Softcat Plc

Salesforce Data 360 (formerly Data Cloud)

- Cloud Software - Digital Marketplace - Lot 2b

- G-Cloud 15

CONTACT INFO

Dominic Clark

psitq@softcat.com

Table of Contents

Salesforce Data 360

1. Service Description

Salesforce Data Cloud [Former name] is now [Data 360]. You may continue to see the former name in our application and documentation.A data cloud is a centralised, cloud-based platform designed to store, integrate, and manage massive volumes of data from various sources. It offers the scalability, flexibility, and processing power required to handle the complexity of modern business data, structured, semi-structured, and unstructured.

Unlike traditional on-premise databases or data warehouses, a data cloud leverages the architecture of the public or private cloud, enabling real-time data access and analysis throughout your organization. A data cloud can break down data silos and create a unified view of your business. The ultimate goal is to transform raw, dispersed information into trusted, actionable insights. With a data cloud, you can prepare your data for advanced analytics, artificial intelligence (AI), and agentic AI.

1.1. Service Overview

For Salesforce Users

Traditional BI tools are designed for complex analysis, but are built for an analyst. Einstein Analytics is specifically designed for front office sales, service or marketing users. Einstein Analytics closes the gap between the business user and insight to have a data-driven conversation.

Mobile Self-Service

Business users freely explore and drill through data to get answers, just like how we do a Google search. And users get the answers they need on any device – on the desktop, or on the phone.

Cloud Trust & Speed

Einstein Analytics native to Salesforce. This means you can unlock customer insights right from your CRM, without having to switch systems. It also means you can embed insights directly onto Salesforce objects, such as Accounts or Contacts. Additionally, it's built on Salesforce's #1 trusted cloud platform, which means it's scalable, secure, and enterprise ready. You'll get the same innovation, 3 releases a year of new features.

Service Analytics

Discover insights for every channel and customer conversation. Automatically populate dashboards with Service Cloud data, and gain visibility into your business with omni-channel, chat, activity, and backlog analysis. Then, embed customer service dashboards anywhere in your CRM to increase visibility within the company and promote collaboration.

Reach faster resolutions and create happier customers when you explore and act on trends in every channel. With Service Analytics, agents and managers can create tasks, escalate cases, or even open opportunities from within the Salesforce apps you already use, on any device.

Deliver personalised service with predictive CSAT-based existing Salesforce data and trends. Every employee can stay ahead of business concerns by identifying product issues, customer issues, and churn risks with data-driven insights. And reps resolve issues faster and get recommendations for the next best actions.

Easily drill into customer profiles and case history with Service Analytics. Optimise agent productivity and case queues to deliver better customer experience. Create cases, update case history, and collaborate with your team directly from the Service Cloud console you already use. Since it's not structured like legacy business intelligence solutions, you can quickly diagnose service issues and find the best answers quickly.

Sales Analytics
Sales Analytics App delivers a complete view of customer data, pipeline health, best practices for closing deals, and more for Sales Cloud customers.

Easily share best-practice dashboards among your entire sales team. Power your weekly pipeline call, quarterly business review, and performance metrics using Sales Analytics templates prepopulated with your Sales Cloud data and preconfigured with key performance indicators (KPIs).

Sales managers can finally track year-over-year business performance and get a better view of the big picture. Review pipeline movement and risk indicators, and forecast metrics to identify behaviours that drive sales.

Create or update records and objects with newfound insights or answers, right within the app — instantly communicate your findings and collaborate on next steps through Chatter, on any device.

1.2. Service Features

- Take action on insights into any business process in Salesforce
- Easily build, customise, and extend to meet your business needs
- Make smarter decisions with AI-powered discovery, predictions, and recommendations
- Quickly explore millions of data points
- No-code & Low-code declarative configuration, Open API integration, Training included.
- Descriptive, Diagnostic, and Prescriptive analysis
- Built on the proven Salesforce Platform facilitating agile development
- Open, API first service for easy integration (SOAP and REST)

1.2.1. Data 360 Database Management System Capabilities

1. **Relational Database Management Systems** - Data 360 does support relational capabilities through its lakehouse architecture with structured data storage and SOQL querying. However, it's built on a modern lakehouse foundation rather than traditional RDBMS.
2. **Low-Code Database Management Systems** - Data 360 provides declarative, low-code capabilities through the Salesforce Platform with drag-and-drop data modeling, visual flow builders, and point-and-click configuration.
3. **Navigational Database Management Systems** - Data 360 doesn't implement hierarchical or network-style navigational database models. It uses modern lakehouse architecture with relational and document models.
4. **Fixed Record Database Management Systems** - Data 360 supports flexible schema evolution and dynamic data models, not fixed record structures. The architecture specifically enables schema evolution through Apache Iceberg table format.
5. **Object-Oriented Database Management Systems** - While Data 360 has object-like structures through its Customer Data Model (CDM), it's not a true OODBMS. It provides object abstraction through the Salesforce Platform layer, making it an Object-Relational abstraction on top of a big data stack.
6. **Multivalue Database Management Systems** - Data 360 doesn't support traditional multivalue database capabilities like nested repeating groups or multivalued attributes in the classic sense. While Data 360 can ingest complex JSON arrays, it typically flattens or normalizes this data into a relational structure for harmonization.
7. **Non-Schematic Database Management Systems** - Salesforce allos you to ingest unstructured data to Data 360. Refs https://help.salesforce.com/s/articleView?id=data.c360_a_unstructured_data_connect_overview.htm&type=5 . unstructured data in Data 360 to ground Agentforce agents, generative AI, analytics, and automation use cases with business-specific data that delivers deeper insights for your users and customers. Data 360 supports several connectors and file formats for unstructured data.

Unstructured data comes in many forms. Consider these common examples:

a. Text documents - Case notes, reports, internal memos, handwritten notes, invoices, business cards, and emails that contain detailed descriptions or conversations
b. Multimedia file- Audio files from customer service calls, podcasts, video recordings from training sessions, tutorials, meetings, images from product reviews, infographics, scanned documents, and medical imaging
c. Social media content- Comments, posts, or messages containing customer feedback, opinions, or sentiments
d. Web content - Blogs, articles, and forum posts that discuss products, services, or customer experiences
e. Chat logs - Conversations from live chat support, often containing nuanced questions, complaints, and preferences
f. Sensor and IoT data - Data from wearable devices or machinery that doesn't follow a standard format, such as raw accelerometer data
g. Customer feedback - Survey responses, feedback forms, or reviews written in free text

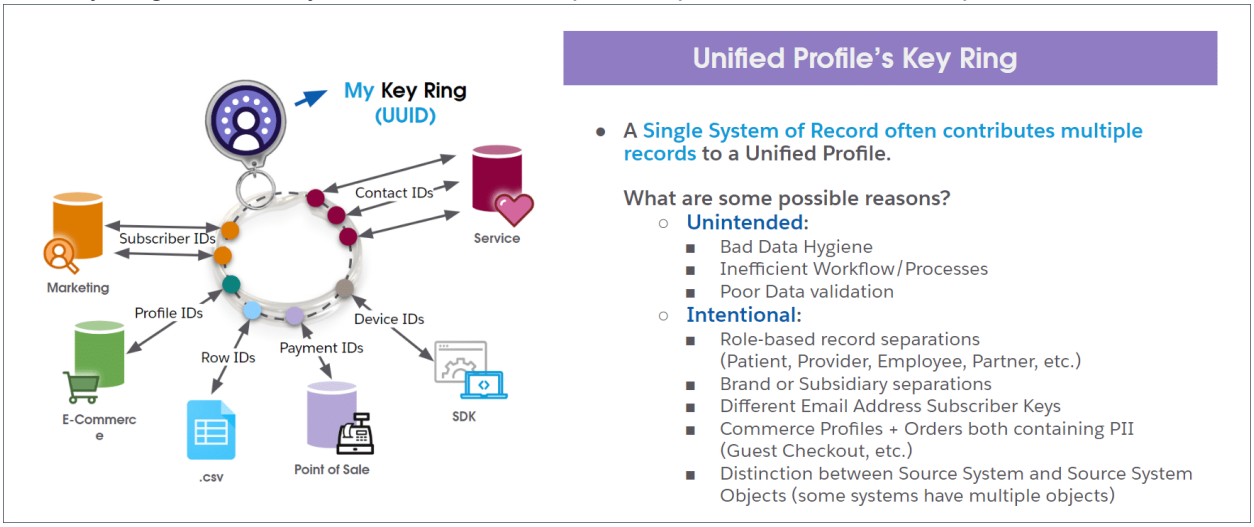
See Unstructured Data Connectors (https://help.salesforce.com/s/articleView?id=data.c360_a_unstructured_data_connect.htm&type=5) and Data 360 Limits and Guidelines (https://help.salesforce.com/s/articleView?id=data.c360_a_limits_and_guidelines.htm&type=5).

Using unstructured data in your AI, automation, and analytics workflows can help you make informed decisions and improve customer experiences.

In summary, while Data 360 supports semi-structured data ingestion, it requires mapping to structured schemas for processing and querying.

8. Document-Oriented Database Systems - Data 360 supports document storage through MongoDB as a warm store for Data Graphs (DGs) and can handle JSON/semi-structured data.

9. Key-Accessible Database Systems - Data 360 provides key-value access patterns through its object store capabilities backends for fast lookups. For "Hot" storage—specifically for the "Key Ring" and Identity Resolution, where rapid lookups of Unified IDs are required.



Refs (Rethinking the Golden Record: The Advantages of Data Cloud's Unified Profile https://admin.salesforce.com/blog/2025/rethinking-golden-record-advantages-of-data-cloud-unified-profile)

10. Graph Database Management Systems - Data 360 supports graph-like relationships through Data Graphs (DGs) but isn't a native graph database like Neo4j. It provides relationship modeling within its document store.

11. In-Memory Shared Data Managers - Data 360 implements distributed caching as a "hot store" with ~2ms latency backed by object storage for extreme low latency access. Data 360 leverages Apache Spark for in-memory processing of massive datasets during segmentation and calculated insights. This allows for high-speed "real-time" data processing that traditional disk-based systems cannot match.

12. Data Lake Management Systems - Data 360 is fundamentally a lakehouse platform combining data lake and warehouse capabilities, using Apache Iceberg table format with Parquet columnar storage on cloud object stores. Refs Guide to Data Lakehouses (https://www.salesforce.com/data/what-is-a-data-lakehouse/#:~:text=A%20data%20lakehouse%20is%20a,insight%20and%20value%20from%20it.)

Data lakehouses help you manage massive volumes of data and act on it—fast. As CIOs look to consolidate apps, streamline workflows, and become more efficient, data lakehouses can make a significant impact on their bottom lines. See how you can future-proof your customer personalization and automation efforts with data lakehouses.

What is a data lakehouse?
A data lakehouse is a modern architecture that gathers all of your organization's unstructured, structured, and semi-structured data and stores it at a low cost—while keeping it highly accessible to users. A data lakehouse essentially takes the best features of data warehouses and data lakes, bridging the gap between structured and unstructured data in one system.

For example, imagine a marketing team analyzing campaign performance. A data lakehouse allows them to combine structured data (e.g., sales figures) with unstructured data (e.g., customer sentiment from social media) to build more personalized campaigns—all without switching between systems.

Some data lakehouses benefit from a "zero-copy principle," which allows IT teams to avoid the need for data copies and cumbersome extract, transform, and load (ETL) tools to improve compute performance. The end result is less time, less effort, less cost, and less latency involved in not just managing information, but quickly getting insight and value from it.

Data lakehouse vs. data lake vs. data warehouse
Let's take a closer look at how data lakehouses use the best of data lakes and data warehouses.

Data warehouse
A data warehouse can house large amounts of data that has already been processed. Data warehouses are very good at storing and applying business analytics to structured data (such as numbers and addresses). But they require time-consuming ETL tools to import data from other systems of record.

Data lakes
A data lake is a pool of raw data you can use to centrally store data in its raw form. Data lakes were built to capture the vast (and continually growing) wealth of unstructured data, such as social media posts, images and audio files. Extracting useful insights often requires data science skills.

Data lakehouse
A data lakehouse combines the best features of data warehouse and data lake technology while overcoming their limitations. Data lakehouses make it much faster and easier for you to extract insights from all of your stored data, no matter what format it is in or how large its volume. You get the low-cost, flexible storage of a data lake with the data management, schema, and governance of a data warehouse.

Why your business can benefit from a data lakehouse
The volume of data businesses generate is growing at an unprecedented rate. Organizations handle petabytes of data across hundreds of systems, whether it's customer interactions or IoT

sensor data. In fact, the average business uses 976 applications to track customers, according to recent studies.

The challenge? Each of these applications creates its own siloed version of customer records, leaving businesses with fragmented insights. We're talking 976 versions of one customer when only one will do. These silos slow decision-making, increase operational costs, and limit innovation.

How can a data lakehouse help your business?
By unifying structured and unstructured data in a single system, a data lakehouse offers benefits that help organizations work faster, smarter, and more efficiently.

Here's how a data lakehouse can help your business.

Scale and flexibility, with structure and schema: Handle growing volumes of data—both organized records and unstructured content—without the limitations of traditional systems. Organize data as needed to make it ready for analysis or AI.

Reduce silos and improve operational efficiency: Integrate data from across hundreds of apps and systems, breaking down barriers and providing a unified view. This speeds up decision-making and eliminates duplicate efforts.

Lower costs: Store raw and processed data affordably by separating storage and computing. Need more storage? Add it without paying for extra computing power. Need more processing? Scale it without unnecessary storage costs.

Enable real-time insights: Get up-to-date insights without the delays caused by moving and reformatting data between systems. For example, sales reps can act on the latest customer interactions to close deals faster, and marketers can adjust campaigns on the fly.

Support for advanced analytics, AI, and agentic AI: Use unified customer profiles to power predictions or AI-based tools—all from one system. No need for expensive engineering or extra platforms.

Data lakehouse vs. investments in data solutions
Your existing solutions can stay put. There's no need to "rip and replace" when adopting a data lakehouse.

Thanks to their open data protocols, data lakehouses can integrate easily with legacy apps and systems, whether they're pulling in first-party ad data, business intelligence (BI) tools, or proprietary AI models. You can then begin to phase out obsolete data management tools that require a lot of care and feeding on your timetable. Like any powerful technology, a data lakehouse should adapt to changes in your business strategy — not box you in.

Data lakehouse security and compliance
Businesses can drastically simplify data governance and compliance without slowing the pace of innovation. We've seen this as a top concern for many of today's IT and business leaders, according to our IT & Business Alignment Barometer.

Data lakehouses can consolidate multiple systems for data management into one platform — reducing the amount of data spread across systems and the number of hands data travels through. They allow you to exert more control over security, authorization levels, and more, thanks to their open schema.

As an IT leader, you can implement role-based access, so that marketing teams only have access to segmentation data, and sales teams only have access to order data. You can also audit who's requesting data from the lakehouse, from where, and from which functions.

Real-World Use Cases for Data Lakehouses

A data lakehouse can make your data work smarter—no matter what industry you're in.

Healthcare: Bring together patient records and data from connected devices such as wearables to improve diagnostics and care. Hospitals can use AI tools to predict readmission risks and act faster to better serve patients.

Finance: Spot fraud faster by analyzing transaction data and customer call records side by side. Banks can reduce losses by identifying suspicious activity in seconds instead of hours.

Manufacturing: Combine sensor data from factory equipment with supply chain data to predict delays or shortages. Automakers can prevent bottlenecks and keep production running smoothly.

Experience the power of a data lakehouse

You have a massive volume of information, and every decision you make relies on the quality and accessibility of your data. A data lakehouse simplifies the way you store, manage, and use your data so that the right people can use it when they need it most.

Make the most of your data with the right data platform

The right data platform gives you the freedom to unify your data, personalize customer experiences, and inform sharp decisions. A data lakehouse is a great choice if you want to make your data work smarter for you. And because your data will continue to grow, take a look at Data 360, a hyperscale engine inside Salesforce that fuels intelligent decisions and agentic AI.

1.2.2. Data 360 Architecture

Refs (https://architect.salesforce.com/fundamentals/data-360-architecture)

Context
Data platforms have been evolving for over three decades. Initially, the industry was dominated by on-premise, centralized, and structured (mostly relational), operational/OLTP databases. This expanded to include data warehouses OLAP/Big Data platforms that were primarily used for analytical processing and remained relational and centralized. Cloud storage drove distributed architectures like data warehouses, lakehouses, and disaggregated storage. However, operations platforms and analytical platforms stayed separate. Today, cloud computing and the AI revolution are foundationally changing the data platform architecture.

Enterprises already invest in mature Big Data platforms like Snowflake, Databricks, BigQuery, and Redshift. But these platforms serve as data silos. Customers aren't deriving business value out of their data because the data can't be acted on directly inside the business flows and applications. These solutions lack generative Agentic AI processing and can't deliver data access in real time, so they're unable to deliver AI-driven personalization at the moment of customer engagement and other industry-leading features.

The future of data platforms is characterized by unified, flexible, accessible, and open data infrastructure. This new architecture is built on modern compute and storage trends—GPUs, large memory, NVMe SSDs, and cloud storage—to integrate with cloud computing and AI. They're able to deliver real-time insights, power autonomous decision-making, and drive real-time applications. This includes the rise of agentic AI, predictive AI, analytics, real-time high scale OLTP databases, data lakes, and lakehouses. These modern data platforms are designed for simplicity, scalability, agility, performance, security, availability, and cost efficiency.

Key Trends in Next Generation Data Platform Architecture
The following data trends drive the next generation data platform architecture.

- **AI, Machine Learning, and Analytics at the core**: The rise of Agentic AI will foundationally change data platform development, deployment and usage/access. Agentic AI will understand conversation/query intent, plan, generate workflows, and automate decision making. Agentic (short and long term) memory is constructed from conversation history for personalizing agent planning and

decisions, real-time conversation modeling and personalization support critical in data platforms. Agents will assist in automating operational “abilities” like data governance (i.e. security, compliance, trust), performance (i.e. auto scaling for concurrency, throughput, and latency), failover and availability, and observability and maintenance. AI-powered analytics, forecasting, natural language processing (NLP) for analytics Q/A, and analytics on unstructured data (text like PDFs, images, audio, video) will be standard, allowing enterprises to derive deeper insights from diverse data sources.

- **Decentralization of Data but Unified Data Access**: Agents need enterprise data to derive insights and to make decisions, and to automate business activities. Data is inherently decentralized in enterprises, in disparate applications and data platforms. But it is not easy to seamlessly unify the silos across different business units within the enterprise and with partners outside the enterprise. Unifying data involves data sharing, either via ingestion from sources or federating with data sources; raw data from data preparation, harmonization, and modeling for analytical and AI processing; storage and management of data at scale for efficient access with low CTS; and data access via various query and analytics mechanisms and tools, deeply integrated with the underlying storage and data access platforms
- **Cloud-based Open Lakehouses**: Cloud-based Big Data (OLAP) platforms are converging on adopting open file formats (Parquet) and table formats (Iceberg) enabling data federation (data in) and sharing (data out).
- **Unstructured Data Processing**: With the emergence, advancement, and adoption of generative AI, enterprises are beginning to derive valuable insights and business value out of the enterprise corpus of data constituting large volumes of text documents, audio transcripts, video recordings, and other media. Unstructured data processing, including chunking, vectorization, semantic search, and knowledge graphs, make these insights possible. Techniques like RAG (retrieval augmented generation) and CAG (cache augmented generation) are becoming mainstream drivers of fast and agentic search across the corpus of data.
- **Knowledge Management**: Knowledge goes beyond just the raw content itself (documents, articles, videos). It represents augmentation of that content by deriving meaning, curating metadata, and placing it in context to develop a shared understanding of content across an organization or enterprise. Knowledge itself is generally structured. Knowledge management involves content management, knowledge extraction, representation through models like graphs, and navigation.
- **Rich Data Access**: Rich data access means that data, analytics, and AI tools must be accessible to a variety of personas including end users, business users, admins, and analysts. Accessibility is achieved through mechanisms such as ensemble querying (with relational, keyword, and semantic query), natural language to SQL (NL2SQL) querying, real-time access, and so on.
- **Real-Time Processing**: Agentic applications make real-time decisions based on current state and based on new events, personalizing responses, and actions, which requires accessing, processing, and acting on real-time data. Real-time

processing requires up-to-date data (data latency) and interactive access (access latency). Such data and access latency require the underlying data platform to support up-to-date data access from operational and analytical stores, low latency access (point lookups and query) processing, high data scale, and high throughput.

- **Data Security, Governance, and Residency**: Agentic and conversational AI simplifies application UI, allowing anyone, from consumers to employees to other AI agents, to interact with applications conversationally using spoken or written natural language. The valuable customer and personal data that must be stored and modeled for Agentic applications must be secured and governed with well-defined access and sharing policies. Increasingly, many customers must comply with regulations that require data residency in their own country or region, especially those in government or doing work with governments.

Salesforce Data 360 is architected for the future addressing these data trends. Data 360 is a cloud-native, metadata-driven data platform that unifies siloed data across the enterprise, allowing organizations to store, model, and process their data to enable analytics, AI, machine learning, and Agentic applications.

This document is an essential guide for enterprise architects and CTOs. It details the architecture, capabilities, design principles, and use cases of Data 360. It introduces the fundamentals of Data 360 architecture as a primer, followed by a series of deep dives into its key differentiators such as interoperability with existing data platforms, including multi-org strategy, security, governance and privacy, real-time activation, and data Clean Rooms.

Design Principles

Salesforce Data 360 is engineered around a core set of principles that make enterprise data operational, trusted, and real-time.

- **Openness and Interoperability**: Built for multi-cloud ecosystems. Federates with data platforms like Snowflake, Databricks, BigQuery, and Redshift without duplication, extending Customer 360 while preserving existing investments.
- **Storage-Compute Separation**: Scales storage and processing (batch, streaming, and interactive) independently. Delivers elasticity and efficiency for high-volume, high-performance workloads.
- **Multi-Model Storage and Processing**: Supports structured and various unstructured data types like text, image audio, and video. Provides efficient storage, real-time and batch processing, extensible indexing, unified search, querying and analysis.
- **Metadata-Driven Design**: Applications are defined by metadata rather than code. Metadata is treated as a first-class asset, enabling unified governance, flexibility, and deep integration into the Salesforce platform.
- **Real-Time Hybrid Processing**: Supports low-latency queries and instant decision-making, along with batch processing and analytical workloads.

- **Intelligent and Active Data**: Continuously ingests, analyzes, and pushes insights directly into business workflows. Powers no-code, low-code, pro-code and AI-driven automation with the most current context.
- **Governance and Privacy by Design**: Lineage, access control, residency, data encryption and compliance are built in. Trust and regulatory confidence are reinforced at every layer.
- **One-to-Many Tenancy**: A centralized Data 360 Org serves as the single Source of Truth for Customer 360, seamlessly supporting multi-org salesforce environments widely used by Salesforce customers.

These principles ensure Data 360 makes data open, intelligent, and actionable in real time.

From Principles to Capabilities
Salesforce Data 360 is a modern data platform built on design principles that address current data trends. Its architecture capabilities ensure enterprise data is trusted, unified, and actionable in real time, aligning with its guiding principles.

- **Cloud-Native Foundation**: Runs on Hyperforce, deployed on Hyperscalers (like AWS), with immutable, microservices-based infrastructure. Provides elastic scaling, zero-trust security, continuous delivery, and global compliance.
- **Salesforce (Core) Metadata-driven**: Metadata is designed, modeled, and stored as Salesforce metadata enabling immediate use by ALL Salesforce applications. Such metadata is stored in a fully ACID-compliant RDBMS. Ensures governance, lifecycle consistency, and deep integration with Salesforce Lightning Platform.
- **Lakehouse Storage**: Built on Apache Iceberg and Parquet, combining data lake scale with warehouse governance supporting schema evolution, time travel, and high-volume updates. Data 360 stores, models, and processes structured and unstructured data with extreme-scale storage with modern open Standards, and with rich transformation and data processing capabilities for batch and event-driven workloads.
- **End-to-End Data Pipeline with flexible ingestion**: Covers the full lifecycle—ingest, prepare, model, unify, analyze, and activate—reducing reliance on fragmented point solutions. Supports batch, near real-time, and streaming with 270+ connectors and MuleSoft. ELT-first approach enables fast data availability with downstream transformation flexibility.
- **Enterprise Data Interoperability with Open frameworks and Federation**: Unifies silo'd data across the enterprise with bidirectional Zero Copy federation with Snowflake, Databricks, BigQuery, and Redshift avoiding data migration or duplication.
- **Data Classification, Modeling, and Organization**: Data 360 organizes data as raw ingested data, cleaned and stored data, and data modeled conforming to a common information schema known as SSOT (Single Source of Truth). Such SSOT objects form the basis for defining Semantic Data Models (SDM) and other curated and application-specific models.

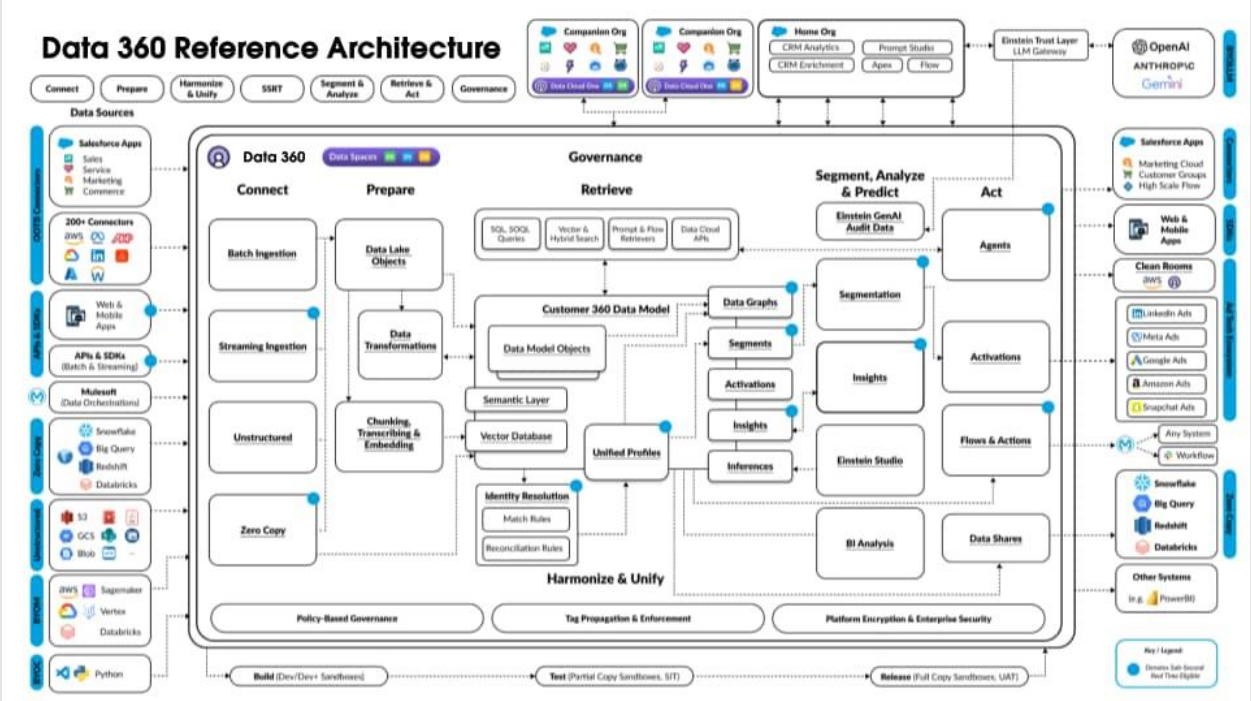
- **Built-in Semantic Data Modeling** for extensible analytics with open semantic query APIs, driving Tableau Next and enabling application-specific analytics.
- **High Performance SQL query engine** supporting a unified Data 360 SQL query across structured, unstructured, and graph data.
- **Low-Latency Data Stores**: Key-value storage for hot data with millisecond response times. Enables personalization and event-driven scenarios in real time. Collects and Processes customer engagement data in real-time. Unifies identities, interactions, and conversations into a single, trusted Customer 360 Profile and Context graph.
- **Unstructured Data Processing Pipelines** for flexible and extensible support for unstructured data storage, Chunking, Embedding generation (vectorization), Metadata extraction (augmentation), Summarization, Indexing, Knowledge extraction, Intelligent document processing, Short and Long term (conversation) memory creation, and so on.
- **Native Keyword, Vector, and Hybrid Indexing** for accurate and efficient unstructured data access like Fast and Agentic search, RAG, Knowledge extraction, Agentic memory derivation, etc.
- **Profile, Personalization, Context services** for enabling AI/ML and Agentic applications.
- **Built-in governance and security** at every layer for lineage tracking, data masking, data residency, and zero-trust security ensuring compliance and trust.
- **Elastic Compute Fabric**: Kubernetes-native, multi-tenant compute fabric. Runs Spark for distributed processing and Hyper for SQL workloads. Scales elastically across diverse jobs and support isolation for running untrusted code.

All of this runs on Hyperforce, Salesforce's cloud foundation. Hyperforce provides:

- **Zero-trust security** with encryption, isolation, and least-privilege policies.
- **Resilience** through multi-region deployments. While Salesforce Data 360 benefits from Hyperforce's multi-region resilience and platform-level fault tolerance, true enterprise-grade disaster recovery (DR) demands a broader architecture similar to any data platform with key capabilities: replayable ingestion pipelines, replication, and metadata-driven rehydration across all dependent ecosystems.
- **Observability** with monitoring, metrics, and tracing built in.
- **Automated scale and FinOps awareness** for efficiency without cost overflow.

Data 360 does not replace existing enterprise investments. Instead Data 360 makes the data you already have trusted, governed, and actionable—delivering real-time, AI-driven engagement where it matters most. In short, Salesforce turns all enterprise data including external data as (Salesforce) metadata-driven objects and enables Agentic applications for discovery, decision-making, and for taking actions.

The following figure illustrates the Data 360 Reference Architecture:



Let us consider a hypothetical Agentforce Loan Agent layered on Data 360 to describe an example architecture flow. Say the Loan Agent is a customer-facing agent where customers (consumers) apply for loans and get instant loan approvals.

Data 360 performs these steps as scheduled, preparing data for use by the Loan Agent.

- Data 360 ingests structured Customer Account data from CRM and stores it in the data lake.
- Data 360 processes unstructured company loan and financial policy data.
- Data 360 federates Personal data from an external data source such as Snowflake.
- Data 360 transforms and models ingested and federated data.
- Data 360 builds and maintains the profile data graph.

Each time a customer applies for a loan, these actions are performed.

- A customer signs into the Loan Agent, which begins a customer session in the real-time layer. The customer's unified profile is pulled into the real-time layer.
- The customer completes a loan application by providing the required information.
- The customer uploads Financial documents (like tax returns, investments, bank statements) to Data 360 for unstructured data processing.
- Uploaded data is chunked and vectorized (embedding generation), and indexes (keyword and vector) are created.

- Next, the customer fills out the loan application document and uploads it. Data 360 extracts the loan amount and duration in real time.
- The Loan Agent retrieves relevant financial data using Data 360 query and hybrid search over profile and other pre-created indexes.
- The Loan Agent activates an Approval Agent with loan data and other financial profile data to make the loan approval decision.
- The Loan Agent responds to the customer with a decision.
- This entire interaction between the customer and the Loan Agent is also captured and stored in Data 360.

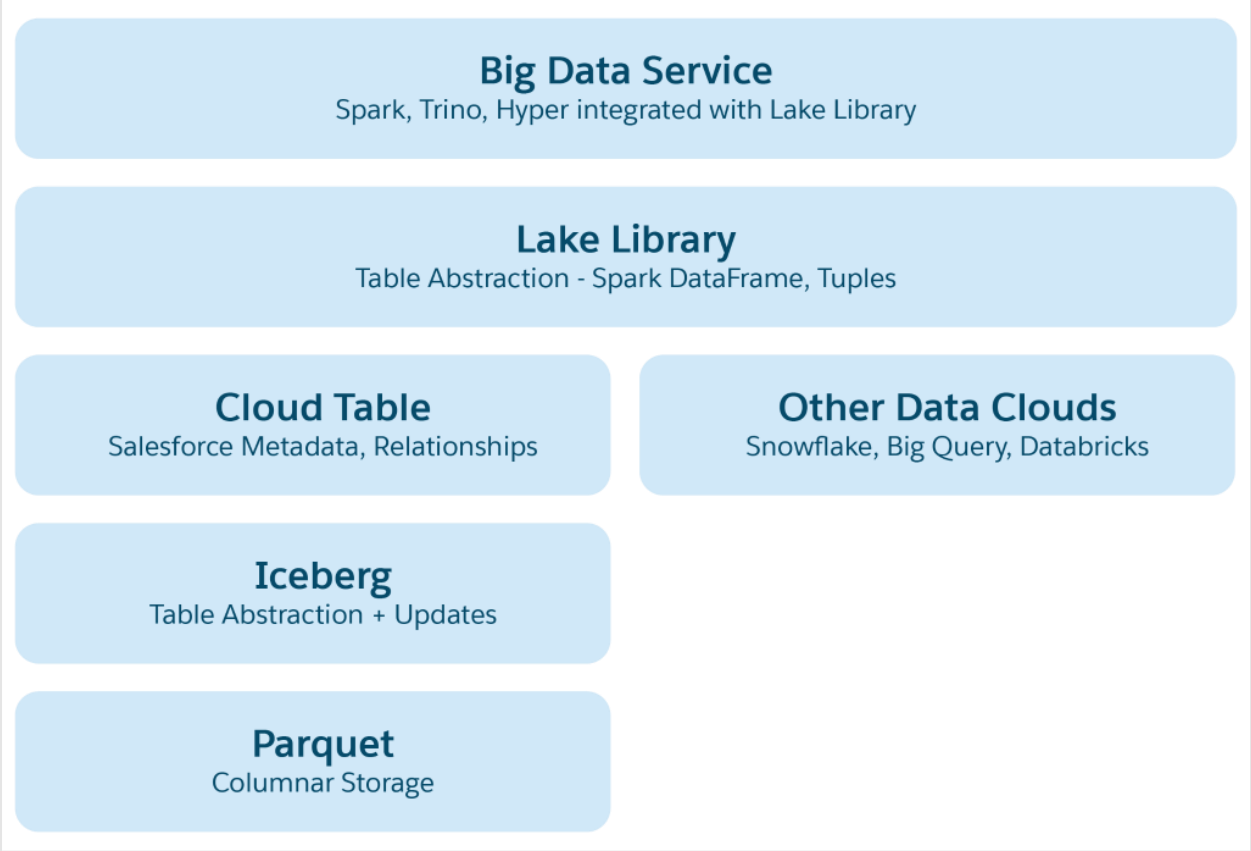
The above example provides an overview of Data 360 architecture components used to build an Agentic application like a Loan Agent. In the next section we describe the Data 360 architecture layers and components.

Data 360 Architecture
In this section, we will delve into the foundational building blocks of Salesforce Data 360, beginning with its robust storage model and subsequently exploring the mechanisms for connecting, ingesting, and preparing data. We will then examine how structured and unstructured data is stored, modeled, and processed, culminating in an understanding of its harmonization, unification, retrieval, and intelligent activation capabilities.

Data 360 Storage
Salesforce Data 360 is built on a tiered but integrated storage model that combines the strengths of a lakehouse with real-time storage. The lakehouse layer provides scalable, cost-efficient storage for large volumes of historical and batch data, enabling advanced analytics and machine learning use cases. Real-time storage, on the other hand, is optimized for low-latency access and high-frequency updates, ensuring that customer interactions, profiles, and engagement signals are always current. Together, these tiers work seamlessly, allowing data to move fluidly between historical and real-time contexts—delivering both depth and immediacy in a unified data foundation for personalization, AI, and activation.

Lakehouse
Data 360 features a native lakehouse architecture based on Iceberg/Parquet, designed to handle large-scale data management and processing for batch, streaming, and real-time scenarios supporting both structured and unstructured data, crucial for AI and analytics applications.

In cloud-based data lakes like Azure, AWS, or GCP, the fundamental storage unit is a file, typically organized into folders and hierarchies. Lakehouse enhances this structure by introducing higher-level structural and semantic abstractions to facilitate operations like querying and AI/ML processing. The primary abstraction is a table with metadata that defines its structure and semantics, incorporating elements from open-source projects like Iceberg or Delta Lake, with additional semantic layers added by Data 360.



Abstraction Layers in Lakehouse:

1. **Parquet File Abstraction**: At the base, storage consists of data lake files (for example S3 in AWS or Blob in Azure) in Parquet format. Data for a source table is stored across multiple partitions as Parquet files, with each table being a collection of these files.
2. **Iceberg Table Abstraction**: Tables are organized as folders, with data partitions stored as Parquet files within these folders. Modifications to a partition result in new Parquet files as snapshots. Iceberg manages a metadata file for each table, detailing schemas, partition specs, and snapshots.
3. **Salesforce Cloud Table Abstraction**: Building on Iceberg, this layer adds semantic metadata such as column names and relationships, along with configurations like target file size and compression. It abstracts tables across various platforms like Snowflake and Databricks, shielding Data 360 applications from underlying storage platform specifics.
4. **Lake Access Library**: This library provides access to the Salesforce Cloud Table, handling both data and metadata, and abstracts the underlying storage mechanisms for application developers.
5. **Big Data Service Abstraction**: This includes processing frameworks like Hyper for querying, and Spark for processing across any cloud table platform.

6.

Real Time Storage
To support real-time analytics and agentic applications, Data 360 augments the Lakehouse big data storage with Low Latency Store. Data 360 Real-Time Layer processes real-time signals and engagement data in memory. However, since the memory-based storage capacity is limited, all data cannot fit and processing may not be done in real-time. Data 360 adds a low latency store (LLS) to remove such limitations, enabling scalable real-time processing.

The low latency store is a petabyte scale NVMe (SSD) storage layer on the Lakehouse. Not all data needs to be kept in the low latency store. It is a durable cache. Most data eventually makes it to the Lakehouse for long term persistence. In-session data in the real time layer can be flushed to the low latency store for subsequent fast access. For example, in an Agentic conversation, recent messages can be processed in memory; older messages can be flushed to the low latency store. If a previous conversation is required, it can be accessed within a few milliseconds from the low latency store. NVMe based storage allows for large amounts of data to be stored and accessed at millisecond latencies. Data may make its way to the Lakehouse cloud storage for long term persistence. In addition, data from Lakehouse that is required for real time processing or to augment real time experiences is fetched and kept in the low latency store. For example, customer Profile Context is pre-fetched or brought from the Lakehouse and cached in the low latency store. Also, any lakehouse objects and other objects that are required for real-time processing during in-session processing can also be cached in the low latency store.

Data 360 low latency store enables the Real Timer layer on a true storage hierarchy with memory (SSD) Lakehouse storage layers, with data seamlessly migrating between these layers. We discuss the Data 360 Real Time layer later in this document.

Data Organisation
Salesforce Data 360 is engineered to standardize, harmonize, and activate all customer data—structured and unstructured—following a rigorous lifecycle that transforms raw input into a unified, current data model.

Data Lifecycle and Object Types
The lifecycle focuses on taking various external data inputs and structuring them into persistent, modeled objects. Modeled data can be harmonized into Customer 360 unified profiles.

Raw Ingested Data and Initial Transformations
The process begins with raw data ingested as-is from source systems (CRM, Marketing, files, etc.). This includes full data loads and continuous change events (deltas), which are managed and merged with persistent data to maintain a current state.

Inline transformations (e.g., trim, normalize, concatenation) are immediately applied during ingestion to ensure preliminary data quality and cleanliness.

Data Lake Objects (DLOs): The Persistent Layer

DLOs (Data Lake Objects) form the core persistent storage layer within Data 360. They store the clean, transformed data and serve as the organized, long-term repository for all customer information.

Advanced data transformations (e.g., joins, aggregations, calculated insights) are applied to source DLOs to produce new, highly curated derived DLOs.

Data which is made available through Zero Copy Data Federation is represented directly as DLOs.

Unstructured Data and Metadata Organization
For unstructured content (like text, media, documents), Data 360 incorporates the data by extracting and persisting its structured metadata within specific DLOs called Unstructured Data Lake Objects (UDLOs).

These specialized DLOs function as directory tables, providing a map to the physical location and context of the unstructured assets. This capability allows architects to seamlessly relate the metadata of unstructured data with the rest of the structured customer data, enabling unified querying and harmonization.

Data Model Objects (DMOs): The Harmonized Layer

DMOs (Data Model Objects) represent the final, harmonized and structured data layer.

They are created by mapping DLO fields (from source, derived, and unstructured metadata DLOs) to the standard Customer 360 Data Model.

The DMO layer acts as the single source of truth for all customer data, enabling unified profile creation, segmentation, and activation across the broader ecosystem.

Data Space: The Logical Container
A data space is the fundamental logical container for organizing all data and metadata within Data 360, including all DLOs (structured and unstructured) and DMOs. Data spaces offer a secure, isolated environment for data processing and modeling.

Data spaces act as logical and governance boundaries, enabling internal multitenancy by separating data for distinct entities such as business units, regions, or brands—while maintaining enterprise-wide visibility, lineage, and compliance, serving as the basis for defining coarse grained access control.

Isolation within data spaces is enforced at multiple layers of the platform:

- **Data-level isolation**: Each DLO/DMO belongs to a single data space, ensuring that queries, transformations, and object mappings cannot cross data space boundaries unless explicitly authorized.
- **Access control integration**: Permission sets are natively tied to data spaces, allowing control over read, write, and administrative operations. This ensures that

only authorized users and services can access objects, insights, and activations within a data space.

- **Governance and audit**: All operations within a data space are logged with enterprise-grade audit trails, enabling traceability for compliance, stewardship, and regulatory reporting.

Access and permissions are managed through Permission Sets, ensuring granular visibility, controlled updates, and prevention of cross-domain data leakage. By integrating data space boundaries with Data 360's security and governance architecture, architects can confidently implement both centralized and decentralized governance strategies while maintaining consistency across multiple clouds and business domains.

Compute Fabric
Data 360 compute fabric provides a unified layer for managing and executing all big data workloads, simplifying underlying infrastructure complexities. Its core component is the data processing controller (DPC).

DPC is a comprehensive, multi-workload data processing orchestration service that provides Job-as-a-Service (JaaS) capabilities across diverse cloud compute environments. It abstracts infrastructure complexity and unifies job execution for frameworks like Spark (EMR on EC2 and EMR on EKS) and Kubernetes Resource Controller (KRC) workloads. By serving as a centralized control plane gateway, DPC orchestrates, schedules and monitors jobs across multiple data planes, ensuring reliability, scalability, cost efficiency and a consistent developer experience.

The need for DPC stems from the limitations of directly interacting with native cluster management systems like EMR.

Infrastructure and Cloud Abstraction
While EMR offers APIs for clusters, tasks, and steps, it still burdens client teams with critical infrastructure management tasks such as provisioning, scaling, performance tuning, and cost optimization. DPC addresses this by offering a simplified, platform-level API for job submission. It supports automatic failure handling, retries and dynamic load balancing. Provides cost efficiency through binpacking, spot and graviton based nodes. Provides strong security with TLS, PKI, IAM isolation and automated patching. Manages Spark and EMR runtime version upgrades to deliver performance improvements, security patches and feature enhancements.

Moreover, DPC provides a unified, cloud-agnostic interface for submitting and managing data jobs, abstracting the complexities and proprietary APIs of the underlying cloud substrate (AWS, future providers). This ensures that client teams interact solely with a common Data 360 API based job submission interface that abstracts away the complexities of underlying resource managers like Kubernetes and YARN. This allows client teams to submit Spark jobs through a simple, unified API without needing to manage pods, node pools or Spark cluster configurations directly.

Manually tuning Spark parameters requires specialized skills, and incorrect configurations can lead to slow job execution. The DPC team centralizes this expertise, providing optimized configurations to prevent common performance issues. This specialized team continuously

integrates knowledge from the open-source community to ensure optimal performance across all workloads managed by the controller.

DPC is not limited to Spark; it supports a wide array of workloads. These include:

- Real-time processing workloads
- Event delivery for Data Actions feature
- Management of Milvus (the vector database for unstructured data indexing)
- Low-latency storage infrastructure

DPC also leverages the **Kubernetes Resource Controller (KRC)** framework, which supports workloads like Trino for Query, Event Delivery for Data Actions, Data Extraction Jobs for connectors, and Real-time processing. For all KRC workloads, DPC provides central Job-as-a-Service capabilities, handling compute provisioning, deployment, and management at a high-level job abstraction.

JaaS Benefits and Architecture
The Job-as-a-Service model, provided by DPC, ensures a cost-effective and resilient job processing pipeline.

Users provide simple cluster specifications, focusing on required CPU, memory, storage, instance counts, and Min/Max cluster counts and tags for cluster matching. DPC then automatically manages abstract infrastructure details, including selecting optimal VM SKUs, managing Instance Fleets, determining the ratio of Core vs. Task nodes, and managing On-Demand vs. Spot instances based on inputs. It also handles EMR and component version management and upgrades without downtime.

Crucially, DPC inherently supports multitenancy, designed to understand and enforce Data 360 tenancy boundaries and resource separation. It also ensures Security and Compliance by enforcing Salesforce-certified machine images, managing service-specific IAM roles, and guaranteeing encryption both in transit and at rest. For routing and capacity control, job-to-cluster matching is managed using Cluster Tags, and capacity-based routing uses a maximum job concurrency setting to effectively control resource utilization.

The Cloud Agnostic Client Experience is a core benefit, as the complexity of underlying cloud environments is hidden from client services, allowing them to focus purely on business logic. This achieves the goal of Cloud Provider Abstraction. Finally, DPC enables easy Usage and Cost Tracking, allowing cluster utilization and costs to be segmented by service for accurate accounting. Overall, DPC follows a pluggable architecture that allows new execution engines (e.g. Flink, Ray) and cloud substrates (GKE/Dataproc) to be integrated seamlessly without exposing underlying infrastructure details to users. This design decouples the control plane from the execution layer, ensuring a consistent API and operational experience regardless of the backend

Data Processing in Data 360: From Raw to Actionable
Data 360 refines and enriches raw data, bridging the gap between raw information and actionable

business consumption. It complements the data object lifecycle by preparing complex data for sophisticated activation and analysis. Data 360 supports various processing types, including Batch and Streaming Data Transforms, Batch and Streaming Calculated Insights, Unstructured Data Processing, and Identity Resolution. To enable these diverse operations efficiently, especially in near real-time and across massive datasets, a sophisticated mechanism is required to handle data changes effectively.

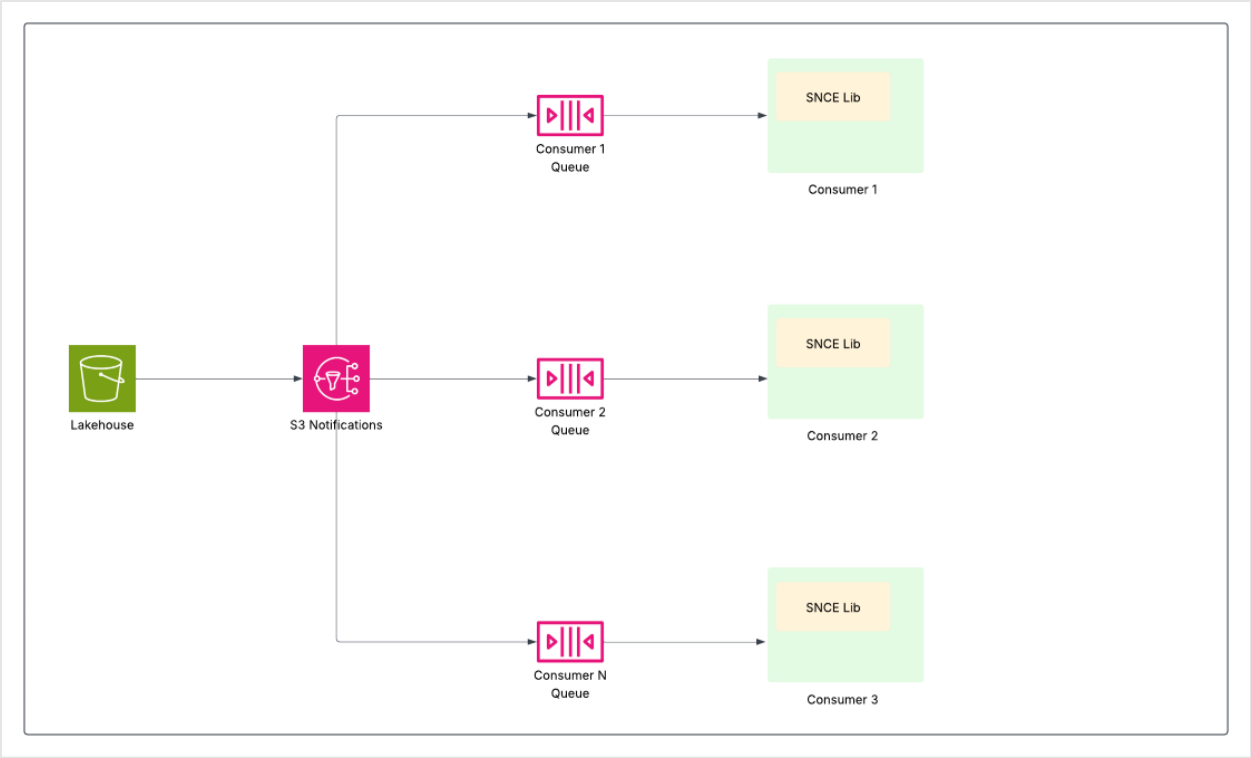
Powering Incremental Processing

To achieve efficient, near real-time data processing—especially with terabyte-sized tables and millions of potential updates—Data 360 needed a breakthrough. It required a way to notify systems precisely when data changes and then efficiently identify what data got changed so that only relevant updates are processed and only when they are updated. This challenge led to two complementary innovations: Storage Native Change Events (SNCE) to notify when something is changed and Change Data Feed (CDF) to identify what changed.

Storage Native Change Events (SNCE)

SNCE has fundamentally changed Data 360 into a reactive and incremental data platform. This shift involves moving from actively polling the data lake to passively monitoring for atomic commit events, using a standardized event format and a high-throughput message delivery system.

Every successful write operation (INSERT, UPDATE, DELETE) to an Iceberg table culminates in an atomic swap of the table's current metadata file pointer in the catalog. The underlying object storage layer (say S3) is configured to emit a native notification event (like an S3 event) whenever a new metadata snapshot is written to the table's directory.



The SNCE library offers a standardized method for consuming these events, and it can also enrich them with snapshot metadata upon request.

This enables downstream data pipelines—such as streaming calculated insights, identity resolution, and segmentation—to subscribe and act only when data has changed, significantly boosting efficiency by avoiding costly full-table scans.

Change Data Feed (CDF)
Building upon SNCE, the Change Data Feed (CDF) provides a streamlined mechanism to consume and incrementally process the changes.

CDF leverages Iceberg snapshots to efficiently generate the stream of changes. Critically, Data 360's optimized Iceberg writer computes and persists the changes as part of the write operation itself, making CDF generation highly efficient and minimizing additional overhead. This allows processing jobs (like streaming transforms or streaming calculated insights) to selectively process only the altered records avoiding the expensive snapshot diff computation.

This incremental strategy provides several benefits for large datasets, including cost savings, reduced latency, and improved efficiency. It enables features such as streaming transforms and incremental identity resolution, which in turn lead to faster insights, more predictable system loads, enhanced performance, and lower operational expenses.

Connect, Ingest, and Prepare

Data 360 offers robust ingestion capabilities with native support for Salesforce products, ensuring seamless data flow. For external sources, it provides extensive connectivity through over 270 connectors, APIs, SDKs, and MuleSoft. Additionally, Data 360 features zero-copy federation, allowing for BI and analytics without data duplication.

Data 360 Connector Framework
The Data 360 connector framework (DCF) is the foundation for most of Data 360 connectivity. It enables ingestion, federation, and egress through a unified architecture. DCF defines the standards for building and managing connectors, from the UI for setup and administration to metadata persistence, data extraction, and delivery into the Lakehouse or via live queries against external sources. It also supports private connectivity options (such as private links, VPNs, and secure tunnels) to ensure enterprise-grade data security and compliance when connecting to customer or partner environments. By providing a consistent approach across all connectors, DCF empowers Data 360 to connect seamlessly into the broader ecosystem by ensuring extensibility, reliability, and secure integration.

Data 360 Connectors
Data 360 provides robust connectivity to a vast ecosystem of data sources, supporting both native Salesforce products and numerous external systems. This extensive connectivity is crucial for unifying siloed enterprise data and enabling AI/ML and Agentic applications.

Data 360 offers over 270 connectors natively or through MuleSoft, APIs, and SDKs to support its end-to-end data pipeline capabilities with batch, streaming, or real-time ingestion. These connectors can be broadly categorized by the type of source system they integrate.

Salesforce Native Connectors
These connectors ensure seamless and native data flow from Salesforce products.

Examples include connectors for Salesforce CRM, Data Cloud One , Marketing Cloud Engagement, Marketing Cloud Account Engagement, and B2C Commerce.

External Applications and SaaS
Connectors for various business applications and cloud services allow data ingestion from external software platforms.

Examples include Adobe Marketo Engage, Microsoft Dynamics 365, Mailchimp, and Airtable.

Databases and Data Warehouses
Data 360 connects to a variety of relational and cloud-based data storage platforms.

Examples include Amazon Redshift, Amazon DynamoDB, Amazon RDS (MySQL, PostgreSQL, Oracle), Google BigQuery, and Microsoft SQL Server.

Cloud Object Storage and File Systems
These connectors integrate with hyperscaler storage solutions for both structured and unstructured data.

Examples include Amazon S3, Google Cloud Storage (GCS), and Azure Blob Storage.

Streaming and Messaging Services
Connectors that handle continuous, real-time data streams are critical for event-driven scenarios and real-time processing.

An example is the Amazon Kinesis Connector.

Integration Platforms
The MuleSoft Anypoint Connector extends Data 360's reach by integrating it with a wider range of applications and databases via the Anypoint Exchange.

Unstructured Data and Cloud Object Storage Connectors
These connectors are critical for ingesting and referencing unstructured data (data that lacks a pre-defined model) to power generative AI capabilities.

All of these connectors are built on the Data 360 connector framework providing consistent experience.

Data Transforms
Data transformation is a fundamental architectural component in Data 360, designed to clean, enrich, and shape raw ingested data into normalized, actionable data assets aligned with the Customer 360 data model. This process is essential for data harmonization, quality improvement, and ensuring data is ready for downstream use cases like profile unification, segmentation, and activation. Transformations leverage both source data lake objects (DLOs) and Data Model Objects (DMOs) as input, producing the results to new DLOs or DMOs respectively.

Data 360 provides two primary transformation paradigms to address different data velocity requirements: batch data transforms and streaming data transforms.

Batch Data Transforms
Batch Data Transforms are designed for the high-volume processing based on a defined schedule or on-demand trigger. This engine is optimized for handling complex, resource-intensive restructuring operations.

The Batch Transform process is configured using a visual, low-code pipeline canvas that enables users to define multi-stage transformation logic. This engine uniquely supports complex restructuring operations vital for canonical data model alignment: data structuring and normalization. This includes pivoting (decomposing denormalized records into multiple normalized records) and flattening (restructuring hierarchical data, like JSON, into structured tables). The system’s execution mode supports both full synchronization (processing all records) and a highly efficient incremental processing mode. The incremental mode significantly reduces processing time and resource consumption by only processing records that have changed since the last successful run. Batch transforms are ideal for tasks where real-time updates are not essential, such as periodic aggregations and complex data restructuring.

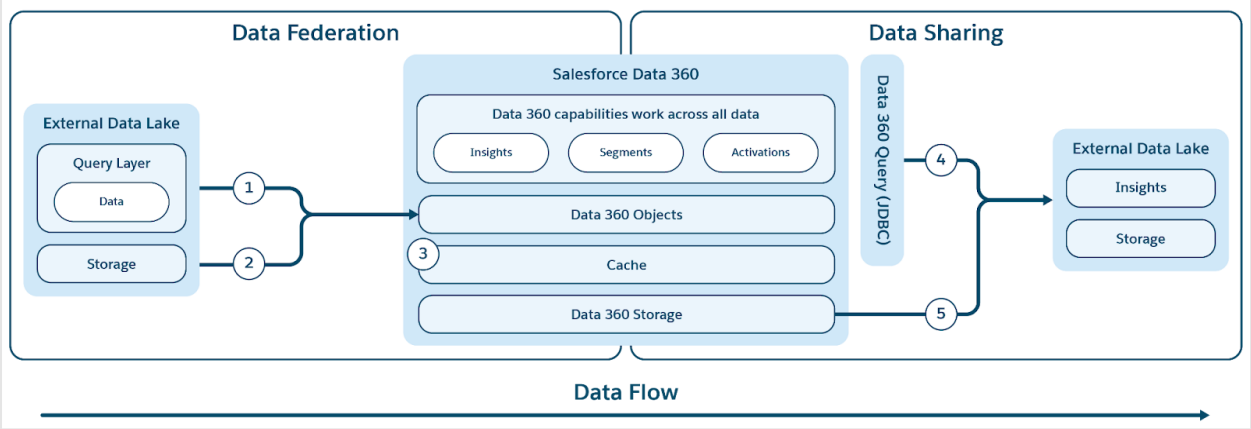
Streaming Data Transforms

Streaming data transforms process data continuously and incrementally in near real-time as it flows into the system, making them essential for low-latency use cases.

The primary interface is an SQL-first approach, where transformations are defined as a SQL SELECT query that continuously executes against the incoming stream of record changes. This engine supports core transformation functions, including data cleansing and Standardization (e.g., validating PII and standardizing data formats) and data enrichment and merging (using Joins and Unions). Critically, it supports Streaming Lookup JOINs to enable real-time data enrichment and lookups against static or slowly changing reference data, ensuring instant profile updates. To optimize cost-to-serve, the architecture employs a High-Density (HD) job design, which packs multiple streaming transformation definitions for a single tenant into a single underlying compute job, maximizing resource utilization. Streaming transforms are essential for use cases like event monitoring, immediate personalization, and real-time profile updates.

Zero Copy Data Federation and Data Sharing

Data 360 revolutionizes data management by supporting Zero Copy federation and data sharing, which eliminates the need to move or duplicate data. This capability allows users to seamlessly and directly access data from diverse external sources and share data with external environments, significantly reducing complexity, lowering storage costs, and ensuring all decisions are based on the freshest, most reliable information.



Data Federation

Data 360 supports zero-copy federation with external data warehouses (Snowflake, Redshift), lakehouses (Google BigQuery, Databricks, Azure Fabric), SQL databases and many other sources. Its abstraction layers enable direct querying of external data without duplication, reducing ingestion time, storage costs, and ensuring up-to-date information.

Data 360 simplifies access to external and federated data by providing a unified metadata layer that abstracts both Salesforce and external objects. This enables the entire Salesforce platform and its applications to seamlessly use this data.

Data 360 supports both file and query-based federation, with live query and access acceleration as shown in the figure.

Labels (1) and (2) illustrate Data 360's query (including live query pushdowns) and file-based federation for accessing data from external data lakes/warehouses/data sources; and label (3) highlights acceleration of federated access from external data lakes/data sources.

Query Federation
The core of Data 360's federation capability lies in its query federation layer, which manages the complex process of accessing external data and performing intelligent query pushdowns (illustrated by label 1). Data 360 connects to and retrieves data from sources using the JDBC protocol, enhanced with additional logic for improved efficiency. The Query Federation Layer is responsible for understanding and translating different SQL dialects, figuring out the most optimal part of the query to be pushed down to external systems for efficient processing, retrieving the results, and performing any necessary further processing to derive final insights.

Caching (Query Acceleration)
For enhanced utility, Data 360 provides an optional acceleration feature for its federated capabilities.

When Acceleration is activated, Data 360 caches the federated data to achieve faster access and lower costs—as it avoids repeated, direct access to external sources. This cache is treated as an acceleration layer and is incrementally updated to quickly reflect any changes in the external source data, ensuring the accelerated view remains near real-time.

File Federation
Data 360 supports file-based federation (illustrated by label 2) for accessing data from external data lakes and sources. The technical basis for this zero-copy capability relies on standardization: the underlying data must be in the Apache Parquet file format and use the Apache Iceberg tabular format. Data 360 can federate into any source that exposes an Iceberg REST Catalog (IRC) for metadata and storage access, ensuring seamless, governed access to files residing outside the platform.

With file-based federation, Data 360 compute handle all data processing because they directly access the underlying storage. This eliminates the need for query pushdown and managing various SQL dialects, which are often required with query-based federation.

In addition to this, Zero Copy capability also extends to unstructured data sources like hyperscaler storage solutions (S3/GCS/Azure storage), Slack and Google Drive, which can be accessed by Data 360's unstructured processing pipelines.

Data Sharing
Data 360 facilitates both query-based and file-based sharing of data it manages with external data lakes and warehouses (illustrated by labels 4 and 5 in the original figure context).

Query-Based Sharing
For query-based data sharing, Data 360 exposes a JDBC driver using which external engines

and applications can get secure access to the data. This mechanism allows external systems to connect, authenticate, and execute live queries directly against the data within Data 360.

File-Based Sharing (Data Share & DaaS)
The primary mechanism for file-based sharing involves two concepts: data share and data share target, which leverage DaaS (Data as a Service) API.

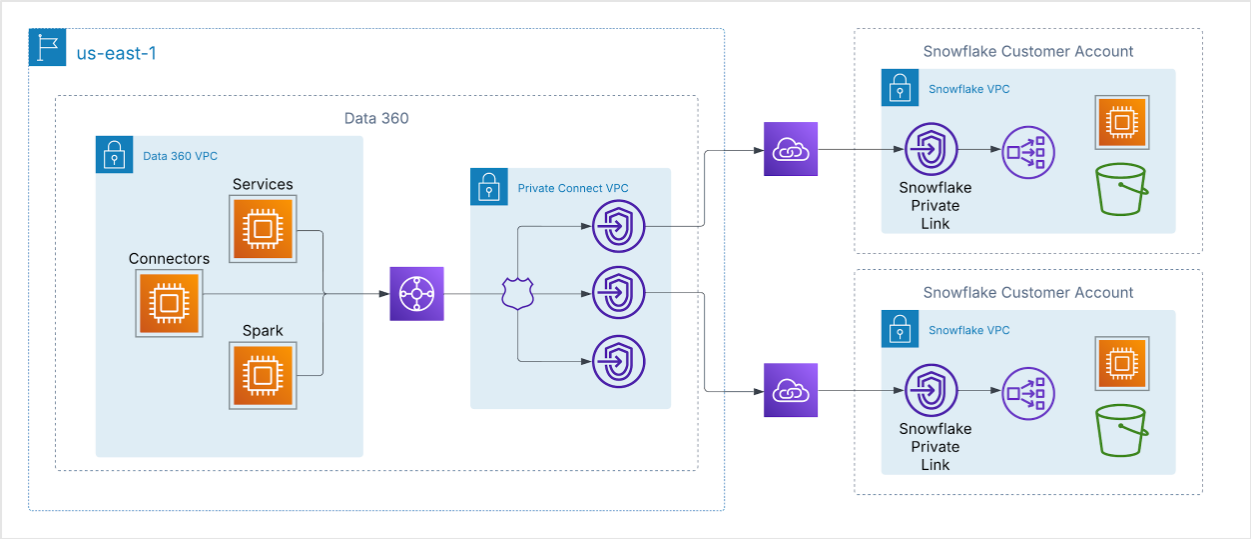
- Granular Control: The data share concept allows customers to precisely define what objects (DLOs, DMOs, CIOs etc.) are shared externally, preventing unintended data exposure.
- Secure Targeting: It also controls the data share target, ensuring data is only made available to explicitly authorized external environments, accounts, or partner organizations (e.g., sharing with a specific Redshift or Databricks instance).

DaaS API provides a secure and governed interface for external engines to consume data. It grants access to both the essential metadata and the underlying table storage while preserving all the Data 360 semantics. This ensures external engines access the data in a consistent and meaningful context in a secure way.

Private Connectivity
Many security-sensitive customers, especially large enterprises, regulated industries, and public sector organizations, restrict all internet access to their data lakes as part of their security posture. This policy, while essential for compliance and risk reduction, also prevents Salesforce Data 360 and Agentforce from connecting to those environments over the public internet.

Most of these data lakes are deployed in hyperscaler environments such as AWS, Azure, or Google Cloud. Because Data 360 itself runs on AWS, accessing customer data lakes hosted on a different cloud provider requires a cross-cloud network connection. Without a secure, private connectivity option that bypasses the public internet, customers are often unable or unwilling to adopt Data 360 or Agentforce for use cases that rely on those data lakes.

Private Links



To address this, Data 360 supports private, network-level connectivity to customer-managed data sources across clouds. On AWS, this is enabled through AWS PrivateLink, which allows Data 360 to connect directly to customer-provisioned endpoints, either within their own accounts or within third-party data lake environments (e.g., Snowflake), without traversing the public internet.

This architecture ensures that all traffic remains entirely on the AWS backbone, using private IP addressing and non-routable network paths, thereby satisfying strict security and compliance requirements while enabling seamless access to customer data.

Private Interconnect
For customers with multi-cloud architectures, Data 360 is extending private connectivity beyond AWS through cross-cloud interconnect support. This enables secure, backbone-only network paths from Data 360 to data lakes and services hosted in Azure or Google Cloud, preserving the same principles as AWS PrivateLink — private IP addressing, non-public routing, and zero internet exposure.

Customers can choose between two deployment models:

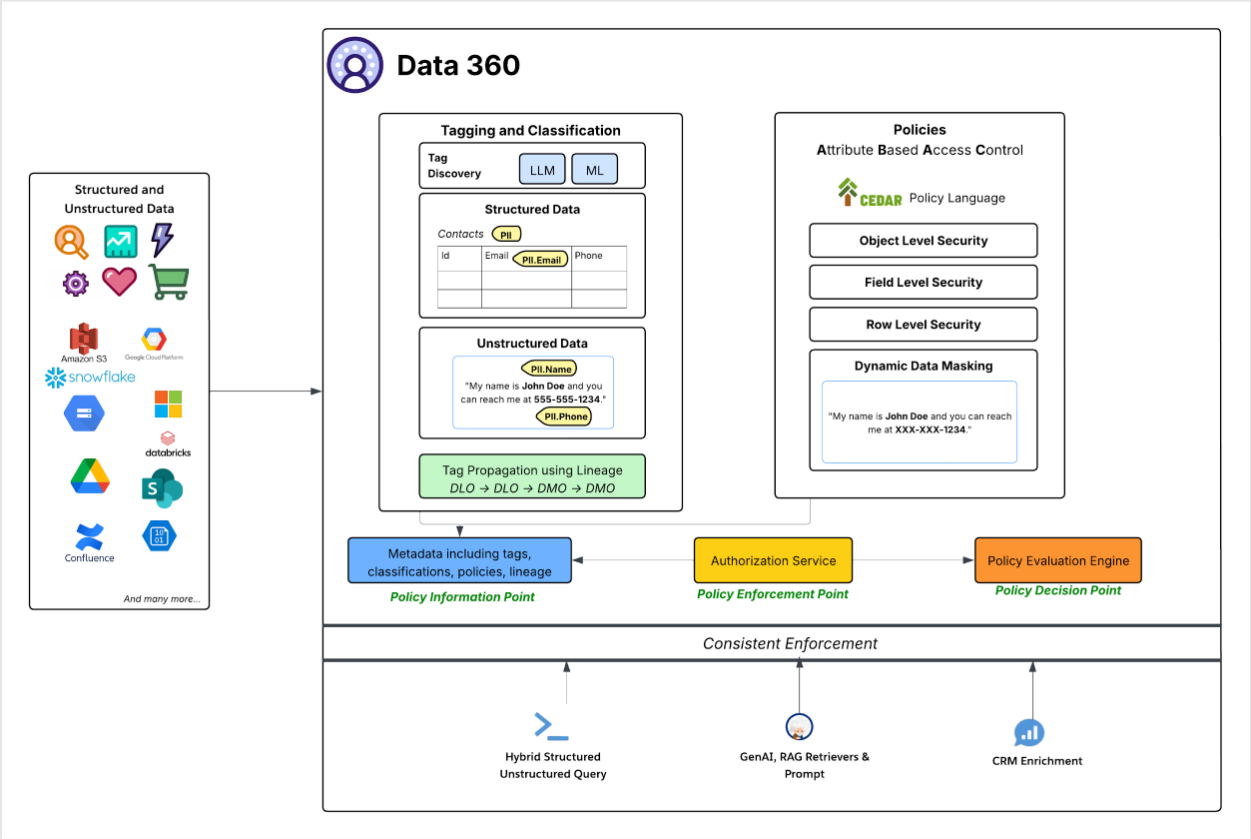
- **Customer-Managed Interconnect**: Integrate existing private circuits such as Azure ExpressRoute, Google Cloud Interconnect, or Equinix Fabric directly with Data 360's VPCs.
- **Salesforce-Managed Interconnect**: Use a fully managed, turnkey connection where Salesforce provisions and operates the cross-cloud link, exposing private endpoints in the target cloud.

In both models, the experience is consistent: Data 360 services connect to external data sources across hyperscalers as if they were local, enabling secure ingestion, activation, and querying without traversing the public internet.

Data Governance in Data 360: A Foundation of Trust and Control

For enterprise architects, robust data governance is not merely a compliance checkbox but a foundational pillar for building trusted, scalable, and actionable customer intelligence. Salesforce Data 360 is architected with a comprehensive governance framework that ensures data quality, security, and adherence to regulatory mandates across its entire data lifecycle.

Data 360 functions as a centralized governance hub, ensuring that all data—from raw ingestion to activated insights—is managed with integrity and control.



Advanced Access Control: Why Attribute-Based Access Control (ABAC)
While dataspace provides coarse grained access control to determine access to all objects within a dataspace, ABAC based policies provide fine grained access control to individual objects, fields and rows within a dataspace. Data 360 has adopted Attribute-Based Access Control (ABAC) as its core authorization model for fine grained access control. This strategic choice provides superior flexibility and scalability compared to traditional Role-Based Access Control (RBAC), particularly crucial for dynamic, complex enterprise environments with vast amounts of data and varied access needs. ABAC allows access decisions to be based on attributes of the user (e.g., department, role, location), the data (e.g., PII, sensitivity, data space), and the environment (e.g., time of day), rather than just pre-defined roles. This enables highly granular and contextual access policies that adapt as data and user attributes change.

- **CEDAR Policy Language**: At the heart of Data 360's ABAC implementation is the use of the CEDAR policy language. This purpose-built, formal policy language provides a precise and verifiable way to define complex authorization rules, ensuring that policies are unambiguous and can be evaluated consistently at scale.

Governance Architecture: Policy Information, Enforcement, and Decision Points
The governance system within Data 360 adheres to a standard, robust ABAC architecture:

- **Tagging, Classification, and Policy Authoring (Policy Information Point - PIP)**:
    - Data 360 provides automated Tagging and Classification mechanisms, leveraging LLM (Large Language Model), and ML (Machine Learning) to identify sensitive data categories (e.g., PII.Email, PII.Phone, PII.Name) and other purpose built taxonomies (PHI, FinancialData) in both Structured Data (e.g., Contacts table) and Unstructured Data (for ex from Google Drive).
    - Crucially, Tag Propagation occurs along Data Lineage (DLO -> DLO -> DMO), ensuring that classifications automatically follow data transformations and derivations, from raw ingested data to the harmonized DMO layer and through derived data created from process definitions.
    - Finally, the policy authoring pane provides a simple experience to leverage the data and user attributes to define dynamic access rules for an organization.
    - This enriched metadata (including tags, classifications, policies, and lineage) feeds into the Policy Information Point (PIP).
- **Authorisation Service (Policy Enforcement Point - PEP)**:
    - The Authorization Service acts as the Policy Enforcement Point (PEP). It intercepts all data access requests from various consumption layers (Hybrid Structured/Unstructured Query, GenAI RAG Retrievers & Prompt, CRM Enrichment) and consults the Policy Decision Point to determine if access is permitted.
- **Policy Evaluation Engine (Policy Decision Point - PDP)**:
    - This engine serves as the Policy Decision Point (PDP). It takes the access request context from the PEP, along with policy definitions (in CEDAR), and attributes from the PIP, to make an authoritative access decision.

Key Policy Enforcement and Integration

- **Granular Security Policies**: Policies defined in CEDAR enforce various security levels, including:

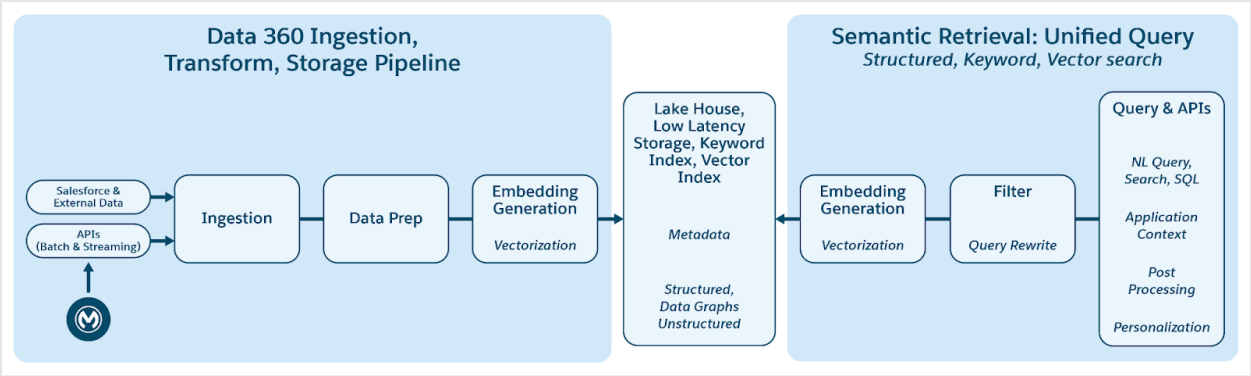
    - **Object Level Security**: Controlling access to entire DLOs or DMOs based on tags associated with these objects.
    - **Field Level Security**: Restricting access to specific sensitive fields within an object based on tags.
    - **Row Level Security**: Filtering data on specific objects to show only relevant rows based on user attributes.
    - **Dynamic Data Masking**: Dynamically mask certain data (based on tags) at the point of access, without altering the underlying data. This ensures sensitive information is protected while still allowing broad utility. This applies to masking fields in structured data as well as content in unstructured data.
- **Consistent Enforcement**: The entire ABAC framework ensures consistent enforcement of policies across all Data 360 consumption patterns, whether it's direct data querying, retrieval for Generative AI applications (RAG), or enriching Salesforce CRM experiences via Related Lists, for example.
- **Deep Integration with Salesforce Platform**: Data 360's governance capabilities are defined and administered directly within the Salesforce core platform. This integration allows administrators to manage access policies, user identities, and attribute management using familiar Salesforce tools, ensuring a unified and consistent governance layer across the entire Salesforce ecosystem.

By building this sophisticated ABAC framework with CEDAR policies, Data 360 provides architects with an unparalleled level of control and flexibility, ensuring that customer data is not only actionable but also secure, compliant, and trustworthy across the enterprise.

Data 360 Encryption: Bring Your Own Key
Across industries, organizations are placing heightened emphasis on end-to-end data security to ensure protection against data leakage, unauthorized access, tampering, or destruction. Most data platforms, including Data 360 today, provide encryption at rest using a vendor-managed encryption key. However, enterprises (particularly those in regulated sectors) are increasingly mandating customer-managed encryption capabilities for both data at rest and in transit.

This model allows companies to control their own encryption keys, ensuring that even in the highly unlikely event of a platform-level breach or unauthorized access, the data remains cryptographically protected. Without the customer’s proprietary key, no entity (including the platform provider) can decrypt or reconstruct the data, thereby maintaining full confidentiality and control.

Unstructured Data Storage and Processing
Data 360 supports storage and management of structured (tables), semi-structured (JSON), and unstructured data seamlessly across data ingestion, processing, indexing, and query mechanisms. Data 360 supports various unstructured data types beyond text, including audio, video, and images, broadening the scope of data handling and analysis. The figure below illustrates the two sides of grounding (ingestion and retrieval).



Data 360 manages unstructured data by storing it in columns as text or in files for larger datasets. It supports data federation for unstructured content, which allows for the integration and management of data from multiple sources.

The data is then prepared and chunked, embeddings are generated, and processed for keyword indexing and vector indexing. Data 360 hosts multiple out-of-the-box and pluggable models for chunking and embedding generation. Data 360 supports automated and configurable transcription of audio and video content for subsequent processing and indexing. Search service is used for keyword indexing. For vector indexing, Data 360 supports both native indexing (with Hyper) and also leverages vector databases like open-source Milvus. Data 360 also integrates with Salesforce Search platform to support keyword indexing on unstructured data. Such integrated multi-modal indexing in Data 360 powers search on any unstructured data, as discussed in the Agentic Enterprise Search section later in the document.

For retrieval, Data 360 provides APIs for search. Our Hyper-based unified query facilitates ensemble queries across structured, keyword index, and vector indexes, maintaining strict visibility and permissions, thus enhancing RAG and Search.

Extensible Pipelines
Data 360's unstructured data indexing pipeline is designed as a modular, extensible architecture comprising five core stages:

1. Parsing
2. Pre-Processing
3. Chunking
4. Post-Processing
5. Embedding

All the stages also support LLM based processing which allows customers to come up with the custom prompts. Both the pre-processing and post-processing phases can include multiple sequential steps, allowing complex transformations to be composed flexibly. Each stage is entirely metadata-driven, enabling seamless configuration and extension without code changes.

Examples of pre-processing include operations such as noise elimination, language normalization, and image understanding (optical character recognition and captioning), while post-processing stages may include metadata enrichment, semantic grouping, or advanced techniques such as Raptor chunking.

The pipeline fully supports Data 360 Code Extension, allowing customers and internal teams to plug in custom logic at any stage. The code extension components are lightweight Python functions whose lifecycle—execution, scaling, and failure handling—is fully managed by Data 360. This approach ensures that innovation and domain-specific processing can be introduced rapidly while maintaining operational consistency and governance across the platform.

Context Indexing
For setting up RAG with unstructured processing, two critical factors are key:

1. **Rapid Iteration**: The ability to quickly validate with sample test queries.
2. **Persona-Specific Content**: The capacity to configure content tailored to the consuming persona.

Context Indexing is a user-friendly tool designed to address both of these aspects. This interactive UI is powered by a Real-Time (RT) pipeline that executes all five previously outlined stages. The pipeline utilizes GPUs when necessary for tasks such as embedding generation and Optical Character Recognition (OCR). Furthermore, it allows customers to quickly test the RAG pipeline with an Agent before deploying the configuration for comprehensive content processing.

Document AI
Data 360 Document AI enables reading and importing unstructured or semi-structured data from documents like invoices, resumes, lab reports, and purchase orders. This feature supports ad hoc interactive processing as well as bulk batch processing. This is a key capability that enables business process automation for our customers. This is powered by artificial intelligence including LLMs and ML models

Enterprise Knowledge
Enterprises possess vast amounts of knowledge spread across disparate systems like wikis, file shares, content management systems, internal databases, and more. This fragmentation makes it difficult for employees (especially service agents and sales reps) and customers to find relevant information quickly and efficiently. Key problems include: Lack of a single, unified search experience across all knowledge sources; Inconsistent presentation and rendering of content from different sources; Lack of access governance to sensitive information scattered across systems; and difficulty in leveraging authoritative knowledge source within core business workflows (e.g., attaching relevant articles to a Case).

Enterprise Knowledge represents content that has been curated, either manually or automatically, from the broader pool of enterprise data. Manual curation involves deliberate actions, such as creating Salesforce Knowledge Articles or developing knowledge within external systems which is then ingested. We envision automated curation that utilizes processes, like Salesforce agents and transforms, that run over ingested data to generate refined, curated layers, potentially mixing

structured and unstructured content. Whether curated manually or automatically, internally within Salesforce or externally before ingestion, the outcome is value-added content distinct from raw data.

The Enterprise Knowledge Hub solution leverages Data 360 capabilities for :

- **Ingestion and Storage**: CRM Connector is ingesting Salesforce knowledge articles, and data connector framework (DCF) unstructured connectors ingest raw content and metadata from external sources. The content is ingested into source-specific unstructured data lake objects (UDLOs) mapping to the content on SFDrive (or source in case of zero copy).
- **Harmonization & Structuring**: The Harmonization Pipeline processes UDLO and file data, performing cleaning, normalization, enrichment (NLP, etc.), PII masking, and transformation into the harmonized intermediate format, stored in SF Drive and a harmonized UDLO (HUDLO) that maps to it.
- **Indexing**: Unstructured Pipeline (UDS) is triggered over the harmonized content and Search Indexes are configured for each HUDMO.
- **Consumption**: Consuming applications include search, retrieval, rendering, and linking to business objects like Case. Engagement by consuming applications is collected to provide usage analytics (like clicks, reviews, etc.)

Calculated Insights
Calculated insights (CIs) in Data 360 allow customers to define and generate aggregated metrics from their data. These metrics are then used for timely customer engagement, analysis, segmentation, and activation. The aggregated data computed by CIs is written to Lakehouse and represented as a calculated insight object (CIO).

There are two main types of calculated insights:

- **Batch Calculated Insights**: Designed for complex, high-volume data aggregation, where metrics can be computed periodically (e.g., daily or weekly).
- **Streaming Insights**: Provide the ability to generate metrics and actions from real-time event data, enabling immediate, low-latency customer engagement.

Calculated insights are defined on data model objects (DMOs) and can also be defined on other calculated insight objects. The calculated insights service manages the orchestration of both batch and streaming jobs.

Both batch and streaming insights computations use Spark. The key difference is that streaming insights utilise Spark Structured Streaming, while batch CIs are executed using periodic, scheduled batch Spark jobs. For cost efficiency, the calculated insights service groups CIs to be computed together in the same batch CI job or streaming CI job, based on factors like dependencies and overlap of source data objects.

SNCE and CDF play a significant role in Streaming Insights computation

Identity Resolution
Identity resolution is responsible for transforming disparate data from multiple sources into a single, comprehensive Unified Profile.

It is important to understand that a unified profile isn't a "golden record", and that identity resolution doesn't pick winning values or override any existing data while unifying profiles. Unified profiles serve as a set of keys that unlock your source data by identifying all matching records that relate to the same entity, within one data source or across many sources. With this information, you can select the right source system data to use for a given business use case.

Identity resolution can consolidate a variety of record types, including Individuals, Accounts, and Households. It can also be used to match leads to existing accounts. The unification process is essential for achieving a complete Customer 360 view and driving personalized, real-time engagement across both B2C and B2B scenarios.

The identity resolution pipeline is built on a highly scalable, cloud-native framework designed to handle massive volumes of data continuously. The process involves three key stages, relying on a powerful search index to manage the matching process:

- **Matching (Candidate Selection)**: The goal of the match process is to search for records that may belong to the same entity. Records are analyzed against a customizable set of rules, each of which contain a set of criteria that define what data to match at which level of strictness. To efficiently retrieve potential matches from the data store, the system generates indexes to find likely matching records using two techniques:
    - **Blocking Keys**: A blocking key is a value generated from a record's data and match rules (like the first few letters of a name, normalized phone number, etc) to group potentially similar records together. Each record has multiple blocking keys that are indexed and stored as an inverted index, ensuring the system only performs detailed comparisons on small groups of records, rather than across the entire dataset.
    - **Locality Sensitive Hashing (LSH)**: For match rules with fuzzy matching, hashes are generated based on embeddings from trained models.
- **Deep Matching**: After the candidate selection step creates smaller groups of potential matches, the system begins a more detailed comparison. In this stage, AI models and advanced algorithms analyze each pair of records to calculate a probabilistic match score. This score quantifies the likelihood that two records refer to the same entity by intelligently comparing fields that often contain misspellings, variations, or formatting differences.

- **Clustering and Unification**: Once matching records are identified from the candidates, they are grouped into a cluster. This process critically includes resolving transitive matches. For example, if Record A matches Record B, and Record B matches Record C, all three are linked into the same cluster even if A and C were never directly compared. These complete clusters form the foundational structure of the Unified Profile. This clustering process ensures that all related source records are correctly linked under a single, persistent identifier.
- **Reconciliation**: Data values from all clustered source records are evaluated using defined Reconciliation Rules (e.g., *Most Frequent*, *Most Recent*, or *Source Priority*) to that populate the resulting Unified Profile with an excerpt of profile data. Reconciliation doesn't overwrite any existing data, as all source data is available using the keys linked to the unified profile.

Identity Resolution Types
The architecture supports the resolution of multiple entity types to fulfill a variety of use cases.

- **Individual Matching**: Focuses on creating the Unified Individual profiles, which link all known personal identifiers (emails, phone numbers, loyalty IDs, cookies) to a single person.
- **Account Matching**: Focuses on creating the Unified Account profiles, which link data about accounts. When matching on company names, the engine uses a finely tuned model when fuzzy matching.
- **Household Matching**: Extends the matching logic to aggregate Unified Individual records into groups of related individuals.
- **Cross Entity Matching**: Apart from unification, identity resolution also creates links between profile objects by using the same match rules. For instance, a Lead can be linked to an Account using fuzzy matching on Account Name.

Near Real-Time Processing
To ensure that the Unified Profile is always current, the identity resolution engine operates with a near-real time architecture. This cloud-optimized architecture is designed for continuous processing, achieving fast processing times. While processing speed varies depending on how source data is received, small batches of changes can be processed by identity resolution as often as every 15 minutes.

Source Profile Linking
The system maintains identity link objects that map every source record ID to its corresponding Unified Profile ID. This foundational data structure allows the engine to efficiently track relationships and quickly propagate changes and updates to the Unified Profile, ensuring that customer experiences—such as website personalization, next-best-action recommendations, and segmentation—always leverage the freshest available customer data.

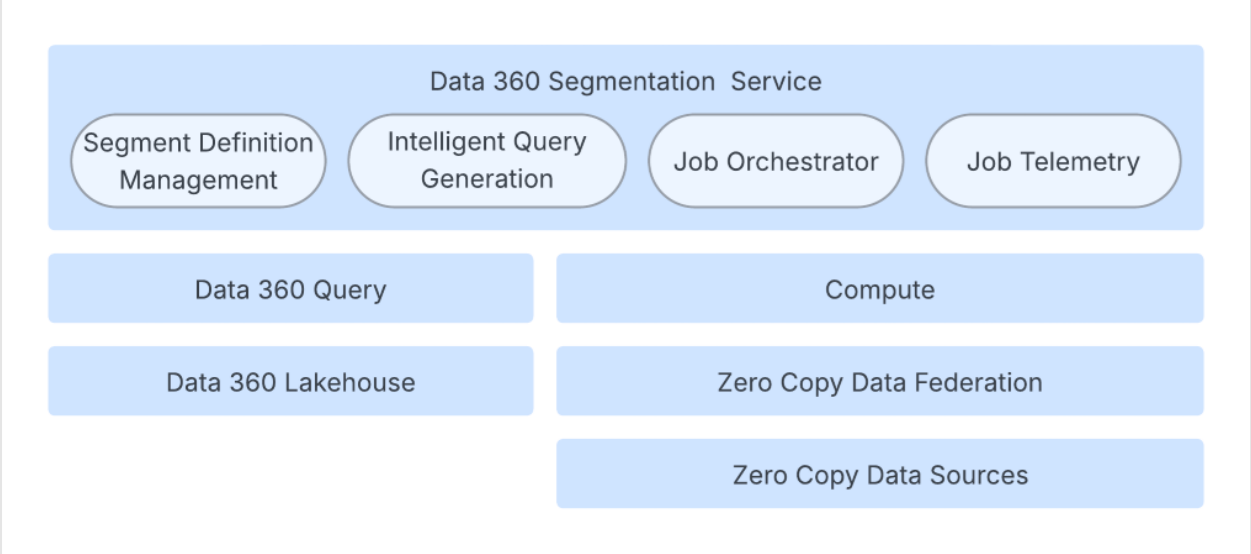
Segmentation
Segmentation is the core process of transforming unified customer profiles into actionable audiences. This capability is critical for powering personalized experiences across marketing, commerce, and service channels. The Salesforce Data 360 Segmentation platform is designed

for large-scale operations. It manages intricate metadata, working with a data model comprising thousands of objects and relationships. The platform supports complex rules, aggregation-based filters, and window-based ranking, all while ensuring fast and reliable computation at petabyte scale.

Segmentation Types and Cadences
Data 360 supports various segment types to meet distinct business requirements for speed, complexity, and hierarchy:

1. **Standard Segment**: The primary, batch-processed segment type. It publishes on a customizable schedule, with a Standard Publish cadence of a minimum of 12 hours up to 24 hours, or a faster Rapid Publish cadence of 1 to 4 hours, which is optimized for recent engagement data.
2. **Real-Time Segment**: This segment completes on-demand in milliseconds for immediate action based on recent events and profile data. It is highly optimized for instant personalization but cannot utilize exclusion criteria or nested segments.
3. **Waterfall Segment**: A hierarchical structure of subsegments used to prioritize a customer into a single, most valuable segment if they qualify for multiple criteria.
4. **Nested Segment**: This allows for reusing an existing segment as a filter for a new, more specific segment (a refinement of a base segment), inheriting the schedule of the parent segment.

Segmentation Engine: Architecture Fundamentals
The segmentation engine operates on a robust, cloud-native architecture that ensures speed, scale, and resilience.



The core process is managed by a job orchestration service that controls the segment's lifecycle, generating the necessary job config and triggering the execution. This orchestration layer maintains state and metadata in a sharded database for scalability.

Although Data 360 Query handles segmentation count computations, the Spark compute layer is responsible for calculating actual Segment membership. The Spark application executes Spark SQL queries on extensive customer data. This data can reside in the Data 360 Lakehouse, external systems via Zero Copy data federation, or a combination of both.

The system is highly optimized through intelligent query generation, which refines the underlying Spark SQL query. This includes techniques like intelligent partition pruning to minimize data scanning and the elimination of redundant sub-expressions. To ensure service reliability, the architecture features adaptive resource management that dynamically adjusts compute resources based on workload size and complexity. Furthermore, SLO Adherence is proactively managed with adaptive durations and retry logic. For a fast user experience, accelerated segment counts use a sampling-based approach to provide rapid size estimates during segment creation, avoiding a full query run.

Finally, a deep focus on observability and root cause attribution is maintained through comprehensive Spark execution metrics and automated classification of errors (e.g., customer-side vs. system issues), significantly reducing diagnosis time and ensuring a highly resilient data platform.

Activation
Activation is the critical final step in the Data 360 lifecycle. Its core function is to transform static, segmented, unified customer profiles into actionable and enriched data and deliver this data to internal and external endpoints (such as Marketing Cloud, Commerce Cloud, and Adtech platforms). This process is designed to trigger personalized customer journeys and near real-time interactions. It supports advanced capabilities like related attributes, activation membership filtering, consent filtering, limiting, and ranking.

Activation offers three distinct methods for external delivery and channel compliance:

- **Batch Activation**: Designed for high-volume, scheduled operations such as large-scale email campaigns and advertising audience refreshes. Data is delivered by staging it to Secure Internal Buckets (Cloud Object Storage) or via Secure File Transfer, followed by an API ingest process initiated by the target system. Batch activations can utilize special refresh mode - incremental - to reduce volumes sent to and processed by Salesforce partners.
- **Streaming Activation**: Optimized for near real-time use cases requiring event-driven automation. Delivery is achieved through direct API calls sent to the destination endpoint.
- **Activation-Triggered Flows**: This highly platformized channel provides a no-code/low-code approach for integrating Audience Data with hundreds of customer API-enabled engagement platforms. Upon activation completion, Data 360 populates an Audience DMO, which then triggers a High Scale Flow. The Flow

engine subsequently consumes the audience data and utilizes platform capabilities like External Services and Mule Outbound destinations to make calls to the final API-based destination. This method significantly reduces the time required to onboard new activation targets.

Activation utilizes the same patterns as segmentation for job management, distributed execution, and monitoring. This includes the principles of the job orchestration service for lifecycle management and the compute layer (Spark) for processing, and relies on job telemetry for performance observability and service level objective (SLO) adherence.

In addition to those, activation has -

Activation Target Management oversees the secure connections, credentials, and configurations for all destination endpoints. It guarantees that data formats and security protocols are standardized, ensuring dependable outbound delivery to various platforms, including Marketing Cloud, Adtech partners, and other external applications.

Activation tailors the payload for specific targets. For Salesforce Marketing Cloud, this includes Business Unit (BU) Aware Filtering, Multi-EID support, and cross-pollination controls.

Communication Governance acts as a gatekeeper, ensuring data usage and communication are compliant with customer preferences and legal requirements. Centralised consent model unifies all customer preferences, from global opt-outs to channel- and purpose-specific consent, and are stored on the Unified Individual Profile. During execution, the platform strictly enforces these policies by using exclusion filtering to automatically remove non-consenting individuals from the final payload. Furthermore, the system applies contact point selection rules to ensure it uses the single, most compliant, and preferred contact point for the intended channel before any data is transmitted. This enforcement mechanism is secured by the underlying governance framework, which employs protective measures like dynamic data masking and access controls to secure sensitive data fields throughout the activation process.

Unified Query: Seamless Access to All Customer Data
The true value of a unified data platform lies in its ability to provide effortless, consistent access to all its data assets, irrespective of their origin or structure. Salesforce Data 360's Unified Query capability is precisely engineered to deliver this, abstracting the underlying complexities of diverse data stores to provide a single, powerful query interface.

The Unified Query layer offers sophisticated access for diverse consumption patterns:

- **Hybrid Structured & Unstructured Querying**: It provides extensive SQL support to seamlessly query both structured data and the structured metadata of unstructured data. This is enhanced by operator extensibility via table functions, enabling specialized search across text, image, and spatial types.
- **Accelerated Performance with Hyper**: Leveraging Hyper, a high-performance, in-memory engine, Data 360 accelerates complex analytical queries and interactive dashboards, providing near-instantaneous responses over massive datasets.

- **Unified Approach for AI & Personalization**: This unified access is crucial for generating targeted and personalized results, directly facilitating more precise LLM responses using Retrieval Augmented Generation (RAG) by grounding AI models in rich, enterprise data.
- **Integration with Downstream Consumption**: It serves as the foundational data access layer for UI-driven experiences, robust APIs, Generative AI workflows, and CRM enrichment, seamlessly connecting data to activation.

By providing a single, intelligent, and highly performant query interface, Data 360's unified query empowers architects to build agile, data-driven applications that fully leverage their complete spectrum of customer information.

Data Actions
Data 360 is an active platform that supports the activation of pipelines in response to data events. For instance, a significant event, such as a drop in a customer's account balance can trigger a Salesforce Flow to orchestrate a corresponding action. Similarly, updates to key metrics, such as lifetime spend, can be automatically propagated to relevant applications.

Data Actions continuously monitor incremental data for changes using storage native change events (SNCE) and change data feed (CDF). This data is evaluated against customer-configured action rules, such as threshold monitoring or state changes. When these rules are met, a data action event is generated. This event is enriched with additional information (e.g., customer loyalty status) and immediately sent to its configured destination, such as Salesforce Flow or an external application, to trigger business orchestrations.

Customer Data Platform
Data 360 supports native CDP (Customer Data Platform) features including advanced identity resolution capabilities and creating unified individual identifiers and profiles along with comprehensive engagement histories. This platform is adept at handling both business-to-business (B2B) and business-to-consumer (B2C) frameworks by supporting identity resolution and identity graphs that utilize both exact and fuzzy matching rules, as described above. These identity graphs are enriched with engagement data from various channels, which helps in building detailed profile graphs with valuable analytical insights and segments.

A key concept that supports the customer profile is the data graph. Data 360 offers an enterprise data graph in JSON format, which is a denormalized object derived from various Lakehouse tables and their interrelationships. This includes a "Profile" data graph created by CDP that encompasses a person's purchase and browsing history, case history, product usage, and other calculated insights, and is extensible by customers and partners. These data graphs are tailored to specific applications and enhance generative AI prompt accuracy by providing relevant customer or user context. The real-time layer of Data 360 utilizes the Profile graph for real-time personalization and segmentation. We envision modeling Agentforce Context that encompasses Conversations, Sessions, and Agentic memory as data graphs.

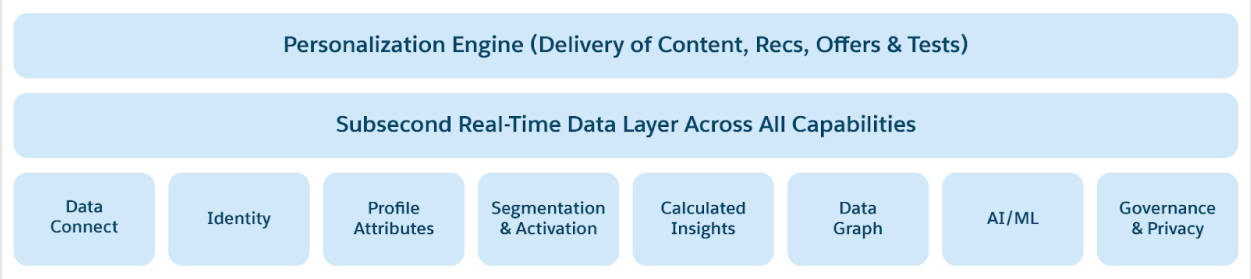
Additionally, the CDP enables effective segmentation and activation across different platforms such as Salesforce's Marketing Cloud, Facebook, and Google. It processes customer profiles in

batch, near real-time, and real-time, which allows for immediate decision-making and personalization. This functionality enhances interactions in both B2C and B2B scenarios, ensuring that businesses can respond swiftly and accurately to customer needs and behaviors.

Real Time Layer

Data 360's real time layer is backed by the Data 360 CDP and extends its concepts for real-time use cases. Data 360's real-time layer is engineered to process events such as web and mobile clickstreams, visits, cart data, and checkouts at millisecond latencies, enhancing customer experience personalization. It continuously monitors customer engagement and updates the customer profile from Customer 360 with real-time engagement data, segments, and calculations for immediate personalization.

For example, when a consumer purchases an item on a shopping website, the real-time layer quickly detects and ingests this event, identifies the consumer, and enriches their profile with updated lifetime spend information. This allows for the personalization of their experience on the site in sub-seconds. Additionally, this layer includes capabilities for real-time triggering and responses, enabling immediate actions based on customer interactions.



The Sub-Second Real-Time platform powers this transformation through several key features:

1. **Real-Time Data Graphs**: A Customer 360 profile is built using a denormalized graph that includes key objects and fields most relevant to brands. These data graphs enable real-time data processing and deliver actionable content and insights within milliseconds
2. **Real-Time Ingest and Transform**: Ingest user events and profile in milliseconds from web and mobile sources.
3. **Real-Time Identity Resolution**: Merge customer profiles across devices, unifying unknown and known users instantly.
4. **Real-Time Calculated Insights**: Calculate metrics like lifetime value (LTV) or user visit history in milliseconds to enable Personalization or Offers for Web, ChatBot or Service Agent.
5. **Real-Time Segmentation**: Segment audiences on the fly, personalizing messages and interactions in real-time.
6. **Real-Time Actions**: Empower brands to evaluate every user engagement and take action via Salesforce Flow or other relevant communication channels.

In Data 360, we built a new real-time platform with real-time pipeline, low latency storage, and a sub-second data processing layer. As fast interactive data is ingested from Web and Mobile channels, it goes through a series of rapid processes.

Our Web & Mobile SDKs and real-time APIs collect data from web/mobile applications (in future Agentic interactions) and send it to our beacon server. This data is then routed to the Real-Time Layer for millisecond processing and the Lakehouse Layer for integration with batch/streaming data. The Real-Time layer processes the incoming real-time data in the context of a user profile (anonymous or logged in) like updating user total spend or lifetime value, etc. for in-session, real-time personalization. The Real-Time Layer is backed by main memory and NVme (SSD) storage storage for storing real-time data and customer profiles. Once the data is in the Real-Time Layer, it goes through the following processes before getting updated into Real-Time Data Graph:

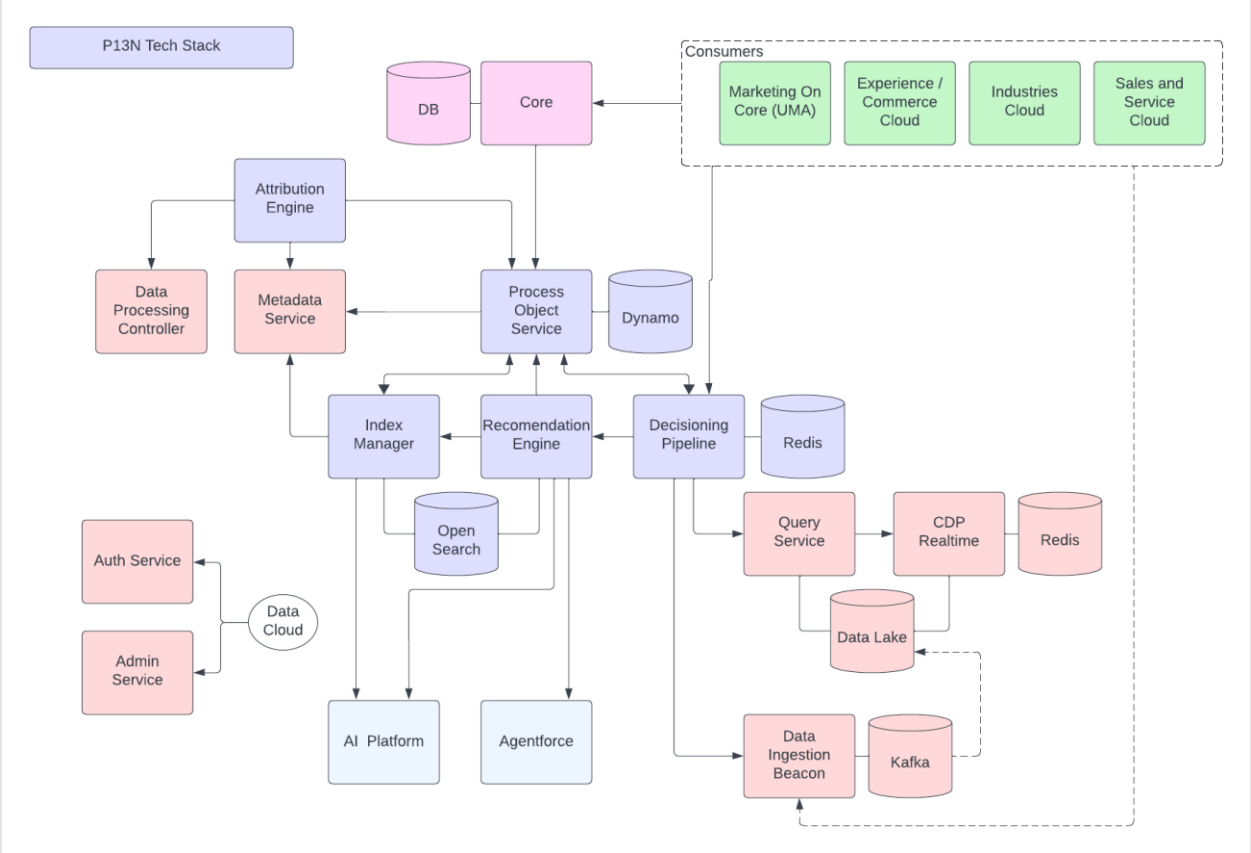
- **Simple Ingest and Transformations**: The data is ingested and transformed for further processing.
- **Identity Resolution**: Exact match rules are applied to unify profiles with all existing match rules sets, so data-aware specialists don't have to create new identity resolution rules sets specifically for real-time.
- **Computed Insights**: Every engagement is evaluated, simple computations such as sum and count in milliseconds are run, and data is updated in the Real-Time Data Graph.
- **Real-Time Segments**: Every engagement data is evaluated to determine if it meets the criteria for defined real-time segments, and users are added to qualifying segments in milliseconds.
- **Real-Time Actions and Triggers**: Every engagement is evaluated against defined rules to trigger actions to a range of targets in real-time when rules are met in milliseconds.
- **Real-Time Data Graph and API**: Real-time data graph, which also includes a real-time API, allows brands to retrieve updated JSON-format data for each user, ensuring that all customer interactions are informed by the freshest data.

Personalisation
Personalisation is all about knowing who to target, when and where to deliver relevant content and recommendations, what to say and what frequency. The Personalization Services Platform is the orchestrator of what decisions are made to optimize goal achievement through personalized experiences.

The **Personalisation Services** delivers the following capabilities:

- Consistent set of models and ways to interpret profile, activity, and asset data in Data 360
- Platform integrated Experimentation (A/B/n, Multi-Arm Bandit)
- Integration of Goals at design time (configuration), ML training time and runtime (ML inference)

- B2C scale real-time and batch interaction support (anonymous users, high volume real-time/interactive external, high volume internal batch)
- Analytics driven through Data 360
- Patterns to integrate AI model and service from other parties (internal and external)
- Integration into the Core metadata ecosystem (PLATE characteristics)
- OOTB implementations of high-value AI-driven use cases (Recommendations/decisions with various ML algorithms including contextual bandits for promotion/content selection, product recommendations, pricing decisions, etc.)

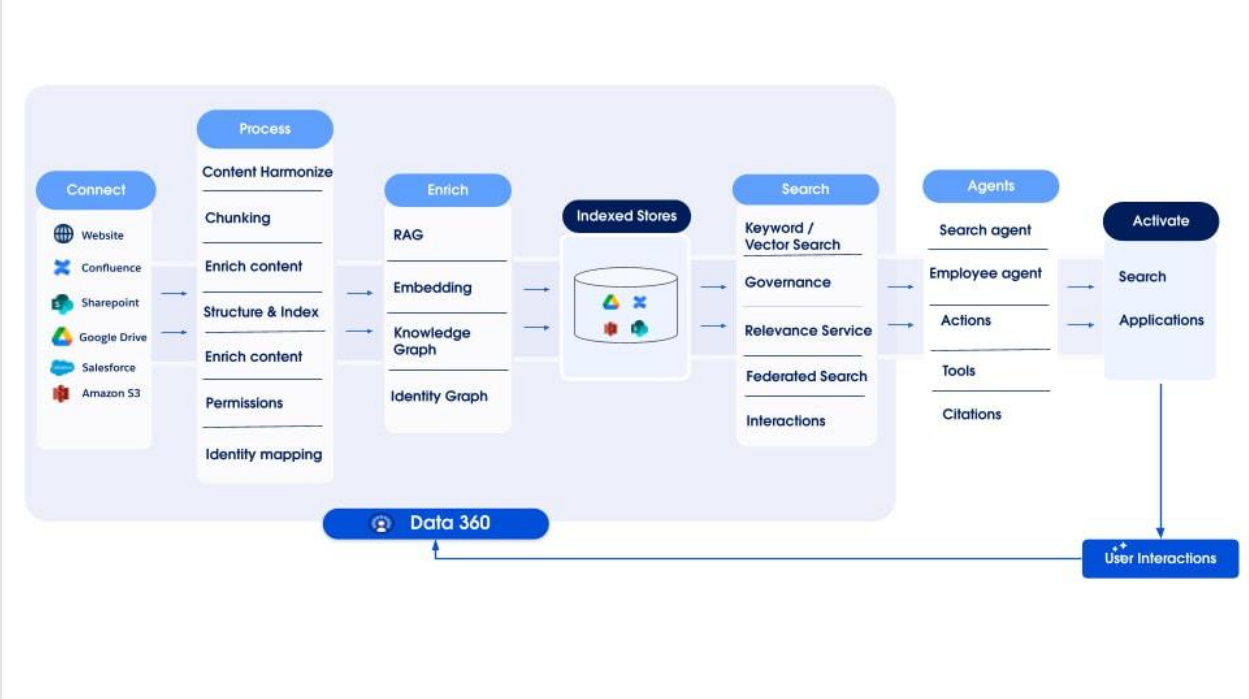


- **Decisioning Pipeline**
  - Servicing external requests for personalization decisions including profile augmentation, experimentation, and recommendations.
- **Recommendation Engine**
  - Runtime servicing of rule or ML based recommendations.
- **Index Manager**
  - Manages/Orchestrates workflow for asynchronous processes including ML trainings for recommendation models

- **Process Object Service**
    - Responsible for syncing personalization metadata between core and off-core
- **Attribution Engine and Experimentation**
    - Analytics attribution and experimentation of personalization recommendations

Agentic Enterprise Search

Data 360 is designed as a robust and rich platform specifically for supporting the emerging agentic experiences. We achieve these capabilities by building upon various existing Data 360 services and through deep integration with Agentforce.



Leveraging Data 360's Core Strengths

Our approach to Agentic Enterprise Search is founded on the following principles:

- Enterprise data is kept in siloed services or stores, with secure permissions needed for access. The ability to access this data and process it while maintaining the source permissions is vital to ensure trust.
- Cross-ranking and relevance across the complete corpus of data enable better results, which in turn can provide better context for agentic experiences.

To deliver these experiences, Agentic Enterprise Search is built on top of these key architectural components:

- **Connectors**: The broad set of connectors available in Data 360 enables one to access and ingest data from a wide variety of sources.
- **Unstructured Data Processing**: This is foundational for handling non-tabular content, enabling the system to derive meaning and context from diverse data.
- **Governance**: Ensuring enterprise-grade security, compliance, and access controls for all data consumed by search. Support for source visibility permissions ensures that the data is accessible only to authorized users, for both simple search and agentic experiences. To ensure fast retrieval, security permissions are evaluated and enforced natively by the search backends at the earliest stage of data access.
- **Unified Retrieval Layer**: To address the challenge of siloed data, the connectors feed into a comprehensive Unified Retrieval layer. This layer provides a single access point to all data, whether it remains in external systems accessed via Federated Search or is managed natively through advanced indexes for zero-copy and ingested data.
- **Intelligent Query Understanding**: Before retrieval, the system uses AI-powered mechanisms to interpret user intent. In addition to embedding representations of the query for semantic vector matching, it can rewrite and expand keyword-based searches to improve accuracy.
- **Hybrid Search and Advanced Querying**: To find the most relevant information, the platform uses multiple strategies in parallel. Hybrid search provides precise keyword matching with semantic vector search on optimized chunks of data, while Full-record search simultaneously retrieves entire documents. Both are combined to ensure both semantic relevance and complete content coverage.
- **Hierarchical Ranking**: After data is retrieved, a multi-stage hierarchical ranking architecture scores, merges, and re-ranks the results from every source and method. This process creates a single, unified list that surfaces the most pertinent information for the user or agent.

Agentic Search Applications

Generative AI is shifting the primary consumer of enterprise search from human users to Large Language Models (LLMs). Data 360 Search is engineered from the ground up to serve both. It is optimized to handle the longer, more complex queries from agents and return the rich, contextual results necessary for programmatic consumption and reasoning loops. At the same time, the system can handle the shorter, often ambiguous queries typical of human users, providing features like snippets and highlighting for rapid assessment in a UI.

The ultimate delivery of agentic search experiences combines both approaches:

1. **Direct Search Results**: application can display a traditional list of ranked results by using a metadata-driven API built on the Data 360 unified search foundation.
2. **Agentic, multi-turn conversational answers**: agentic answers are implemented through native integration with Agentforce. This conversational experience is driven by a primary agent that orchestrates actions and queries, delegating all information retrieval tasks to a specialized, internal Search Agent.

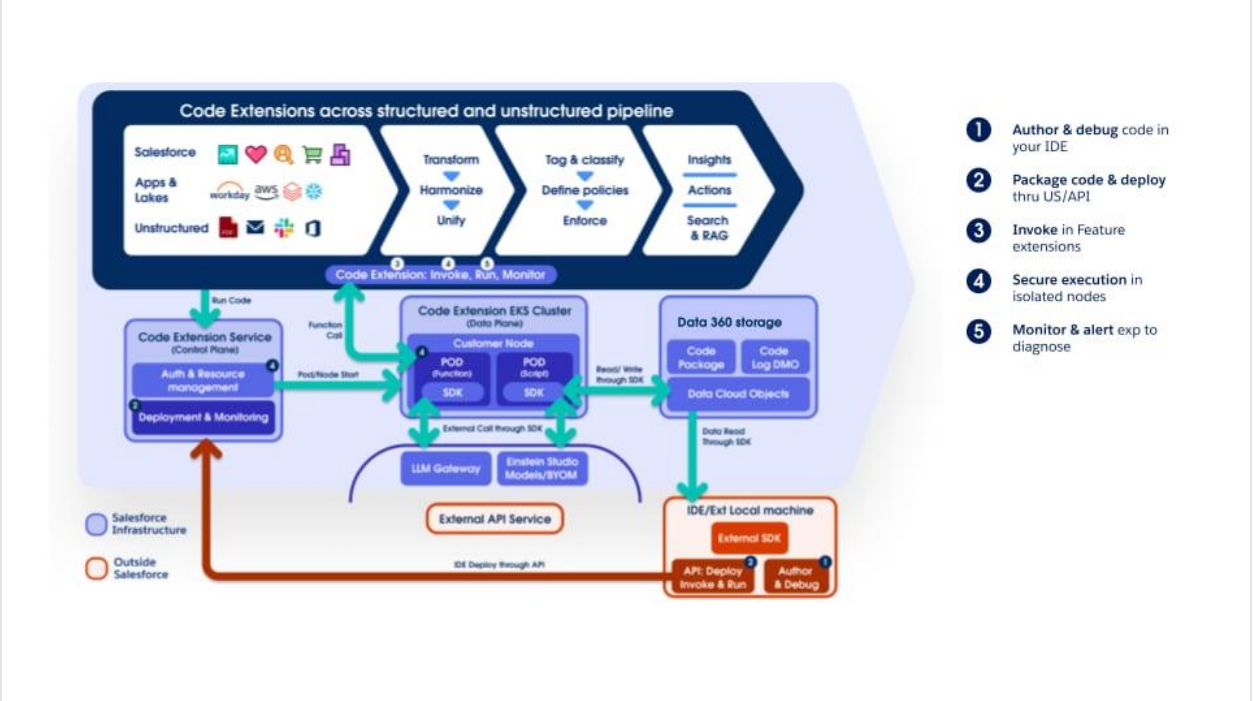
This specialized Search Agent is optimized for enterprise information retrieval. It uses a reasoning loop to formulate and execute parallel searches to explore different facets of a user's request. It utilizes a powerful set of tools, including the Data 360 unified search for all data types and structured query languages to retrieve precise data from tables and entities.

Through this architectural synthesis, Data 360 empowers the creation of highly intelligent, context-aware, and actionable agentic enterprise search experiences.

Extending Data 360 with Custom Code
Extensibility is a key feature in the Salesforce platform. Code Extension provides extensibility in Data 360, allowing pro-code users to run custom Python logic directly within the Data 360 environment, supplementing its rich declarative and low-code functionalities. Using Code Extension, users can securely extend Data 360 core capabilities like Transforms and Unstructured Data Pipelines (custom chunking).

Our design for Code Extension prioritizes flexibility, security, efficiency, and a streamlined developer experience. It supports two primary execution models, each tailored for specific architectural needs:

- **Script Model**:
    - **Purpose**: For comprehensive custom logic requiring direct interaction with the Data 360 Lakehouse.
    - **Functionality**: Customers write full Python scripts using the Code Extension SDK, enabling read and write access to the Lakehouse via SDK APIs. Ideal for custom/complex data preparation, or bespoke data manipulation.
    - **Isolation and Security**: While scripts access the Lakehouse, their execution is confined to a secure, isolated environment within the Data 360 runtime, preventing interference with other processes or unauthorized system access.
- **Function Model**:
    - **Purpose**: Analogous to a serverless function, for modular, stateless computations invoked from existing Data 360 pipelines (e.g., custom chunking in an Unstructured Pipeline).
    - **Functionality**: Customer-provided functions take input, compute, and return output.
    - **Isolation and Security**: These functions are designed for strict isolation; they do not have direct Lakehouse access. Their execution is sandboxed, stateless, and resource-constrained, making them suitable for focused, stateless processing steps, ensuring security, predictable execution, and minimizing blast radius.



Both the Script and Function models aim to securely execute customer code, preventing one tenant's code from affecting others or gaining unauthorized access to other tenants' data, Salesforce resources, or external resources. This security is achieved through a layered (defense-in-depth) architecture. This architecture provides an isolated execution environment for each tenant's custom code, incorporating various guardrails. These include logical isolation at the Kubernetes (K8s) level, network isolation, runtime sandboxing, and least privilege permissions, all complemented by operational monitoring and incident response readiness for detection and response.

Developer Experience & Operational Visibility
To support a robust development lifecycle, Code Extension offers:

- **External Authoring and Debugging**: Developers can craft and debug Python code in familiar environments like VSCode, leveraging the SDK.
- **Flexible Deployment**: Custom code can be packaged and deployed using SDK utilities, the Data 360 UI, or API, enabling integration into CI/CD.

Operational Logs: Access to detailed execution logs provides transparency and aids troubleshooting in production.

By offering these secure and flexible code extension capabilities, Data 360 empowers architects to tailor the platform to their most unique and complex data processing requirements, truly solidifying its role as an extensible enterprise data fabric.

Extending Data 360 by Bringing Your Own AI Models
As enterprises accelerate AI adoption, most maintain heterogeneous ML ecosystems — including Amazon SageMaker, Google Vertex AI, and custom Python-based environments — hosting models that drive mission-critical predictions, such as credit risk scoring, churn propensity, product recommendations, and next-best-action decisions.

Traditionally, integrating these external models into Salesforce required bespoke API layers, ETL pipelines, or middleware orchestration, introducing data duplication, governance overhead, latency, and operational complexity—challenges that conflict with a unified, compliant, and real-time Customer Data Platform (CDP) vision.

Bring Your Own Model (BYOM): Delivered via Einstein Studio in Data 360, addresses these challenges by enabling direct invocation of externally trained models within Salesforce workflows, Apex logic, and automation tools—without moving or replicating data. Through zero-copy federation, Data 360 acts as the governed single source of truth, exposing harmonized Customer 360 data for inference at external endpoints. Prediction outputs flow back in real time, powering business processes with scalable intelligence.

BYOM effectively closes the gap between data science and operational execution by decoupling model development, governed data, and consumption layers. It preserves platform independence, reduces integration complexity, accelerates AI deployment, and maintains governance over sensitive data.

The architecture operates as follows: Data 360 provides a unified customer 360 data foundation, while Einstein Studio orchestrates connections to external ML platforms (SageMaker, Vertex AI, or custom endpoints). External models execute inference in real time, batch, or streaming modes. Salesforce layers—Flow, Apex, and Query APIs—consume outputs to deliver personalized, automated, and analytical insights across Sales, Service, Marketing, and Industry Clouds.

From an enterprise standpoint, BYOM delivers:

- **Data integrity and governance**: Eliminates uncontrolled data copies and enforces policy compliance.
- **AI democratization**: Makes complex models accessible to non-technical users via Salesforce tools.
- **Time-to-value acceleration**: External models activate immediately within Salesforce processes.
- **Scalability and hybrid architecture support**: Enables multi-cloud deployment of AI workloads.
- **Future-ready AI architecture**: Supports composable AI strategies, decoupling data, model, and consumption layers for operational agility.

Bring Your Own LLM (BYO-LLM): Offers the same extensibility mechanism but for generative models. By enabling direct invocation of external LLMs we allow customers to use them in the Agentforce Platform in place of Salesforce-provided models. For enterprises BYO-LLM allows:

- Access to fine-tuned models
- Integration of models not currently provided by Salesforce
- Usage of models in customer-provided accounts

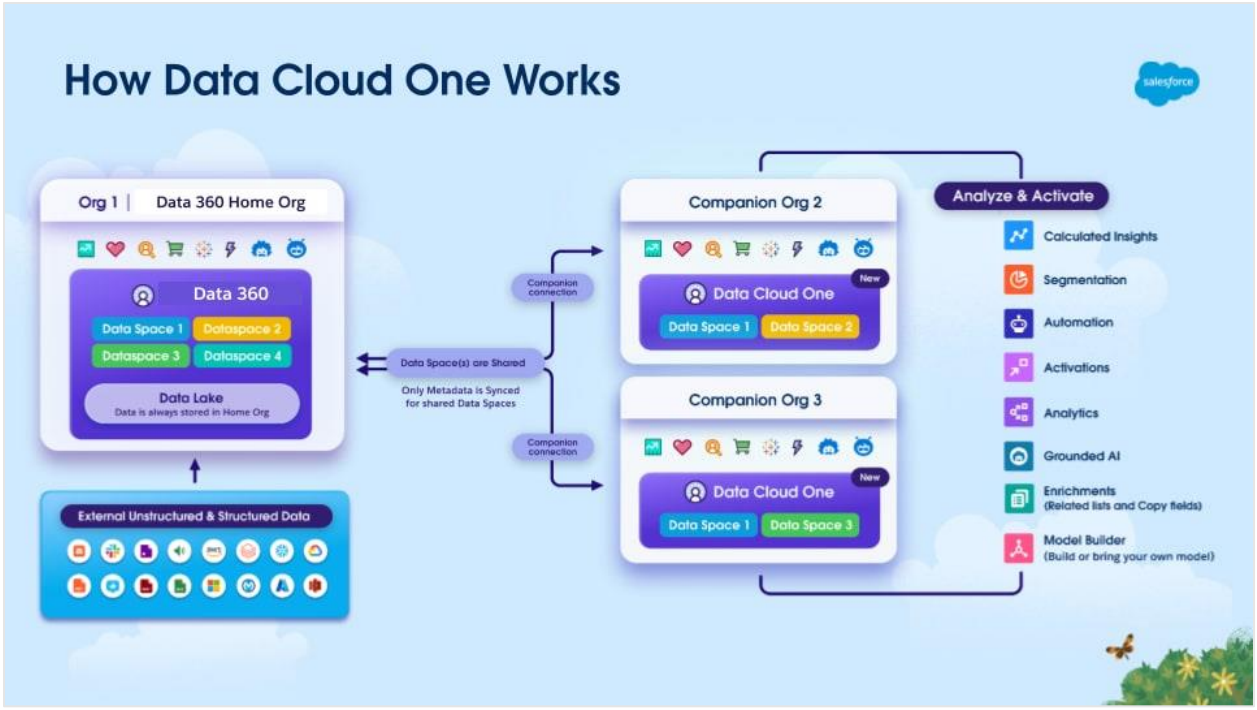
Data Collaboration
Modern enterprises operate within a complex data landscape characterized by two major architectural challenges:

1. **Intra-Enterprise Fragmentation**: Large organizations frequently utilize multiple Salesforce Orgs (often segmented by region, business unit, or historical acquisition) and numerous other data systems. This fragmentation creates internal data silos, making it impossible to establish a single, trusted, and unified view of the customer for real-time engagement across the business. The challenge is to unify this data without physically consolidating or duplicating it across all systems, ensuring governance remains intact.
2. **Cross-Enterprise Collaboration**: Companies often need to share data with partners and suppliers for joint marketing, measurement, and business intelligence. The challenge is enabling this collaboration while protecting sensitive, proprietary data and adhering to privacy regulations like GDPR and CCPA, as well as competitive barriers.

Salesforce Data 360 addresses these challenges with a zero-copy, trust-by-design framework built on the principle of sharing access and insights rather than moving or duplicating data.

Data 360 Collaboration Mechanisms
Salesforce Data 360 addresses data fragmentation and collaboration challenges with Data Cloud One, Data Sharing between Data 360s, and privacy-first Data Clean Rooms. These solutions unify customer data, enable secure data exchange, and provide privacy-preserving insights. With a zero-copy, trust-by-design approach, organizations can unlock data potential for real-time engagement, enhanced partnerships, and intelligent decision-making. Each of these data collaboration options serve different purposes.

Internal Enterprise Enablement with Data Cloud One
Data Cloud One is the foundational architectural solution for enterprises operating multiple Salesforce orgs. Its purpose extends beyond simple data sharing; it is designed to establish a single, trusted customer view and enable full Data 360 platform capabilities across the entire organization.

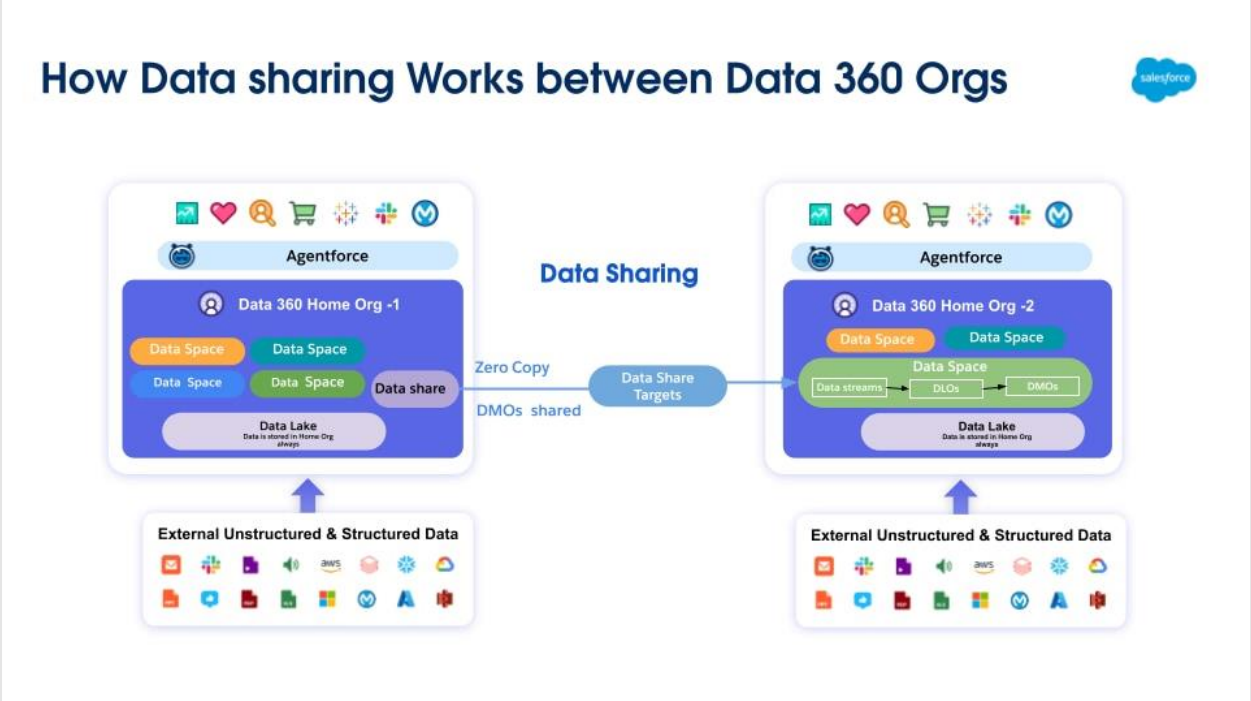


This mechanism centers on a designated Home Org Data 360 instance, which serves as the central authority for data management and building the unified customer profile. The Home Org is the org that Data 360 is provisioned on. A Data Cloud One connection is established between Data 360 and other Salesforce orgs, called Companion Orgs. As part of the Data Cloud One connection, Data 360 shares one or more of its data spaces with each companion org, providing access to both data and metadata in each shared data space. This is achieved through a zero-copy federation model and cross-org metadata synchronization.

Data Cloud One also allows Companion Orgs to leverage the Home Org's Data 360 instance for their own activation, personalization, and intelligence needs. This strategy is essential for eliminating internal data fragmentation and ensuring all business units activate against the same, governed, and unified customer profile, maximizing the ROI of the core Data 360 implementation.

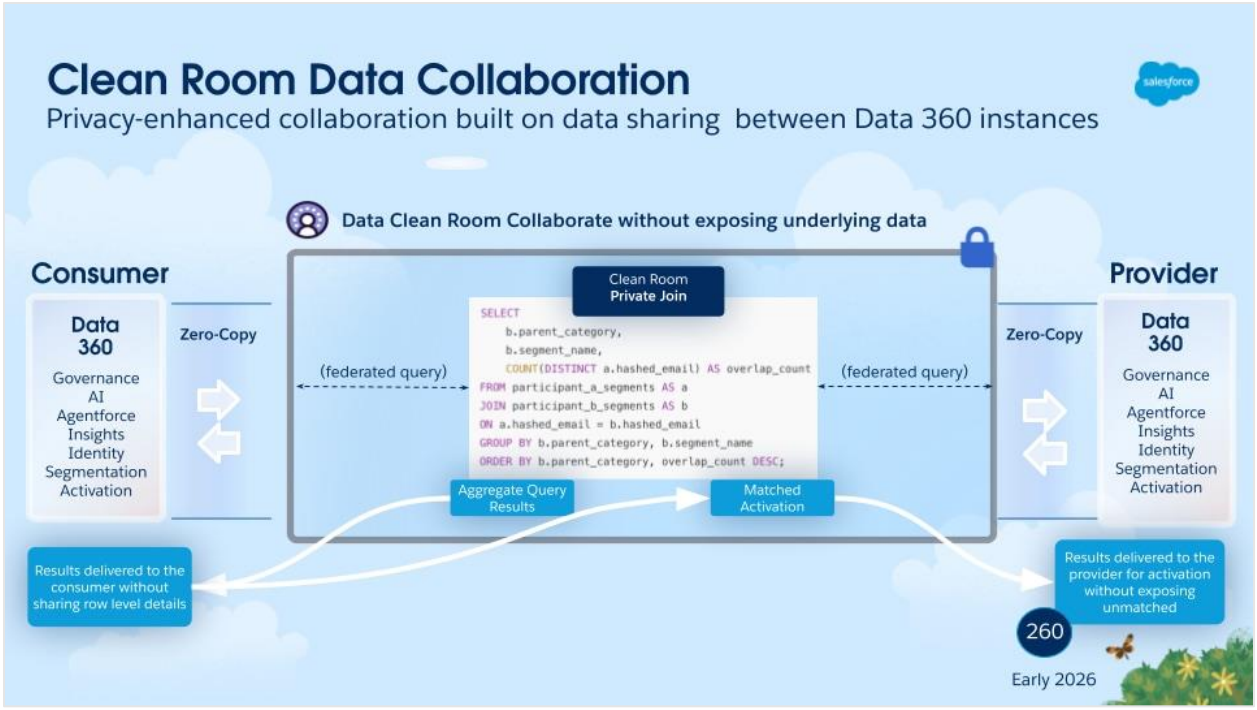
Data Sharing Between Data 360 Orgs

For distributed internal environments (where full centralization is not feasible) and for collaboration with trusted external partners, Data 360-to-Data 360 data sharing connects independent Data 360 instances.



This zero-copy sharing model establishes a connection between separate Data 360 tenants provisioned on different Salesforce orgs for the purpose of secure exchange of data objects (DLOs, DMOs, and CIOs). Once connected, the entire data object becomes accessible in the recipient Data 360. The recipient Data 360's administrator can then set governance rules to manage user access to this data.

Privacy-First Collaboration with Data 360 Clean Room Collaboration
When collaboration requires the highest levels of privacy and compliance, or when competitive concerns forbid the sharing of raw data, Data 360 Clean Rooms are architecturally mandated.



Architecturally, Data 360 Clean Room Collaboration is built on top of the Zero-Copy Sharing framework used by Data 360-to-Data 360 data sharing, but with additional layers of governance and computational constraints applied. The Data 360 Clean Room provides a secure, controlled compute environment where parties can join their data sets based on anonymized keys. Its core purpose is to allow joint analysis and insight generation without exposing the underlying proprietary data. The environment enforces strict, programmable rules such as minimum aggregation thresholds and non-exportable identifiers. These rules ensure that only approved, privacy-enhancing, and aggregated insights are derived and shared. This makes Clean Rooms essential for use cases like cross-platform campaign measurement and sensitive audience overlap analysis.

Conclusion: Salesforce Data 360: The Indispensable Data Backbone for the Agentic Enterprise

Data 360 is engineered as an intelligent, extensible, and trusted data fabric necessary to power next generation AI. Its architectural design addresses the problem of fragmented data, enabling organizations to unify, process, and activate all customer data at scale, using Zero Copy Data Federation to ensure efficiency.

Its robust Data Organization establishes a harmonized view from DLOs (including unstructured data) to DMOs, all secured within partitioned data spaces. Data 360's versatile data processing capabilities—including batch and streaming transforms, calculated insights, unstructured data processing, and identity resolution—are powered by the incremental SNCE and CDF architecture, ensuring efficient, near real-time processing and significant cost savings.

Extensibility is provided by the Code Extension architecture, securely enabling custom Python logic via Scripts or isolated Functions for unique requirements. Moreover, a comprehensive data

governance framework, built on attribute-based access control (ABAC) with CEDAR policies, guarantees granular security, dynamic data masking, and consistent enforcement across all data consumption. This culminates in sophisticated segmentation and activation capabilities, translating unified customer profiles into dynamic, multi-channel engagement strategies with real-time responsiveness.

Crucially, Data 360's ability to unify vast, diverse data, provide real-time context via its unified query (including hybrid structured/unstructured search and Hyper acceleration), and enforce stringent governance is paramount for powering intelligent AI agents. It provides the necessary trusted, fresh, and relevant data to ground Generative AI workflows (RAG) and to fuel the multi-turn, action-oriented capabilities of Agentforce-powered agents, ensuring they operate with precision and accuracy.

By providing a future-proof platform where source data is transformed into actionable insights, Data 360 serves as an indispensable architectural foundation for organizations building Agentic customer experiences. It's the vital backbone that turns customer data into the sophisticated, personalized customer experiences driving success for today's modern organizations.

1.3. Service Benefits

By improving data accessibility, accelerating insights, and optimizing costs, a data cloud can drive growth and operational efficiencies. Moving to this modern data architecture lets you focus on strategy and innovation rather than the maintenance of complex data infrastructures.

- Unified Customer View
  Data clouds can integrate all kinds of data, including operational, social, transactional, and IoT device data. If your customer data is spread among disparate systems, integrating it in a data cloud can give you a single, cohesive customer profile. This single source of truth of unified data about your customers helps you understand them and their preferences and serve them better.

- Scalability and Flexibility
  The cloud infrastructure allows you to scale storage and computing resources up or down easily, based on demand. This flexibility in handling sudden spikes in data volume means you won't need major upfront hardware investments anytime the data volume increases in your data cloud.

- Real-Time Data Processing
  Modern data clouds support high-speed ingestion and processing of streaming data. This capability is critical for acting on timely insights, such as personalizing a website experience while a customer is actively browsing.

- Cost Efficiency
  The pay-as-you-go model of cloud services transforms capital expenditure (CapEx) into operational expenditures (OpEx). This reduces your total cost of ownership compared to maintaining complex on-premise infrastructure.

- Enhanced Security and Compliance
  Leading cloud providers offer robust security protocols and automated compliance features. This helps protect sensitive information and ensures adherence to regulations like GDPR or HIPAA, simplifying governance for your organization.

1.4. Main components & Functions of this service

A data cloud can serve every function of a modern enterprise. Below are some common use cases.

Marketing Personalisation and Segmentation
A data cloud can be a valuable tool if your goal is to deliver hyper-personalized experiences to your customers. Marketers can combine behavioral data (website views, cart abandonment, liking a social media post) with demographic and transactional data to create precise, real-time customer segments. A retailer, for instance, can use the cloud to identify a customer browsing a specific product line and trigger a personalized email offer or display a tailored website banner in the moment.

Sales Performance and Forecasting
Unified customer data can be analyzed in a data cloud with AI or advanced analytics so as to generate more accurate sales forecasts. Advanced algorithms acting on unified historical data and real-time market signals can predict future performance with greater precision.

Service and Support Optimisation
A data cloud can provide immediate access to the full context of a customer's journey, regardless of the touchpoint (phone, chat, social media) to human and AI service agents. An agent can resolve issues faster and with more empathetic, informed responses, significantly improving customer relationships. Companies can also analyze service patterns to identify product defects or knowledge gaps proactively, leading to process improvements that reduce overall support costs.

2. Information Assurance

Salesforce holds a suitably scoped ISO:27001 certificate that covers the majority of its core services, please refer to the Salesforce Security, Privacy and Architecture (SPARC) documentation here for further product specific information. In addition, for country and product specific compliance please refer to this link for current status.

Salesforce.com EMEA Limited and its affiliates are committed to achieving and maintaining customer trust. Integral to this mission is providing a robust security and privacy program that carefully considers data protection matters.

In the UK Salesforce has completed an external CHECK Security assessment performed against its core services by a CREST member organisation, Salesforce also holds a current Cyber

Essentials certificate, again completed by a CREST member organisation and can provide a response to the NCSC 14 security principles upon request.

In accordance with the EU Data Protection Directive and implementing national legislation, the Salesforce Processor BCR (Binding Corporate Rules) is intended to provide an adequate level of protection for Personal Data during international transfers within the Salesforce Group made on behalf of Customers and under their instructions, further information is available here https://secure.sfdcstatic.com/assets/pdf/misc/Salesforce-Processor-BCR.pdf

Salesforce adheres to the principles of the EU-U.S. Privacy Shield framework with respect to personal data submitted by Salesforce's customers in reliance on the Privacy Shield to the following online services: Sales Cloud, Service Cloud, Force.com, Communities, Chatter, Site.com, Database.com, Analytics Cloud, Financial Services Cloud, Health Cloud, Heroku, Pardot and Configure, Price, Quote (CPQ) https://www.salesforce.com/assets/pdf/misc/privacy-shield-notice.pdf

In addition, we currently hold the following certifications:

Geographical Recognition

- TRUSTe Certified Privacy Seal
- Japan Privacy Seal from the Japan Information Processing Development Corporation
- (JIPDC)
- TUV Certificate
- FedRAMP
- Cyber Essentials
- Privacy Shield

Global Audit Compliance

- ISO 27001, 27018
- SSAE 18/ISAE 3402
- SOC-1
- SOC-2
- SOC-3 (SysTrust)
- PCI-DSS
- Cyber Essentials

Learn more here: Salesforce Services Trust and Compliance Documentation

Relating to the Government Security Classification (GSC) scheme and the National Cyber Security Centre (NCSC), Salesforce has various Public Sector customers, holding a variety of data at OFFICIAL and OFFICIAL Sensitive levels. Further Salesforce has various documents to assist organisations assessing the Security posture of Salesforce such as a response to the NCSC 14 Security principles. Please contact us for up-to-date information and access to additional information.

3. Data Backup / Restore and Disaster Recovery

Redundancy and Scalability
The Salesforce service is built for high availability across 3 availability zones within the region. Each availability zone includes the following:

- Multiple network carriers for customer connectivity
- Multiple ISPs for customer transit and internal replication
- Multiple dedicated connections for DR/BCP
- Redundant cloud-based networking infrastructure
- Web, Application, API, Cache, Search, Index, Query and Batch servers are load balanced and live within multiple availability zones within the region, with failover capabilities
- Database servers are replicated both within the same availability zone for performance reasons, and across 3 availability zones to ensure availability.

Extensive use of high-availability servers and network technologies, and a carrier-neutral network strategy, help to minimise the risk of single points of failure, and provide a highly resilient environment with maximum uptime and performance. The Salesforce Services are configured to be N+1 redundant at a minimum, where N is the number of components of a given type needed for the service to operate, and +1 is the redundancy. In many cases, Salesforce has more than one piece of redundant infrastructure for a given function.

Salesforce provides Enterprise class recoverability through replication of data across three distinct availability zones.

Disaster Recovery
Production data centres are designed to mitigate the risk of single points of failure and provide a resilient environment to support service continuity and performance. The Covered Services utilise secondary facilities that are geographically diverse from their primary data centres, along with required hardware, software, and Internet connectivity, in the event Salesforce production facilities at the primary data centres were to be rendered unavailable.

Salesforce maintains a disaster recovery plan that supports a robust business continuity strategy for the production services and platforms. Our platform is configured in a high availability mode and redundant fashion, and our Global Product Operations team monitors the software around-the-clock to ensure top performance. Product-specific information is available in the Security, Privacy and Architecture (SPARC) documentation for the respective services available on our website (https://www.salesforce.com/company/legal/trust-and-compliance-documentation/), and customers can view detailed policy documentation at https://compliance.salesforce.com/en/disaster-recovery-bcp.

Backup
All networking components, network accelerators, load balancers, Web servers, and application servers are configured in a redundant configuration. All Customer Data submitted to the Covered Services is stored on a primary database server with multiple active clusters for higher availability. All Customer Data submitted to the Covered Services is stored on highly redundant carrier-class

disk storage and multiple data paths to ensure reliability and performance. All Customer Data submitted to the Covered Services, up to the last committed transaction, is automatically replicated on a near real-time basis to the secondary site and backed up to localised data stores. Backups are verified for integrity and stored in the same data centres as their instance. The foregoing replication and backups may not be available to the extent that Customer has managed packages that are uninstalled by Customer's administrator during the subscription term because doing so may delete Customer Data submitted to such services without any possibility of recovery.

To help customers routinely back up their data, Salesforce offers several native options that are available for no additional cost to customers. Salesforce provides tools like Data Loader and the API as a method for customers to manually restore their data. It is important to note the order in which data is restored, so that relationships and the connection to related records can be preserved.

The following documentation and options are available to customers as a method of backing up their data:

- Data Export Service: Manual or scheduled exports of your data via the UI. Export Backup Data from Salesforce: https://help.salesforce.com/s/articleView?id=xcloud.admin_exportdata.htm&type=5..
- Data Loader: Manual on-demand exports of your data via the API. Export Data: https://help.salesforce.com/articleView?id=exporting_data.htm&type=5.
- Report Export: Manual on-demand exports of your data via reports. Export a Report: https://help.salesforce.com/articleView?id=reports_export.htm&type=5.
- Salesforce Backup and Restore product: https://www.salesforce.com/platform/data-backup-recovery/

Restore
Salesforce offers several native options at no additional cost to help customers routinely back up their data. Depending on your edition, your org can generate and export backup files on a weekly or monthly basis. Individual users can view and restore their deleted records from the Recycle Bin, where all deleted records reside for 15 days before being deleted from the system. Admins also have access to an org-wide Recycle Bin to restore or permanently delete records during the 15-day window. Salesforce also supports manual restore via Data Loader and API.

As an add-on solution, Salesforce Backup provides additional backup and restore capabilities. You can configure and maintain automatic backups, make a backup on demand, and quickly find and restore data. You can control the order in which data is restored to preserve relationships and connections to related records.

You can find additional information and best practices for backup and recovery at http://sfdc.co/BackupAndRecovery.

4. On-Boarding

Describe how easy and simple it is for a customer to On-Board to our service

4.1. Deployment

Salesforce is a SaaS (Software as a Service) solution and users access the tools via a web browser. Therefore, no software or hardware deployment is necessary. Implementation takes the form of assigning usernames and passwords to staff and (if needed) completing webinar-based training on the application.

4.2. Configuration and Customisation

The service provides many configuration options, including integration with custom applications. Details can be found here

5. Off-Boarding

Off-boarding from the service is as simple as exporting and downloading your data if you choose to do so. We provide your data in an industry standard readable format to make it as easy as possible for you to migrate to another service if you wish to do so.
For further details of data extraction and removal, please refer to section 5.2 below.

5.1. Termination

Details of termination options and implications are contained in our terms.

The Supplier would never terminate except for buyers' material breach of the Call-Off Contract and the terms.

5.2. Data Extraction and removal

With a commercial model that requires customer trust to be at the core of everything we do, Salesforce also helps ensure that our customers can exit our service with open and transparent processes and clearly defined commercial considerations. As such, should for any reason a customer wish to cease using the Salesforce service then outlined below are some of the key technical considerations for ending the subscription service and extracting both the customer's data and the technical investment made in the platform. Further detail relating to an exit strategy can also be discussed on understanding the nature of any potential solution

- Within 30 days post contract termination, customers may request return of their respective Customer Data submitted to the Covered Services (to the extent such data has not been deleted by Customer). Salesforce shall provide such Customer Data via a downloadable file in comma separated value (.csv) format and attachments in their native format. Note that Customer Data your organisation submits to Einstein Analytics instance groups for analysis is derived from other data to which your organisation has access, for example, data stored by your organisation using Service Cloud, Sales Cloud, third party applications, etc.
- Data export is provided at no additional charge. Full details are contained in our terms.

Salesforce understands the intentions of G-Cloud, and is keen to support customers adoption of cloud technology, however that should also include knowing how to leave a given cloud technology. Salesforce has an exit strategy paper that can be used to help plan a path away from Salesforce. Notwithstanding that Salesforce has one of the lowest attrition rates in the industry, we remain committed to your success. Continued investment is how we sustain 3 upgrades every year and a constant delivery on new innovation in every update.

6. Pricing Summary

Comprehensive pricing details can be found in our separate pricing document.

The service has various pricing levels and bandings based on type of service required and the volume needed. Please refer to pricing documents for detailed pricing.

Resource Tagging

Currently, Data Cloud and Digital Wallet do not explicitly support FOCUS (FinOps Open Cost and Usage Specification) resource tagging out-of-the-box.

However, Salesforce does support and how it relates to FOCUS principles:

What Salesforce Currently Supports

Resource Tagging Capabilities:

- Data Cloud Resource Tags for tracking consumption by specific resources
- Standard consumption tags by feature, agent, action, usage type
- Custom tagging (available February 2026) for department and cost center attribution

Resource references that link usage to specific sources and API names

Digital Wallet Cost Transparency [2]:

- Near real-time usage tracking with comprehensive consumption breakdowns
- Billing category, units consumed, multiplier applied, total credits consumed
- Data export capabilities through Data Cloud DLOs (Data Lake Objects)
- Custom reporting via Data Cloud reports and Tableau Next analytics

Current Export & Reporting Capabilities:

Data Export Options [3]:

- Export Digital Wallet data like any other Data Cloud object
- Data Cloud reports export with increased export limits
- Query editor access to billing usage DLOs
- TenantBillingUsageEvent and related consumption data streams

Available Data Fields [4]:

- Billing category
- Usage type and multipliers
- Resource ID and API Name
- Consumption timestamps and units
- Entitlement and transaction data

FOCUS Alignment:

While Salesforce provides extensive consumption tracking and export capabilities, the data format is not explicitly mapped to FOCUS standard dimensions like:

- focus:service_owner
- focus:billing_currency
- focus:resource_type
- focus:cost_category

Customer Implementation Path to match those:

- Export raw consumption data from Digital Wallet DLOs
- Transform the data to map Salesforce fields to FOCUS dimensions
- Custom reporting layer that presents data in FOCUS-compliant format
- Integration with FinOps platforms that expect FOCUS formatting

7. Service Constraints

Here are some key considerations we wish to highlight; a full list is contained in our Supplier Terms.

Usage Limits

- Services and content are subject to usage limits, including, for example, the quantities specified in order forms.
- A user's password may not be shared with any other individual.

If you exceed a contractual usage limit, we may work with you to seek a reduction in your usage so that it conforms to that limit. If, notwithstanding our efforts, you are unable or unwilling to abide by a contractual usage limit, you will execute an order form for additional quantities of the applicable services or content promptly upon our request.

Customisation and configuration of the service is independent of the underlying infrastructure; upgrades do not impact any changes you may have made. As a result of this independence it is typical for customers to customise their service to their own requirements. As the service is SaaS based, Salesforce is aware of which users are using which features etc., therefore we are aware of any potential impact a feature deprecation would have on customers and can work with them should there be a need for a feature deprecation.

8. Service Management

Your success is our success

Customer success is a top priority for Salesforce. Every customer gets a Standard Success Plan for online support and training. Our most successful customers take advantage of our Premier Success Plans to achieve an 80% higher return on their Salesforce investment. Large enterprise customers can benefit from Signature Success, our highest level of service to support their most critical business demands.

Premier and Signature Success packages can be purchased through G-Cloud specialist cloud services.

Standard Success
Every Salesforce customer gets a Standard Success Plan for online support and training. Our Standard Success Plan, included with each licence, provides:

- Success Communities to share with other customers
- Guided Journeys on how to use Salesforce
- Circles of Success Interactive Events
- 12/5/365 online case submission
- Response in two business days
- Trailhead online training
- “Getting Started” online training catalogue

Standard Success is for companies that need standard guidance in getting started with Salesforce. If you need a faster response, 24x7 support coverage, and/or a comprehensive training solution, we recommend our Premier Success Plans.

Premier Success
Our most successful customers take advantage of Premier Success to achieve an 80% higher return on their Salesforce investment.

The Premier Success Plan provides specialised guidance whether you have how-to questions, experience technical issues, need troubleshooting, or want to increase the value you get from Salesforce.

Benefits include:

- 24/7 online and phone support by senior support analysts
- Expert coaching sessions
- Reviews of your platform health and business value
- Developer support
- Expert Assistance

Specialised Guidance

- A discount on all Trailhead Academy courses and certifications
- Quicker response times than Standard Success. When critical issues arise, our skilled support engineers respond within one hour.

Expert Assistance
Expert coaching sessions are specialised engagements designed to help you get more value

from Salesforce products. With Premier Success, you can attend webinars on a specific topic and watch coaching videos and then have an individual follow-up session to dive deeper.

We also offer personalised sessions with Salesforce experts to help you overcome obstacles and drive long-term success. More than 200 options cover a range of needs and interests across Salesforce products.

To get real-time answers to your questions, Premier Success includes live Q&A sessions with Salesforce experts, addressing topics from adoption and how-tos to best practices. Specialised technical support is also available for admins and developers to troubleshoot custom code issues.

Specialised Guidance
To help ensure that you see continuing success with Salesforce, Premier Success includes personalised guidance and insights. Through periodic reviews and check-ins, we evaluate your platform health and value maturity. These technical and business reviews assess key areas of platform performance, prioritise areas for growth, and set and track progress against your targets with quantifiable success metrics.

Signature Success
Extend your team with Signature Success. Rest assured knowing our certified experts are here to help you maintain your Salesforce solution.

Signature Success is the highest level of support from Salesforce and provides a high-touch experience led by a named expert who acts as an extension of your team. Customers with Signature Success benefit from increased performance and productivity through all the features of Premier Success plus:

- 15 minutes initial response for critical issues.
- The fastest response times from our most skilled support engineers.
- A designated Technical Account Manager.
- Customer Success Score for actionable insights and recommendations.
- Technical Health Reviews.
- Proactive Monitoring and Key Event Management.

Technical Account Manager (TAM)
Account management by a named champion. An assigned TAM provides consistent advocacy and guidance with a deep understanding of your business. These highly experienced technical experts provide support case oversight and escalation, weekly meetings, and tailored solution guidance.

Proactive services
Signature Success includes 24/7 monitoring that is tailored to your configuration. Your TAM coordinates with Proactive Services Engineers experts and helps ensure you prevent or mitigate potential issues identified, before they can create business disruption. If anything does go wrong, you can expect early alerts, remediation, and reviews to eliminate recurring root cause patterns.

Available at all times and offers the fastest case response times, including 15 minutes for severity 1 issues via phone or chat through our Help site.

Accelerators
Accelerators are quick, personalised work sessions that solve specific Salesforce challenges. The list of accelerators available varies depending on the type of plan you chose; there are no accelerators available with the standard plan. A customer can have unlimited accelerators as long as only one is being delivered at a time.

Learn more: https://www.salesforce.com/services/success-plans/overview/
There is no financial recompense model for not meeting service levels.

9. Customer Responsibilities

Here are some key responsibilities we wish to highlight; a full list is contained in our Salesforce Supplier Terms.

Management
Using our online tools, you will need to manage your usage of and access to the service, for example user accounts, installed applications, sites, etc.

Compliance
In using the service, you need to ensure your compliance with our terms, which include:

- Being responsible for the accuracy, quality and legality of your data and the means by which you acquired your data. Using commercially reasonable efforts to prevent unauthorised access to or use of the service, and notify us promptly of any such unauthorised access or use.

Usage Restrictions & Information Security
You must:

- Ensure that only information of an appropriate security classification is placed into the service.

You must not:

- Use the service to store or transmit infringing, libellous, or otherwise unlawful or tortious material, or to store or transmit material in violation of third-party privacy rights.
- Use the service to store or transmit malicious code.
- Interfere with or disrupt the integrity or performance of any service or third-party data contained therein.
- Attempt to gain unauthorised access to any service or content or its related systems or networks.
- Permit direct or indirect access to or use of any service or content in a way that circumvents a contractual usage limit.

10. Client-side Technical Requirements

Salesforce is a SaaS (Software as a Service) solution and users access the tools via a web browser. Therefore, no hardware or software installation is necessary.

Browser requirements
Salesforce supports a range of popular browsers. Learn more: Supported browsers

Internet access
Salesforce is designed to use as little bandwidth as possible, so that the service performs adequately over high-speed, wireless and mobile Internet connections.

While average page size is on the order of 90KB, Salesforce supports compression as defined in the HTTP 1.1 standard to compress the HTML content before it is transmitted as data across the Internet to a user's computer. The compression often reduces the amount of transmitted data to as little as 10KB per page viewed, due to the lack of image content. The site was designed with minimum bandwidth requirements in mind, hence the extensive use of colour coding instead of images. Our average user also is known to view roughly 120 pages from our site per day. However, it is best to measure any page that has been customised, especially if Visual Force components have been added to the page, to get an accurate measurement of the page size.

Our application is stateless; therefore, there are no communication requirements in the background once the page loads like traditional client server applications e.g. Outlook, therefore, once the page loads there are no additional bandwidth requirements until a user queries or writes information to Salesforce. Further information here
https://help.salesforce.com/articleView?id=000004958&type=1

11. Planned Maintenance Windows

When maintenance is scheduled, Salesforce publishes the dates and times of the maintenance windows on trust.salesforce.com. Premier Alerts are sent via email when the maintenance windows are posted to trust.salesforce.com. In the event of planned maintenance that requires customer action in advance, such as updating network settings in preparation for additional login pools, Salesforce endeavours to communicate via email to system administrators of your organisation months prior to the maintenance. If emergency system maintenance is required, customers may be notified less than one week in advance.

There are two types of maintenance at Salesforce:

- System maintenance is for sustaining the security, availability, and performance of the infrastructure supporting Salesforce services.
- Release maintenance is for upgrading Salesforce services to the latest product version to deliver enhanced features and functionality. There are three different kinds of release maintenance: major releases, patch releases, and emergency releases.

Major release maintenance dates and times are posted on trust.salesforce.com approximately one year before the release date. Major release maintenance occurs three times per year.

Patch releases and emergency releases are used to deliver scheduled and ad hoc application fixes and are typically seamless to customers. Whenever possible, patches and emergency releases are deployed during off-peak hours and without downtime. You can see our preferred maintenance schedule at https://help.salesforce.com/apex/HTViewSolution?id=000176208&language=en_US or learn more at http://trust.salesforce.com

12. Training

We provide webinar-based and instructor led training for the majority of our services. Details can be found in our separate listing for Salesforce Training also listed on the Digital Marketplace. Further information can also be found here: https://www.salesforce.com/services-training/training_certification/training-by-cloud.jsp

13. Data Centre Locations

For the Salesforce Services (services branded as Force.com, Site.com, Sales Cloud, Service Cloud, and Chatter), if a new 'Org' of the Salesforce Service is provisioned for a Customer with a billing address in the UK, Customer Data in that 'Org' is stored in data centres in England.

In addition, Salesforce may store information in data centres located outside of EEA, such as identifying information, relating to the Customer's instances(s) of Salesforce Services and users for the purposes of operating the Salesforce Services, such as facilitating the login process and the provision of customer support.

Such identifying information as provided by the Customer in its provision of user accounts shall only include the following personal data about users: first and last name, email address, username, phone number, and physical business address.

For other services which are not Salesforce Services, Salesforce uses various data centres located throughout the world. The location of Salesforce data centres for all services is specified in theSalesforce Infrastructure and Subprocessors Documentation, available under the following link - https://www.salesforce.com/company/legal/trust-and-compliance-documentation/

14. Performance

Full details are contained in our terms.

14.1. Service Levels

We will use commercially reasonable efforts to make the services available 24 hours a day, 7 days a week, except for:

- Planned downtime (see section 13).
- Any unavailability caused by circumstances beyond our reasonable control, including, for example, an act of God, act of government, flood, fire, earthquake, civil unrest, act of terror, strike or other labour problem (other than one involving

our employees), Internet service provider failure or delay, non-Salesforce application, or denial of service attack

.

14.2. Incident Response Time

The incident response time varies from 15 minutes to two days depending on the level of support selected. Details can be found in section 8

14.3. Incident Updates

Incident update intervals vary based on the level of support selected, ranging from every 30 minutes to longer periods. Details can be found in section 8.

THE CONTENT OF THIS DOCUMENT IS FOR INFORMATIONAL PURPOSES ONLY AND A CUSTOMER'S RIGHTS AND REMEDIES ARE GOVERNED EXCLUSIVELY BY THE G-CLOUD 15 CALL-OFF CONTRACT AND SUPPLIER TERMS. ANY RESPONSE TIMES OR CUSTOMER OUTCOMES STATED HEREIN ARE INDICATIVE ONLY.

`