

The Data Flywheel:

Building momentum
by putting your data
to work

Take a comprehensive approach to
getting the most value from your data,
one project at a time



Data may be your most valuable business asset

What do Airbnb, Lyft, Snapchat, Tinder, and Peloton all have in common?

Apart from being successful startups founded in the last 5-10 years, all of these companies have built their businesses on the foundation of maximizing the value of data. In other words, they have figured out how to "put their data to work."

And that's critical, because data has transformed from a cumbersome, expensive-to-store asset into a core component of many successful business models. Some have gone so far as to say that data has replaced oil as the world's most valuable resource.¹

Migrating to AWS drives results



Uses data to personalize search and recommend results. Says that AWS is helping it to prepare for more growth.



Can now manage up to 8X more riders during peak times. Gains customer insights that power its shared-ride product. Leverages scalability to store GPS coordinates for all rides.



Tinder | **Snap Inc.**

Stores user profiles and serves up relevant ads through data, both growing from zero to hundreds of millions of active users in a short time.



Gets the high elasticity, low latency, and real-time processing needed to power its live leaderboard and deliver customizable rider data to its community of virtual bike riders.

¹ <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

How did we get here?

Over the last 5-7 years, three key trends have significantly increased the value of data:

- 1 Connected devices, apps, and systems now generate more data than ever before
- 2 The cloud has reduced the cost of storage; it's now practical and cost effective for businesses to store all of their data on the cloud
- 3 The cloud provides pay-as-you-go, on-demand compute, allowing organizations to more easily analyze and gain insights from data

We all know that data volumes are exploding. The good news is the cloud dramatically reduces the cost of storage, meaning your organization can now store every relevant piece of data, even as those stores grow to massive volumes. And, it can use a variety of analytics approaches to derive better insights from that data, driving better decisions and, ultimately, giving you a competitive edge.

In many ways, connected devices and the cloud have democratized data, enabling companies of every size to use data in innovative ways that spark business growth. Organizations today are using data to determine when and how to develop and release new products, create opportunities for new revenue streams, where to automate manual processes, how to earn customer trust, and more. All of these decisions fuel innovation and drive your business forward.

But what's the best model—the best strategy—for analyzing this data and applying the resulting insights in a structured way? In this e-book, we'll introduce you to the Data Flywheel, a comprehensive approach to managing your data strategy.

The flywheel—as popularized by author Jim Collins—is a self-reinforcing loop made up of a few key initiatives that feed and are driven by each other to build long-term business momentum. AWS built on this concept to create the Data Flywheel: a holistic framework that applies this strategy to help you derive the greatest value possible from your data.

Before we get into the specifics of the Data Flywheel, let's take a look at some leading businesses that are using the philosophies behind the flywheel to achieve data-driven success today.



How today's companies are finding success with data

FORTNITE

Fortnite is a massively popular online, multiplayer game. The game has become a phenomenon, now attracting more than 250 million players.

Epic Games uses AWS to drive real-time, data driven decisions that continuously improve gaming for over 10 million concurrent players, resulting in higher levels of player engagement during in-game events—such as the world's largest concert.

Fortnite is a free-to-play game where players attempt to be the last person standing in a 100-player "Battle Royale." Epic Games needed a way to process and analyze over 100PB of data (125M events/min) ingested from game clients and game servers to understand and adapt to player engagement.

Epic Games turned to AWS for an Amazon S3 data lake, Amazon EMR, Amazon EC2, and Amazon Kinesis, in addition to Amazon ElastiCache, Amazon DynamoDB, and Amazon RDS.

With AWS, Fortnite is continually able to prove the viability of its micro-transaction model and maintain its position as one of the most popular and successful games on the market, with a portion of that success led by real-time, data driven decisions.

[Watch the video >](#)



Peloton sells stationary bikes with 22" touch-screen monitors where riders can watch and interact with live and on-demand classes. The company has built a community of over a million riders who are connected to each other and instructors—all in the comfort of their own homes.

Peloton's business model is based on a subscription service. Apart from buying the equipment, users purchase a yearly or monthly subscription to access the live and on-demand classes. The renewal rate of that subscription is entirely dependent on one factor: user engagement.

One of the primary tools Peloton uses to drive user engagement is a capability it has built into its application called the "leaderboard."

The leaderboard lets users see the other riders currently using the application, showing their overall rank, comparisons to their own personal bests, and total power output—all in real time. This helps create an environment where riders can encourage each other, compete with each other, and compete with themselves.

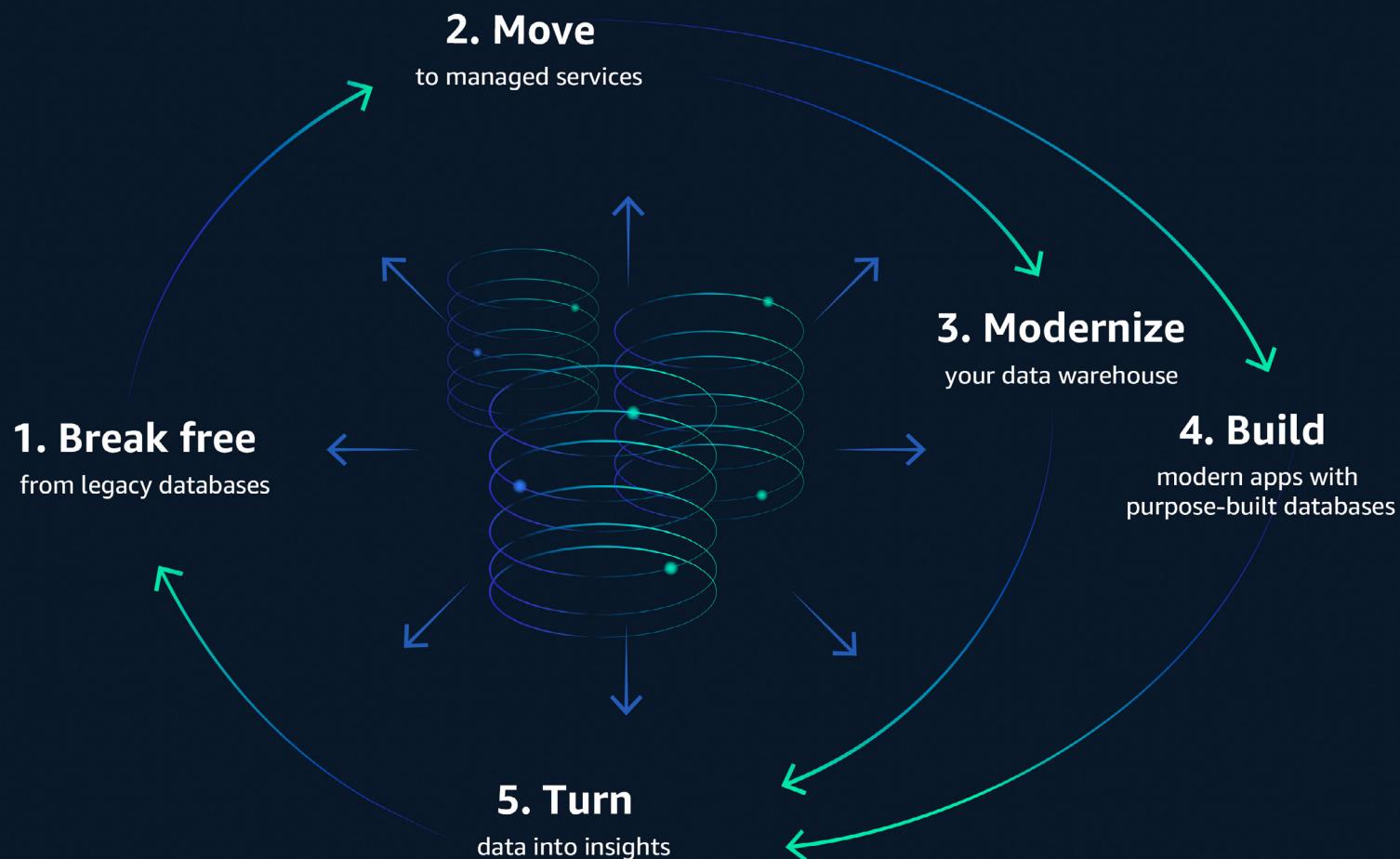
For the leaderboard to work, it must update tens of millions of data points—simultaneously and in real time. Peloton uses Amazon ElastiCache, a specialized, purpose-built in-memory cache database, to achieve this.

Peloton's user base exceeded one million riders within its first five years. With the leaderboard and other data-driven innovations enabling user experiences that turn customers into loyal, long-time subscribers, Peloton's rapid success offers further evidence of the transformative power of data.

[Watch the video >](#)

Get the most from your data with the Data Flywheel

Now that we have established the value of data, we'll explore the value of a self-sustaining business momentum through the Data Flywheel strategy. So how do you get started? The Data Flywheel consists of five core components. In the next sections, we'll explore each of these components individually.

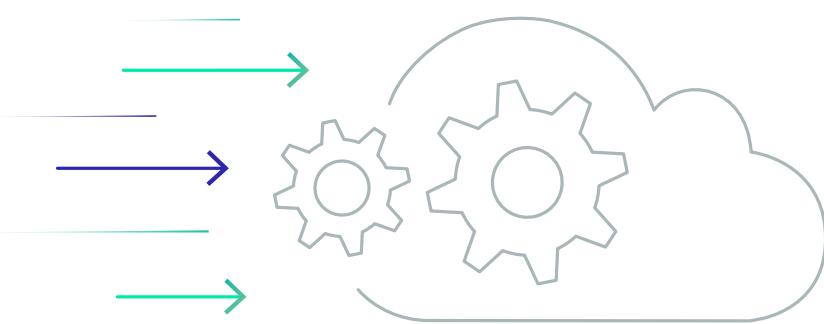


1 Break free from legacy databases

The world of databases has long been a challenging place for most organizations. Legacy databases are expensive, proprietary, create lock-in, offer punitive licensing terms, and come with frequent audits. Because of these issues and others, many companies are moving as fast as they can to open source alternatives like MySQL, PostgreSQL, and MariaDB. But while the simplicity and cost effectiveness of open source databases are certainly attractive, these businesses still need the performance and availability of high-end commercial databases. In short, they need the best of both worlds.

That's why AWS built [Amazon Aurora](#), a MySQL and PostgreSQL compatible relational database built for the cloud that provides the benefits of both high-end commercial databases and open source databases. Amazon Aurora offers 5x the performance of standard MySQL and 3x the performance of standard PostgreSQL, optimum availability and reliability, and hardened security—all at 1/10 the cost of a commercial-grade database.

Businesses are also leaving legacy relational databases in favor of purpose-built options to gain massive scale and speed. [Amazon DynamoDB](#) is a key-value and document database that delivers single-digit millisecond performance at any scale.



You might choose to move some databases to AWS first, as is, to immediately get the scalability, reliability, and ease of management that the cloud provides. This gives you the time you need to migrate any legacy code and exhaust the remaining term of your license agreements with other commercial database vendors. For these databases, you should consider using [Amazon Relational Database Service \(RDS\)](#)—which we'll discuss in more detail in the next section.

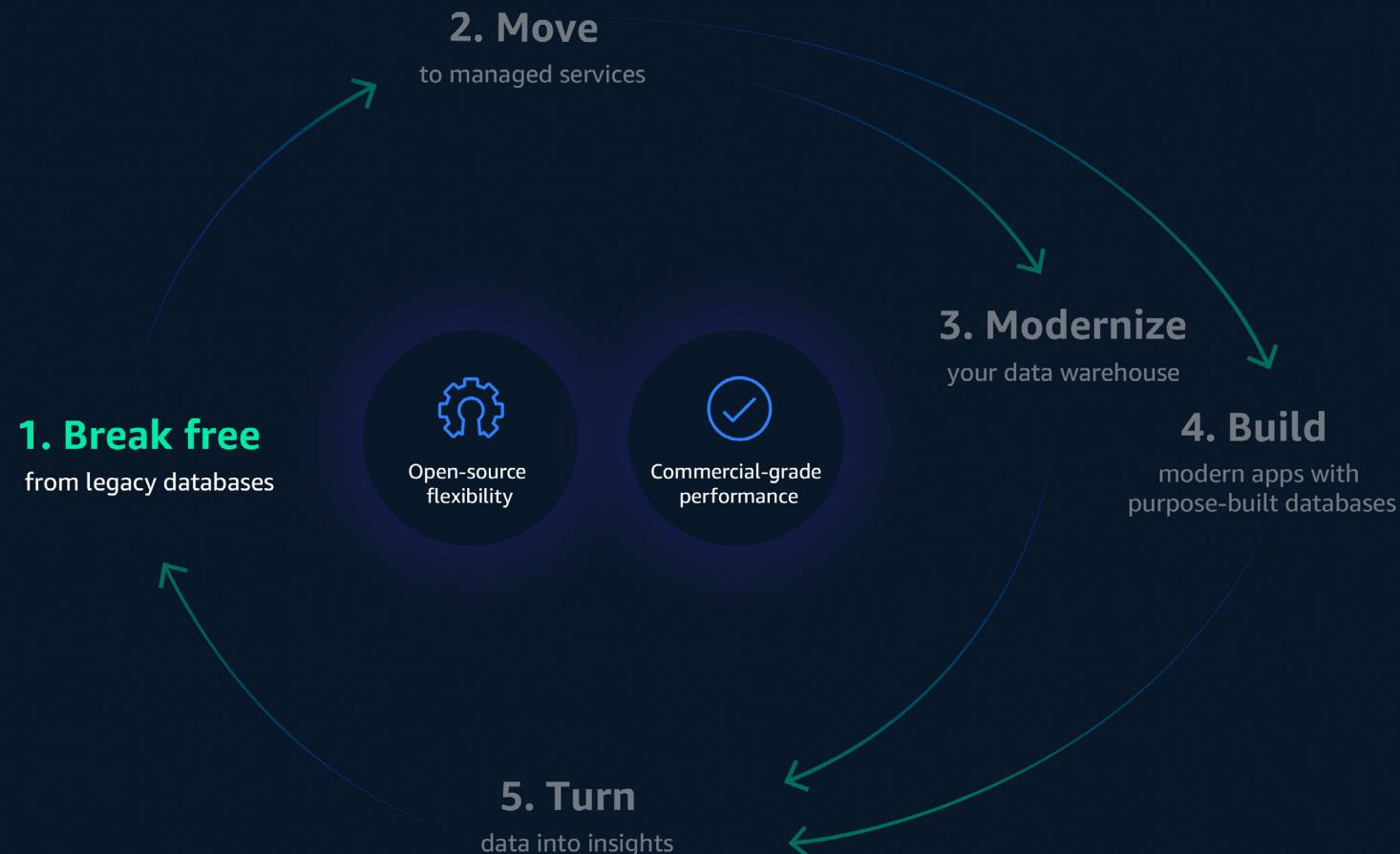
Migrating to AWS drives results

amazon

Reduces costs by more than 60%, improves performance, and gives admins more time to innovate after migrating its thousands of on-premises databases off Oracle to AWS

DOW JONES

Improves performance and reduces costs by more than 50% by modernizing their data storage with Amazon Aurora



2 Move to managed services

Managing your databases and analytics infrastructure can be tedious and time-consuming. Moving to fully managed services on AWS allows you to spend time innovating and building new apps, not managing infrastructure.

Why fully managed databases?

Organizations often begin their move to managed services by migrating to [Amazon Relational Database Service \(RDS\)](#), a fully managed solution that can run your choice of database engines—including open source engines as well as Oracle and SQL Server. Amazon RDS improves database scale and performance and automates time-consuming administration tasks such as hardware provisioning, database setup, patching, and backups.

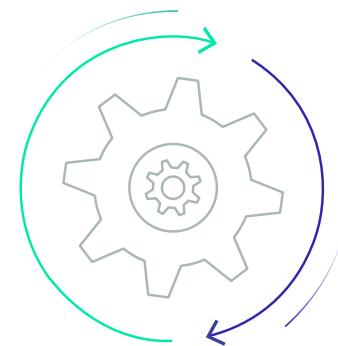
Amazon RDS is:

- **Easy to administer:** Easily deploy and maintain hardware, OS, and database software and use built-in monitoring
- **Performant and scalable:** Scale compute and storage with a few clicks and minimal downtime
- **Available and durable:** Get automatic Multi-AZ data replication and automated backup, snapshots, and failover
- **Secure and compliant:** Encrypt data at rest and in transit; take advantage of industry compliance and assurance programs

Move all your workloads to the cloud

To get even more benefits, move both your relational and non-relational databases to fully managed services in the cloud.

- **Relational databases:** Migrate from expensive Oracle and SQL Server databases to Amazon Aurora. Migrate from standard MySQL and PostgreSQL to Amazon RDS.
- **Non-relational databases:** Move document- and key-value stores to [Amazon DynamoDB](#) and document databases like MongoDB to [Amazon DocumentDB](#).
- **Apache Cassandra databases:** Migrate and run workloads using [Amazon Keyspaces](#), a fully-managed, serverless, Apache Cassandra-compatible database service that's highly available and scalable.



Why fully managed analytics?

Like their database counterparts, fully managed analytics services from AWS reduce costs and give you more time to dedicate to innovation. They also enable dynamic scalability, faster processing, easier visualizations, higher availability and resiliency, and stronger security.

- **Hadoop and Spark:** Move on-premises Hadoop and Spark deployments to [Amazon EMR](#) for time and cost savings
- **Operational analytics:** Elasticsearch, Logstash, and Kibana (ELK) on-premises can move to [Amazon Elasticsearch Service](#) for time and cost savings
- **Real-time analytics:** Apache Kafka deployments can move to [Amazon Managed Streaming for Apache Kafka \(MSK\)](#) to prepare, load, and analyze data streams into data stores and analytics tools for immediate use

More time, more savings, more innovation

Switching to fully managed databases and analytics services in the cloud allows you to offload tedious and time-consuming tasks to your cloud provider—which means you no longer have to spend time on the undifferentiated heavy lifting of common database and analytics administration.

Going the fully managed route will boost productivity and efficiency, but no matter who's managing them, your databases, analytics, workloads, and applications will all benefit from the agility and elasticity of the cloud, which allows you to store and process data on-demand. No more figuring out what data you need and what data to throw away—in the cloud, you can seamlessly scale your workloads up and down and pay only for what you consume.

Once you're taking advantage of the scalability of the cloud and running fully managed databases and analytics solutions, you can start focusing more time on building applications that deliver the performance and high availability your business needs and your customers demand.

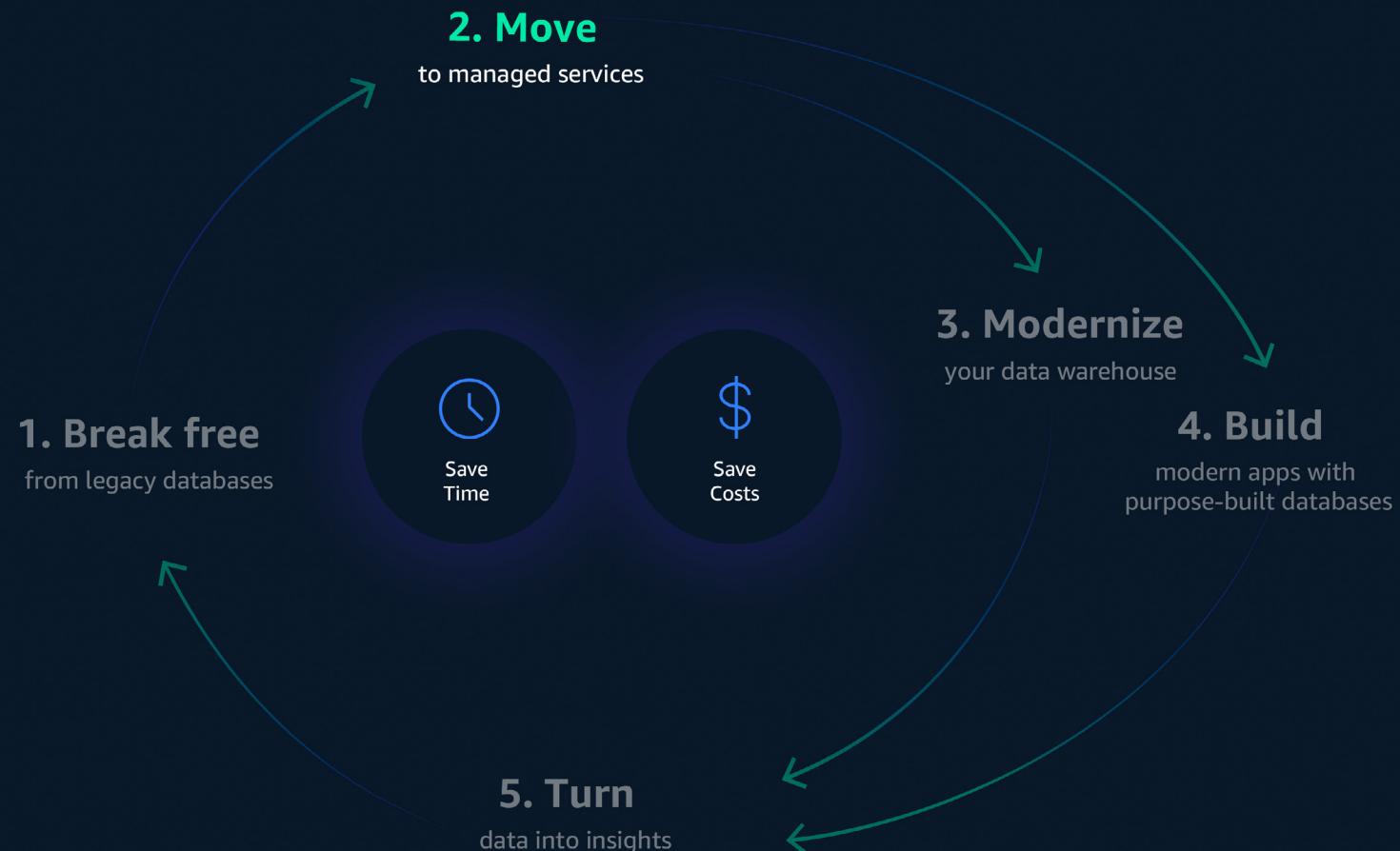
AWS fully managed services drive results

AUTODESK.

Transforms its "good enough" app into a much greater product, with real-time metrics, faster error detection, and shorter mean time-to-recovery using managed analytics services from AWS

FANDUEL

Moves critical workloads to AWS using Amazon Aurora, achieving almost 100% uptime



3 Modernize your data warehouse

Traditional approaches to data warehousing can't keep up with the needs of the modern business. That's because these strategies are:

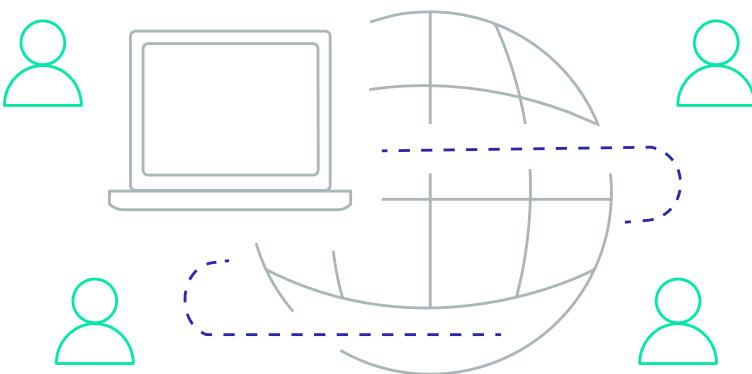
- **Hard to scale:** You'll have to purchase and install larger, more powerful hardware every time you reach storage and compute capacity limits
- **Too slow:** Transforming and moving all of the data you want to analyze into a data warehouse limits the insights you can generate and slows down the ability to answer new questions about your business
- **Expensive:** You'll need to purchase from old-guard database providers, which are expensive, proprietary, and impose punitive licensing terms
- **Rigid:** Traditional data warehouses can't suitably accommodate new data types generated by websites, mobile apps, and internet-connected devices
- **Siloed:** Older, siloed architectures don't give you the flexibility needed to use powerful analytics concepts and solutions, like data lakes or Hadoop
- **Complex:** Only a small set of individuals have access or the necessary skills to analyze data stores in older data warehouses

To give your data flywheel the push it needs to achieve self-sustaining momentum, you'll need to adopt a modern, cloud-based data warehousing strategy—one that allows you to store, process, and analyze more data, faster and more efficiently.

Discover the most popular and fastest cloud data warehouse

Modernizing your data warehouse with [Amazon Redshift](#) gives you the performance, scale, and deep integration you need to get the most value from your data.

Amazon Redshift is the most popular cloud data warehouse, with tens of thousands of customers. It easily integrates with your existing data architecture, allowing you to query petabytes of data across your data warehouse, data lake, and operational databases. Amazon Redshift is also 3X faster and 50% less expensive than other cloud data warehouses.



Modernizing data warehouse strategies with AWS drives results

nielsen

Ingests, processes, and reports
significantly more data to its clients each day

EQUINOX

Uses a modern data lake architecture to enable powerful analytics and foster experimentation and innovation



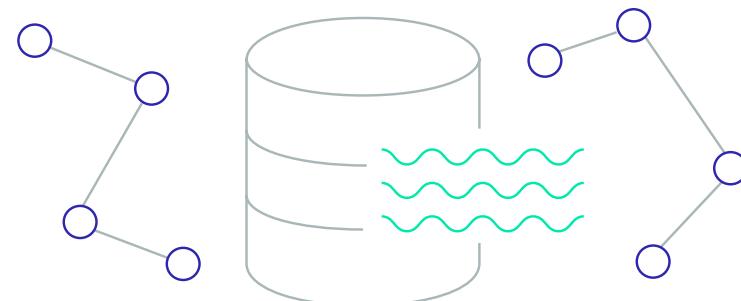
4 Build modern apps with purpose-built databases

Building high performing, scalable, and available applications that will grow with your business requires a new set of rules. You'll need to change your database strategy and leave complex, monolithic designs behind in favor of highly distributed apps comprised of microservices. While these may sound like two separate imperatives, they're actually two parts of a larger whole. Embracing newer purpose-built database types makes the shift to microservices easier, allowing your developers to use the right tool for the right job when building modern apps.

Modern database types for modern app requirements

Modern applications place a new set of requirements on databases. Today's apps need databases to scale from terabytes to petabytes of data, support millions of concurrent users (often around the globe), and deliver performance with millisecond to microsecond latency. They also need to be built with scalability, economics, and instant API developer access in mind. And many of them will need separate solution sets and architectures for multiple user interfaces across web, devices, and IoT.

Due to these more complex requirements, the one-size-fits-all approach of using relational databases for all app development must evolve. Relational databases have been the cornerstone of every enterprise application for the past two decades, and they are still essential—in fact, they are still growing.



But with the rapid growth of data—not just in volume and velocity, but also in variety, complexity, and interconnectedness—a “relational only” approach no longer works.

Using the right purpose-built databases for the job can more easily provide the real-time performance—with little to no downtime—that modern applications need when scaling to meet these new requirements. Like Airbnb, Lyft, and so many others, businesses of all sizes are using purpose-built databases today to build scalable and globally distributed applications.

Types of purpose-built databases include:

- Relational databases for transactional applications
- Key-value databases for internet-scale applications
- In-memory data stores for caching and real-time workloads
- Graph databases for building applications with highly connected data
- Document databases to store semi-structured data as documents

- Time series databases for collecting, synthesizing, and deriving insights from time-series data
- Ledger databases for when you need a centralized, trusted authority
- Wide column databases for fast management of massive amounts of data

With these purpose-built databases, your business can innovate more rapidly to build modern applications. These applications take full advantage of your data, using it to create new capabilities, improve existing functions, and enhance customer experiences.

The process for building modern applications will differ greatly based on business needs and use cases. But in most instances, modern application design will involve these three fundamental considerations:

- 1 Breaking down complex applications into highly distributed microservices
- 2 Choosing the appropriate purpose-built database type
- 3 Examining every application component to ensure proper architecture and scalability

AWS purpose-built databases:

							
Relational: Amazon Aurora and Amazon RDS	Key value: Amazon DynamoDB	Document: Amazon DocumentDB	In-memory: Amazon ElastiCache	Graph: Amazon Neptune	Time series: Amazon Timestream	Ledger: Amazon QLDB	Wide column: Amazon Keyspaces (for Apache Cassandra)

AWS purpose-built databases drive results



Uses Amazon DynamoDB to store user search history for consistently fast lookups even at internet-scale, Amazon ElastiCache to store session states in-memory for faster site renderings, and Amazon RDS as its primary database for transactional workloads



Uses Amazon DynamoDB to achieve 24,000 reads/second with its 31 million item cloud databases, Amazon Aurora for transactional user data, and Amazon ElastiCache for instant access to common words and phrases



5 Turn data into insights

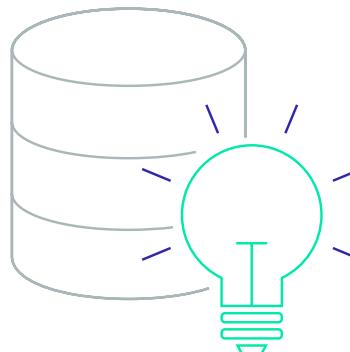
Your data is in the cloud—now make it work for you. Forget silos and incomplete analytics. Today's businesses gather smart, accurate insights faster through data lakes, cloud analytics, and machine learning and empower end-users to see and visualize their data from any device or application.

Data is growing exponentially

Data is a difficult beast to tame. It's growing exponentially, coming in from new sources, becoming more diverse, and is more and more difficult to process. If you are building modern applications, those applications generate new data on customer patterns and preferences that must also be captured, stored, and processed.

In today's world, traditional, on-premises data warehouses and big data solutions struggle to produce the analytics your business needs. These traditional solutions are usually:

- **Siloed:** With data trapped in silos, seeing the big picture is nearly impossible
- **Delayed:** Analytics based on yesterday's data arrive too late to gain competitive advantage today



- **Expensive:** When performing analytics costs more than the value of their insights, nobody wins
- **Cumbersome:** You might get an analytics platform running, but few will likely be able to use it efficiently
- **Limited to specialists:** Making more data-driven decisions means democratizing analytics

With all that in mind, how can your business possibly capture, store, and analyze its data at the near-real-time rate needed to remain competitive?

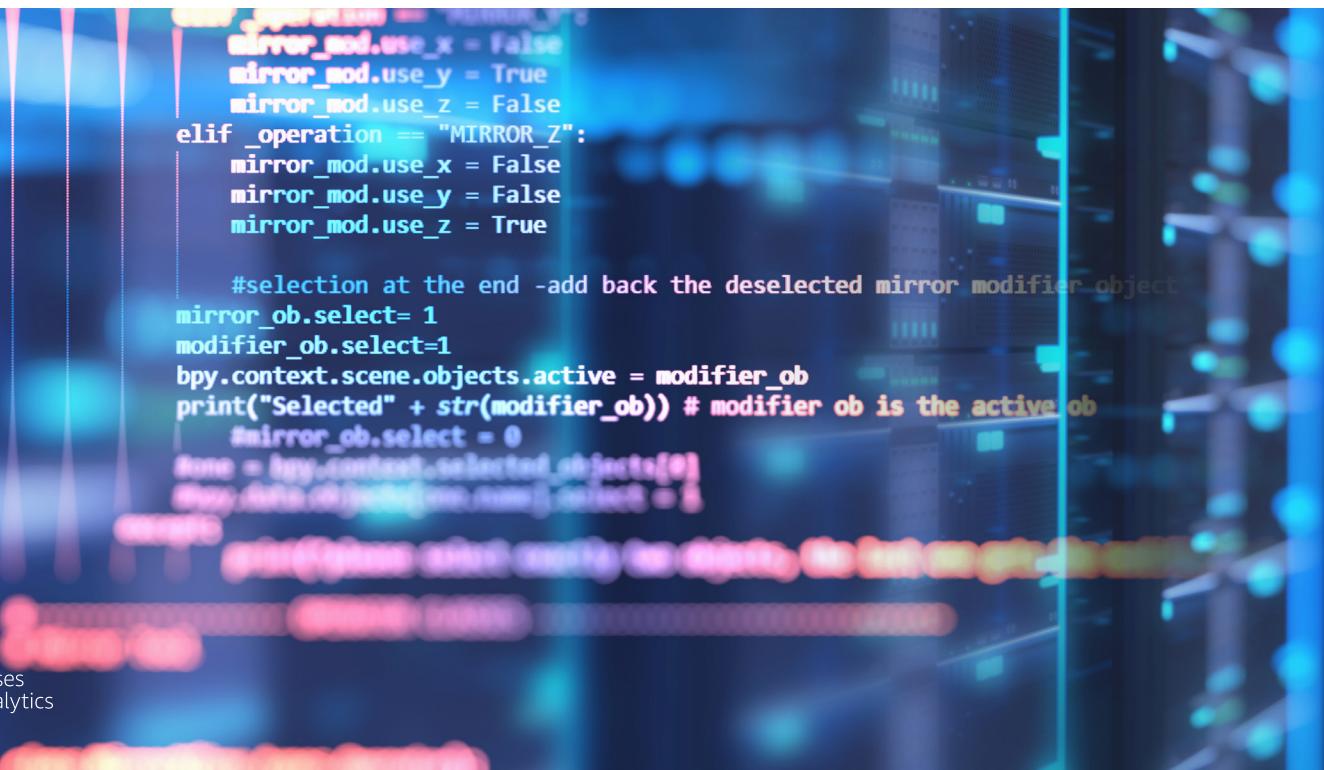
Building a secure data lake is challenging

Manually creating and implementing a fully productive data lake generally requires months of tedious, complicated work. You'll need to set up storage to handle massive amounts of data, collect and organize data from various sources, clean the data to prepare it for use, configure and enforce complex security policies, and find ways to make the data easy to locate. Thankfully, there's a faster, easier solution.

Data lakes on AWS are different

[AWS Lake Formation](#) simplifies the lake creation process and automates many steps, enabling you to set up a secure data lake in days, not months. AWS Lake Formation provides:

- **Faster setup:** Accelerate and automate the moving, storing, cataloging, and cleaning of your data
 - **Broader security:** Enforce security policies across multiple services
 - **More insights:** Empower your analysts and data scientists to gain and manage new insights



AWS data lake and analytics solutions

AWS provides the most comprehensive, secure, and cost effective portfolio of services for every step of building a data lake and analytics architecture. These services include data migration, cloud infrastructure, management tools, analytics services, visualization tools, and machine learning.

Big data processing

Amazon EMR: Process vast amounts of data using the Spark and Hadoop frameworks—easily, quickly, and cost effectively. This managed service supports 19 different open-source projects, with managed EMR Notebooks for data engineering, data science development, and collaboration.

Data warehousing

Amazon Redshift: Run complex analytic queries against petabytes of structured data—at less than a tenth the cost of traditional solutions. Includes Redshift Spectrum, which runs SQL queries directly against exabytes of structured or unstructured data in S3, without the need for unnecessary data movement.

Streaming and real-time analytics

Amazon Managed Streaming for Apache Kafka (MSK): Fully managed, highly available, and secure service that makes it easy for you to build and run applications that use Apache Kafka to process streaming data.

Amazon Kinesis: Collect, process, and analyze streaming data—such as IoT telemetry data, application logs, and website clickstreams—as it arrives in your data lake. Respond in real-time instead of waiting until all your data is collected.

Operational and log analytics

Amazon Elasticsearch Service: Search, explore, filter, aggregate, and visualize your data in near real-time. Enable fast, simple application monitoring, log analytics, clickstream analytics, and more. Get easy-to-use APIs alongside the availability, scalability, and security that production workloads require.

Recommendations

Amazon Personalize: Improve customer engagement with machine learning-powered product and content recommendations, tailored search results, and targeted marketing promotions. No prior machine learning experience necessary—you provide an activity stream from your app, and the solution will identify what is meaningful, select the right algorithms, and train and optimize a customized personalization model.

Predictive analytics

Amazon SageMaker: Easily build, train, and deploy predictive machine learning models. Get everything you need to connect to your training data, select and optimize the best algorithm and framework, and deploy your model on auto-scaling clusters of Amazon EC2.

Visualizations

Amazon QuickSight: Easily deliver insights to end-users, whether you are building interactive dashboards for your organization, or embedding analytics in your applications or websites with fast performance, machine learning insights, serverless architecture, and pay-per-session.

Get more insights with machine learning

[Machine learning](#) can provide better customer experiences at a scale that was not possible before. With machine learning, your business can gain valuable insights into its data with ease and speed.

While machine learning has yet to reach its full potential, we have hit a tipping point—the cloud has made machine learning available to businesses of every size.

Machine learning for all

AWS has the broadest and deepest set of machine learning and AI services for your business. AWS is focused on solving some of the toughest challenges that hold back machine learning from being in the hands of every developer and data scientist. You can choose from pre-trained AI services for computer vision, language, recommendations, and forecasting; Amazon SageMaker to quickly build, train, and deploy machine learning models at scale; or build custom models with support for all the popular open-source frameworks.

While these innovations enable developers and data scientists of all skill levels—even those with no prior machine learning experience—to create sophisticated models and systems, many businesses still struggle to know where to get started.

In nearly all cases, starting small will maximize your chances of machine learning success. The first step in your machine learning journey should be to identify and develop a single pilot project. This will give you a path to learn, innovate, and expand. All machine learning pilots will be different, but we suggest following these basic steps to begin:

- 1. Start with data strategy:** A good starting point is to collect, prepare, and store relevant data.
- 2. Identify the business problem:** Find your best use cases by considering how machine learning can improve customer experiences, increase efficiency and productivity, and drive innovation.
- 3. Choose your project:** Evaluate possible pilot projects focused on your priority use cases, then determine what tools, skills, and budget are needed.

AWS data lake and machine learning services drive results



Reduces data retrieval time from months to minutes and extends data access to more users

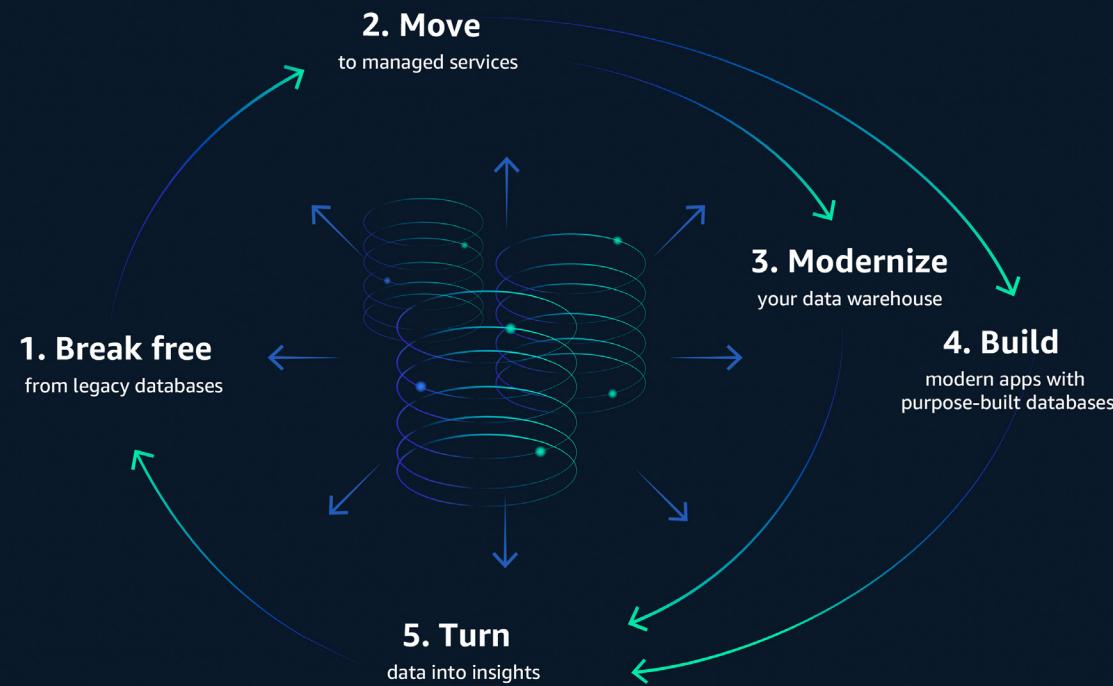


Increases self-service data querying and visualization



Delivers personalized experiences that drive engagement and reduce returns





Spinning the Data Flywheel

The Data Flywheel begins and ends with data—and with the integration of new data continuously fueling it as it spins. Modern apps, analytics services, and machine learning systems create more data that can be stored and managed in the cloud through data warehouses and data lake architectures. That allows your organization to develop new capabilities and apps through the use of purpose-built databases, gain new insights, and deliver new products. This is a stepwise, repeatable process which must be run project-by-project—like turning a flywheel, building momentum with each turn.

One of the most unique characteristics of the Data Flywheel is that no “one thing” powers it, and organizations that search for such a fundamentally basic solution will likely lose their way. The Data Flywheel moves by many components acting in concert, equating to a whole that’s greater than the sum of its parts.

As you work to develop the components outlined in this eBook—implementing the most relevant technologies and procedures at each phase—your flywheel will begin to generate perpetual business momentum. The flywheel will help put your data to work, automating data management and analytics and giving you and your team the time needed to make smarter, data-driven decisions—while directing more resources toward the business objectives that matter most.

Want to learn more about how the Data Flywheel can help your organization achieve perpetual business momentum? Visit our interactive Data Flywheel site to dig deeper into each of its components.

[Interact with the Data Flywheel >](#)