

Platform for Hypercustomization on LeanXcale

Table of contents

Introduction.....	2
Use cases.....	2
LeanXcale disruptive technology	3
How to integrate LeanXcale	4
Benefits.....	4
About LeanXcale	5

Introduction

In the old days, we were used to being addressed by the waiter by our name and be served our favorite drink personalized to our taste. The shop assistant knew what kind of suit or dress we can pay for and our preferences in terms of style, brands, and colors. The hairdresser just used to ask us: “As usual?” We were used to getting a highly personalized service.

e-Commerce has brought better prices thanks to the economy of scale with the convenience of ordering from home. However, this high customization grade got lost.

In recent years, a movement to recover this high-quality personalized service has appeared:

- Publishers do not sell just inventory anymore, but to anonymous clients with known preferences.
- Advertisers only display their ads to high-probability interested people in their customers' products.

Anyhow, companies still have a long way to go:

- They need to leverage all the power of all the first-party data that is spread around the company: CRM, helpdesk, internal databases ...
- Companies must speed up their learning process about real-time customers to identify their level of interest and overall if they are in-the-market.
- Adjust the offer to the customer preferences, mood, and behavior, predict with the power of machine learning what the customer wants next and offer it: product, information, or just user experience.

In summary, companies need to leverage this asset to enable a new hypercustomized experience where the ad invest is done on the right message, to the right person, and at the right moment.

Use cases

A solution with these characteristics has several technical challenges from the database management point of view:

It has to track all the customer's behavior on the web. The traditional method to do this is through a JavaScript tag that informs about all the user interaction. These events are stored for later analysis. Current online traffic and the granularity in the tracking means high insertion rates.

Because the window time from the moment a user clicks a link to access a new URL until the ad is displayed is less than 200 seconds, latency plays a vital role in a platform for hypercustomization. Website customization generates similar situations.

Machine learning frameworks should easily access the information stored in the database. These frameworks provide the intelligence level that enables understanding customers, predicting behavior, and ultimately increasing sales.

These kinds of platforms need to have a full picture of the situation that is more straightforward. It means calculating KPIs, aggregation, and time-series analysis.

Analytics and visualization tools frequently need a data warehouse. A typical pattern is to use ETLs (high cost of development and maintenance) to copy overnight information from operational databases to the data warehouse, resulting in analytics written over yesterday's stale data. **Having access to real-time operational data and removing ETLs can mean a big difference** overall in a low margin scenario.

LeanXcale disruptive technology

LeanXcale was founded on the idea of database technical excellence, trying to sort out all the problems that enterprise databases have. This philosophy is in LeanXcale's DNA, and it is embodied in all database aspects, being the origin of more than 12 disruptive technologies. The most important ones for this scenario are:

Event Ingestion: LeanXcale provides a dual key-value/SQL interface that enables ingesting of data very fast and efficiently through the key-value API, thus saving the overhead of SQL processing. This approach can result in significant performance improvement when the workload has a high data ingestion as happens with websites' tagging. It also provides the ability to ingest data faster by using a large cluster to ingest the logs or, for example, CRM's data of one year in seconds.

Time-series optimization: Our innovative bidimensional partitioning enables storing KPI time series at fast rates by maximizing cache efficiency. This partitioning allows having as a primary key the most convenient column(s) while having a lot of data locality in the updated data blocks, keeping only the latest data ingested in memory. It optimizes cache management and at the same time provides optimal search of a related customer time series of all the users' information in a particular time window.

Real-time analytics: Ad hoc analytical queries are supported over the operational data thanks to the blended OLAP+OLTP. No ETL is required anymore, with the consequent time- and money-saving benefits. However, the main benefit of these capabilities is that it enables real-time analytics and the capacity to take decisions based on the most recent information rather than yesterday's data.

Aggregation online: Semantic concurrency control enables generating aggregates in real time as data are ingested. In this way, most of the aggregation analytics queries

required by and the hypercustomized platform become single-row costless queries, highly reducing the footprint of the hardware needed to perform real-time analytics.

How to integrate LeanXcale

Most of the platforms use several databases. Each of them is used to face a specific challenge.

- Key-value for high-rate insertions.
- Data warehouse for analytics and, through BI, for visualization.
- Data lakes for machine learning modeling.
- Operational SQL databases for operational data.

This approach, beside the fact that it is ubiquitous, has several issues: TCO and time-to-market are very high, data have to be moved across databases, and real-time data are not exploited. Data lakes are not transactional, so data scientists must wait until the whole loading is completed. In addition, when there is a failure during the loading that is unnoticed, it results in generating a wrong model based on corrupted data, and it is a very long and painful process to trace back what caused the wrong model. In other words, coherence of data in data lakes is compromised and this is not acceptable as they take a more relevant role in enterprise architecture.

LeanXcale can play the role of all these types of databases while at the same time saving all ETLs/data movements and offering an operational data lake.

Because LeanXcale is transactional at any scale, and thanks to its integration with Machine Learning (ML) frameworks, it can be used as an operational data lake for creating ML models always based on the most recent data. LeanXcale offers JDBC interfaces (integration with Spark, C3IoT) and Apache Arrow (TensorFlow, Pandas, Spark) for these frameworks.

Benefits

LeanXcale allows creating an architecture that enables the processing of millions of events per second making them available in real time for any workload, including real-time analytics, and integration with ML frameworks to enable forecasting, clustering, and classifying, at any scale.

This versatility allows a low time-to-market of a new product or service, functionalities, or changes because developers only need to develop and learn a single technology based on the easy-to-learn SQL.

An architecture based on LeanXcale also means that dev-ops teams only have to maintain a single database engine with no data movement (ETLs don't need to be developed and maintained) reducing the Total Cost of Ownership.

LeanXcale allows creating a hypercustomization platform to leverage the value that provides all the data scattered and siloed in the company: CRM, ERP, Helpdesk, internal databases, website ... This strategy offers benefits at any level of the funnel.

At the prospection level, it reduces the CAC because users only visualize according to their preferences ads with a customized message to them when they are more open to clicking.

At the touch level, it improves the conversion rate, thanks to the in-depth knowledge about the customer, which allows having all the site adapted to the end-user preferences, making it much more effective.

At the conversion level, it increments the ARPU because the customers find more easily the product they like and by promoting up-selling and cross-selling strategies.

LeanXcale facilitates making more money by proving real value based on their interest and preferences to your customers.

About LeanXcale

LeanXcale was founded by some of the best researchers in the field of distributed scalable databases. This group has been enriched with a group of engineers from several industries and a select group of advisors: Glenn Osaka, PayPal advisor (Peter Thiel & Elon Musk), and the guru of databases, Dr. Patrick Valduriez, stand out among them.

Resources

Visit www.leanxcale.com for more information or contact us at info@leanxcale.com.

Free trial (<https://www.leanxcale.com/trial>)

Documentation and drivers (<https://www.leanxcale.com/company-resources>)

Whitepapers and videos (<https://www.leanxcale.com/company-resources>)

Get a demo (<https://www.leanxcale.com/get-a-demo>)

Talks (<https://www.leanxcale.com/talks>)

Blog (<https://www.leanxcale.com/blog>)