

TWOO.COM

CUSTOMER SUCCESS STORY

With over 30 million users, Twoo.com is Europe's leading social discovery site. Twoo runs the world's largest scale-out SQL deployment, with 4.4 billion transactions a day—without a full-time DBA.

ClustrixDB has seamlessly supported the exponential growth in Twoo users and new application features, expanding from 24 to 168 cores over the past 3 years. With this kind of scale-out simplicity designed for cloud, ClustrixDB is not your run-of-the-mill legacy SQL database.

Read on to learn more about Twoo's massive workload—2 Terabyte tables with seven-way joins and complex analytics—and the benefits of using a scale-out SQL database solution.

With over 11 million unique visitors monthly, Twoo.com has grown into one of the largest social discovery sites and most popular network of its kind.

What makes Twoo special, and sets it apart from its competitors, is that it's free to use and features innovative matchmaking algorithms that bring you closer to the people you want to meet. The site offers an enormous number of active and real members; 3 million active users every single day. It is available in 38 languages and 200 countries, with support for desktop and mobile devices.

A Platform for Hyper-Scale Growth to Millions of Users

When Facebook first saw rapid growth, very few scalable database options existed. The Facebook operational database was built and still runs on top of a highly customized sharded MySQL environment with an abstraction layer added in the middle. Separately they built Cassandra for inbox search. Websites such as Facebook and Twitter rolled out their own custom infrastructure because they had no choice.

Twoo.com, on the other hand, has been able to build and grow entirely on a single scale-out Clustrix database, enabling sophisticated application and user queries. The distributed database industry has evolved quickly over the last few years. Innovative companies now have the option of using proven, off-the-shelf database platforms that don't require in-house database engineering—so they can keep all of their engineering talent focused on application innovation. With ClustrixDB, Twoo.com never had to hardcode their application for database sharding, making the core simpler and cheaper to maintain. And the customer service provided by ClustrixDB engineers made sure that Twoo.com could run their business with five-9s of availability.

Twoo Takes the New Path

The Twoo.com team was passionate about avoiding a sharded database deployment. Massive Media, the company behind Twoo.com, has another website called Netlog. To scale Netlog's database in 2007, Massive Media decided to partition, or shard it by language. As the database continued to grow, this strategy proved insufficient and they were forced to shard by language and user ID, which required a very large in-house development effort to maintain and optimize the database layer. In addition, the Netlog development team was limited in the types of queries they could efficiently execute in this architecture. It was a maintenance headache.

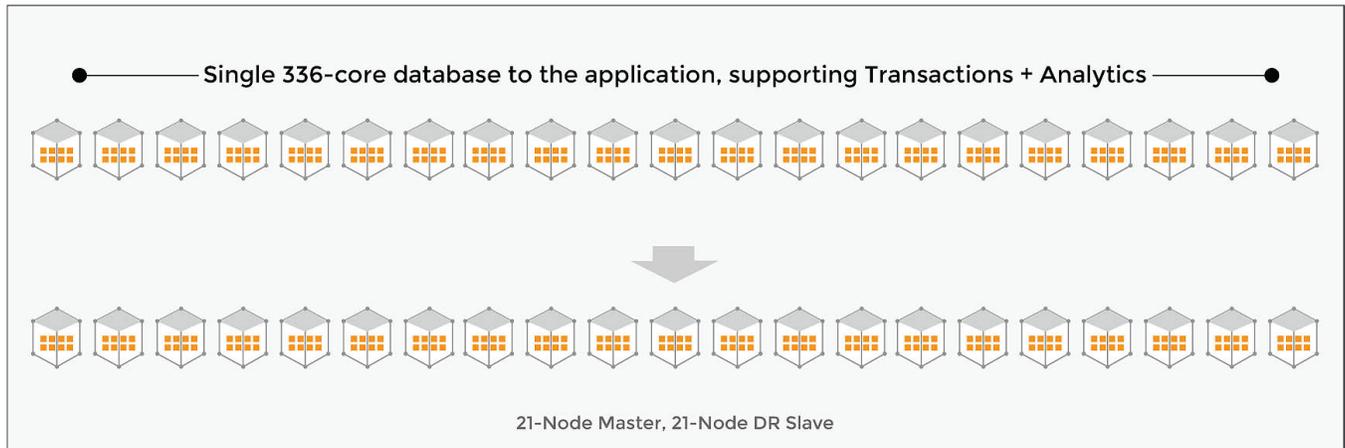
“ There is always some risk, but a trust relationship needs to build. We really liked Clustrix since it is building ground-breaking technology.”

Twoo.com started out on a single MySQL server and quickly ran into scale issues. At that time ClustrixDB was being tested as a slave, but when Twoo's MySQL server could not handle the load and crashed, they switched over completely to ClustrixDB. By 2010, the product was in production and they have not looked back since.

World's Largest Scale-Out SQL Deployment

Twoo.com supports millions of complex user interactions every day, resulting in billions of transactions per day. It is also a data-intensive application, with petabytes of read/write data volume each month.

Twoo.com runs entirely on a single ClustrixDB database cluster. It is the largest scale-out SQL database in the world. Twoo.com has a master cluster and a slave cluster for disaster recovery.



**30 MILLION
USERS**

**4.4 BILLION
TRANSACTIONS
PER DAY**

**70,000
TRANSACTIONS
PER SECOND**

The Twoo Application

The goal of the Twoo application is to be “the most fun way to meet new people near you.” It allows users to log in and discover interesting people who match their interests. Twoo.com is not just a dating site, since it also enables users to find friends and develop relationships. With the site’s high number of users and fast growth, the application has challenging requirements.

Twoo runs the entire website and stores all their business-critical data on ClustrixDB. In addition, they have logging data that goes to Hadoop, where offline analytics and reporting are performed, albeit with a multi-hour delay.

Primary Data Profile

Core to the Twoo.com site is the user information, which is updated very frequently. The site also includes the text and messaging system allowing users to communicate with one another. To support the matchmaking algorithm, the application uses many multi-billion row tables. Naturally these tables have an intense query load, and schemas change as the company’s algorithm evolves. User_contactlist recently had an online schema change for a table that is nearly 2 terabytes in size and ClustrixDB handled it easily.

Complex Analytics Queries & Recommendations

Ad hoc analytic queries are performed routinely in ClustrixDB. Some queries can be challenging. For example, a male may be interested in finding females with certain characteristics between 30 to 35 years of age, but we can only show him females who are also looking for males who are 33 years of age.

This type of complex query needs split-second response time and is typical of social applications. It presents a more complicated problem than is dealt with by traditional e-commerce sites such as Amazon. For example, a bookstore can easily suggest that users who bought book_1 also buy book_2. Clearly book_2 is not going to decide it's not interested in customer A!

An example query (bottom) to get the top messages in your inbox as you log in has a 5-table join, including the chat_message table with 4.7 billion rows and the chat_conversation table that is more than 300GB in size. (Right) This example query looks up the list of friends of a user; it involves a 7-table join including uxxx_cxxx, a table that is 1.9TB in size.

```
SELECT SQL_CALC_FOUND_ROWS cc.fromunread,
    cc.datechanged,
    cc.toid AS inboxUserId,
    cc.todatelastsent,
    cm.fromid,
    cm.toid,
    cm.messages,
    cm.message,
    cm.fromid AS lastmessagefromid,
    cm.premium,
    cn.typeid,
    cn.options,
    cn.fromdeleted,
    cn.todeleted,
    ccl.locked
FROM chat_conversation cc
JOIN chat_message cm
    ON cm.messageid = cc.lastmessageid
JOIN user_join uj
    ON uj.userid = cc.toid
LEFT JOIN chat_notification cn
    ON cn.messageid = cm.messageid
LEFT JOIN chat_conversation_lock ccl
    ON ( cc.fromid = ccl.fromid
        AND cc.toid = ccl.toid )
WHERE cc.fromid = <a number>
    AND cc.fromdeleted = '0'
ORDER BY cm.premium DESC,
    cc.datechanged DESC
LIMIT 0, 10
```

```
SELECT uc.contactid,
    IF(u.firstname IS NOT NULL, u.firstname, Trim(uc.name)) AS name,
    IF(u.birthdate IS NOT NULL, Floor(( Curdate() -
u.birthdate ) / 10000), NULL) AS age,
    uc.email,
    uc.tool,
    uc.externaluserid,
    uc.avatarthumb,
    IF(u.gender IS NOT NULL, u.gender, uc.gender) AS gender,
    uc.datesync,
    u.userid AS userid,
    u.userid AS userID,
    IF(ub.user_id IS NOT NULL, 1, 0) AS hasblocked,
    IF(f.userid IS NOT NULL, 1, 0) AS isfriend,
    u.avatar AS photoID,
    up.photow AS parentWidth,
    up.photoh AS parentHeight,
    Ifnull(upd.rotation, 0) AS rotation,
    Ifnull(upd.cropx, 0) AS cropX,
    Ifnull(upd.cropy, 0) AS cropY,
    Ifnull(upd.cropw, 0) AS cropW,
    Ifnull(upd.croph, 0) AS cropH,
    up.biguid,
    up.thumbuid
FROM uxxx_cxxx uc
LEFT JOIN usxx_eeee ue
    ON uc.eeee = ue.email
LEFT JOIN user u
    ON u.userid = ue.userid
LEFT JOIN user_photo up
    ON up.photoid = u.avatar
LEFT JOIN user_photo_detail upd
    ON upd.photoid = up.photoid
LEFT JOIN user_blocked ub
    ON ub.user_id = uc.userid
    AND ub.other_user_id = u.userid
LEFT JOIN friends f
```

Strict Low-Latency Requirements

A critical requirement for Twoo.com is that every page must render within 800 milliseconds. Given the network latency, resources (such as images and javascript), and application work, the total database access time must be less than 40 milliseconds. ClustrixDB is up to the challenge with average query times under 5 milliseconds.

Other Options

Sharding MySQL and NoSQL are other options that companies struggling with scale typically consider. These options are often inferior. Here's why these options didn't work for Twoo.

Why Clustrix is Preferable to Sharding

“We spent a lot of time on optimizing for performance and scale. Now we can spend those resources better.”

With ClustrixDB, database-wide ad hoc queries can be run easily to mine the live database for useful trends—a feature that was not possible in a sharded architecture. In their previous web property Netlog, the company needed multiple DBAs to maintain the sharded infrastructure. Now they don't have any dedicated DBAs, only application developers.

Toon Coppens, CTO and co-founder of Twoo, said, “It does not make sense to build a new site using the old system of sharding MySQL. Pre-ClustrixDB, we spent a lot of time on optimizing for performance and scale. Now we can spend those resources better.”

Why NoSQL Was Not an Option

“NoSQL databases didn't provide any real-time analytics capabilities.”

Twoo really cares about SQL and ACID guarantees. In addition to their free user tier, Twoo provides users with premium for-fee functionality. This level of service requires the data on their end to be correct and consistent all the time with no issues.

When first considering the database options, Twoo looked at MongoDB, Redis, and other NoSQL data stores, but the company could not tolerate any inconsistencies or irregularities in the data store. They also found that tool support for the NoSQL databases was lacking. Tools such as phpMyAdmin show data in an understandable way, but the company was not able to find such tools for the NoSQL databases.

To make matters worse, NoSQL databases don't provide any real-time analytics capabilities, as they offer no support for joins and other SQL constructs.

Twoo's Journey with ClustrixDB & Its Rapid Growth Since 2010

“With ClustrixDB, we have not run into scaling issues anymore. As we need capacity, we just add nodes and see linear growth. We have seamlessly gone from a 3-node cluster to a 21-node cluster.”

—Nicolas Van Eenaeme, CIO of Twoo's parent company Massive Media

Growth

The design decision to use ClustrixDB enabled very rapid development, and the Twoo.com website went from concept to launch in less than six months—from December 2010 to May 2011—with less than 30 developers. The

Twoo.com website experienced greater than 1,000% growth from June to December 2011. The site started on a single MySQL server and quickly ran out of capacity. Since switching to ClustrixDB, they have been able to handle growth in traffic simply by adding more nodes. Now they're pushing the limits of the MySQL replication protocol as they run the highest-volume MySQL replication stream on the planet to their disaster recovery cluster.

Business Impact

On the rapid growth experienced by Twoo.com, Van Eenaeme said, "Growth as a web 2.0 company was challenging because in 1.5 years it became the largest dating site." When asked if he believes ClustrixDB helped the company grow, he added, "It definitely helped us grow faster. Our company wants to build the website and its features and not worry about the database or database development. We don't have a database lab like Clustrix has. We also didn't want to hire a team of DBAs. Clustrix has many high-traffic customers. If you'd ask us again, we'd happily go down this path. We don't regret it one bit since then."

ClustrixDB is designed to be self-managing, and ClustrixDB Insight provides a rich view into the database. The rich built-in statistics inform both the user and the automation inside ClustrixDB. ClustrixDB intelligently distributes data across the cluster. Tables are divided into slices and each slice has a replica. Indexes are internally treated like tables and get their own distribution. The patented Rebalancer running in the background makes sure that the data and load are always evenly distributed. New nodes are added with a single command or click and the Rebalancer will move data to the new node. If node or disk failure occurs, the Rebalancer will regenerate lost copies and move them to other nodes automatically. If the cluster is under stress or some event needs attention, the database will send an email to notify the user and will optionally alert the Clustrix support team.

Clustrix

Clustrix provides the leading scale-out SQL database engineered for the cloud. With ClustrixDB you can build innovative business critical applications that deliver real-time analytics on live operational data with massive transactional volume. Our exceptional customer service team supports more than one trillion transactions per month across a range of industry segments including Ad Tech, e-commerce, and social analytics. Clustrix customers include AOL, engage:BDR, MedExpert, Photobox, Rakuten, Symantec, and Twoo.com. Headquartered in San Francisco, Clustrix is funded by HighBAR Partners, Sequoia Capital, U.S. Venture Partners, Don Listwin, and ATA Ventures. ClustrixDB is available as a free community edition for developers, a software download that runs on any cloud, and on the AWS marketplace.

To learn more about Clustrix, visit us at www.clustrix.com

