VOLTDB

Six Use Cases of **Real-Time Decisioning**

Table of Contents

Introduction

As humans, we are hardwired to want things now. Instant gratification impulses are a powerful force, and the current real-time economy reflects this. Transactions between business parties are digitized, increasingly automated (M2M communication, IoT, Artificial Intelligence, and Machine Learning) and completed in real-time.

According to the Psychology of Waiting survey conducted by VoltDB, when asked "what does real-time mean," over 70% respondents said "immediately".

The real time economy fueled by the need for instant gratification is leading a big surge in vendors trying to create a new breed of applications that use real-time decisioning. Next generation in-memory database technology enables real-time decisioning, allowing businesses or organizations to instantly derive value from vast volumes of data.

This eBook will provide six use cases for real-time decisioning showcasing real-world examples of how businesses are making the most of their high-velocity data.

Leveraging Machine Learning for Real-Time Fraud Prevention

Fraud is everywhere and ever present. According to the LexisNexis True Cost of Fraud 2016, "the average volume and value of fraudulent transactions has risen sharply since last year. The level of fraud as a percentage of revenues has also inched upwards."

There is every reason to believe this trend will continue or accelerate given the status quo. This is made worse by the growing number of transactions on the edge, an increase in the number of card swipes/minute or second, and an increase in mobile and e-commerce fraud.

According to Javelin Strategy & Research's E-Commerce Payment Acceptance Optimization Study, commissioned by Barclaycard, more merchants are focused on acceptance than on prevention. Focusing on acceptance leads to increased revenue, but it also leads to fraud repayment and reputation losses. However, when you focus on fraud, it can lead to problematic false positives. Making correct decisions, and reducing false positives, is crucial for both attracting and keeping customers and merchants.

But throwing more resources at the problem is not good enough. More fraud attempts are successful, and manual fraud detection is getting overwhelmed by the increased number of transactions, even with more money being funneled into detection efforts. The reality is that the costs of traditional fraud detection measures are rapidly inflating, while, at the same time, fraudsters are constantly innovating to stay ahead of countermeasures and detection. Underlying this struggle between fraudsters and companies is one simple fact: fraud detection is just fixing it after the fact, not preventing it in the first place. Therefore, as time goes on, fraud detection efforts have less and less actual value.

Changing the Game From Detection to Prevention

By changing your fraud system from a post-transaction detection system to a proper in-transaction prevention system, you can reduce operating costs, reduce false positives, and stop fraud as it happens.

When choosing a fraud solution, or the underlying technology for one, it's important to keep some key requirements in mind. For in-transaction processing, you need technology that is both fast and scalable, preferably on commodity hardware. You need to process thousands of card swipes, NFV taps, and online payments per second, with a time budget of only milliseconds for fraud detection.

In addition, you need to make sure your fraud solution is robust and resilient. On top of existing regulations regarding financial data, you need to make sure that when your system is making a decision, it's based on the correct data. As transaction tempo increases, eventually consistent data is not quick enough; you need immediate consistency. In addition, losing data due to nodes going down is not acceptable.

Huawei

Huawei is a leading global information and communications technology solution provider as well as a consumer electronics giant. With their FusionInsight platform, powered by VoltDB, they providing fraud analysis for leading banks.

Huawei chose VoltDB because they needed a system that could perform thousands of queries per financial transaction with low latency (<50ms) and high throughput (>10k tps). Huawei uses VoltDB to monitor 10 thousand complex transactions per second with 99.99% of transactions finishing in less than 50ms. Huawei applies hundreds of rules and scoring checks to each transaction in milliseconds, moving fraud detection from weeks to real-time transactions.

Huawei and VoltDB provided 10x better performance than traditional fraud detection and resulted in more than a 50% reduction in fraud cases and over $15M/year saved from fraud loss.

Enhancing Customer Experiences with Hyper-Personalization

In general, people want personalized products, services, and experiences. It's one of the big use cases driving the development and interest in machine learning. Customers who are engaged with personalized content or experiences are more likely to make a personal connection with that product or service, and continue to use it.

Personalization has been around for a while, but the requirements for successful personalization have become more stringent. Companies increasingly need to deliver more relevant, more personalized customer experiences in short amounts of time. To meet modern and future demands, any personalization must address specifically the needs or tastes of each customer in real-time.

Increasing Challenges of Personalization

As people grow accustomed to current levels of personalization, they will demand higher levels of personalization. Practically, this means handling more data in shorter amounts of time. A real-time application or service must predictably and actively engage customers with highly personalized experiences.

For example, instead of taking a look at what articles a user has read recently and making suggestions based on those, a hyper-personalization system could take those, as well as location history, shared social media information, and more to make the best recommendation possible — for hundreds of thousands users at once.

With this increase in personalization scale and speed, it's easy to miss an important component: accuracy.

FT.com

By providing fast access to user's behavioral data, VoltDB enables FT.com to alter each user's experience. They store various pieces of user event data in VoltDB and use this to make changes to different pieces of the site.

For example, for users with ad blockers, FT.com can expose this data via VoltDB and show different UI components on the site. Another good example is providing personalized recommendations to users. The myFT feature of FT.com allows users to curate their own version of FT.com to include topics that they're most interested in.

For new users, it can be tricky to provide relevant recommendations, as they're not following anything, which makes doing any type of collaborative filtering hard. By storing user page views in VoltDB, the myFT team have used data to make some recommendations on what topics users might want to follow based on the articles they have been reading.

While using VoltDB to power real-time personalization, FT.com decided to also use VoltDB to power their headline A/B testing. Each headline is posted for about an hour, and the system determines the most popular one and uses it for future readers. In addition, the data stored for deeper analytics on disk using redshift. While A/B testing has been a mainstay for business analytics, with a large volume of articles and readers, real-time A/B testing was the best way to ensure FT.com content gets to the largest audience possible.

In addition to UI changes, recommendations, and A/B testing, VoltDB is used by FT.com in their new marketing and journey automation project, named Envoy. This new internal product allows teams to build automated journeys for users and have control over various messaging touch points. Previously, there was no central control over how they messaged users, not only on site, but across email, SMS, phone, and more. Envoy handles all sorts of messaging, including non-payment, advertising, and personalized recommendations.

Envoy provides this functionality and is hooked into VoltDB in a couple of ways. The data that flows into Envoy via VoltDB allows FT.com to enrich

events passing through the system with the extra user data that was held in VoltDB. If, for example, during a specific user's journey, that specific user needs to see an on-site message, this information is pushed into VoltDB to allow it to be exposed in a similar way to the previously mentioned onsite UI changes.

By leveraging VoltDB, FT.com is able to engage, retain, and gain hundreds of thousands of subscribers and even more readers with personalized content, messaging, and journeys. Envoy has proven successful with a 25% message click-through rate. With an increase in payment recovery by 4 percentage points and millions of monthly unique users, FT.com has seen a four-times return on investment. "Real-time decisioning," says one FT.com representative, "provides 5% revenue that we would not see without VoltDB."

Increasing Player Retention and ARPPU with Real-Time Analytics

Gaming is the undisputed king of mobile applications. Even though games only accounted for 40% of the total mobile app downloads for 2017, a total of 80% of app spending in 2017 was generated by mobile games. Mobile game spending even out performed PC gaming by 2.3x AND console gaming by 3.6x last year. While these numbers suggest a gold rush for mobile developers, things are not as profitable as they may seem.

This explosive growth in mobile gaming means that average revenue per paying user (ARPPU) is growing year over year. However, the top performing games, such as Candy Crush and Toon Blast, are taking the largest piece of the pie. According to the 2016-2017 GameAnalytics Mobile Gaming Benchmark, from January through September 2017, the top 16% of mobile games have increased ARPPU by $15/user, totalling an overall ARPPU of $50 — which is impressive success. However, there is a significant gap between these top performing games and the average game, which is stuck at an ARPPU of $7 (a gap of around $43). And not only are average performing games seeing minimal growth overall, but the gap is worsening, not improving.

So it's no surprise that retention rates are showing a similar story. Strong retention benchmarks should be a least 15% for day 7 (D7) retention. However, the data is showing that even the most successful games are struggling to meet these percentages, as the top 16% of games still seem to fall short at 12%. And for the majority of games, this proves to be even more challenging; with retention rates hovering around 4%, they more closely resemble the poorest performing games than the top 16%.

Common Monetization Strategies and Their Limits

Regardless of how successful a game is, there are a number of common monetization strategies that are familiar and implemented by most gaming companies. The most obvious approach is to get more users — to do that, companies are increasing their UA spend with the thought being that more users will drive more money. However, with a cost per install (CPI) of $2 for iOS, an ARPU of $0.95 isn't going to cut it. Even though paid UA is seeing 4x the success of organic UA, the average game is still only generating half of their CPI back in the first month of play.

In order to not only see a positive ROI for UA spend, but also generate revenue on top of that, companies taking this approach will need to maintain competitive retention rates for D30 and on. But with dismal D7 retention rates, the average game can't survive on this strategy alone.

Many teams have also leveraged more data-driven monetization strategies. Examples of these typically focus on players with the highest LTV to drive revenue through more targeted ads and paid premium offers. However, these strategies also rely heavily on successful retention rates; if a player never comes back, they'll never spend. And with D1 retention rates struggling to get past 50%, that's a significant amount of opportunity lost.

Beating Retention Loss

If you could use the data that you already have to drive retention rates instead, you'd be able to create new opportunities for monetization (rather than focusing on strategies that are compensating for low retention rates). But you'll need to act quickly: an average game's session length is

only around 5 minutes, so in-game decisions need to be made faster than ever before.

To successfully achieve this, you'd need to be able to take in data on a per player basis, analyze the information, apply any existing context that you have on the player, and make a decision on which next best action will be most successful — and you need to do all of this while that player is still engaged.

By positively influencing players at the right moment and within their current session, you can create a more personalized experience on a per player basis that will increase engagement and drive long-term retention. An example that we've seen our customers implement with adaptive gameplay, is being able to fine-tune difficulty on a per player basis. This ensures that each session has the optimal balance of success — not boring and challenging but not frustrating.

This capability is also being applied to in-app offers, where not only is the timing important, but so is providing the right offer. Repetitive and irrelevant offers can negatively impact UX (especially for first time users) and are often ignored. However, an offer that is presented when the player has the greatest motivation to buy — like when a player is stuck on a level or has failed a number of times in a row — will have a greater acceptance rate. And players at that point in play will be more willing to do things like watch an ad to gain a competitive advantage.

King

With VoltDB, King.com got a boost in in-game purchases of 30%, had 7 quarters of growth in a row, and increased the average session to 35 minutes per day. Proven contextual offers that are presented in real-time (within milliseconds) have a higher acceptance rate (in some cases as high as 257% more) than when compared to offers that are delayed or presented at the next session.

DeltaDNA

Game developers and publishers produce an enormous amount of data, and need a system that can both perform in-depth player analysis on the data streaming from their games and take action within a quarter of a millisecond.

DeltaDNA uses the VoltDB database to run its real-time game analytics and marketing platform to meet the need of these developers and publishers. Using DeltaDNA, customers are able to track, anticipate, and respond to player behavior in real-time.

DeltaDNA used VoltDB to help customers increase player retention by 200% and revenues by 40% by providing micro-personalization of each player's experience.

Meeting the Data Demands of 5G and Next-Generation Telco Networks

With the ever-nearing move to 5G and the impending explosion of industrial IoT use cases, data is going to play a very important role in the quality of service assurance and security. In order to be competitive in this industry, telcos need extremely fast and reliable access to their data. They use this real-time access to provide software for major carriers that support critical services for operations support systems (OSS) and business support systems (BSS).

As the carriers continue to evolve, they are creating new services and revenue streams, improving their services, and reducing costs. One of the ways they are doing this is by moving to Network Function Virtualization (NFV) and Software Defined Networks (SDN). Both of these allow telcos to be agile and respond as quick as their customers — even when it's a device. However, to successfully implement SDNs and NFV, telcos need a real-time platform.

Similarly, the evolution of policy and charging rules function (PCRF) from 4G to 5G involves millions of devices, millions of network slices and big operational

decisions. CSPs' 5G plans will get underway this year with the ultimate goal of getting to Phase 2 of 5G in the future. The goal is to get a 5G deployment where every layer of the communications stack is virtualized with far greater agility than ever before. This is a lofty, but achievable, goal.

However, getting to Phase 2 will involve an expanded role for PCRF to support many more functions ranging from network slicing to fixed mobile convergence, dynamic backhaul policies for industrial IoT use cases and assured quality of service across many different types of use cases. In addition, policies will need to be able to be adjusted dynamically by both operators and their subscribers in order to provide the flexibility needed by new services. Further to this, policies for security and deep packet inspection will be required in addition to policies for OTT partnerships and for successful multi-party services.

Openet

Openet selected VoltDB as the backbone of their OSS/BSS solutions, because they needed a cloud-deployable, transactional database that can flexibly handle high-volume data streams for service providers to monitor and leverage in real time. VoltDB provides the performance of in-memory, the scalability of NoSQL, and the transactional consistency of traditional relational databases.

Traditional database systems were simply too slow to ingest data, analyze it in real-time, and make decisions at the rate required. With VoltDB, Openet now offers transactional, database-oriented applications against data feeds that were previously limited to stream processing methods because of scale.

Since Openet solutions are always inline in a service provider's call path, Openet required latencies for its transactions to be less than 20 milliseconds, so performance and scalability were major requirements.

"Our solutions are primarily deployed by Tier 1 and Tier 2 operators worldwide, so we need a virtualized database platform that can provide elasticity while supporting ease of operations," said Oisin Loftus, former VP of Global Product Development at Openet. "VoltDB not only meets the latency requirements of

our customers but also simplifies deployments with Atomic, Consistent, Isolated, Durable (ACID) properties and built-in high availability for risk-averse service provider customers, and offers the performance and scalability necessary to provide real-time control of network resource consumption. VoltDB offers the TCO, performance, and scalability we need while enabling us to handle fast data and the real-time feeds of service provider traffic."

Mahindra Comviva

Mahindra Comviva transforms big data into useful insights to align the right message with the right people at the right time, on the right channel, and in the right way. Mahindra Comviva works closely with mobile operators to significantly increase revenue opportunities by presenting tailored, relevant and timely offers to customers.

To differentiate its offering and help operators take advantage of real-time analytics and decisioning, Mahindra Comviva selected VoltDB's in-memory scale-out SQL database as the core of its Mahindra Comviva Real-time Event Decisioning (ERED) platform.

By using VoltDB, Mahindra Comviva was able to create a fast data solution that requires 3 milliseconds for the ingest-analyze-decide journey. In addition, their system is able to deliver customized offers to subscribers in fewer than 250 milliseconds.

With Mahindra Comviva, mobile operators can now achieve a measurable ROI in real-time personalization. Real-time offers were proven to be 1.5 to 2.5 times more successful than near real-time offers. In one use case, personalized real-time offers delivered a 253% increase in acceptance. In addition, data bundling increased by 50% and billing increased by $30k/mo.

"We conducted a global search and selected VoltDB because we believe it's the world's fastest and smartest in-memory database," says executive head of Mahindra Comviva. "It met our core requirement for sub-250 millisecond response time. In terms of transactions per second, it's highly scalable and delivers the performance we need. We're happy we selected VoltDB as the underlying architecture for our real-time layer."

Achieving a Single Source of Truth in Financial Services

Financial services have to deal with a large and fast incoming stream of data. With the increase of e- and m-commerce, this stream is only going to get larger and faster. To compound all this data, financial institutions also have to comply with an ever-growing web of regulation. Besides fraud management, there are a number of use cases within financial services that suit real-time decisioning.

SLA Management

Implementing real-time policy and service level agreement (SLA) management requires the ability to ingest real-time data feeds, analyze events in real time, and implement granular policies immediately. To process at this speed and without latencies requires them to automate real-time policy decisions and SLA compliance. In order to do this, the database must support moving transaction processing closer to the data. Companies throughout the financial services industry rely on real-time decisioning for real-time policy and SLA management to ensure compliance and meet SLA requirements.

Trade Reconciliation

With trades happening within milliseconds, and security prices changing just as fast, exchanges need to be able to deliver real-time and consistent information to their trading firms. Exchanges partner with trading firms that have strict SLAs for the time it takes for a trade to be executed. Even the smallest delay in the time an order can be executed can cost a firm money when securities prices move up or down in millisecond bursts. In addition, any real-time decisioning needs to be 100% accurate to ensure proper reporting.

Regulatory Compliance

Regulations require financial institutions to prove all databases and replicas are the same, with audited consistency across different data sources. In addition, institutions must comply with the SEC's National Best Bid & Offer regulation.

Portfolio/Risk Management Portfolio managers and risk managers need to monitor their risk exposures and market conditions in real-time to maximize their performance in fast-moving financial markets. Managing risk in real-time requires sophisticated technology that enables aggregating, monitoring and analyzing multiple high-velocity data sources and event streams with accuracy.

In order to stay be compliant in real-time, financial services need to use a real-time solution capable of ingesting and aggregating real-time streaming market data, referencing data, counterparty risk data, credit risk data etc., and then convert it into insights in milliseconds with immediate consistency to help investors make time-sensitive investment and risk management decisions in real-time.

Kline

Kline are the developers of Sphera, a sophisticated software suite that enables direct access to financial markets and/or brokers, trade assets, and receiving real time updates from the markets. A typical Sphera user can manage up to many hundreds of requests per second. Each request requires a wide range of checks, ranging from database centric ones (including ones requiring more than a single query), to algorithmic ones.

When Kline was looking to expand Sphera's capacity, they chose VoltDB. VoltDB was integrated into their ecosystem to leverage the characteristics of an in-memory database and cut down the time required to complete the database centric procedures, which were a significant bottleneck in the previous disk-based implementation. In addition, they were able to maintain the advantages related to the ACID consistency and the chance to keep working with standard SQL, to which their developers are already accustomed.

Another advantage of the VoltDB implementation is the ability to scale with minimal effort to different machines, to adapt to the requirements of each customer in terms of cost, performance and continuity of service. The latter point was by far the most sought-after, and VoltDB implements natively all the facilities to enable it without any delay.

Finally, a multiple-hosts VoltDB instance can also be configured as a high availability cluster, in which the failure of at most a predetermined number of nodes does not cause the loss of the cluster functionality. This is especially important for large customers that cannot afford even the slightest downtime.

In parallel tests compared to the old system, VoltDB was able to provide massive improvements in speed. For inserting and updating orders, VoltDB provided an average of 67-times faster processing than the old system, bringing results down from hundreds of milliseconds to single-digit milliseconds. By implementing real-time decisioning with VoltDB, Kline was able to significantly speed up Sphera without compromising consistency.

Market Venue	Operation	Old Time [ms]	New Time [ms]	Improvement [times]
Market 1	Insert new order	197	3	65,6
	Update order	201	3	67
	Cancel order	2	<1	3
Market 2	Insert new order	135	2	67,5
	Update order	150	2	75
	Cancel order	1	<1	4
Market 3	Insert new order	262	4	65,5
	Update order	262	4	65,5
	Cancel order	2	<1	5

Evolving Your Infrastructure for the New Realities of IoT

The Internet of Things (IoT) has been projected to revolutionize just about every industry. Analyst firm Gartner predicted that around 6.4 billion IoT devices would be in use worldwide in 2016, and a McKinsey report estimates that IoT has the potential to represent around 11% of the world's economy by 2025. These figures alone show the transformative impact that IoT is having and will have.

Preparing for the IoT revolution will be a major challenge in the near future. Most current infrastructures just do not meet the stringent requirements to power IoT applications. You will have to deal with millions, if not billions, of sensors sending data to your data centers each second. That data needs to be processed in real-time to make real-time decisions and get maximum value out of your applications (while it may be easy to imagine an IoT application that does not need to be processed in real-time, real-time capabilities enables most of the exciting IoT applications).

With the deluge of data IoT applications will provide, making large amounts of real-time decisions quickly - in real-time - will be critical. However, it is crucially important that those decisions are also correct. Imagine a health-related IoT device (a heart monitor, for example). Ensuring that data and decisions regarding this device are immediately correct could be the difference between life and death.

Smart Meters

A number of electric utilities are using operational databases to collect real-time data from IoT smart electric and water metering systems. These applications require real-time decisioning capabilities.

Smart metering platforms typically provide meter readings to the IoT data management infrastructure every 15 minutes. Usually meters are associated with some kind of a concentrator—a device in a sensor network that collects data flowing from separate sensors or meters, batches the information, and provides it to the data management infrastructure.

Once that data has arrived, there are a number of different rules that need to be applied. Industry-specific validation, error checking, and estimation rules need to be applied. For example, if a meter reading is lost, the system might want to interpolate the value between the last two events. The goal is to be able to guarantee that a reading isn't obviously corrupt, and that it's a value that is valid.

With a number of other relatively straightforward validation processes, being able to supply or execute these validation processes in near real-time improves operational efficiency, makes it clear when data is being corrupted or lost, and also allows interesting operational applications to be developed as a benefit of the real-time infrastructure.

For example, if the system hasn't received readings from a set of meters over the last two reporting periods, it's important to understand if those meters are associated with one concentrator or are distributed across a number of concentrators. This understanding might indicate two different operational problems that need to be resolved in different ways.

At the same time, as this data is being collected into a real-time intraday repository or operational system, you can start to write real-time applications that track real-time pricing and consumption, and then begin to manage data or smart metering grids in a more efficient way than when data is only available at the end of day. However, the billing infrastructure that's calculating total utilization and generating the eventual bill to the consumer is still expecting data in a bulk fashion.

This system doesn't expect data to trickle in over the course of the day; rather, it expects the traditional format of data to be provided at the end of the day or at the end of some longer period. In this case, the system needs to be able to collect that intraday data, apply the events, rules, and triggers, and then at the end of the billing period, gracefully dump that to the billing system as an input in the time period that it expects. The same process applies if the utility wants to capture all of this data to a historical system for long-term offline analytics, exploration, and reporting.

A number of smart metering system uses VoltDB for fast ingest of events, the application or the ability for a rules engine to access real-time data to support real-time analytics that might trigger alerting/alarming to other operational applications, to buffer data for export to an end-of-day billing system, and then finally, to become an ingest point to an offline storage system or a nearline storage system like Hadoop.

Conclusion

These are six use cases where VoltDB has been used for real-time decisioning. These businesses are making the most of their high-velocity data. Is your business ready for real-time decisioning?