# SCYLLA

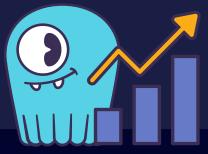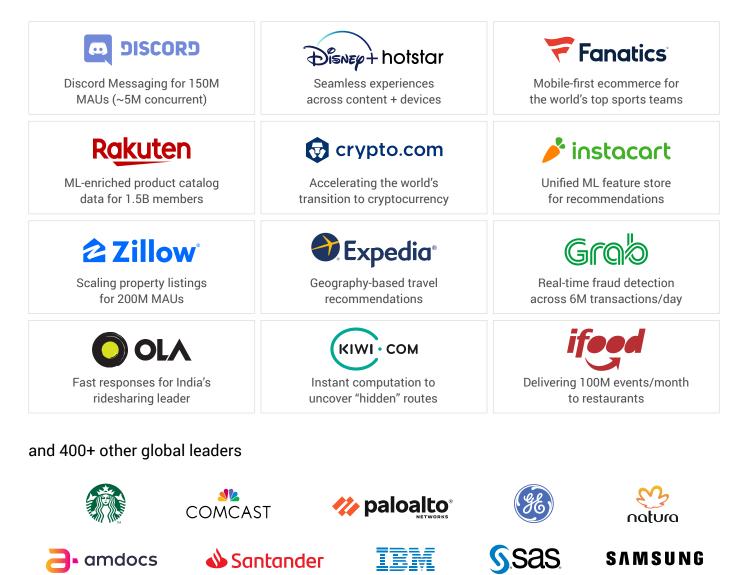# Why and How Gamechangers are Adopting ScyllaDB on AWS

**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**<br>Discord Messaging for 150M MAUs (~5M concurrent) | **Disney+ hotstar**<br>Seamless experiences across content + devices | **Fanatics**<br>Mobile-first ecommerce for the world's top sports teams |
| **Rakuten**<br>ML-enriched product catalog data for 1.5B members | **crypto.com**<br>Accelerating the world's transition to cryptocurrency | **instacart**<br>Unified ML feature store for recommendations |
| **Zillow**<br>Scaling property listings for 200M MAUs | **Expedia**<br>Geography-based travel recommendations | **Grab**<br>Real-time fraud detection across 6M transactions/day |
| **OLA**<br>Fast responses for India's ridesharing leader | **KIWI·COM**<br>Instant computation to uncover "hidden" routes | **ifood**<br>Delivering 100M events/month to restaurants |

and 400+ other global leaders

Starbucks    COMCAST    paloalto NETWORKS    GE    natura

amdocs    Santander    IBM    SAS    SAMSUNG

Let's take a look at why and how 4 of these gamechangers decided to modernize their database architecture with ScyllaDB running on AWS, and what that move is enabling them to achieve.

# Disney+ Hotstar

# 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** | **300M** | **25.3M** |
|---|---|---|
| paid subscribers | monthly active users | 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, 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.

# Zillow®

# Scaling background and real-time workloads with speed and accuracy

*"No one notices that we're processing the entirety of Zillow's property and listings data to correct some data issue or change a business rule. We can process the entirety of our data with no performance hit to real-time data."*

- Dan Podhola, Principal Software Engineer at Zillow

| **243M** | **10.2B** | **$4.3B** |
|:---:|:---:|:---:|
| monthly users | annual site visits | revenue Q1 2022 |

**The company:** As the most-visited real estate website in the United States, Zillow and its affiliates offer customers an on-demand experience for selling, buying, renting, and financing with transparency and nearly seamless end-to-end service.

**Database use case:** The database team is responsible for processing property and listing records, mapping those to common identifiers, and translating messages into a common interchange format that can be shared across teams. They are also responsible for deciding what's best to display. A listing processor service uses ScyllaDB as a storage layer, taking messages from two queues, publishing to a Kinesis stream, and backfilling messages to an S3 bucket.

**Previous database:** Apache Cassandrah

**Why change:** Zillow's Cassandra database couldn't keep up with out-of-order messaging as data volumes grew. They have a highly threaded application and receive two different message queues from two different producers. Messages can be received out of

order, simply by the nature of the queues, but the application cannot go 'back in time' to reorder messages. For example, a message may be received that a property has changed price, followed by a message from a different data source that the property had sold. If the sold message is processed before the price change, problems ensue.

**New approach:** The Zillow team was able to solve the problem without resorting to application locks, or even using ScyllaDB's lightweight transactions (LWTs). According to Podhola, "We provide the ScyllaDB write timestamp and employ a couple other tricks to provide correct and consistent data to our consuming services and avoid doing transactions."

**Results:** According to Podhola, the real benefit of ScyllaDB is its ability to scale in support of both background processes and real-time, user-facing workloads. On just three nodes, they can scale up to 35 instances and process over 6500 records per second, while also running their real-time workloads.

# Fanatics®

# 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 at Fanatics

| $27B | $5B | 1K |
|---|---|---|
| valuation | annual revenue | 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.

# Grab

## 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** | **6M** | **2.8M** |
|---|---|---|
| users | trips per day | 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 AND AWS

ScyllaDB's NoSQL database was architected from the ground up to take advantage of modern multi-CPU, multi-core public cloud servers such as those provided by AWS's Amazon Elastic Compute Cloud (EC2). ScyllaDB's high system utilization optimizes performance as well as TCO by taking full advantage of powerful infrastructure such as Amazon EC2 I4i Instances.

ScyllaDB Enterprise and ScyllaDB Cloud are listed in the AWS Marketplace for both public and private offers. ScyllaDB Cloud is certified to run as a fully managed on-premises extension of AWS, known as AWS Outposts, compatible with Amazon DynamoDB. Moreover, ScyllaDB is Graviton Service Ready, allowing you the flexibility to run your workloads on these instances known for their price/ performance advantage.

**Additional Information on ScyllaDB + AWS**

SCYLLADB.COM

**SCYLLΛ**

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

**Israel Headquarters**
11 Galgalei Haplada
Herzelia, Israel