



# The Data Foundation for the Age of AI

How Data Cloud unifies data to power AI-driven customer experiences



# Designed for the Future of Intelligent, Data-Powered CRM Experiences

For businesses today, it's not just about how much data you have, but how effectively you use it. This is especially true at a time when AI-driven solutions are disrupting traditional business paradigms.

**PwC estimates that AI could generate more than \$15 trillion for the global economy by the end of the decade.<sup>1</sup> Data Cloud is made for this moment.**

More than a data management solution, Data Cloud is an evolutionary leap that enables transformative AI experiences and higher degrees of customer-centricity. By unifying any data source with your CRM data from the Salesforce platform, Data Cloud allows business leaders to unlock deep, actionable insights.

To deliver these experiences, **Salesforce and AWS are expanding their strategic partnership**, bringing AWS AI services to Salesforce's Einstein Trust Layer, and providing Data Cloud with seamless, zero-ETL access to any AWS data service, as well as the ability to execute data processing tasks using AWS compute services. And to make it easier to realize this combined value, Data Cloud and other Salesforce offerings are now available for purchase through the AWS Marketplace.

This e-book will show how Data Cloud unifies and harmonizes huge amounts of diverse business data with CRM data from the Salesforce Einstein 1 Platform, providing businesses with an unparalleled view of their data universe and a sturdy foundation for better customer experiences powered by AI.



**New: Purchase Data Cloud and other select Salesforce products through AWS Marketplace Private Offers.**

<sup>1</sup>[pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html](https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html)

# Harmonizing Data at Petabyte Scale

As data from other sources grows in volume and variety, we wanted to make it easier to combine CRM data from the Salesforce platform with all the other data a business needs to form a comprehensive portrait of its customers. To unlock the power of advanced analytics, machine learning (ML), and generative AI, we first needed to enable the transactional database underpinning the Salesforce platform to interact with data stored in other environments – everything from web interactions to product purchases to log files and beyond.



**Before Data Cloud**, engineering teams would laboriously move CRM data to data warehouses or data lakes to analyze it or build machine learning models with data from other sources. Not only was this time-consuming, but the stored data was now outside the secure and compliant Salesforce trust boundary. Bringing the data back into Salesforce was also a challenge and a resource drain, involving additional vendors and processes like reverse ETL, all of which shifted focus away from the customer.

## That changes with Data Cloud.

Because Data Cloud is built on AWS and integrates with data stores like S3, it can connect and prepare data without the need for ETL, unless data engineering teams choose to transform the data.

We also leveraged Apache Iceberg, “an open source high-performance format for huge tables,”<sup>2</sup> to design Data Cloud. Salesforce engineers contributed code to the Iceberg project and forged the

ability to combine transactional CRM data with data from other sources.

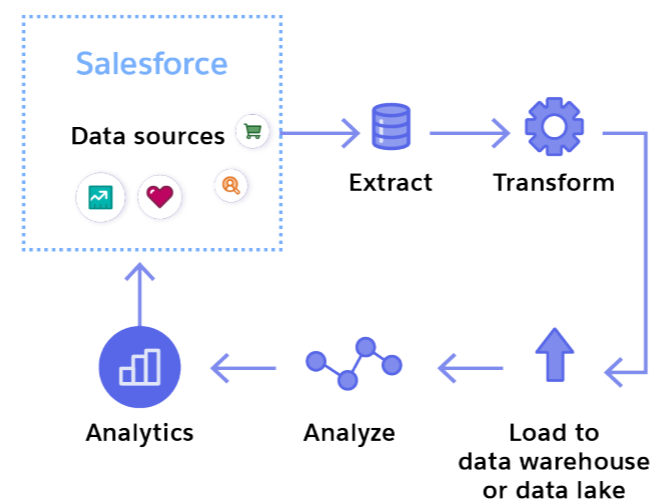
These can include transactional data, customer behavior data, IoT device data, and even unstructured data like social media posts or customer support chat logs. By pulling all this information into one place, Data Cloud allows for the creation of a single source of truth, enhancing decision-making, the relevance of AI models, and overall business efficiency.



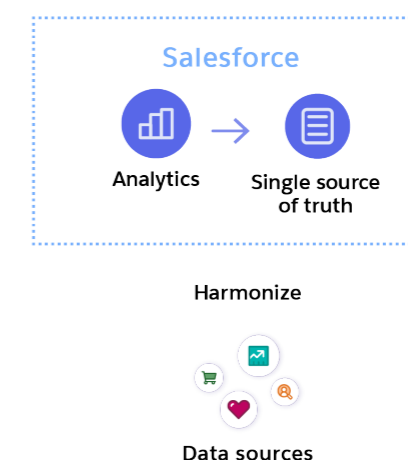
**Data Cloud empowers every business and every employee to unlock a consistent view of the customer, then harness the power of AI to drive personalized actions anywhere.”**

**MURALIDHAR KRISHNAPRASAD,**  
EVP OF SOFTWARE ENGINEERING,  
DATA CLOUD AND EINSTEIN AI

## Before Data Cloud



## With Data Cloud



<sup>2</sup> iceberg.apache.org

# How Data Cloud Works

## Data Sources

## Connect & Prepare

## Harmonize

## Unify

## Activate

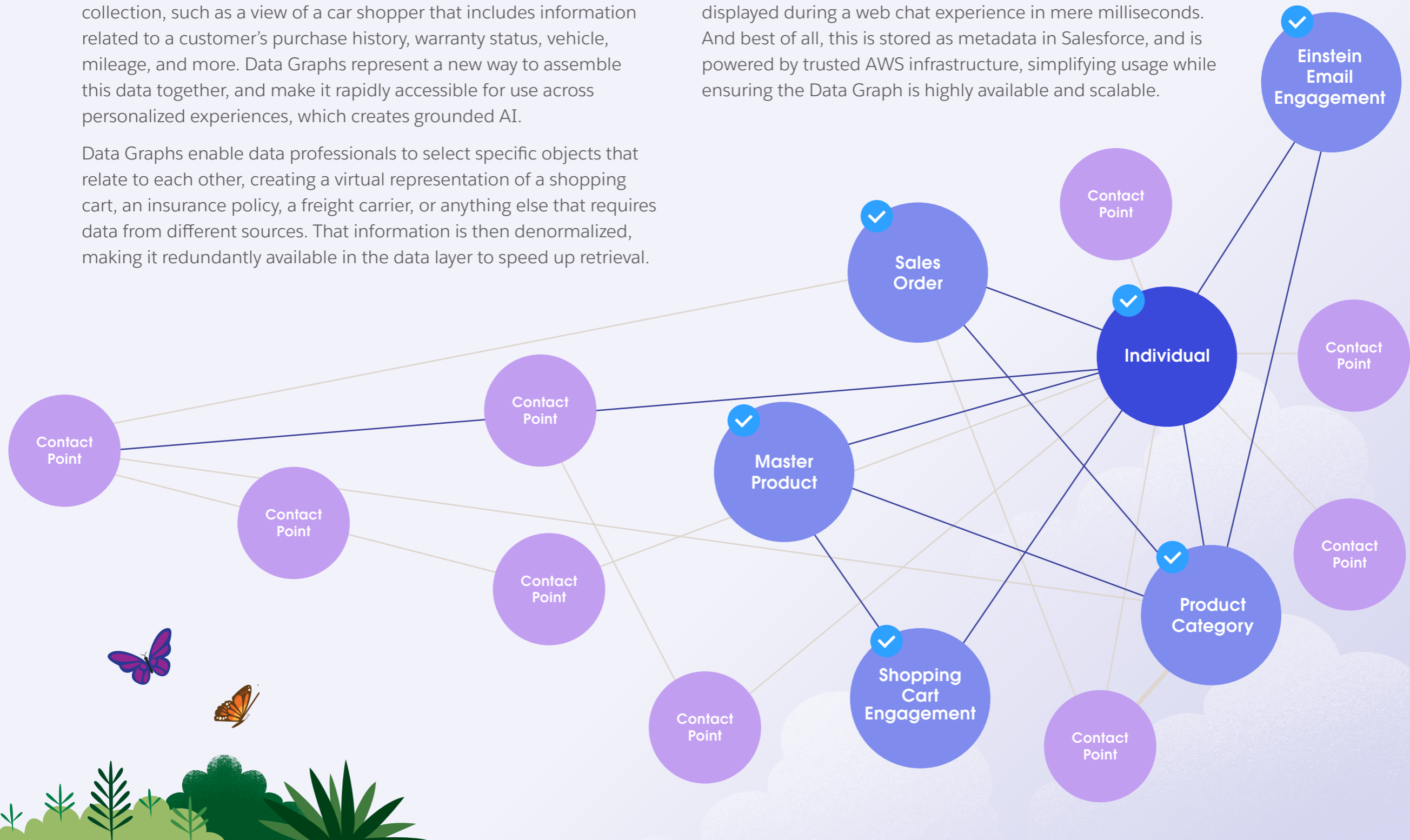


# Create Custom Collections of Data with Data Graphs

It's not enough to collect data into one place—sometimes teams will need to assemble data from different sources into a meaningful collection, such as a view of a car shopper that includes information related to a customer's purchase history, warranty status, vehicle, mileage, and more. Data Graphs represent a new way to assemble this data together, and make it rapidly accessible for use across personalized experiences, which creates grounded AI.

Data Graphs enable data professionals to select specific objects that relate to each other, creating a virtual representation of a shopping cart, an insurance policy, a freight carrier, or anything else that requires data from different sources. That information is then denormalized, making it redundantly available in the data layer to speed up retrieval.

By reducing the latency involved in retrieval, the data in the Data Graph can be brought into a grounded AI prompt or displayed during a web chat experience in mere milliseconds. And best of all, this is stored as metadata in Salesforce, and is powered by trusted AWS infrastructure, simplifying usage while ensuring the Data Graph is highly available and scalable.



# Bring Your Own ML Model or LLM with Einstein Copilot Studio

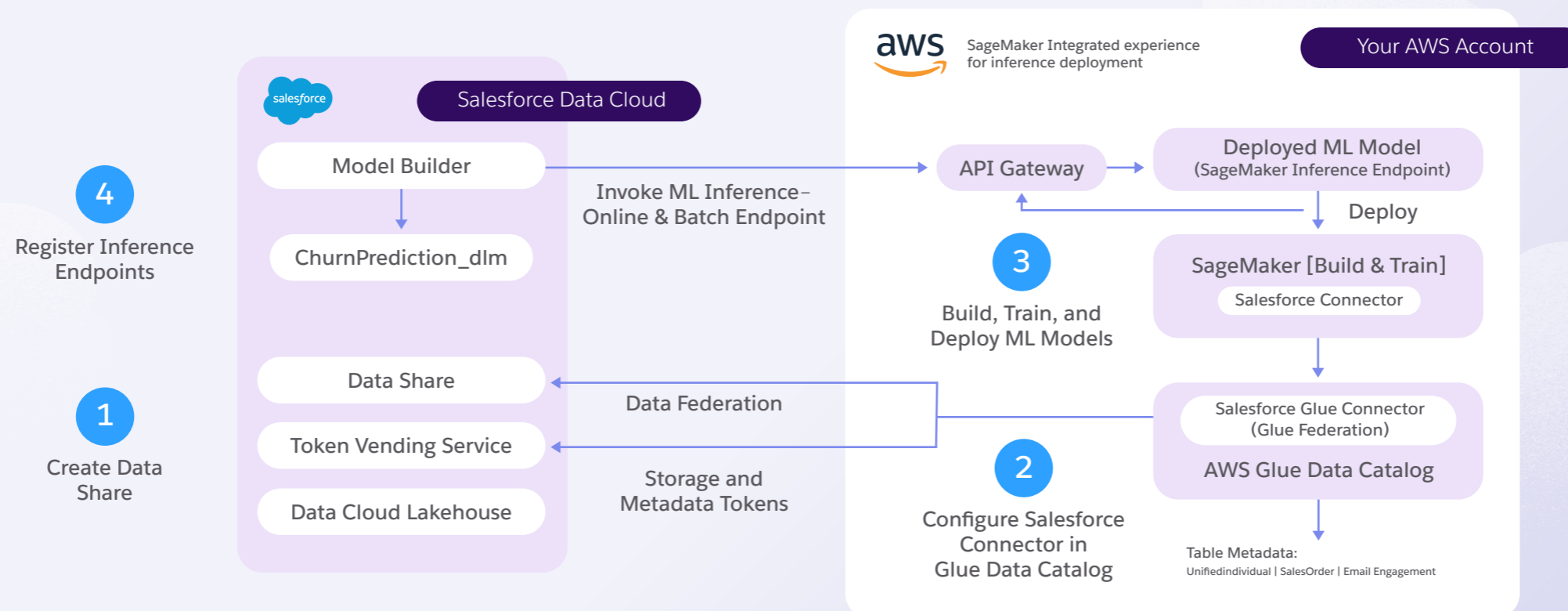
Amazon SageMaker is one of the world's most popular machine learning platforms, and Salesforce Data Cloud is often the center of data gravity for a company's customer engagement strategy. So what happens when a business wants to use that valuable customer context in Data Cloud to power custom, predictive AI models?

Historically, this would be accomplished through an ETL job, but with Model Builder, a new product available within Salesforce's Einstein Copilot Studio suite of AI solutions, any data professional can now access their Data Cloud data directly in Amazon SageMaker with zero ETL.

This means that without the usual lift and shift headaches, a data science or machine learning engineering team can use their preferred ML frameworks and algorithms—such as PyTorch and TensorFlow, or LightGBM and XGBoost among many others—to create their own custom AI models that are effortlessly trained on Data Cloud tables.

Once these models are trained, they can be served up back in Salesforce through a secure API endpoint, registered within Model Builder. The model can then be activated in the flow of work, within a Salesforce Flow automation, or as a prediction stored in a view. In every use case, the predictions being used are rooted in real-time data within Data Cloud.

Beyond predictive machine learning, Einstein Copilot Studio will also connect to models hosted on Amazon Bedrock, enabling AWS customers to create custom prompt templates that can be sent to their preferred FMs. This will enable customers to securely use techniques such as retrieval-augmented generation (RAG) with models from leading AI companies including AI21 Labs, Amazon, Anthropic, Cohere, Meta, and Stability AI.



# Apache Iceberg: Open Data Fuels Open Innovation

In the evolving landscape of AI-first business, open data has a critical role to play. Data Cloud breaks down data silos and provides access to a broad spectrum of data. Open data promotes transparency, fuels innovation, and enables informed decision-making. That's why Salesforce and AWS invested in Apache Iceberg.

## Here are some additional benefits of open data access:

**Consistency:** The use of Apache Iceberg as a table format ensures that every user sees a consistent view of the data, regardless of updates or modifications. This feature, known as snapshot isolation, is critical in a CRM context where different team members could be accessing and modifying data at the same time.

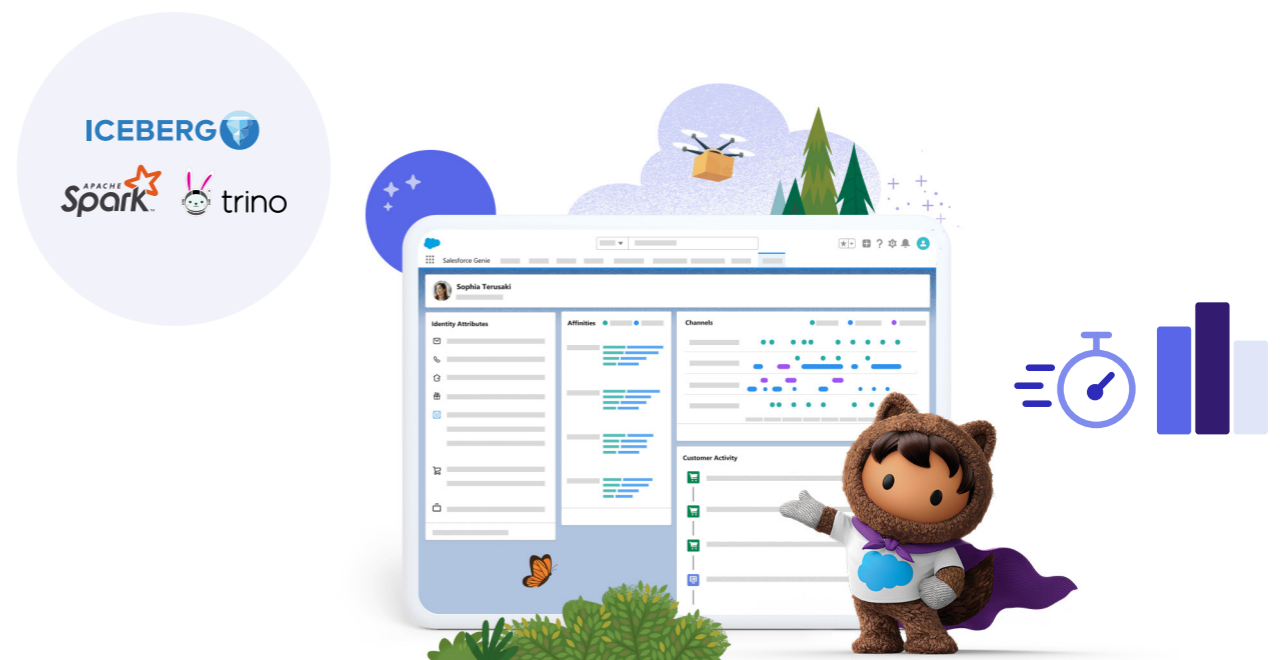
**Durability:** Data Cloud supports schema evolution, meaning that the structure of the data can change over time without breaking existing workflows. This enables CRM customers to modify their data structures as needs change.

**Interoperability:** Data Cloud pushes Salesforce metadata to the level of Iceberg tables, enabling zero-ETL data sharing. This means customers can virtually access data from other sources through zero copy federation, as well as extend their Salesforce data to other systems like Amazon SageMaker, enabling custom machine learning without an added ETL lift.

**Scalability:** Data Cloud can handle huge datasets, scaling easily as demand grows. This is particularly important for large CRM customers who may have extensive customer databases – like a recent customer who brought in over three trillion rows of data in one week.

**Efficiency:** In Data Cloud, tables are designed to optimize data access patterns, allowing for quicker query responses. Here, Salesforce engineers contributed back to the Iceberg project, enabling Iceberg tables to support streaming reads. This enabled engineers to achieve a paradigm where tables can operate as events, and events can operate as tables, making Data Cloud a truly “active” data platform.

**Flexibility:** Data Cloud is compatible with a variety of data computing frameworks, like Apache Spark and Trino, providing flexibility in data processing and analysis. This lets customers bring the tools they want to use for analysis and insight.



# The Data Lakehouse Is Open for Business

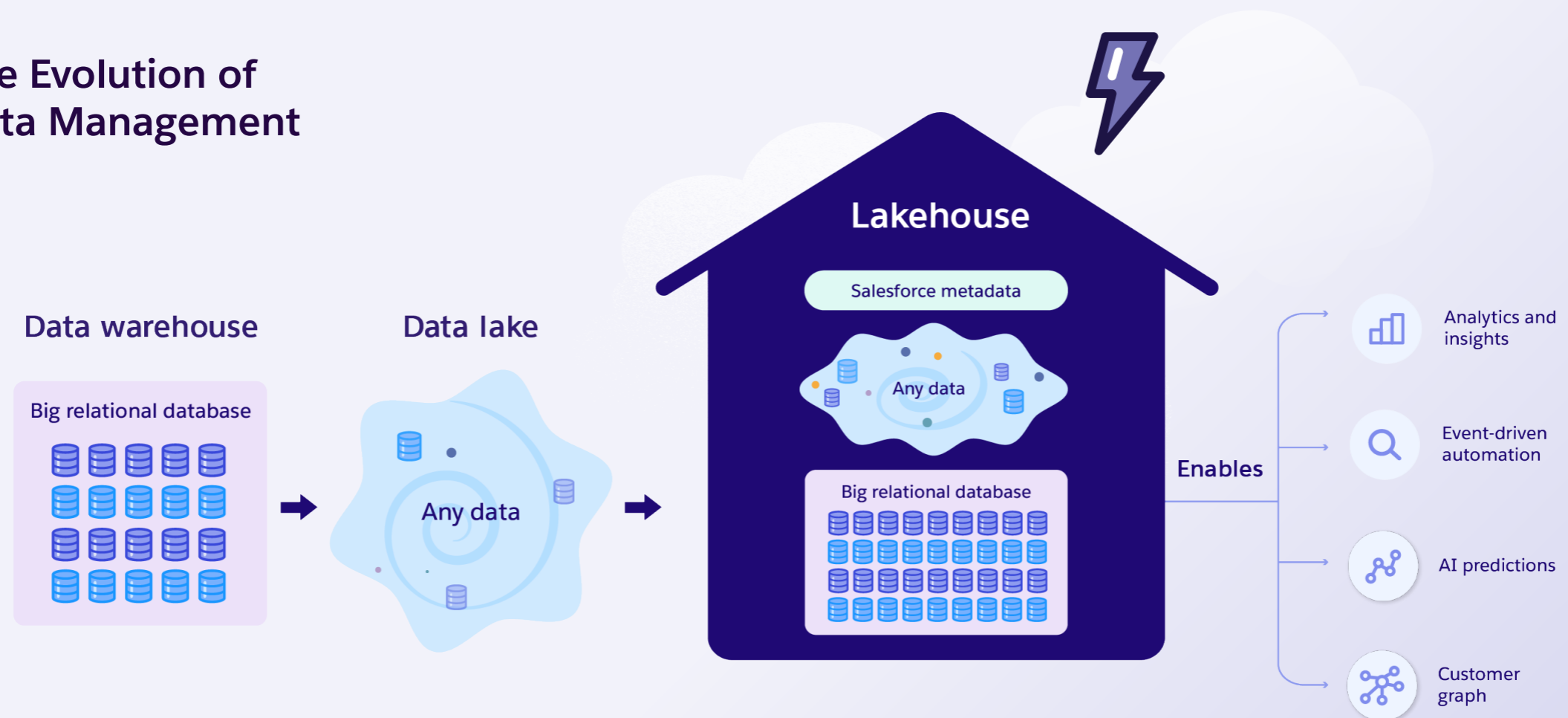
We built Data Cloud on an active data lakehouse because it is an open data management architecture that combines the flexibility, cost-efficiency, and scale of data lakes with the data management, queryability, and the ACID transactions of a data warehouse.

With an active data lakehouse, businesses can leverage the flexibility of data lakes to ingest any data type. This is crucial for businesses dealing with vast amounts of diverse data, including text, images, audio, video, log files, and others.

Additionally, the data lakehouse provides scalability, handling data volumes from a few gigabytes to an exabyte (1,000 petabytes).

Perhaps most importantly, AI and machine learning require vast amounts of high-quality data. The data lakehouse, with its ability to handle diverse and gigantic datasets, is an ideal environment to feed these models.

## The Evolution of Data Management





## CUSTOMER SUCCESS

## Innovation in Action: Delivering Great Customer Experiences Daily

Casey's is the fifth-largest pizza chain and the third-largest convenience store in the United States. The retailer operates more than 2,400 locations in 16 states and records upwards of 700 million guest transactions annually.

With digital increasingly becoming the "front door" for modern retail, Casey's turned to Salesforce Customer 360 and Data Cloud to help the company organize and marshal its data for better customer experiences.

Today, Casey's has 13 million customer profiles in Data Cloud, over 150 profile attributes, and over six billion engagements captured in a single, unified record of each customer.

Before Salesforce, Casey's had a loyalty program that collected data, but it was scattered across disparate systems, and not fully utilized. The retailer would send the same message to 300,000 customers a month. Now, Casey's can engage customers with 200 million personalized messages to more than five million loyalty program members every month.



**[I'm] sitting on probably one of the cleanest first-party data sets out of any of my peer sets in the industry."**

**ART SEBASTIAN, VICE PRESIDENT OF  
EXPERIENCES AT CASEY'S**

Using Data Cloud, Casey's  
is able to capture:

**13 million**  
customer profiles

Over  
**150**  
profile attributes

Over  
**6 billion**  
engagements

**a single, unified record  
of each customer**



# AWS and Salesforce: A Partnership for an AI-First Economy

Our partnership with AWS is set to redefine the landscape of data management and artificial intelligence (AI).



Salesforce, the world's #1 CRM, has long excelled at creating solutions that help businesses put customers at the center of everything they do. With Data Cloud, we've extended our prowess beyond conventional CRM boundaries, harmonizing diverse data types and harnessing the raw potential of data.

On the other side of the partnership, AWS has built a reputation as a powerful, flexible, and scalable cloud services provider, powering millions of businesses worldwide.

Data Cloud combined with AWS's expansive infrastructure creates a powerful platform, transforming the way businesses manage and use their data. But it's not just about providing a reliable and scalable platform for data management; the partnership also brings AI capabilities to the forefront.

With the integration of Salesforce's Einstein, Data Cloud, and AWS's advanced machine learning services, businesses can access powerful AI tools that can parse through massive datasets in Data Cloud, identify patterns, make predictions, and provide actionable insights. This fusion empowers businesses to leverage AI as a practical tool to drive decision-making and spark business growth.



**Our partnership with Salesforce signifies more than a corporate alliance. It represents the commitment of two cloud pioneers to redefine data management, harness the power of AI, and shape the future of business in the digital era.”**

**SWAMI SIVASUBRAMANIAN,  
VP OF DATABASE, ANALYTICS AND ML AT AWS**



## Broadening the Horizon of What's Possible

Data Cloud marries the robustness of Salesforce's metadata-first platform with the transformative power of customer-aware AI. Seamlessly blend and manage data from various sources to create a unified, scalable, agile, and versatile ecosystem for AI-powered customer experiences that will grow with your business needs.

[Learn more](#)





The information provided in this e-book is strictly for the convenience of our customers and is for general informational purposes only. Publication by salesforce.com does not constitute an endorsement. Salesforce.com does not warrant the accuracy or completeness of any information, text, graphics, links, or other items contained within this guide. Salesforce.com does not guarantee you will achieve any specific results if you follow any advice in the report. It may be advisable for you to consult with a professional such as a lawyer, accountant, architect, business advisor, or professional engineer to get specific advice that applies to your specific situation.