



Why and How Gamechangers are Adopting ScyllaDB

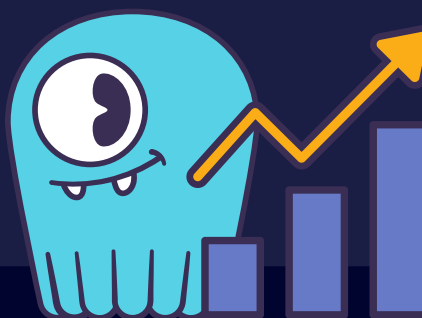
Disney+ hotstar

Fanatics

Zillow

Expedia

DISCORD



Gamechangers are taking advantage of this next tech cycle to disrupt the status quo.

They're ingesting an unprecedented amount of data and tapping it to deliver differentiating user experiences that transform markets and displace legacy leaders.

An increasing number of the companies driving this change trust ScyllaDB to power their data-intensive applications. Like so many of our customers' products, ScyllaDB was born of frustration with legacy approaches. Legacy NoSQL databases were architected for a different age: a different (nascent) cloud, different hardware, different bottlenecks, and even different programming models. By taking a fundamentally different approach to architecting a database, ScyllaDB harnesses the computing power of modern infrastructure.

With ScyllaDB at their core, top gamechangers are powering engaging experiences with impressive speed – and at rapidly increasing scale.

 Discord Messaging for 150M MAUs (~5M concurrent)	 Seamless experiences across content + devices	 Mobile-first ecommerce for the world's top sports teams
 ML-enriched product catalog data for 1.5B members	 Accelerating the world's transition to cryptocurrency	 Unified ML feature store for recommendations
 Scaling property listings for 200M MAUs	 Geography-based travel recommendations	 Real-time fraud detection across 6M transactions/day
 Fast responses for India's ridesharing leader	 Instant computation to uncover "hidden" routes	 Delivering 100M events/month to restaurants

and 400+ other global leaders



Let's take a look at why and how 7 of these gamechangers decided to modernize their database architecture with ScyllaDB, and what that move is enabling them to achieve.



Rapid scaling for seamless viewer experiences

"Every 6 months, we're almost doubling our data. With Redis and Elasticsearch, that resulted in burdensome admin and increased costs."

- Vamsi Subash Achanta, Architect at Disney+ Hotstar

46M

paid subscribers

300M

monthly active users

25.3M

simultaneous views

The company: Disney+ Hotstar, India's most popular streaming service, accounts for 40% of the global Disney+ subscriber base. Disney+ Hotstar currently offers over 100K hours of content as well as livestreams of the world's most-watched sporting events.

Database use case: The database powers the "continue watching" feature, which tracks every show for every user. It allows users to resume an in-progress video on any device. It also serves up the next episode in a series and notifies viewers when new series episodes are added.

Previous database: Redis and Elasticsearch

Why change: Their infrastructure was built on a combination of Redis and Elasticsearch, connected to an event processor for Kafka streaming data. Their Redis cluster held 500 GB of data, and the Elasticsearch cluster held 20 TB. Their key-value data ranged from 5kb to 10kb per event. Having multiple data stores meant maintaining multiple data models, making each change a huge burden. Moreover, with data doubling every 6 months, they were constantly increasing the cluster size, resulting in yet more admin and causing costs to rise dramatically.

New approach: First, the team adopted a new data model that could suit both use cases. Then, they set out to adopt a new database. Apache Cassandra, Apache HBase, Amazon DynamoDB, and ScyllaDB were considered. The team selected ScyllaDB for two reasons. 1) Consistently low latencies for both reads and writes, which would ensure a snappy user experience. 2) ScyllaDB Cloud, a fully-managed database as a service (NoSQL DBaaS), offered a much lower cost than the other options they considered.

Results: The team is now achieving sub-millisecond p99 latencies, and average read and write latencies in the range of 150 – 200 microseconds. Beyond the improved performance, predictable low latencies, and better TCO, they are also relieved of the burden of administrative tasks like backups, upgrades, and repairs. Now they can focus on scaling their business.



Faster geo-targeted recommendations with 30% cost savings

"We no longer have to worry about 'stop-the-world' garbage collection pauses. Also, we are able to store more data per node and achieve more throughput per node, thereby saving significant dollars for the company."

- Singaram Ragunathan, Cloud Data Architect at Expedia Group

750M

monthly site visitors

600B

AI predictions per year

200+

websites in 70+ countries

The company: Expedia is one of the world's leading full-service online travel brands helping travelers easily plan and book their whole trip with a wide selection of vacation packages, flights, hotels, vacation rentals, rental cars, cruises, activities, attractions, and services.

Database use case: The application provides information about geographical entities and the relationships between them. It aggregates data from multiple systems, like hotel location info, third-party data, etc. This rich geography dataset enables different types of data searches using a simple REST API with the goal of single-digit millisecond p99 read response time.

Previous database: Redis and Apache Cassandra

Why change: The team was using a multilayered cache with Redis as a first layer and Cassandra as a second layer, but they grew increasingly frustrated with Cassandra's technical challenges. Managing

garbage collection and making sure it was appropriately tuned for the workload at hand required significant time, effort, and expertise. Also, burst traffic and workload peaks impacted the p99 response time – requiring buffer nodes to handle this peak capacity, which drives up infrastructure costs.

New approach: The team migrated from Cassandra to ScyllaDB without modifying their data model or application drivers. With “just a few tweaks” to their automation framework that provisioned Apache Cassandra clusters, they were able to provision ScyllaDB clusters.

Results: With Cassandra, p99 read latency was previously spiky, varying from 20 to 80ms per day. With ScyllaDB, it's consistently around 5 ms. ScyllaDB throughput is close to 3x Cassandra's. Moreover, ScyllaDB is providing 30% infrastructure cost savings.



Enhanced ordering experiences – with a fraction of the nodes, costs & admin hassle

"We not only reduced TCO, but also reduced the pain that the database engineering team was taking to actually maintain the cluster in a healthy state."

- Niraj Konathi, Director of Platform

\$27B

valuation

\$5B

annual revenue

1K

product vendors

The company: Fanatics is the leading online retailer of officially licensed sports merchandise. They power the ecommerce sites of all major professional sports leagues (NFL, MLB, NBA, NHL, NASCAR, PGA), major media brands (NBC Sports, CBS Sports, FOX Sports) and 200+ collegiate and professional team properties.

Database use case: A range of functions associated with order capture, including orders that are created, carts that are modified, loyalty, promotional codes, order visibility, and rate limiting.

Previous database: MySQL, then Apache Cassandra

Why change: When they moved their on-prem infrastructure to the cloud to support anticipated growth several years ago, Fanatics migrated from MySQL to Cassandra. But Cassandra led to node sprawl requiring a large cluster size, frequent garbage collection pauses, and CPU spikes during compactions — all of which led to timeouts. These issues were so severe that they were driving users to abandon shopping carts and exit their site. Attempting to avoid these issues, they continued scaling out Cassandra

– ultimately creating a giant 55-node cluster. Running such a large cluster size increased EC2 costs, but more importantly, increased maintenance overhead.

New approach: The Fanatics team selected and deployed ScyllaDB for their ordering use case. Their automated provisioning follows regular procedures, such as securing deployments with node-to-node and client-to-node encryption. Their data at rest is also encrypted. With ScyllaDB, their rolling restarts across the cluster now take less than an hour (compared to 5-8 hours with Cassandra). They have ScyllaDB Monitor connected to PagerDuty, and use ScyllaDB Manager is used for data repairs.

Results: With ScyllaDB, Fanatics has reduced node count from 55 to 6 nodes, resulting in a drastic reduction in EC2 expenses, all while virtually eliminating the timeouts experienced by end users. During a recent peak minute, they saw nearly 280,000 IOPs for a solid minute – and registered zero timeouts. This translates to happier customers as well as more productive application teams.



Faster catalog updates by eliminating volatile latencies

"With Cassandra, the latencies were extremely volatile. The exact same queries would perform dramatically differently throughout the same day. This made it really difficult to commit to SLAs."

- Hitesh Shah, Engineering Manager at Rakuten

1.5B

members

1B+

total catalog items

3500

partners

The company: Rakuten allows its 1.5B members to earn cash back for shopping at over 3,500 stores. Stores pay Rakuten a commission for sending members their way, and Rakuten shares that commission with its members.

Database use case: Rakuten Catalog Platform provides ML-enriched catalog data to improve search, recommendations, and other functions to deliver a superior user experience to both members and business partners. Their data processing engine normalizes, validates, transforms, and stores product data for their global operations.

Previous database: Apache Cassandra

Why change: While the business was expecting this platform to support extreme growth with exceptional end-user experiences, the team was battling Cassandra's instability, inconsistent performance at scale, and maintenance overhead. They faced JVM issues, long Garbage Collection pauses, and timeouts – plus they learned the hard way that a single slow node can bring down the entire cluster.

New approach: Rakuten replaced 24 nodes of Cassandra with 6 nodes of ScyllaDB. ScyllaDB now lies at the heart of their core technology stack, which also involves Spark, Redis, and Kafka. Once data undergoes ML-enrichment, it is stored in ScyllaDB and sent out to partners and internal customers. ScyllaDB processes 250M+ items daily, with a read QPS of 10k-15k per node and write QPS of 3k-5k per node.

Results: Rakuten can now publish items up to 5x faster, enabling faster turnaround for catalog changes. This is especially critical for peak shopping periods like Black Friday. They are achieving predictably low latencies, which allows them to commit to impressive internal and external SLAs. Moreover, they are enjoying 2.5x lower infrastructure costs following the 4x node reduction.



Cost-effective scaling to meet unprecedented delivery demand

"One thing that's really relevant here is how fast iFood grew. We went from 1M orders a month to 20M a month in less than 2 years."

- Thales Biancalana, Backend Developer at iFood

60M

orders delivered per month

1.5M

app downloads per year

500K

daily active users

The company: iFood is the largest food delivery company in Latin America. It began as a Brazilian startup, and has since grown into the clear leader, with a market share of 86%. After becoming synonymous for 'food delivery' in Brazil, iFood expanded its operations into Columbia and Mexico.

Database use case: Online ordering. Each order represents about 5 events in their database, producing well over 100M events on a monthly basis. Those events are sent to restaurants via the iFood platform, which uses SNS and SQS. Since internet connections are spotty in Brazil, they rely on an HTTP-based polling service that fires off every 30 seconds for each device. Each of those polls invokes a database query.

Previous database: PostgreSQL, then DynamoDB

Why change: After they hit 10 million orders a month and were impacted by multiple PostgreSQL failures, the team decided to explore other options. They moved to NoSQL and selected DynamoDB for their Connection-Polling service. They quickly discovered that DynamoDB's autoscaling was not fast enough for their use case.

iFood's bursty traffic spikes around lunch and dinner times. Slow autoscaling meant that they could not meet those daily bursts of demand unless they left a high minimum throughput (which was expensive) or managed scaling themselves (which is work that they were trying to avoid by paying for a fully-managed service).

New approach: iFood transitioned their Connection-Polling service to ScyllaDB Cloud. They were able to keep the same data model that they built when migrating from PostgreSQL to DynamoDB. Even though DynamoDB uses a document-based JSON notation, and ScyllaDB used the SQL-like CQL, they could use the same query strategy across both.

Results: iFood's ScyllaDB deployment easily met their throughput requirements and enabled them to reach their mid-term goal of scaling to support 500K connected merchants with 1 device each. Moreover, moving to ScyllaDB reduced their database expenses for the Connection-Polling service from \$54k to \$6K — a 9x savings.



Nonstop operations despite data explosion and datacenter fires

"When these kinds of situations happen, it is priceless to be able to get over it with basically no downtime whatsoever."

- Milos Vyletel, Staff Engineer at Kiwi.com

1.6B

virtual interlining
combos per year

12.3M

flight routes searched
per year

40K

seats booked daily

The company: Kiwi.com is one of Europe's fastest growing companies. They use "Virtual Interlining" to make planes, trains, buses, ferries, and taxis connect, even when the operators don't work together. This uncovers low cost options that other search tools cannot find.

Database use case: Kiwi stores a massive (and fast-growing) number of flight combinations. For example, a single route from Prague to JFK has myriad potential combinations. At the same time, the data itself is constantly updating as airlines modify prices for various flight combinations. The data changes at the rate of 60% of data per day, 80% over 3 days, and 100% every 10 days. Topping off the challenge: Kiwi's precomputation engine requires all of the data to be updated every hour.

Previous database: PostgreSQL then Apache Cassandra

Why change: Kiwi initially built their service on PostgreSQL, attempting to scale through sharding (with 60 database instances and 60 Redis caches). This created a management nightmare. To improve performance and scalability, they migrated to Cassandra. However, Cassandra proved unable to keep pace with their growth, even as the team

added more and more nodes. Their use case required full table scans, but they had to write and maintain a custom scanning service to do this with Cassandra. Additionally, their hourly data updates caused CPU overload and massive latency spikes.

New approach: The team migrated to ScyllaDB without needing to change their applications. They deployed ScyllaDB on three independent locations — every location at least 200 kilometers from another. ScyllaDB enables them to perform continuous full table scans that filter for last-update-timestamp. It can also handle token ranges without overloading. This eliminated many of Kiwi.com's technical issues with Cassandra.

Results: In scenarios where Cassandra required 100 nodes to achieve 40K reads per second, ScyllaDB achieved 900K reads per second with only 21 nodes. This translated to 75% cost savings. Their ScyllaDB deployment also survived the ultimate test: trial by fire. The 2021 OVHcloud fire suddenly took down 10 out of the 30 nodes in their distributed cluster. Kiwi's database kept running seamlessly, with zero data loss, since the remaining ScyllaDB database cluster was capable of rebalancing itself and handling the load.



Stopping fraud and lost business by lowering latencies

"During our busiest periods, rush hour and the weekend, peak load can hit many tens of thousands of operations per second. ScyllaDB handles everything with ease using just a single five-node AWS cluster."

- Brian Trenbeath, Technical Program Manager at Grab

187M

users

6M

trips per day

2.8M

drivers

The company: Grab began as a Malaysian taxi-hailing service and grew into a "superapp" after expanding into food, grocery and parcel delivery, digital payments, lending, and other financial services. Grab is Southeast Asia's largest startup

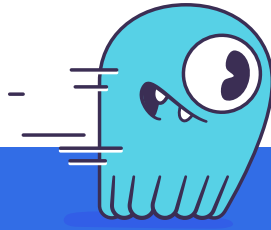
Database use case: Apache Kafka data streaming underlies all of Grab's systems. The engineering teams within Grab aggregate multiple Kafka streams – or a subset of streams – to meet various business use cases. This requires reading the streams, using a powerful, low-latency metadata store to perform aggregations, and then writing the aggregated data into another Kafka stream.

Previous database: Redis

Why change: The Grab development team initially used Redis as its aggregation store, only to find that it couldn't handle the load as their business grew. They started to notice lots of CPU spikes, so they kept scaling it vertically and adding more processing power. Eventually, they decided to explore other options for a low-latency metadata store.

New approach: Grab now aggregates and republishes the Kafka streams using a low-latency metadata store built on ScyllaDB – combining multiple streams in near real-time. Grab's Trust and Safety Engineering team also uses the metadata stored in ScyllaDB to monitor and analyze activity streams for signs of fraudulent transactions.

Results: With more than 6M on-demand rides per day, latency issues could result in missed rides and millions of dollars in losses. ScyllaDB enables Grab to handle peak loads with "very, very impressive" read and write latencies. They estimate that running the same workload on another database would be 3x more costly. Moreover, using a low-latency database for fraud detection helps them stay one step ahead of suspicious activity so they can stop it in its tracks.



Next Steps

If your team is considering moving beyond legacy databases, our technical consultants would be happy to help you assess whether ScyllaDB is a good fit for your use case.

Book a Free Technical Consultation

And here are some other options for moving past your database challenges:

- [Get started with ScyllaDB](#)
- [Take free courses at ScyllaDB University](#)
- [Explore papers, videos, benchmarks & more](#)

ABOUT SCYLLADB

ScyllaDB is the database for data-intensive apps that require high performance and low latency. It enables teams to harness the ever-increasing computing power of modern infrastructures – eliminating barriers to scale as data grows. Unlike any other database, ScyllaDB is built with deep architectural advancements that enable exceptional end-user experiences at radically lower costs. Over 400 game-changing companies like Disney+ Hotstar, Expedia, FireEye, Discord, Crypto.com, Zillow, Starbucks, Comcast, and Samsung use ScyllaDB for their toughest database challenges. ScyllaDB is available as free open source software, a fully-supported enterprise product, and a fully managed service on multiple cloud providers. For more information: [ScyllaDB.com](https://scylladb.com)

SCYLLADB.COM



SCYLLA

United States Headquarters
2445 Faber Place, Suite 200
Palo Alto, CA 94303 U.S.A.
Email: info@scylladb.com

Israel Headquarters
11 Galgalei Haplada
Herzeliya, Israel



Copyright © 2022 ScyllaDB Inc. All rights reserved. All trademarks or registered trademarks used herein are property of their respective owners.