

Question: What is generative AI?

Creates new content and ideas, including conversations, stories, images, videos, and music

Powered by large models that are pretrained on vast corpora of data and commonly referred to as foundation models (FMs)



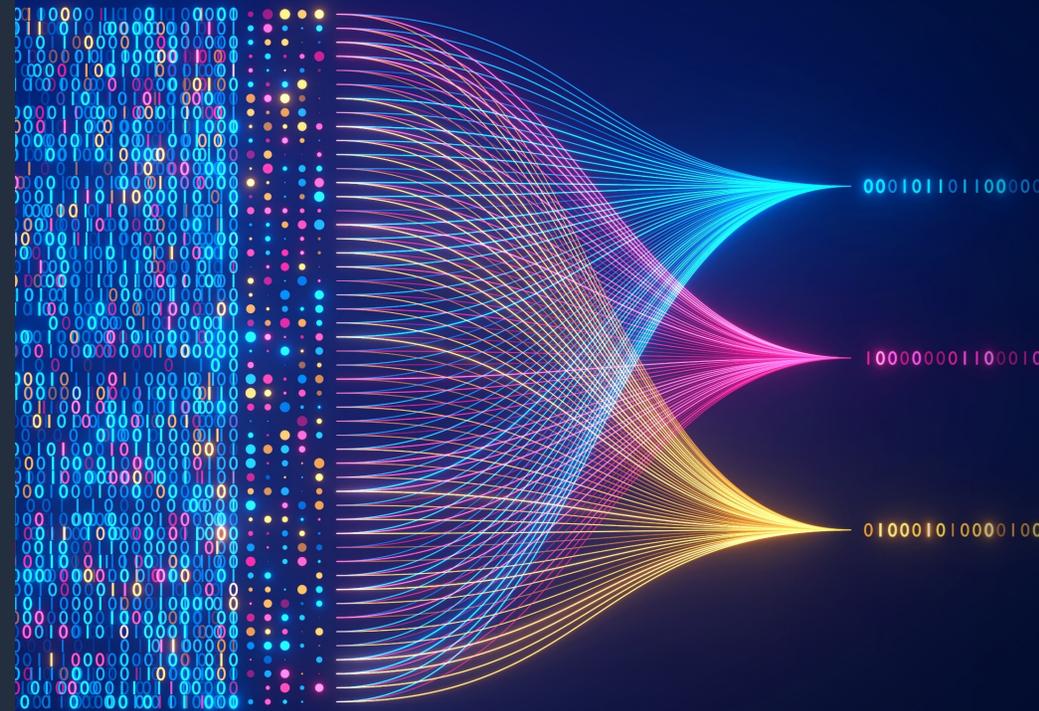
Generative AI is powered by FMs

Pretrained on vast amounts of unstructured data

Contain large number of parameters that make them capable of learning complex concepts

Can be applied in a wide range of contexts

Customize FMs using your data for domain-specific tasks



Generative AI creates significant business value



NEW EXPERIENCES

Create new innovative and engaging ways of interacting with your customers and employees



PRODUCTIVITY

Radically improve productivity across all lines of business, for example **Amazon CodeWhisperer** helped developers complete tasks 57% faster



INSIGHTS

Extract insights and clear answers from all your corporate information, facilitating faster and better decisions



CREATIVITY

Create new content and ideas, including conversations, stories, images, videos, and music

Generative AI can be used for a wide range of use cases

Chatbots

Virtual assistants

AI-powered Contact Center

Personalization

Conversational search

Summarization

Code generation

Data to insights

Writing

Media design

Modeling

Document processing

Process optimization

Cybersecurity

Data augmentation

**Enhance
customer
experience**

**Boost
employee
productivity**

**Creativity
and content
creation**

**Improve
business
operations**



Generative AI is transforming all industries



Financial services



Healthcare and life sciences



Automotive



Manufacturing



Media and entertainment



Retail



Telecom



Energy



Travel and hospitality

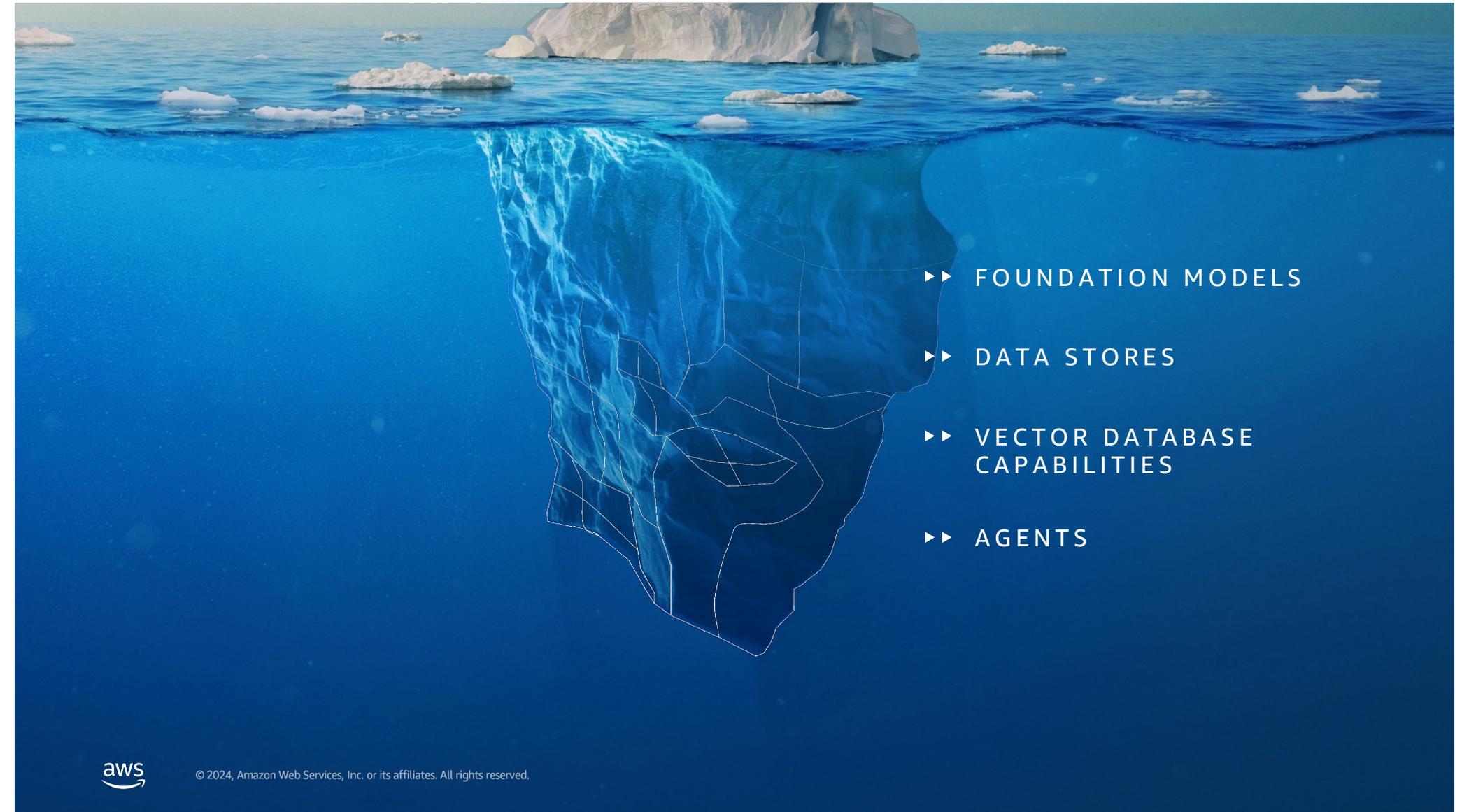


Consumer packaged goods

GENERATIVE AI APPLICATION



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.



▶▶ FOUNDATION MODELS

▶▶ DATA STORES

▶▶ VECTOR DATABASE
CAPABILITIES

▶▶ AGENTS



Your data is your differentiator



Generic
generative AI



Generative AI that knows your
business and your customer

Data explosion requires a modern data strategy

97 ZB of data

created in 2022
(source: IDC)

68% of organizations

Reported that they are
still unable to realize
value from data



Business impact

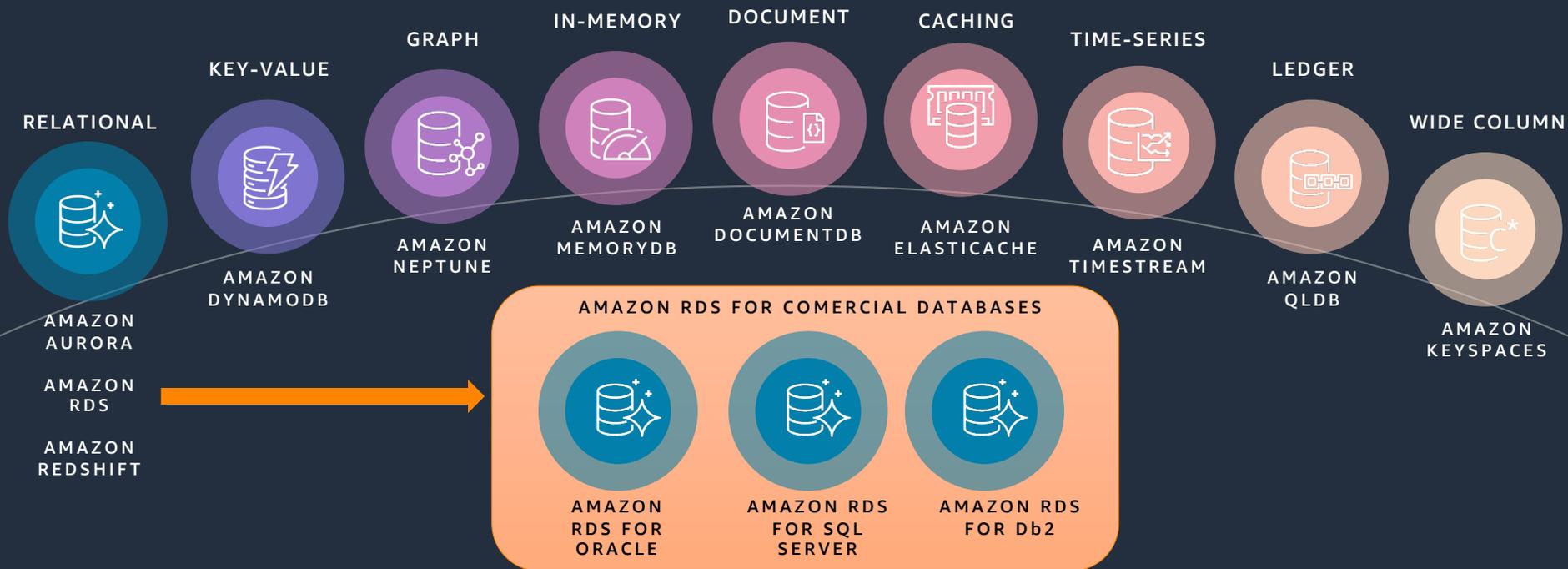
By making **10%** more data
accessible, a typical Fortune
1000 company will see:

**\$65 million increase
in net income¹**

**Data-driven businesses
are growing at an
average of more than
30% annually¹**



Broadest portfolio of purpose-built databases



Amazon Aurora

Designed for unparalleled high performance and availability at global scale with full MySQL and PostgreSQL compatibility at 1/10th the cost of commercial databases



Performance & scalability

- 5x throughput of standard MySQL and 3x of standard PostgreSQL
- Scale out up to 15 read replicas
- Decoupled storage and compute enabling cost optimization
- Fast database cloning
- Distributed, dynamically scaling storage subsystem



Availability & durability

- 99.99% availability with Multi-AZ
- Data is durable across 3 AZs (*customers only pay for 1 copy*)
- Automatic, continuous, incremental backups with point-in-time recovery (PITR)
- Failovers in < 10 seconds
- Fault-tolerant, self-healing, auto-scaling storage
- Global Database for disaster recovery



Highly secure

- Network isolation
- Encryption at rest/in transit
- Supports multiple secure authentication mechanisms and audit controls



Fully managed

- Automates time-consuming management of administration tasks like hardware provisioning, database setup, patching, and backups
- Serverless configuration options



Modernizing leads to maximum innovation velocity and optimal total cost of ownership

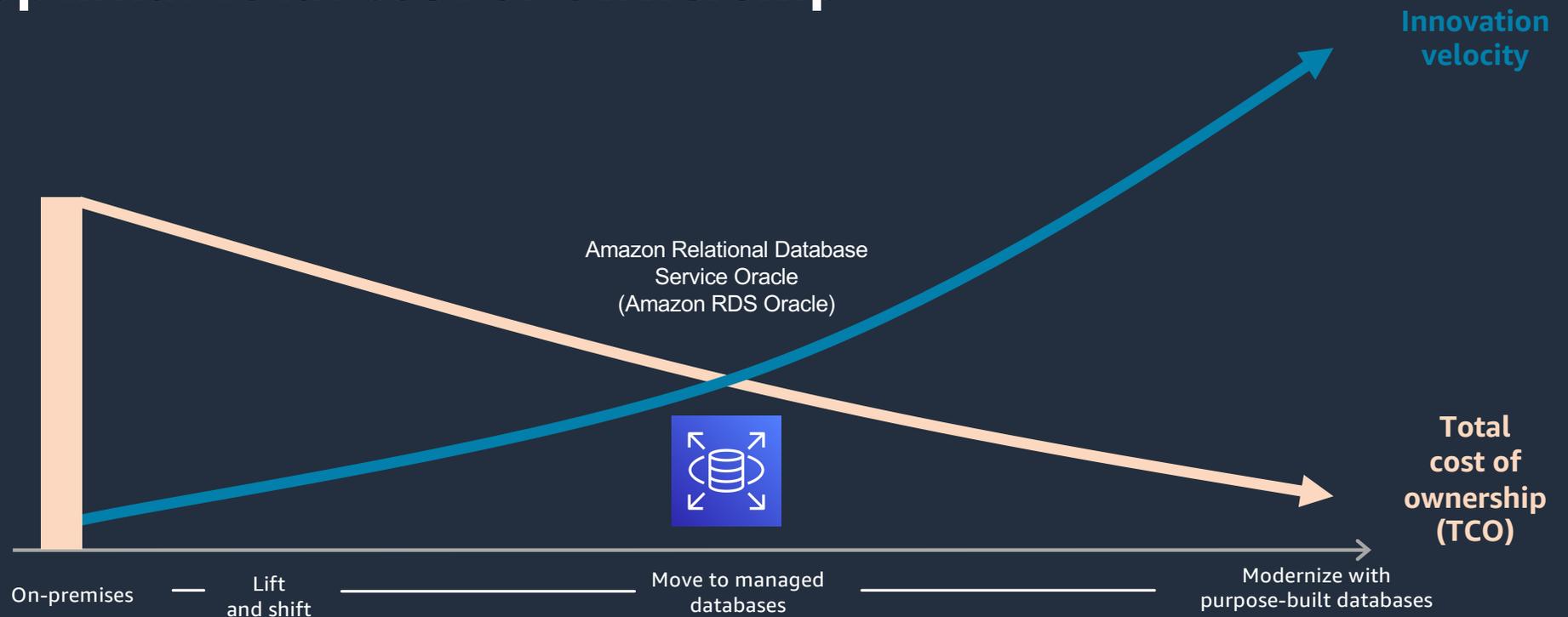
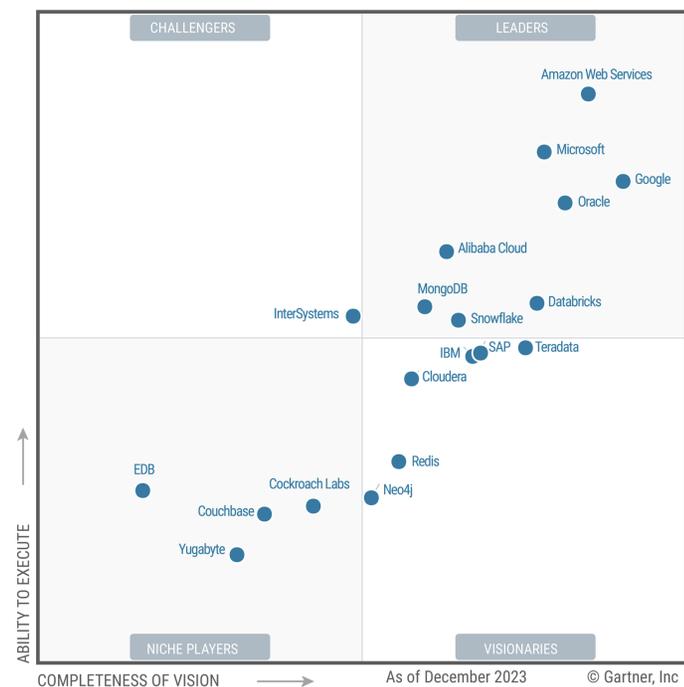




Figure 1: Magic Quadrant for Cloud Database Management Systems



Source: Gartner

AWS named a Leader in Magic Quadrant for Cloud DBMS by Gartner for 9th consecutive year

[Read the report.](#)

Gartner does not endorse any vendor, product or service depicted in its research publications and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose. This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from AWS. GARTNER is a registered trademark and service mark of Gartner and Magic Quadrant is a registered trademark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and are used herein with permission. All rights reserved. The report was previously named Magic Quadrant for Operational Database Management Systems from 2014-2019.

How to provide your data to generative AI applications?

Context engineering
using RAG with
foundation models

Guide foundation models by
prompting with contextually
relevant data (RAG)

Fine-tuning a
foundation model

Fine-tune a foundation model
using your curated, labeled data

Training your own
purpose-built LLM

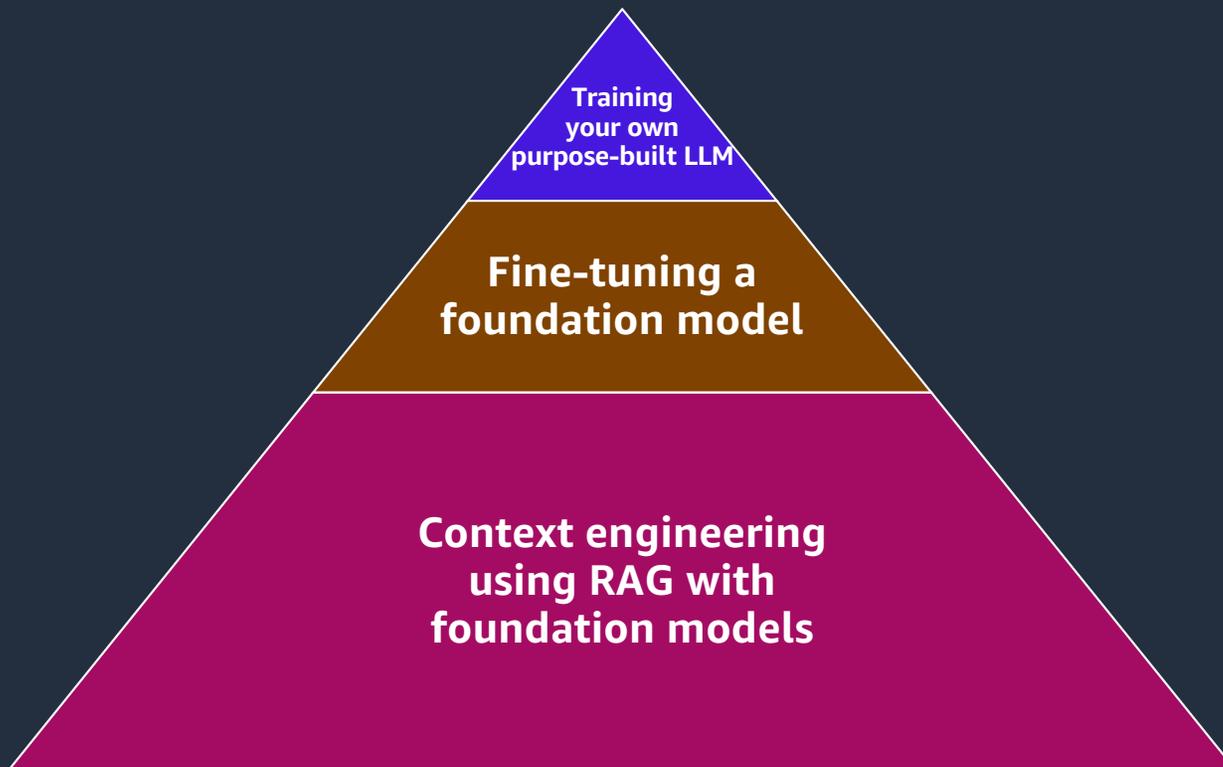
Train a foundation model using
your curated, specialized data

Generative AI implementation patterns

COMPLEX



EASIER



PURPOSE BUILT

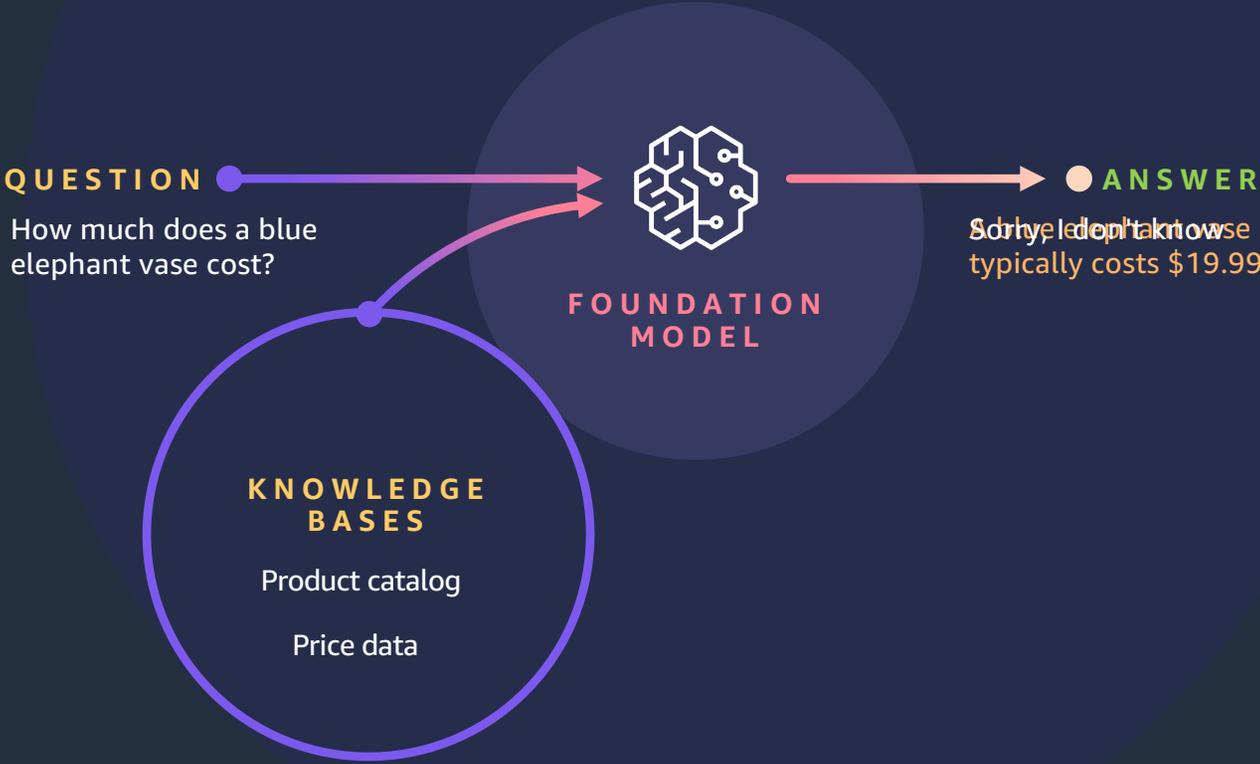


GENERIC MODEL + CONTEXT

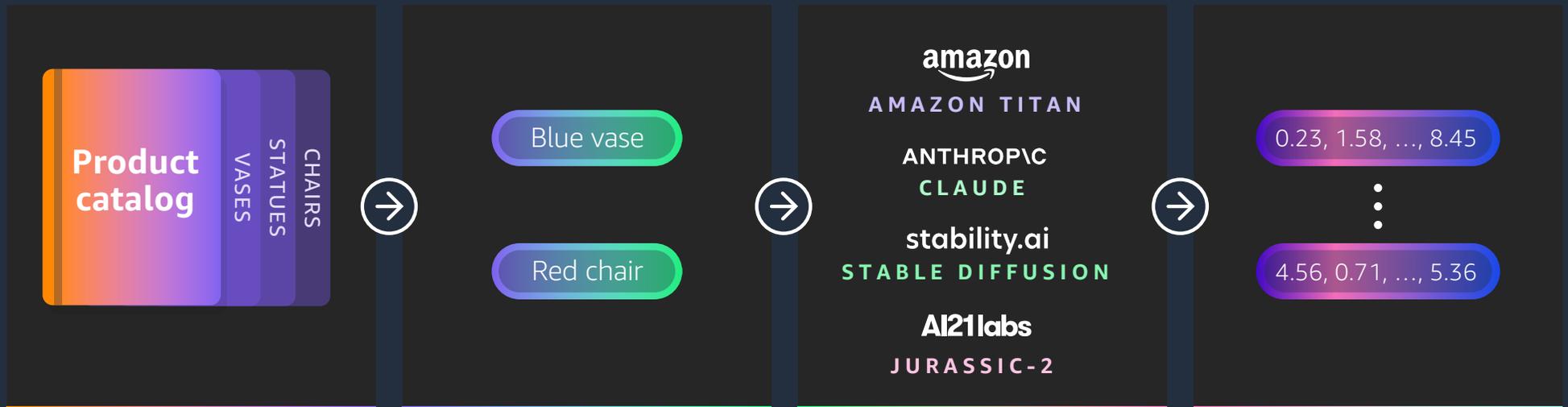


Retrieval Augmented Generation (RAG)

Configure FM to interact with your company data

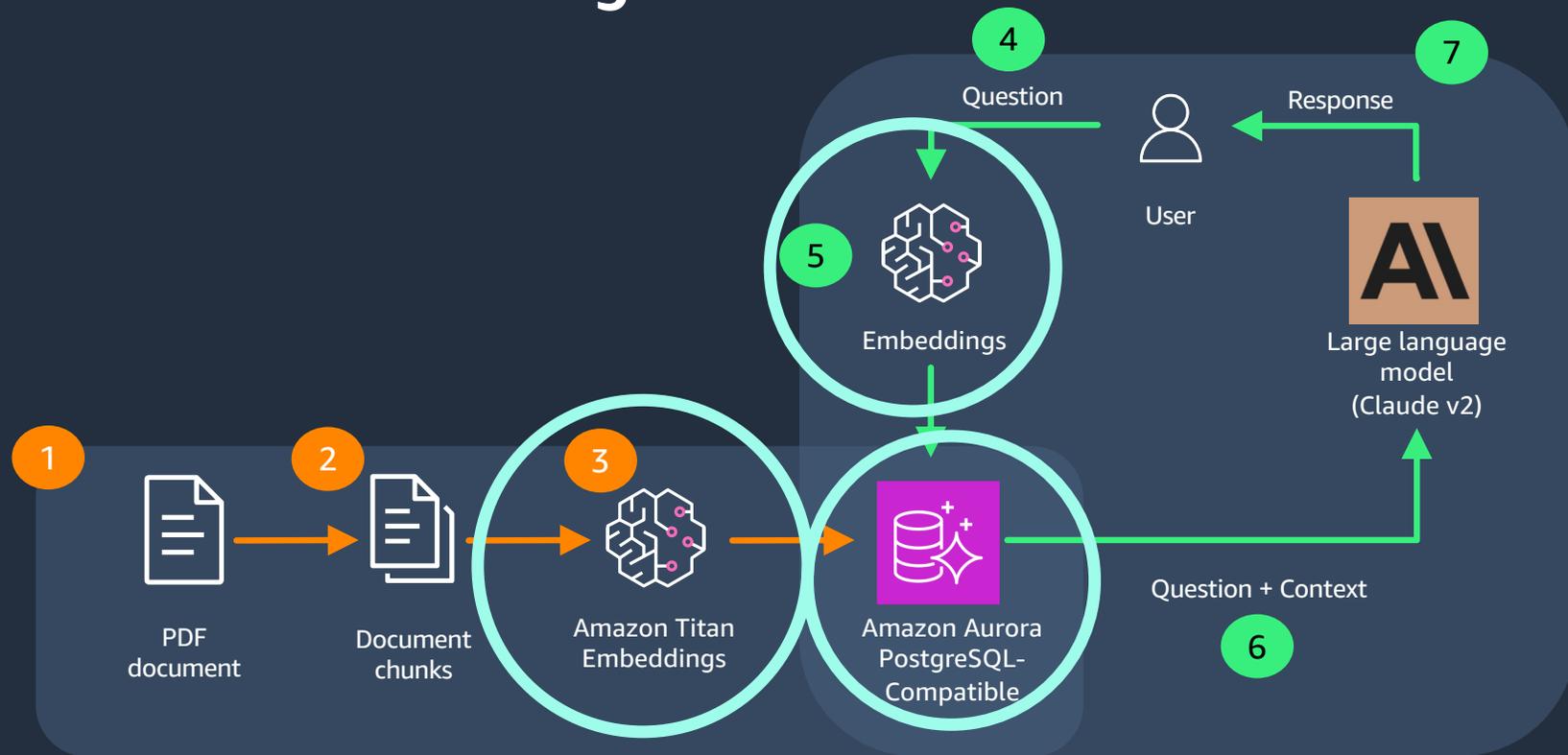


What are vector embeddings?



Unstructured data has to be vectorized into vectors to be used in generative AI applications

How vector embeddings are used



Vector support across our services



AMAZON
OPENSEARCH
SERVICE



AMAZON
AURORA
POSTGRESQL



AMAZON RDS
FOR
POSTGRESQL



AMAZON
MEMORYDB
FOR REDIS



AMAZON
NEPTUNE



AMAZON
DOCUMENTDB



AMAZON
DYNAMODB

VIA ZERO-ETL



Database Optimization and License Assessment (OLA)



Licenses are a powerful driver of migration cost savings

LICENSES ARE A SIGNIFICANT PART OF THE TOTAL COST OF OWNERSHIP

That's why we recommend the **AWS Optimization and Licensing Assessment (OLA)** to all our customers—new and existing. This complimentary program assesses and optimizes your current on-premises and cloud environments, based on actual resource utilization, third-party licensing, and application dependencies.

Analysis of 300 OLAs¹ showed:

45%

reduction in **Microsoft SQL Server** licensing compared to on premises

98%

of customers were able to migrate without additional license investments

3x

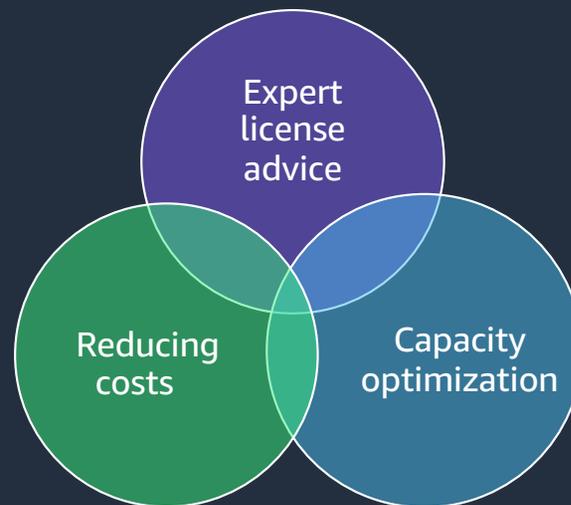
the impact – software licensing costs are 3X the cost of all other costs in the cloud combined

¹ Enterprise Strategy Group, [Reduce Licensing and Infrastructure Cost](#)

What is OLA?

Optimization and Licensing Assessment

- Helps businesses assess and optimize current on-premises and cloud environments based on actual resources and licenses
- Helps you build a migration and licensing strategy to unlock cost savings as you transition to AWS



No charge to the customer



Optimization & Licensing Assessment

Optimize Licensing Assets when Migrating to Cloud

Customer Challenges

Know the Baseline

- What do I have deployed?
- How does my server strategy change in the Cloud?
- Can I leverage my licensing contracts?
- Options for capitalized vs operating costs

Understand Costs

- Pricing scenarios available
- What variables impact cost?
- Bring my licenses or phase out of my perpetual contracts?
- Retire / repurpose hardware & licenses where feasible

Build a Plan

- What do I want to move to the Cloud?
- Which workload do I move first?
- Define project scope / timing
- Understand effort & talent needed
- Will you need Partner support?

Gather Data

- Secure process ensures data privacy
- Collect data from on premises & 3rd party cloud providers
- Access to sophisticated data collection and analytics tools
- Consider dedicated, reserved, spot instance types
- Identify low use & at risk devices

Create TCO

- Recommendations built on actual utilization – “right size”
- Leverage existing licensing assets
- Optimized CPU scenarios
- Competitive cloud pricing options
- Average potential licensing cost reductions ~ **30%**

Accelerate Migrations

- Build move groups based on low risk and high return
- Identify application dependencies
- Prioritize migrations based on results
- Plan roll out schedule aligned with budgets
- Secure APN Partner support
- Leverage Cloud Endure
- Qualify & secure AWS migration funding

OLA Customer Benefits



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.