Brought to you by:

ORACLE | AMD

Next-Gen Data Platforms

for dummies®
A Wiley Brand

Eliminate inefficient data silos and integrations

Maximize the value of data with AI, ML, and analytics

Meet your end-to-end data needs

Oracle and AMD Special Edition

Lawrence Miller

About Oracle

Oracle offers integrated suites of applications plus secure, autonomous services in more than 40 Oracle Cloud data centers around the world as well as customer data centers in more than 60 countries. Our next-gen Data Platform builds on this cloud infrastructure and Oracle's more than 40 years of experience in engineering mission-critical solutions. It provides end-to-end data management capabilities and integrates transaction processing, analytics in data warehouses and data lakes, and AI/ML services into a comprehensive, open, and modular architecture. Oracle Data Platform requires less integration and fewer services than many other approaches while helping customers eliminate isolated data and application silos. With Oracle Data Platform, organizations can improve customer experiences, develop innovative applications in less time, and achieve deeper and more timely business insights. For more information about Oracle (NYSE: ORCL), please visit us at www.oracle.com, on Twitter at www.twitter.com/oracle, and on LinkedIn at www.linkedin.com/company/oracle.

About AMD

AMD is the high performance and adaptive computing leader, powering the products and services that help solve the world's most important challenges. Our technologies advance the future of the data center, embedded, gaming and PC markets. The AMD EPYC™ processors used in Oracle Exadata Cloud Infrastructure accelerate application performance and responsiveness on business-critical workloads, enable organizations to analyze all types of data to fully understand their business, help secure transactions and protect data, and deliver outstanding returns on IT investments. For more information about AMD (NASDAQ: AMD), please visit us at www.amd.com, on Twitter at www.twitter.com/AMD, and on LinkedIn at www.linkedin.com/company/AMD/.

Next-Gen Data Platforms

Oracle and AMD Special Edition

by Lawrence Miller

for dummies®
A Wiley Brand

Next-Gen Data Platforms For Dummies®, Oracle and AMD Special Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

ISBN 978-1-394-20696-4 (pbk); ISBN 978-1-394-20697-1 (ebk)

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

Project Editor: Elizabeth Kuball
Acquisitions Editor: Traci Martin
Editorial Manager: Rev Mengle
Client Account Manager: Jeremith Coward
Production Editor: Tamilmani Varadharaj

Table of Contents

Introduction

As organizations move data and workloads to the cloud, they create new opportunities to increase enterprise value and tear down barriers that restrict data access and stifle innovation. When done right, moving to a public cloud, hybrid cloud, private cloud, or multicloud environment helps businesses operate more efficiently and develop deeper, more timely insights into their data.

A next-gen data platform, built on an open and modular architecture, allows organizations to meet their current needs and add capabilities as required. Inside this book, you'll discover how a next-gen data platform can help your organization maximize the value of its data.

About This Book

Next-Gen Data Platforms For Dummies, Oracle and AMD Special Edition, consists of six chapters that explore the following:

- The business need for a complete data platform, as well as its requirements, architecture, and benefits (Chapter 1)
- Common challenges and limitations of existing data platform solutions (Chapter 2)
- Empowering users to make better and faster decisions (Chapter 3)
- How you can address end-to-end data requirements with Oracle Data Platform (Chapter 4)
- How to increase application performance and resilience with Oracle Data Platform (Chapter 5)
- Important questions you need to ask when evaluating a complete data platform (Chapter 6)

Each chapter is written to stand on its own, so if you see a topic that piques your interest feel free to jump ahead to that chapter.

Foolish Assumptions

Mainly, I assume you have a good understanding of your organization's current data challenges and data platform needs. This means that you are a vice president of IT, chief data officer, senior line of business manager/application owner, development manager, data manager, or senior IT architect or database administrator.

If any of these assumptions describes you, then this is the book for you! If none of these assumptions describes you, keep reading anyway! It's a great book and you'll learn what a next-gen data platform is, how it can benefit businesses, and why you need one for your organization.

Icons Used in This Book

Throughout this book, I occasionally use special icons to call attention to important information. Here's what to expect:

REMEMBER

This icon points out important information you should commit to your nonvolatile memory, your gray matter, or your noggin.

TECHNICAL STUFF

If you seek to attain the seventh level of nerd-vana, then perk up! This icon explains the jargon beneath the jargon.

TIP

Tips are appreciated, but never expected, and I sure hope you appreciate these useful nuggets of information.

WARNING

These alerts point out the stuff your mother warned you about. Well, probably not, but they do offer practical advice.

Beyond the Book

There's only so much I can cover in this short book, so if you want to learn more, go to `www.oracle.com/data-platform`.

IN THIS CHAPTER

- Evolving from data warehouses and data lakes to a next-gen data platform
- Understanding business requirements for next-gen data platforms
- Adopting a comprehensive, open, and integrated data platform
- Recognizing important business benefits with next-gen data platforms

Chapter 1
What Is a Next-Gen Data Platform and Why Is It Important?

This chapter explores the business needs that are driving organizations like yours to consider data platforms, the key requirements they're looking to meet, and the characteristics that define a next-gen data platform.

Understanding the Business Need for a Next-Gen Data Platform

Data is key to success for modern organizations. The right data and analytics delivered to the right people at the right time can enable tremendous business opportunities.

Historically, data used by one application was isolated from data used by other applications. Segmenting data was done to improve predictable performance and because storage costs were high.

These isolated sets of data — known as *data silos* — create their own challenges, most notably that they're extremely inflexible and inefficient. It's difficult to integrate data across silos and to bring new data sources into the workflow. The result is highly complex environments that are costly and challenging to manage and secure.

Data platforms, on the other hand, are intended to meet the end-to-end data needs of an organization. They should provide an integrated environment that accepts many types of data from many sources, support database capabilities for online transaction processing (OLTP) and data warehousing, make vast volumes and varieties of data available in data lakes, and enable data to be analyzed using traditional business analytics, as well as modern artificial intelligence (AI) and machine learning (ML) to build predictive models (see Figure 1-1).

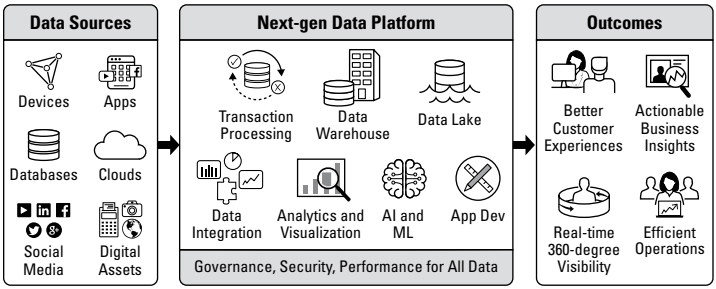

FIGURE 1-1: Key data platform capabilities.

The end-to-end nature of a data platform is important because it allows data to easily flow into it from many different sources so organizations can make decisions using the most current versions of that data. For example, in multi-channel retail environments it's important to manage orders and inventory to eliminate "out of stock" notices. Similarly, manufacturing companies use Internet of Things (IoT) readouts to continuously track and analyze equipment health, use AI and ML to identify potential failures, and perform preventive maintenance to keep production lines running smoothly. There are similar examples in every industry that show the value of making decisions based on real-time data instead of using stale data that's a day or week old.

Unfortunately, many organizations face deeply rooted challenges that make it difficult to build useful and capable data

environments that meet their end-to-end needs. Some of these challenges include the following:

- It's difficult to create a 360-degree perspective for end users and decision-makers because data is siloed and scattered across business units, applications, and clouds.
- Accessing data and moving it between different services is complex and takes time, leading to stale data that can negatively impact results.
- Innovation is stifled because the specialized databases and services used in the current environment don't support the new data types and advanced capabilities such as AI and ML that different lines of business require.
- People who need access to data and analytics often can't access it, can only see part of what they need, don't have easy-to-use tools, or can't reconfigure analyses to address their current questions.
- The high cost of running inefficient data silos and multiple specialized databases means that less funding is available for innovation.
- Applications that serve customers and users around the world need to be available 24/7, but most service providers can't provide the required levels of availability.

Over the past several years, providers have taken different approaches to create a data platform. Some have promoted data warehouses and analytics as a data platform — which is a good starting point, but one that leaves out rich sources of data such as OLTP or semi-structured and unstructured data. This approach requires the use of additional tools to move, cleanse, and reformat the data for analysis in a data warehouse.

Another approach has been to equate a data platform with a data lake that provides a cost-effective way to store "everything" for future analysis. Data lakes don't enforce structure or format, requiring organizations to integrate multiple advanced and, in many cases, custom tools to make sense of the data. Data lakes, usually built on low-cost, cloud-based object storage, don't provide the high performance needed for structured OLTP and data warehouse use cases. In addition, specialized data scientists may be needed to extract value from the data — which doesn't support the business goal of democratizing data access and analysis for everyone who needs it.

Defining Key Requirements for a Next-Gen Data Platform

Today, organizations' needs for data management have evolved and the variety, volume, and velocity of data being generated and analyzed has increased exponentially. Key requirements for a next-gen data platform include the following:

- **Unified, up-to-date data:** Modern applications and analytics combine multiple types of data from multiple sources to improve customer experiences and business insights. Multifaceted teams of end users, data engineers, and analysts need the scale, performance, and consistency offered by data platforms so decisions aren't made using stale or inconsistent data.
- **Versatility and adaptability:** Business requirements change over time, but you don't want to have to reinvent your data platform every time they do. Data platforms need to be versatile enough to handle a wide range of data types and workloads and be ready to solve newly identified business needs.
- **Support for third-party and open-source capabilities:** Some of the most popular analytical tools come from independent software vendors (ISVs) and the open-source community. This broad ecosystem drives innovation and encourages sharing of best practices, so organizations need to be able to easily integrate these capabilities.
- **Integration of AI and ML:** AI and ML technologies are used in many applications of data platforms such as real-time systems and analytics for decision-makers. Data platforms must support pretrained and trainable services to provide maximum value and enable organizations to develop their own AI and ML techniques.
- **Modularity with pay-as-you-go pricing:** Individual workloads may only use a subset of a data platform's capabilities and have variable pricing driven by customer demand or business planning cycles. Paying only for the data platform services that you need and for your actual usage helps keep costs down.

- **Support for multicloud:** Many organizations implement multicloud architectures that combine the most capable cloud services for different use cases. Data platforms need to integrate support for data sources, applications, and services coming from multiple clouds.

Architecting a Next-Gen Data Platform

To address the end-to-end data requirements for an organization's application portfolio and business use cases, a data platform is composed of a suite of services that is

- **Comprehensive:** It has a set of capabilities that meet current and future enterprise-wide requirements. Over time, customers can reconfigure services to meet their ever-changing business needs. It should support traditional and cloud-native applications, virtually any use case (for example, OLTP, data warehousing, analytics, ML, and AI), and with almost any data type including relational, spatial, graph, document, and unstructured data. It is not limited to a single data type or use case.
- **Open:** It has a modular architecture that enables organizations to extend or replace out-of-the-box capabilities with additional ones developed by other suppliers, the open-source community, or internal teams. It is not limited to a single type of deployment and allows data to be easily moved to other platforms or locations.
- **Integrated:** It minimizes the number of separate services, integration points, and data transformations required to meet end-to-end enterprise needs. Its versatile capabilities allow developers to spend more time on innovation and less time on integration. High levels of automation reduce error-prone manual administration and fragmented management found in do-it-yourself (DIY) solutions.

A next-gen data platform's comprehensive suite of capabilities simplifies data management across an enterprise. Its full suite of services increase business agility by reducing IT complexity and bottlenecks, improving data governance and security, lowering data management costs, and democratizing access to data and advanced analytics.

Realizing the Business Benefits of a Next-Gen Data Platform

Next-gen data platforms enable organizations to realize important business benefits that range from improving end-user experiences with compelling, always-available content to enabling decision-makers to gain deeper insights and make better predictions using the most current data from across and beyond the enterprise.

With next-gen data platforms, users throughout the organization have access to the data they need to make decisions through easy-to-use tools and built-in capabilities that bring advanced analytics, AI, and ML capabilities within reach. Democratized access to data and analytics allows organizations to react more quickly to external events and competitive pressures, develop greater competitive differentiation, and make better use of their organization's expertise.

Organizations also benefit from an operational and cost perspective because they can use fewer, more versatile services that use automation to reduce administration, dynamically scale consumption to meet variable demand, and operate 24/7 without downtime.

Key business benefits of a next-gen data platform are summarized in Figure 1-2.

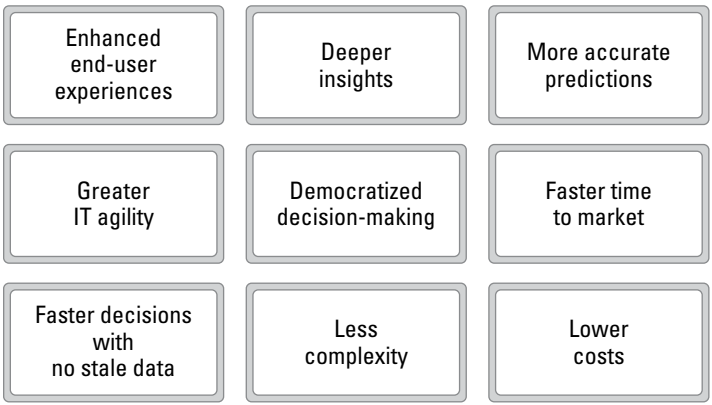

FIGURE 1-2: Business benefits of next-gen data platforms.

IN THIS CHAPTER

- Looking at different data platform approaches
- Working with a single-purpose platform
- Building your own data platform
- Addressing availability, performance, and efficiency challenges

Chapter 2
Recognizing Common Data Platform Speed Bumps

In this chapter, you learn about some of the most common limitations and challenges that organizations encounter with data platforms.

Not All Data Platforms Offer Equal Value

Vendors typically promote their data platforms as "one-stop shops" for managing and analyzing all types of data, but dig a little deeper and you'll quickly discover that data platforms are not created equal. Many have limited functionality — often just analytics with relational data. These are important workloads, but they don't address the needs of organizations to support documents, other types of structured data, and unstructured data. Nor do they meet the diverse needs organizations have for data integration, transaction processing, data lakes, artificial intelligence (AI), machine learning (ML), and analytics.

Organizations can try to create a complete data platform by extending a single-purpose one or by creating their own by combining point services. In both cases, each service adds complexity and requires specialized management and integration expertise.

Neither extending single-purpose data platforms nor building your own creates a compelling, comprehensive data platform. Instead, it's a collection of parts that leads to a difficult to manage "Frankenstack" (see Figure 2-1).

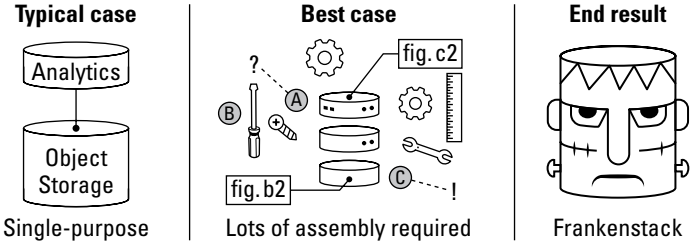

FIGURE 2-1: Common data platform approaches require extensive integration.


TIP

While multiple services will typically be required in a data platform deployment, minimizing the number of services and integration points reduces complexity and costs.

Single-Purpose Data Platform Challenges

Every organization runs multiple applications. Some are transaction oriented, some are analytics focused, and others combine multiple capabilities. Single-purpose data platforms don't address this breadth of use and may not encourage data sharing or generate timely results any better than applications running in data silos.

For instance, what happens if you start with a data platform that only supports analytics and you need to run transaction processing? You add another database service and probably a data transformation service, thereby increasing complexity, driving up cost, and delaying analytics. What if you also need to work with documents? Just add another database, data transformation service, time, and costs. What about training ML models and running them against real-time transactional data? Same answer, you need more, more, and more.

The list goes on and on, and when you combine requirements from across the organization, you end up with a data platform Frankenstack. Moving your workloads to the data platform becomes a daunting task of restructuring data, refactoring applications, and retraining users.

Do-It-Yourself Data Platforms Are Complex

What if you select a data platform approach that offers a smorgasbord of point data and integration services? That would be better, right?

Not really. A do-it-yourself (DIY) data platform assembled from a menu of point services forces you to use many different ones for each workload. If you have enough IT staff, this approach may be tempting, but the integration, complexity (there's a different stack for each workload), and the resulting data and management fragmentation make them difficult to evolve and manage. You spend a lot of time and still have a Frankenstack.

WARNING

Many data platform vendors require organizations to combine numerous single-purpose databases and complementary services to support even simple workloads. For example, Amazon Web Services (AWS) offers 16 different database services and Google Cloud offers 10 of them — and that's just for the database portion of a data platform.

It's easier to have fewer, more capable services than to go down the perilous path of creating custom solution(s) and having to manage them. Challenges of the DIY approach include the following:

- Developing new apps that link point services together or refactoring existing code to fit a different environment
- Maintaining proprietary code after the individual or team that created it is no longer with the organization
- Greater solution complexity, including patching and securing individual services and the complete stack
- The need for time and expert resources to customize and assemble point services into a data platform

As a result, the DIY data platform decision is often a more costly and complex solution. Organizations are increasingly rejecting the "some (or lots of) assembly required" model of data management that both single-purpose and DIY data platforms require.

Instead, these organizations are looking for a pre-integrated, next-gen data platform to manage and extract value from all their data assets. They want a complete solution that requires minimal integration, incurs low setup costs, and consumes resources efficiently, but they also want an environment that is modular so they can incorporate their preferred software components and evolve as needs change.

Not All Platforms Provide the Needed Compatibility, Availability, and Performance

Single-purpose and DIY data platforms are intended to provide performance and availability through scale-out designs. This approach may work for some use cases, but for others it requires extensive application-level coding and workload scripting.

Most organizations have existing application portfolios that are central to their business, that are proven to meet their needs, and that individuals are trained to use. Most organizations won't throw out existing applications just to use a data platform. Instead, they're looking for a data platform that will deliver the performance and availability needed by current workloads and new, cloud-native ones.

Existing applications can benefit from accessing additional types of data and sharing data with other applications through a data platform. However, these applications and the databases that support them, need a data platform with built-in high performance and availability.

Finally, not all data platforms can run where they're needed — particularly if data residency is needed. It's important to look for solutions that can be deployed in different geographic locations.

IN THIS CHAPTER

- Recognizing the integral role of analytics in a next-gen data platform
- Addressing stale data
- Leveraging artificial intelligence and machine learning
- Getting the right data to the right people when and where it's needed

Chapter 3
Enabling Everyone to Make Better and Faster Decisions

In this chapter, you learn how the role of modern business analytics has changed and how a next-gen data platform addresses some common analytics challenges.

Data and Analytics Have Evolved

Analytics is a key part of a next-gen data platform — it's not an afterthought or a "nice to have," but rather an integral part of the design. Analytics has evolved beyond tools for accessing data and rendering visualizations like pie charts, to include complex analyses of multifaceted data with artificial intelligence (AI) and machine learning (ML). Using modern analytics in a data platform is like having a thermostat that automates control of your house's temperature instead of a thermometer that just tells you that it's cold — something you already know. With a data platform, analytics supports everyone in an organization, from business leaders to technical people.

REMEMBER

Analytics is the process of discovering, interpreting, and communicating patterns in data. It helps generate meaningful insights and predictions so people can make better-informed decisions that lead to increased sales, optimized production, less staff turnover, and other improvements that help organizations succeed.

The need for analytics can arise from predictable business cycles (end-of-month or end-of-quarter sales reports) or from ad hoc requests to support real-time decision-making. Ad hoc requests often require looking at data in new ways, placing a heavy burden on IT teams and data scientists who also handle regular business requests. This creates delays for everyone, with results sometimes arriving too late to support time-sensitive business decisions.

Modern analytics has evolved with the variety, volume, and velocity at which data is generated. Analytics within a next-gen data platform can easily scale up and down to meet an organization's needs without burdening it with excessive costs. When combined with self-service access, analytics and AI/ML capabilities no longer need to be relegated to an elite few within an organization.

The End of Stale Data

Data in online transaction processing (OLTP) applications is optimized for rapid access, not business analytics. Additionally, analytics tools may not be allowed to directly access production data due to the risk of slowing down the original mission-critical system. To avoid conflicts, separate data copies are often created and transformed into a format that's optimized for analytics. After the data is in the analytics system, it's combined with information from other internal and external data sources before it can be analyzed and presented in useful ways.

Delays caused by moving, transforming, and integrating data can make it "stale," which means that it represents the state of the business at some time in the past. Stale data percolates through the decision-making process and leads to inaccurate analyses and suboptimal decisions. Using stale data is like deciding to wear a coat based on last month's weather — you might be okay, but you're more likely to be too hot or too cold.

A next-gen data platform includes analytics that enable anyone to analyze the most current information in real time or near real time without time-consuming transformation and integration processes.

The Rise of Self-Service AI, ML, and Analytics

Business expectations for analytics have changed. Simply being data-driven — using historic data in reports for decision-making — is no longer sufficient. Modern organizations must become analytics driven, using AI and ML within analytics to identify new opportunities and predict possible business outcomes.

Many organizations see ML as an "out-of-reach" technology because they don't have the expertise to build ML models, workflows are too complicated, or they don't have powerful enough systems to train and run ML models. However, modern cloud-based analytics environments with built-in ML capabilities have greatly simplified the user experience, making AI and ML accessible to practically any organization.

With AI and ML seamlessly embedded within the analytics environment, rather than as separate tools and processes, anyone can benefit from them without needing to have specialized data science skills. Organizations with data science teams can achieve additional benefits by developing advanced models for distribution within the analytics platform. This increases the efficacy of the resulting analyses and delivers self-service AI and ML to everyone.

Similarly, data engineers can focus on producing data products that support business processes, decisions, and advanced models. Anyone can run ML models from within the analytics environment using these data products and their own data sets with no coding required. From the user's perspective, they're just analyzing their data, and don't need to understand the AI and ML techniques or data engineering that made this possible. This process fundamentally transforms the culture of the organization, helping it become data driven.

TECHNICAL STUFF

ML is a domain within AI that enables applications to automatically learn from data and improve models based on known results that they've been trained with. ML has two key steps. The first step is for data scientists to develop and train models using massive amounts of data (model training is prebuilt or automated in some ML programs). The second step is running those models on new data to make inferences, recommendations, and predictions. Training models requires high-throughput and processor-intensive computing resources to process historical data. In contrast, inferences and predictions often require very fast or real-time processing of smaller amounts of data. Some ML programs also explain their predictions in natural language.

Anyone, Anywhere, All the Time

Historically, many people who needed analytics to do their jobs effectively couldn't access the tools, or predefined reports often didn't answer their questions. As a result, analytics had limited benefits because users couldn't ask their own questions or get answers in a timely manner. Thus, analytics was relegated to providing support of people's intuition or experience-based decisions.

A next-gen data platform provides self-service analytics that work with relational, document, spatial, and graph databases, as well as semi-structured and unstructured data. This enables all relevant information to be considered when making a decision. This way, teams in finance, marketing, human resources, and other departments can run their own analyses using the most up-to-date data available.

REMEMBER

Many people refer to solving this type of challenge as *democratizing* data, but making data available to everyone is only part of the story. You also need to provide easy-to-use tools that enable users to dig deeper into recommendations and forecasts without getting bogged down in the technology of advanced analytics, AI, or ML. Pervasive analytics puts the data into the hands of those who need it, when they need it. This dramatically shortens the data-to-decision cycle.

IN THIS CHAPTER

- Making the Oracle Data Platform work for you
- Getting the most out of data integration services
- Using converged database capabilities
- Simplifying deployments with Oracle Autonomous Database
- Diving into data lakes

Chapter 4
Meeting End-to-End Data Needs with Next-Gen Data Platforms

A data platform should empower organizations to invest more time in creating value from their data estate. It should support your portfolio of applications by running the most demanding, mission-critical workloads, removing barriers to data access, enabling self-service and real-time analytics, and empowering developers so that they can expedite application development and reduce time to market.

This chapter explores the Oracle Data Platform suite of services and how they help customers gain a 360-degree perspective of their business and improve decision-making based on real-time data.

Oracle Data Platform

Oracle Data Platform (see Figure 4-1) is a comprehensive suite of services that help businesses manage data effectively and maximize its value. It is built on the following three guiding principles that help organizations get more value out of their data:

- **Comprehensive:** Support enterprise-wide needs with a full data stack from data integration to transactional databases, data warehouses, data lakes, artificial intelligence (AI), machine learning (ML), and analytics
- **Open:** Extend or replace core data services using a modular architecture
- **Integrated:** Implement diverse workflows with minimal integration and expertise using a unified architecture

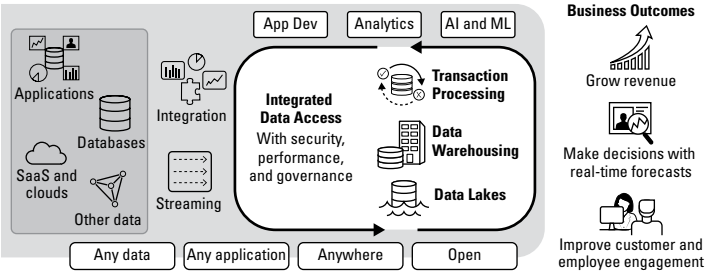

FIGURE 4-1: Oracle Data Platform.

Oracle Data Platform enables organizations to meet their end-to-end data needs while eliminating data silos and avoiding the complexity found in single-purpose and do-it-yourself approaches.

Data Integration Services

Tremendous amounts of data are created across an enterprise that are not where they're needed. Oracle Data Platform includes data integration services like Oracle Cloud Infrastructure (OCI) GoldenGate and Data Integration that automate data movement, transformation, and quality control. They include a broad array of integrations, event streaming, adapters, and low-code customizations that make them easy to use and scale.

Oracle Data Platform's modular approach also makes it easy to incorporate third-party and custom data integration tools, including those from Informatica, into your workflows.

Consolidating Different Data Types and Usage Models

Converged databases support many types of data, including relational, graph, spatial, document, and ledgers, that can be used in different database models. A next-gen data platform needs converged database versatility to reduce the number of services used.

Organizations using Oracle Data Platform achieve several key benefits from converged Oracle Database capabilities, including the following:

- Easier access to and integration of enterprise-wide data through the elimination of data silos
- Developers focus on innovation instead of integration using their access method of choice
- Real-time analytics of transaction processing data with no moving or transformation required
- Less complexity with one database supporting multiple data types and usage models that are managed together

Oracle Database works with unstructured and semi-structured data, making it ideal for combining disparate enterprise-wide data.

Autonomous Database

Oracle Autonomous Database is a fully automated database service that combines Oracle Database with ML-based automation and the high performance and availability of Exadata in OCI. Organizations can easily use it to develop and deploy database workloads regardless of data type, use case, or criticality.

Autonomous Database automatically tunes databases, scales them up to meet spikes in demand, and back down to reduce costs. And because it's a converged database, all data types and database

workloads benefit from higher performance, availability, and consistent security.

Autonomous Database supports two Exadata deployment choices:

- **Shared:** Multiple customers share Exadata compute and storage in OCI with full automation, low consumption costs, and a focus on simplicity and elasticity.
- **Dedicated:** A single customer uses a dedicated Exadata system in OCI, isolating data and workloads from other cloud tenants. It adds more total scalability, performance, and control over database runtime environments.

TIP

Autonomous Database's built-in transaction processing, analytics, ML, and AI capabilities reduce the number of independent services that organizations need to manage their data.

Autonomous Database supports the industry-standard, open-source Delta Sharing protocol so organizations can eliminate stale data and inaccurate results by securely sharing data with applications and services that support the protocol. Autonomous Database provides extensive multicloud functionality with secure access to Amazon Simple Storage Service (S3) or any S3-compliant object storage in other clouds, live SQL connections to leading databases, and prebuilt connectors to more than 100 data sources.

Autonomous Database's versatility and scalability make it the ideal place to integrate data from data lakes and other clouds.

Diving Into Data Lakes

Data lakes are repositories for structured, semi-structured, and unstructured data in any format and any size. Oracle data lake services enable secure data sharing and seamlessly integrate with OCI Data Catalog, Data Flow, Autonomous Database, and other services so users can easily extract value from complex data.

Integration of data lakes and data warehouses, including those using Autonomous Database, allow data from multiple locations, including other clouds, to be accessed with a single SQL query. This lets applications and tools transparently access and process enterprise-wide data and doesn't force users to learn new skills.

IN THIS CHAPTER

- Exploring the performance of Exadata powered by AMD EPYC processors
- Maximizing efficiency and reducing costs
- Increasing resilience for crucial databases

Chapter 5
Maximizing Data Platform Performance and Resilience

Organizations use data platforms to improve decision-making and run critical workflows faster and without interruption. They need to process an ever-growing number of transactions and analyze more data faster and with more complex algorithms. They can't afford either planned or unplanned downtime.

In this chapter, you learn how the Oracle Data Platform uses Oracle Exadata and AMD EPYC processors to deliver high performance and availability for crucial databases, data warehouses, analytics, and in-database machine learning (ML).

Enabling Scalability and Improving Performance

A core element in many Oracle Data Platform deployments is Autonomous Database, which achieves high levels of performance, scale, and availability for diverse databases by running on Oracle Exadata systems in Oracle Cloud Infrastructure (OCI).

Exadata's scale-out architecture separates database servers from intelligent storage and provides unique capabilities that accelerate all database workloads. Exadata delivers substantially lower input/output (IO) latency and faster IO per second (IOPS) rates — which are critical for online transaction processing (OLTP) applications — as well as the fastest database scan rates, which are critical for analytics and in-database ML model training.


TECHNICAL STUFF

Exadata Smart Scan is a unique Exadata feature that offloads SQL processing from database servers to intelligent storage servers. This allows analytic queries and ML training algorithms to run faster than on other systems because the processing happens where the data is located, instead of moving all the data to database servers and processing it there as is required with traditional approaches.

Dedicated Autonomous Database environments in OCI run on Exadata Cloud Infrastructure with database servers powered by AMD EPYC processors. Exadata systems in OCI currently start with 252 AMD EPYC processor cores available in database servers — 2.5 times that of previous systems, but don't worry, you can choose how many to use. This minimum configuration can store 190 terabytes (TB) of encrypted, uncompressed databases on triple-redundant storage to help prevent data loss from component failures. With Oracle's Hybrid Columnar Compression (HCC), it can store up to 1.9 petabytes (PB) in data warehouses.

Exadata's architecture enables the independent scaling of database and storage servers to meet the aggregate needs of virtually any organization. If your initial workloads are small, you can start with a small configuration and grow by adding database servers as OLTP needs increase and storage servers as databases and data warehouses expand. Fully scaled Exadata systems in OCI can have more than 4,000 AMD EPYC processor cores for database

processing with up to 3PB of database capacity and 30PB of data warehouse capacity with thousands of intelligent storage cores for processing (see Figure 5-1).

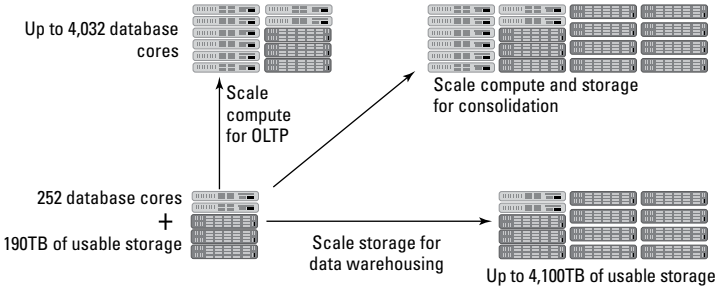

FIGURE 5-1: Oracle Exadata Cloud Infrastructure granular scaling.

Dedicated Exadata systems can simultaneously run the fully managed Autonomous Database and the DBA-administered Exadata Database Service so organizations can increase operational efficiency and performance for all Oracle Database workloads. Autonomous Database is exclusively available in OCI, on hybrid cloud Exadata Cloud@Customer platforms or OCI Dedicated Region, and for multicloud architectures via, for example, the Oracle Database Service for Microsoft Azure.

EXTENDING THE BENEFITS OF THE ORACLE DATA PLATFORM WITH AMD EPYC PROCESSORS

Exadata Cloud Infrastructure in OCI uses AMD EPYC processors, which provide more compute cores, memory bandwidth, and input/output (IO) bandwidth per database server than previous generation systems. AMD EPYC processors help all Oracle Database workloads run faster on Exadata.

Some of the benefits that organizations gain from using Oracle Exadata with AMD EPYC processors include

- **Smoother operations:** With 2.5x the number of processor cores as previous generation Exadata database servers, AMD EPYC

(continued)

(continued)

processors support higher transaction rates with no downtime for scaling, so organizations have greater flexibility when deploying and scaling apps with variable demand — such as for retailing with seasonal peaks.

- **Real-time insights:** Oracle Database's converged capabilities and the performance provided by AMD EPYC processors allow organizations to directly analyze OLTP data in real time, providing critical insights for "analytics-in-the-loop" workloads — such as credit card fraud-detection.
- **Better customer experiences:** Higher transaction rates and real-time analytics enable more responsive user interfaces that include context-sensitive information — such as for location-based services.
- **Deeper, more timely insights:** More cores in Exadata database servers enable broader use of in-database advanced analytics, prebuilt ML algorithms, and customer-created AI/ML models built with Python and R.
- **Faster time to market:** The availability of more database processing cores helps avoid surprise scaling related issues by allowing app development, testing, and deployment on the same scalable infrastructure.
- **Reduced environmental impact:** Running converged databases on servers with more cores increases consolidation rates and reduces the need for stand-alone services. The use of fewer servers, energy efficient AMD EPYC processors, and Oracle's goal of using 100 percent renewable energy in all OCI regions by 2025 — already achieved in Europe — significantly lowers environmental impacts.

Delivering Efficiency and Cost-Effectiveness

Oracle Database's converged capabilities make it easier for applications using Oracle Data Platform to share data because an organization can run all its databases on the same system, no matter what data or workload type they use. Using *database consolidation* increases the value of enterprise-wide data because

it can be easily accessed and combined on a single system. This approach helps organizations reduce the number of platforms and services they need and operate their database fleets more efficiently.

As noted earlier, dedicated Exadata system resources can be granularly scaled to meet the needs of practically any organization. Database services consume resources on these systems in an even more fine-grained way that maximizes performance and minimizes costs. Autonomous Database analyzes currently running queries and automatically scales consumption based on their needs, which combines with per-second billing to deliver pay-per-use economics.

For instance, an online retailer's data platform must support an OLTP environment with spikes in demand based on the time of day. Autonomous Database can automatically raise consumption to meet peak requirements and lower it afterward to reduce cost. Similarly, a large analytics job to evaluate how well promotions are working can cause a surge in resource demand that can be met with autoscaling (see Figure 5-2).

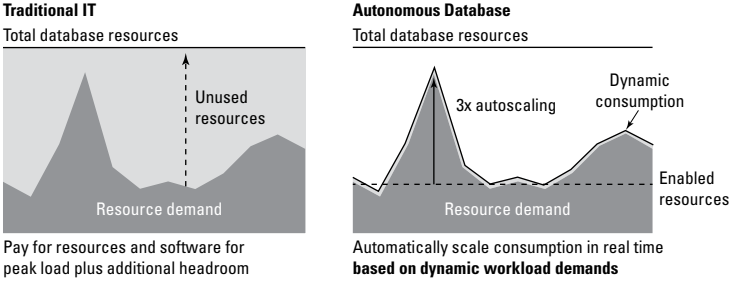

FIGURE 5-2: Autonomous Database autoscaling matches demand.

TIP

Consolidating databases on Exadata systems with AMD EPYC processors provides private pools of dedicated resources to support scalable, high performance so organizations can accelerate decision-making and improve customer experiences. Additional benefits include the following:

- Simpler IT environments because database workloads run on fewer service instances and less infrastructure
- Higher utilization efficiency and lower costs because database workloads use fewer database servers

- Easier access to enterprise-wide data because it's all on the same system
- Easier to achieve consistent, strong database security with a single security model across all databases

Increasing Database Resilience

In today's always-on world, crucial OLTP and analytics workloads need to be available 24/7. They should continue to run while resources are being scaled, if individual components fail, during planned upgrades and maintenance, and when faced with disasters.

Organizations using Oracle Data Platform can benefit from high availability (HA) and disaster recovery (DR) capabilities built into individual services and the Full Stack Disaster Recovery service that orchestrates site-to-site failovers with minimal disruption.

Autonomous Database provides one of the highest service-level agreements (SLAs) for availability of any cloud service. It provides up to a 99.995 percent uptime SLA that covers both unplanned downtime due to failures and planned downtime for upgrades and maintenance — something that other cloud service providers do not cover or meet.

These high SLAs are made possible by capabilities built into Autonomous Database and the Exadata infrastructure it runs on. Autonomous Database monitors thousands of metrics and alarms to identify and resolve issues before they impact users. If component failures happen without warning, Exadata's built-in fault-tolerance enables it to keep running.

Autonomous Data Guard works in concert with Autonomous Database to maintain consistent database copies in another OCI Availability Domain or Region. It is a key part of the proven Oracle Maximum Availability Architecture (MAA), which helps organizations provide the right level of availability for databases with differing criticality.

IN THIS CHAPTER

- Supporting all your data needs and simplifying your IT environment
- Delivering performance, scalability, and high availability
- Maximizing value with a comprehensive open, and integrated data platform
- Automating operations and ensuring robust security
- Democratizing access to data and enabling deep insights

Chapter 6
Ten (or So) Important Questions to Ask

Here are some questions you should ask when selecting a data platform for your organization:

- **Is the data platform comprehensive?** It should support virtually any type of application, data, database, workload, and application style. Being able to support a variety of apps and workloads with a single data platform simplifies your IT environment and saves you money.
- **Can the data platform help minimize the number of different databases, data transformations, and services you use?** Such a data platform further simplifies your IT environment and enables decisions to be made faster and with the freshest data.
- **Does the data platform deliver the performance and scale that your business-critical applications need?**

A data platform that meets your performance and scalability requirements means that customers and internal users are happy, and you save money because you consume and pay for fewer resources.

- **Can the data platform provide the high availability required for your business-critical applications?** A data platform with built-in high availability, disaster recovery, and data protection capabilities leaves fewer things to chance and enables enterprises to keep operating despite system failures or data center outages.
- **Does the data platform provide built-in features, integrations, and automated deployment architectures that reduce do-it-yourself integration to the bare minimum?** A data platform with a broad suite of built-in features, capabilities, integrations, and automations frees developers to spend less time on integration and more time on innovation.
- **Is the data platform open and modular?** A data platform with an open, modular architecture can evolve to meet your changing data sources and workload needs. For example, it should allow you to use only the pieces you want and replace the rest or extend with third-party, open-source, or self-developed capabilities. You should also be able to easily incorporate a data lake, stream data with open-source projects, or govern data with third-party tools.
- **Does the data platform automate operations?** Data platforms that monitor themselves to detect and correct potential errors while automating database tuning and scaling deliver results more reliably, quickly, and at a lower cost.
- **Does the data platform provide a strong security model with defense-in-depth capabilities?** A data platform with fewer stand-alone services reduces the number of security models employed and the risk of security gaps. Defense-in-depth capabilities help identify and mitigate threats across the entire application stack.
- **Does the data platform help make data and decision-making tools available to the people who need them throughout the organization?** A data platform that democratizes access to data and provides advanced analytics and machine learning tools to analyze it helps authorized decision-makers at all levels of the organization develop deeper insights.

Maximize the value of your business data

A next-gen data platform is a comprehensive suite of services that simplifies data management across an organization's portfolio of workloads. It integrates data management capabilities needed throughout the data lifecycle, from transaction processing and data integration through data warehousing and data lakes – all with data governance, security, and high performance. A next-gen data platform goes beyond consolidating data into a single source of truth for an organization. It includes advanced analytics, artificial intelligence, and machine learning so users can extract more value from their existing data; and tools that enable application developers to quickly create new user experiences and value streams.

Inside...

- Use one platform for end-to-end data needs
- Consolidate diverse data and workloads
- Create new data-driven applications
- Deliver superior customer experiences
- Make decisions using real-time insights
- Increase efficiency and lower costs
- Deploy in the public cloud or enterprise data centers

ORACLE | AMD

Lawrence Miller served as a Chief Petty Officer in the U.S. Navy and has worked in information technology in various industries for more than 25 years. He is the co-author of *CISSP For Dummies* and has written more than 200 *For Dummies* books on numerous technology and security topics.

ISBN: 978-1-394-20696-4
Not For Resale

for dummies®
A Wiley Brand

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.