurance companies have used machine learning to automate their underwriting system to determine the risk and premium based on a prospect's personal data, historical data and medical condition data (for life insurance).

With the new underwriting application, companies can expect:

- **Better User Experience**: Non-technical users (like underwriters) can author, test and maintain underwriting rules.

- **Improved Automated Processing**: They often achieve a much higher percentage of automation, such as >90%, of the underwriting process.
- **Cost Saving**: Significant reduction in operation costs, upwards of 70%, in savings.

Complete Rewrite in the Deployment

Though it sounds simple to deploy a machine learning strategy, there is significant development work that needs to be done when implementing the models in production — often complete rewrites of the machine learning models must be done to meet the production requirements of the application, for two primary reasons:

1. **Different Programming Languages**: Most tools that are used to build machine learning models are written in Python or R. However, Python and R are not always the languages that the application is built on. As a result, development teams need to rewrite the models and algorithms in Java/C++.
2. **Designed for Different Purposes**: Most of the tools focus on training the models with high accuracy or to automate the model selection. These tools do not focus on providing high performance or support to a large number of concurrent users during the deployment.

Ideal Platforms To Serve Machine Learning Models

As we have discussed, deploying machine learning models often involves a complete rewrite when deploying onto an application. So what are the needs of the platforms for a successful deployment of your machine learning strategy? Ideally the platform should be able to:

- **Be Interoperable**: Thank Predictive Model Markup Language (PMML), the Portable Format for Analytics (PFA) and other rising deployment standards — they make the machine learning models interoperable among many platforms. Machine learning models built in one platform can be ingested into other platforms to simplify the deployments of the initial models as well the updates.

- **Import Native Libraries**: More and more platforms, such as H2O and mlpack, can import native Java or C++ machine learning, feature extraction and data preparation libraries built on other platforms.

- **Support All Workloads**: Including both application-generated inquiries, which are often highly concurrent, as well as user-issued ad hoc inquiries, usually requiring low latency.

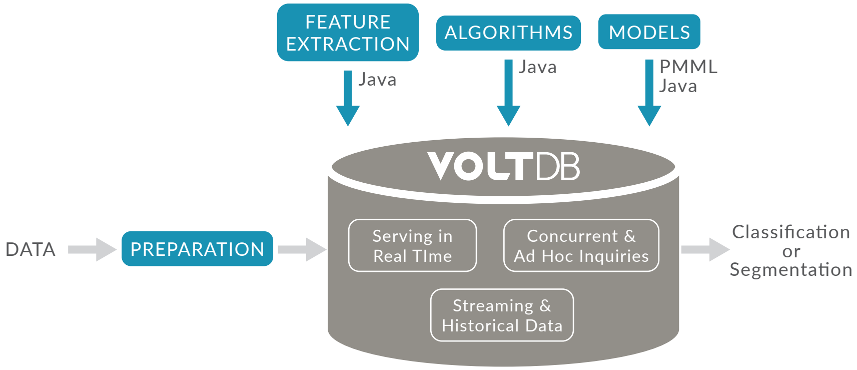
- **Handle Production Complexity**: Run simple and complex business rules for both streaming and historical data, and also handle the updates of the data without complicating the applications.

How VoltDB Helps

VoltDB is an in-memory database that is designed to process highly concurrent queries with low latency. It helps empower the serving of machine learning models in the following ways.

- **Easy to Deploy**: VoltDB can ingest existing machine learning models through PMML or PFA standards. It can also import native Java libraries of machine learning models and algorithms, as well as the Java modules built for data preparation and feature extraction.
- **Decision Making In Milliseconds**: VoltDB provides very low latency of single digits of milliseconds while being able to ingest high throughput performance in the millions of transactions per second. These capabilities will enable VoltDB to serve the machine learning models in real-time and support both application-generated as well as user ad hoc inquiries.
- **Enterprise Production Ready**: VoltDB supports the highest level of ACID compliance — serializable isolation — to ensure there is no data loss with minimal programming within the application. When you have a hardware failure or network hiccup, VoltDB's built-in high availability brings the system back in the shortest time possible.

Example Architecture



The diagram above illustrates an example architecture that shows essential machine learning modules (including feature extraction), algorithms and trained models that are built on separate platforms, and then are ingested and deployed in VoltDB through java and PMML.

Conclusion

Machine learning is a method of data analysis that automates analytical model building. When deploying the machine learning models in VoltDB, the models can be served in real-time to handle continuous customer application interactively with minimal programming within the application.

Just as we discussed with the underwriting applications in this blog, the ability to serve machine learning models within VoltDB can produce immediate decisions making capabilities on your data in both your customer facing and internal applications. Because of this, you are able to maximize the value of your data and take the next step in your data strategy. By automating your business workflows and removing the manual components of the process, your business will operate with greater efficiency, with the ability to operationalize your analytics and capitalizing on the opportunities that afford your business!

Feel free to check out our Technical Overview highlighting the capabilities of VoltDB, and learn more about how market leaders such as Huawei, Openet, and FT.com are using VoltDB to boost their business.

About the Author

Min Xiao, Solutions Architect, VoltDB

Min Xiao is seasoned executive with over twenty years of experience in building high-performing sales and dev teams, forming a bridge between customers and product development, and solving customer problems with repeatable product solutions.

Min's past experience includes early roles at Vertica (employee #10) and Tamr (employee #5). Outside of his role with VoltDB, Min also serves as consultant for NextMile Consulting — a company in which he also founded.