Cohesity Gaia

An executive's guide to accelerating enterprise AI with retrieval augmented generation and secondary data

COHESITY

October 2024

COHESITY

Table of Contents

COHESITY

Introduction

Cohesity Gaia is an advanced generative AI application that uses retrieval augmented generation (RAG) technology to securely manage and process backup data stored in the Cohesity Data Cloud. The primary objective of Cohesity Gaia is to give enterprises in-depth insights into their stored data so they can extract valuable information and strengthen decision-making processes. This document delves into the various technical facets of Cohesity Gaia, encompassing its core services, data management capabilities, and security features.

Cohesity Gaia enables organizations to make smarter and faster business decisions by bringing the power of RAG AI to high-quality enterprise data within Cohesity to unlock business insights.

Cohesity Gaia is part of a leading AI-powered data security and management portfolio. Cohesity protects the world's most critical data workloads and gives organizations a simple way to secure and manage their data across on-premises, cloud-native, and SaaS workloads.

COHESITY

The enterprise imperative for AI

For decades, organizations have invested in a broad range of technologies to use their most valuable asset—data—to unlock transformative business insights. From master data management solutions that rely on rules-based logic to generate clean customer records, to the hype of Big Data and Hadoop, to data warehouses that power business intelligence solutions, these technologies have a mixed record of success.

This uneven track record is often due to a common set of challenges: data silos across the enterprise, poor data quality (including data that's outdated, incomplete, or inaccurate), the need for highly manual processes to consolidate and clean data, and a lack of important metadata or data classification capabilities to transform data into easily searchable and actionable information.

Today, cloud-scale architectures enable transformative insights since enterprise data can be indexed and searched at scale. That's one leap forward. A second leap is AI.

AI has transformed our ability to address many data challenges. It can solve complex tasks that need precise outcomes based on learned knowledge. For example, AI can recognize patterns across large datasets, such as identifying product types in retail using signifiers from large volumes of data and human input to train models. By using AI to drastically minimize the manual overhead to consolidate and classify products accurately, organizations have improved their ability to create tailored web experiences (e.g., retargeting ads and the familiar "products you may like" panel on e-commerce sites).

So why is AI such a big deal for enterprises like yours right now? Let's start with a simple example for how you might answer a business question about a historical legal case with your existing workflows:

1. **Look for a folder:** An employee seeks files related to a specific legal case. After searching through many folders, they pull one related to a particular testimony.
2. **Read documents:** The employee reads the documents found in the folder. They also note any relevant cross-referenced materials from other folders. They repeat steps 1 and 2 until they think they have all relevant documents to generate a comprehensive answer.
3. **Generate an answer:** The employee spends time noting key takeaways from each document and synthesizing all notes into a concise response to the question. The response cites relevant documents. This process can take days or even weeks, depending on the scale.

COHESITY

Now what if you could have an algorithm do all that for you? In recent months, AI technology has advanced so rapidly that it can now generate language and images that make it easier for people to use and interact with machines and data in ways that mimic human cognitive abilities.

Through these advancements, you can get information from business data using natural language questions instead of complex data queries. And with the right technologies, people can receive responses from AI that go far beyond traditional data outputs. In fact, these responses have rich context because the model can search, analyze, and combine different datasets and generate language to answer questions in a human-readable way. The most well-known example is, of course, the experience pioneered by ChatGPT and OpenAI.

Growing interest in AI and what it means to ITOps

In late 2023, Cohesity funded market research to assess the state of the backup and recovery market in terms of AI and ML. Our goal: to uncover the impact of AI initiatives on backup and recovery processes and infrastructure, including the support of broader corporate initiatives. The research surveyed hundreds of IT and data professionals in North America responsible for data protection. The key finding:

Organizations want to use AI to gain business insights, but concerns around data consolidation, data security, data quality, and lack of expertise create barriers to adoption.

When asked what their organization's primary business objectives were for implementing AI, the results were primarily internally focused and included:

- Improve operational efficiency (36%)
- Enhance data analytics and insights (34%)
- Improve decision-making speed and accuracy (30%)
- Improve risk management and compliance (30%)

However, there were still challenges to address before implementing AI initiatives at scale. The research shows that top IT concerns around AI include:

- Concerns about data privacy and security (28%)
- Limited availability of quality data for models (23%)
- Difficulty integrating with existing systems and processes (21%)
- Lack of expertise and talent (21%)

Source: Enterprise Strategy Group, "Reinventing Backup and Recovery With AI and ML," January 2024

COHESITY

While more and more organizations are ready to embrace AI, it's important to note that generative AI and LLMs are only as good as the data that underpins them. Data silos across organizations continue to make it difficult to leverage enterprise data at scale, hindering the ability for AI outputs to be both complete and accurate.

What's more, running LLMs can be extremely compute-intensive and cost-prohibitive. When data is not properly deduplicated, or doesn't have metadata that can improve data retrieval and response generation, response quality goes down and costs continue to go up. This phenomenon has long been described as "garbage in, garbage out."

An underappreciated prerequisite for AI in enterprises, then, is high-quality backup data that's accessible on a modern platform and can be integrated with LLMs.

COHESITY

What retrieval augmented generation is (and why it matters)

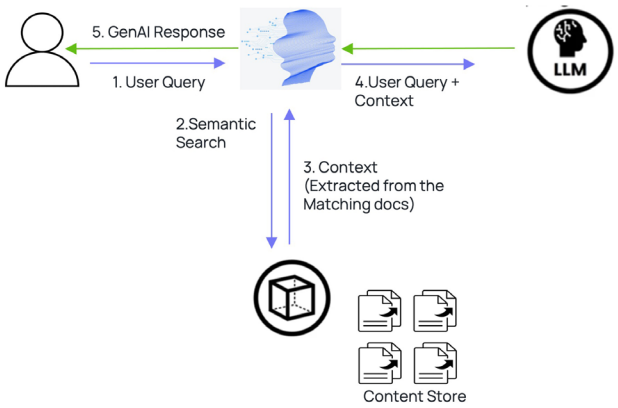
Cohesity Gaia leverages RAG technology to securely manage and process backup data stored in the Cohesity Data Cloud.

There are several technologies that support these technological breakthroughs:

- **Generative AI:** Generative AI uses algorithms to generate new content (written content, images, video, audio, computer code, etc.) based on user input. Unlike earlier versions of AI, generative AI can create new content, like cyber threat analyses presented in a conversational user interface.
- **Large language models (LLMs):** LLMs are learning models trained on vast amounts of data and apply language to generative AI capabilities.
- **Retrieval augmented generation (RAG):** RAG is a natural language processing technique that combines the benefits of retrieval-based and generative-based approaches to improve the quality of text generation tasks, such as question-answering, summarization, and conversational AI. RAG models combine the strengths of LLMs with the ability to retrieve information from multiple sources. RAG enables LLMs to generate more knowledgeable, diverse, and relevant responses.

When coupled with a modern data platform, these AI advancements transform data into knowledge.

COHESITY



Retrieval Augmented Generation architecture

- **What is RAG?** – RAG is the process of optimizing the output of a large language model, so it references an authoritative knowledge base outside of its training data sources before generating a response.

Fig . 1 : What is RAG? – Retrieval Augmented Generation

How does retrieval augmented generation (RAG) work?

RAG integrates retrieval-based techniques with generative-based AI models. **Retrieval-based models** excel at extracting information from preexisting online sources like newspaper articles, databases, blogs, and other knowledge repositories such as Wikipedia or even internal databases. However, such models can't produce original or unique responses.

Generative-based models can generate original responses that are appropriate within the context of what is being asked, but they're not always accurate.

So **RAG models** were developed to combine the strengths and minimize the drawbacks of retrieval- and generative-based models. In a RAG-based AI system, a retrieval model is used to find relevant information from existing sources while the generative model takes the retrieved information, synthesizes all the data, and shapes it into a coherent and contextually appropriate response.

What are the benefits of retrieval augmented generation?

By integrating retrieval and generative AI models, RAG delivers responses that are more accurate, relevant, and original while also using more natural, human language. That's because RAG models can understand the context of the queries and generate fresh, unique replies. It's truly the best of both models. RAG models are:

- **More accurate.** By first using a retrieval model to identify relevant information from existing knowledge sources, the original, human-like responses are more context-rich and up-to-date than responses from a purely generative model.

COHESITY

- **Better at synthesizing information.** RAG can synthesize information from numerous sources and generate fresh responses in a human-like way. This is particularly helpful for more complex queries that require integrating information from multiple sources.
- **Adept at putting information into context.** Unlike simple retrieval models, RAG can generate responses that are aware of the context of a conversation, and are thus more relevant.
- **High quality, relevant results.** RAG can generate high-quality results for your specific needs without resorting to costly and time-consuming training or fine tuning.
- **More efficient.** RAG is more accurate and efficient than LLMs or large scale generative models alone.

COHESITY

Why Cohesity Gaia is such a game-changer

Cohesity was founded with the goal of simplifying data management and modernizing backups and data protection. Our vision was that customers could secure and manage their data, while harnessing insights from that data—all on one platform.

The introduction of Cohesity Gaia ushers in a new era for our customers, harnessing the power of AI to make it possible for our customers to go deeper into their data and glean the learnings and knowledge hidden in the data.

As an AI conversational assistant, Cohesity Gaia brings RAG AI and LLMs to customers' high-quality backup data within Cohesity to unlock data insights. Users can ask natural language questions and receive contextual answers based on enterprise data.

COHESITY

Responsible AI

Cohesity Gaia is powered by Cohesity Turing, a patent-pending collection of AI capabilities and technologies integrated into Cohesity's multicloud data security and management platform, providing both operational and data insights. The foundation of our AI innovations is the concept of "responsible AI," with capabilities and frameworks that enable customers to introduce AI to backup data securely and safely at scale.

Our responsible AI guiding principles include:

- **Transparency:** Protect access to your data with role-based access controls (RBAC) controls. Promote transparency and accountability around access and policies.
- **Governance:** Ensure the security and privacy of data used by both AI models and the workforce—so the right data is exposed only to the right people (and models) with the right privileges.
- **Access:** Integrate indexed and searchable data securely and easily, while ensuring data is immutable and resilient.

In addition, you can:

- **Improve decision-making speed and accuracy with AI-powered conversational applications running on high-quality backup data.**

  With Cohesity, organizations have the unprecedented ability to gain insights from their data with intuitive question-and-answer search experiences. Backup data indexed by Cohesity powers the conversational search solution and contains appropriate metadata that can be used with LLMs to provide human-readable responses. Using authoritative data sources backed up on Cohesity can help ensure more accurate, actionable responses to user or machine queries.

  With Cohesity, customers can get more value out of their backup data and significantly improve time-to-value for AI-powered data insights. Since backup data from Cohesity is AI-ready, organizations can support their first AI-powered conversational search experience in minutes or hours, instead of the weeks or (more likely) months needed to develop custom generative AI applications for enterprise data.

- **Accelerate AI initiatives with a secure, highly-performant RAG AI engine from Cohesity.**

  Inspired by web-scale architectures, Cohesity provides a scale-out, globally deduplicated, highly available platform to consolidate all data, including backups, files, and test/dev copies into a single repository. Backup data used to power RAG AI capabilities from Cohesity doesn't require you to keep an extra copy just for the AI.

  Since Cohesity indexes all backup data, the solution returns context-aware responses in a highly performant way that doesn't use up too much compute power. With Cohesity, organizations can power RAG AI without touching production data or incurring the expense and bother of keeping a full replica of production data elsewhere.

COHESITY

- **Streamline compliance and risk management with the ability to responsibly and securely search enterprise data across time with Cohesity Gaia.**

  End users within the organization can ask conversational questions in Cohesity Gaia related to specific business inquiries such as legal and compliance use cases. At launch, we support legal and compliance use cases that will let those teams ask questions and receive human-readable, actionable responses about their data that contains highly relevant information from historical records, cited documents, or emails.

  Users can prompt Cohesity Gaia for additional details, ask follow-up questions, and dig deeper into answers as if they were speaking directly with a subject matter expert. This is an improvement over current workflows, which can be very time-consuming for IT—as IT would need to produce audit logs, perform data forensics, and work with compliance teams to refine and update data for a potential investigation.

  Our granular role-based access controls (RBAC) for backup data help prevent users from accessing data they don't have permissions for, like sensitive data (patient data/PII, trade secrets, financials, and more). This approach applies to AI, where the AI model only queries data and provides responses that align to users' permissions.

- **Simplify knowledge transfer within your organization with the ability to search across previous work, projects, and departments across time with Cohesity Gaia.**

  Cohesity Gaia can also help in organizational knowledge sharing. Sharing information across an organization can be difficult and cumbersome. As employees join and leave an organization, project, or team, handing off information and transferring knowledge is particularly difficult. Cohesity Gaia can simplify that process by using backups and natural language queries to unlock learnings from previous work, different departments, or projects, creating summaries that can make a transition or knowledge transfer easier for all.

- **Secure your corporate data and keep it private**

  Your own private and secure AI engine won't reveal your corporate data to the rest of the internet. Easily integrate with your enterprise data and infrastructure so your secure and private AI assistant can unlock the knowledge in your data. Keep your data safe with granular role-based access controls (RBAC). Help prevent users from accessing data they don't have permission to access, like sensitive patient data, PII, trade secrets, financials, and more.

COHESITY

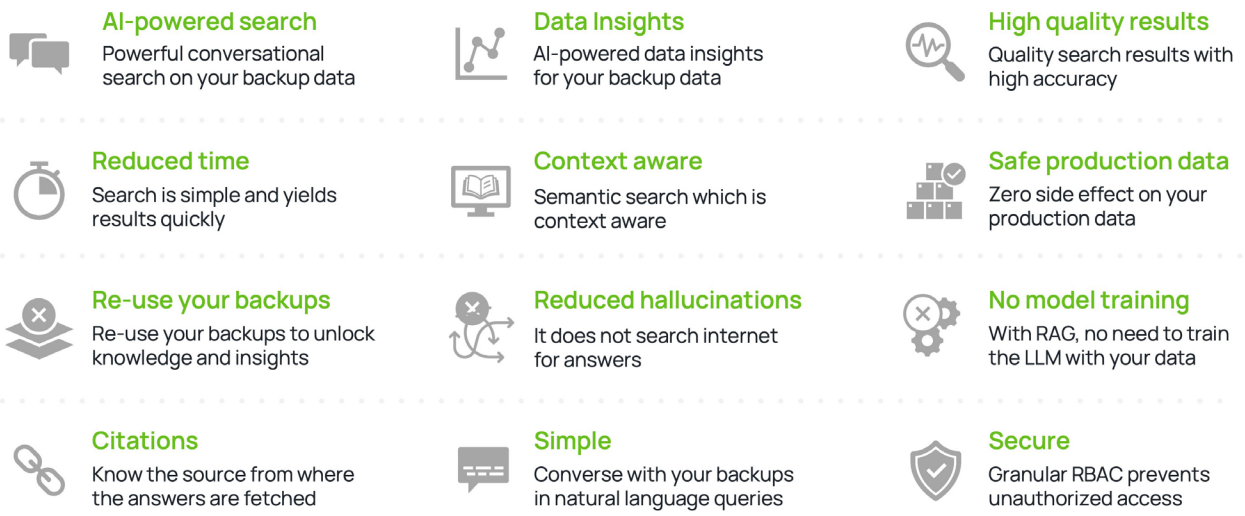


Fig . 2 : Cohesity Gaia benefits

COHESITY

Use cases

At its core, Cohesity Gaia is ideal for any department or job function that searches, retrieves, and summarizes data from your corporate data. Our research shows that roughly a third of time at work is spent doing just this.

Traditionally, searching, retrieving, and summarizing data were very manual processes. While search, in particular, has improved over the years, we're still using complex boolean equations or sensitive keyword searches that may or may not find the information you're looking for. And if you do find what you're after, you still have to manually go into each finding, synthesize the data, and make a judgment call about whether the results are relevant to your question—or not. Then you need to collect the data to include in a summary that you also need to create manually. These processes can take hours, days, weeks, or even months.

With Cohesity Gaia, the entire process can be slashed to mere minutes. You choose the files, docs, etc that you want Gaia to examine—it doesn't matter if you have 10 or 1,000. Then Cohesity Gaia will process those materials so it can rapidly locate the best matches to use in your RAG based queries.

Once indexed, you'll use the chat interface to ask questions using natural language. Cohesity Gaia will look inside the files for answers to your questions, summarize the findings, and link to the files used for easy reference.

This takes just minutes. You only need to evaluate and verify the results.

Because Cohesity Gaia can search across files, documents, presentations, PDFs, and other types of data, it's an excellent solution for departments looking to share data, investigate actions, and track how data is used. Use Cohesity Gaia for:

- Data insights
- Knowledge management
- eDiscovery and information governance augmentation
- Compliance augmentation

Beyond these common use cases, Cohesity Gaia can provide nearly limitless value in a corporate setting.

Fig. 3: Possible Cohesity Gaia use cases

COHESITY

Cohesity Gaia: An inside look

Cohesity Gaia offers a comprehensive solution that caters to the unique requirements of enterprises. Our data security and management platform, the Cohesity Data Cloud, provides the necessary tools to handle diverse data types, ensure data consistency, and maintain a high level of security while enabling powerful analytics capabilities.

Architecture: Cohesity Gaia's architecture consists of a control plane (Gaia-CP) and a data plane (Gaia-DP) that work together to manage and process enterprise data. The control plane is responsible for orchestrating various workflows, managing data models, and providing APIs for user interactions. The data plane is responsible for accessing data stored in the Cohesity cluster. With Cohesity Gaia's Embedding Service (Gaia-ES), enterprises can effectively extract text from various file formats, create semantic indexes on the data, and use these indexes to gain deeper insights into their data.

Infrastructure: Cohesity Gaia is designed with flexibility in mind. Design decisions made today are meant to support the deployment of its components on the cloud, on-prem, or a hybrid of both in the future. For example, Cohesity Gaia's indexing service is designed to abstract away different vector databases from the embedding service, so in the future, enterprises can choose the most suitable vector database technology based on their specific needs and requirements. Cohesity knows that customers will want a choice of infrastructure and services that align with their organization's cost, performance, and security objectives.

Security: Security is a top priority for enterprises, and Cohesity Gaia addresses by implementing fine-grained, specialized RBAC policies. These policies restrict access to Cohesity Gaia APIs, ensuring that only authorized users can access specific data sets. This helps balance data accessibility for RAG applications with the need to protect sensitive information.

COHESITY

Start with Cohesity DataProtect first

Organizations start with data protection from Cohesity and then use their high-quality backup data for Cohesity Gaia's AI-powered conversational questions and responses.

Customers can:

- **Back up data using Cohesity DataProtect:** Protect cloud-native, SaaS, and on-prem data at scale with Cohesity. Our unique backup capabilities:
  - Improve RPO
  - Offer flexible recovery capabilities such as instant mass restore
  - Minimize RTO
  - Make backup data from Cohesity AI-ready
- **Use RAG and LLM on backup data:** The data is applied to a LLM and our RAG AI technologies, creating a unique AI index built on your own data.
- **Get business insights faster:** Start having a conversation with your data. Using common language, ask questions about your data and await responses. Ask follow-up questions, dig deeper into datasets, and let Cohesity Gaia support more in-depth data analysis.

COHESITY

Here's how Cohesity Gaia works

1. Data indexing

- **Data:** Connect your data to Cohesity Gaia in any supported format from Cohesity-managed backups. Initially, Cohesity Gaia will support M365 and OneDrive data from our Cohesity Data Cloud. We plan to support more data sources and targets in future releases.
- **AI processing and indexing:** Cohesity Gaia vectorizes the data, creating a baseline for answering questions on your enterprise data. Cohesity stores the generated vectors in a specialized database optimized for dimensional vector search to provide more contextual-rich responses.

2. Answer generation

- **Use Cohesity Gaia for seamless RAG AI search experiences:** When users ask business questions within Cohesity Gaia, Cohesity uses RAG AI to increase the accuracy of answers by passing relevant retrieved data as context to foundation models. Azure OpenAI LLM is supported at launch.
- **Contextually-rich answers:** Cohesity Gaia delivers generative answers and insights based on questions and your enterprise data.

COHESITY

Architectural overview

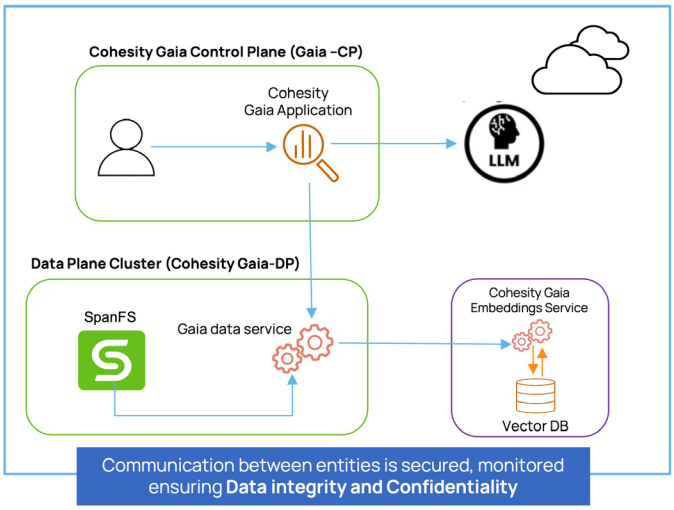
Fig . 4 : Cohesity Gaia under the hood

Cohesity Gaia capabilities

Cohesity Gaia is designed to (1) provide enterprises with a user-friendly interface for accessing backup data through public APIs and (2) create a semantic index on that data. This enables customers to leverage the index for deeper insights into their data, improving their ability to make informed decisions. By integrating with LLMs, Cohesity Gaia delivers an efficient and secure data management solution for organizations.

VectorDB storage and access

A vector database is used by Cohesity Gaia to store generated text embeddings for a dataset. It resides on a Cohesity-managed container service. Access to the vector database is controlled and monitored, ensuring data confidentiality and integrity.



Cohesity Gaia Control Plane

- Central management and orchestration component
- Manages logical and physical states of the application

Cohesity Gaia Data Plane

- Serves as data access layer for stored data

Cohesity Gaia Embedding Service

- Store/Manage generated text embeddings for data sets

Fig . 5 : High level view of core services and architecture

Gaia Control Plane (Gaia-CP)

The Gaia Control Plane (Gaia-CP) serves as the central management and orchestration component of the Gaia Service. It oversees the logical and physical states of the system, handles customer dataset indexing, and manages the scaling of various services based on their performance and usage. Gaia-CP is deployed as a microservice within the Helios Control Plane, ensuring a secure, efficient, and unified management framework for the entire Gaia Service ecosystem.

Gaia Data Plane (Gaia-DP)

The Gaia Data Plane (Gaia-DP) functions as the data access layer for the system, providing a comprehensive set of APIs for data indexing, retrieval, and manipulation. Gaia-DP is designed to be deployed as a managed service on backup as a service (CCS) or, in the future, customer-managed Cohesity clusters, ensuring seamless integration with existing data management infrastructure and streamlined operations.

The resulting set of files from the search filter are passed to the Gaia Embedding Service which lives logically next to the DataPlane cluster, but in separate infrastructure.

COHESITY

Data Access Service in Gaia-DP

The Data Access Service within Gaia-DP is responsible for reading and processing data from various objects, providing a versatile set of APIs for browsing, accessing, and streaming the logical data associated with a given object snapshot. This service is designed to be highly generic and adaptable, enabling it to process and handle a wide array of data types, formats, and structures efficiently and effectively.

Gaia-Indexer in Gaia-DP

The Gaia-Indexer service in Gaia-DP plays a crucial role in managing and coordinating indexing requests, as well as distributing and executing indexing tasks across the cluster. Working in tandem with the Gaia-data-accessor service, the Gaia-Indexer is responsible for locating, streaming, and processing content for indexing purposes, ensuring efficient data processing and accurate indexing results. This service employs a leader-worker architecture, with the leader component overseeing indexing requests and task distribution, while the worker components execute the actual indexing tasks on the data accessed by the Gaia-data-accessor service. This architecture enables the Gaia-Indexer to efficiently manage indexing workloads, optimize resource utilization, and deliver highly accurate indexing results for the Gaia Service.

Gaia Embedding Service (Gaia-ES)

The Gaia Embedding Service (Gaia-ES) serves as a crucial component in the Gaia system, providing a comprehensive abstraction layer for various machine learning (ML) models and the vector database within the platform. Gaia-ES is designed to handle numerous tasks, including text extraction, embedding generation, and document similarity retrieval. Comprising a set of microservices deployed on Kubernetes, Gaia-ES ensures a scalable, secure, and efficient solution for data management.

Text Extraction in Gaia-ES

To effectively extract text from a diverse range of document types, such as PDF, PPT, DOCX, and others, Gaia-ES employs Apache Tika, a powerful open-source text extraction tool. This implementation ensures seamless compatibility with numerous data formats, allowing Cohesity Gaia to process and manage an extensive array of datasets.

In addition to Apache Tika, Gaia-ES may also use in-house built parsers for specific use cases. This hybrid approach allows for greater flexibility and adaptability in handling various data sources and formats.

Vectorization and Storage in Gaia-ES

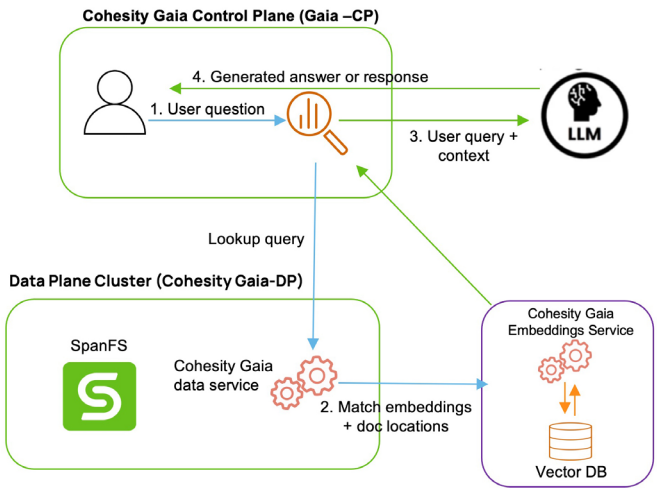
One of the core responsibilities of Gaia-ES is to generate embeddings for text documents and store them in the vector database. This process involves converting high-dimensional text data

into a lower-dimensional space that accurately represents the semantic meaning of the content. As a result, Gaia-ES facilitates efficient clustering and retrieval of similar documents based on their semantic relationships.

To achieve this, Gaia-ES uses state-of-the-art ML models and algorithms to generate embeddings that effectively capture the essence of the text data. By using popular ML models from the Hugging Face library, Gaia-ES ensures high-quality embeddings that yield accurate and meaningful results.

Gaia-ES can also interact with a vector database to store the generated embeddings. Given the emergent nature of vector databases, Gaia-ES implements a robust abstraction layer to manage various vector database collections and is expected to seamlessly interact with different vector stores in future releases. This level of abstraction will allow Cohesity Gaia to handle scalability, performance, or operational constraints that may arise with individual vector databases.

In conclusion, the Gaia Embedding Service (Gaia-ES) serves as a vital component in the Cohesity Gaia platform, providing advanced text extraction, vectorization, and storage capabilities. By employing cutting-edge technologies and sophisticated techniques, Gaia-ES ensures a robust, efficient, and scalable solution for managing and gaining insights from vast quantities of data stored within the Cohesity Data Cloud.



High Level Application Flow

1. User interacts with the Cohesity Gaia application, providing a query or request.
2. The Cohesity Gaia Embeddings Service, together with the Cohesity Gaia data service, retrieves relevant data by matching embeddings and documents based on the user's query.
3. The user's query and context is forwarded to the LLM.
4. The LLM generates an answer or response based on the user's query and context and provides it to the user through the Cohesity Gaia application.

Fig . 6 : Data insights Cohesity Gaia query workflow

COHESITY

Fine-grained RBAC policies

Cohesity Gaia incorporates advanced, fine-grained RBAC policies that precisely govern access to the Gaia APIs. These specialized policies ensure that only authorized users can access and manage the stored data, effectively mitigating the risk of data exposure or unauthorized access. The RBAC system is designed to accommodate various levels of access and privileges, offering granular control over user permissions and actions within the Cohesity Gaia environment.

Encryption and key management

Cohesity Gaia employs state-of-the-art encryption methods, using AES-256 encryption to secure all data, both in-transit and at-rest. The encryption keys are managed by a robust key management system (KMS), which provides secure storage, management, and distribution of cryptographic keys. The KMS is designed to handle key lifecycle events, such as key rotation and key revocation, ensuring that data remains protected even as encryption technologies evolve.

Identity and access management (IAM)

Cohesity Gaia implements a sophisticated IAM framework to manage user access, authentication, and authorization. This comprehensive system includes support for RBAC, single sign-on (SSO), and multifactor authentication (MFA) mechanisms. By leveraging these advanced IAM tools, Cohesity Gaia ensures secure access to stored data while offering seamless integration with enterprise identity management solutions.

Secure communication

Cohesity Gaia prioritizes data security during transit by utilizing modern encryption methods, such as mutual Transport Layer Security (mTLS) and HTTPS. These secure communication channels ensure that data remains confidential and protected while being transmitted between components and services within the Cohesity Gaia system. Additionally, Cohesity Gaia's communication infrastructure is designed to support regular security updates and patches, maintaining a high level of protection against potential vulnerabilities.

COHESITY

Next steps with Cohesity Gaia

Getting started with Cohesity Gaia is simple:

- Cohesity Gaia integrates directly with the Cohesity Cloud Services. The Cohesity Cloud Services is required to use Cohesity Gaia.
- Cohesity offers a full-featured 30-day trial for qualified requests.
- Cohesity Gaia is separately licensed, based on queries and index storage.

If a customer runs out of storage space or queries, they can simply add additional license packages or a la carte storage or queries.

- Cohesity Gaia is compatible with Microsoft M365 mailboxes and OneDrive cloud services. Microsoft M365 mailboxes, OneDrive instances, and/or NAS storage arrays are required to use Cohesity Gaia. We plan to support additional data sources in the future.

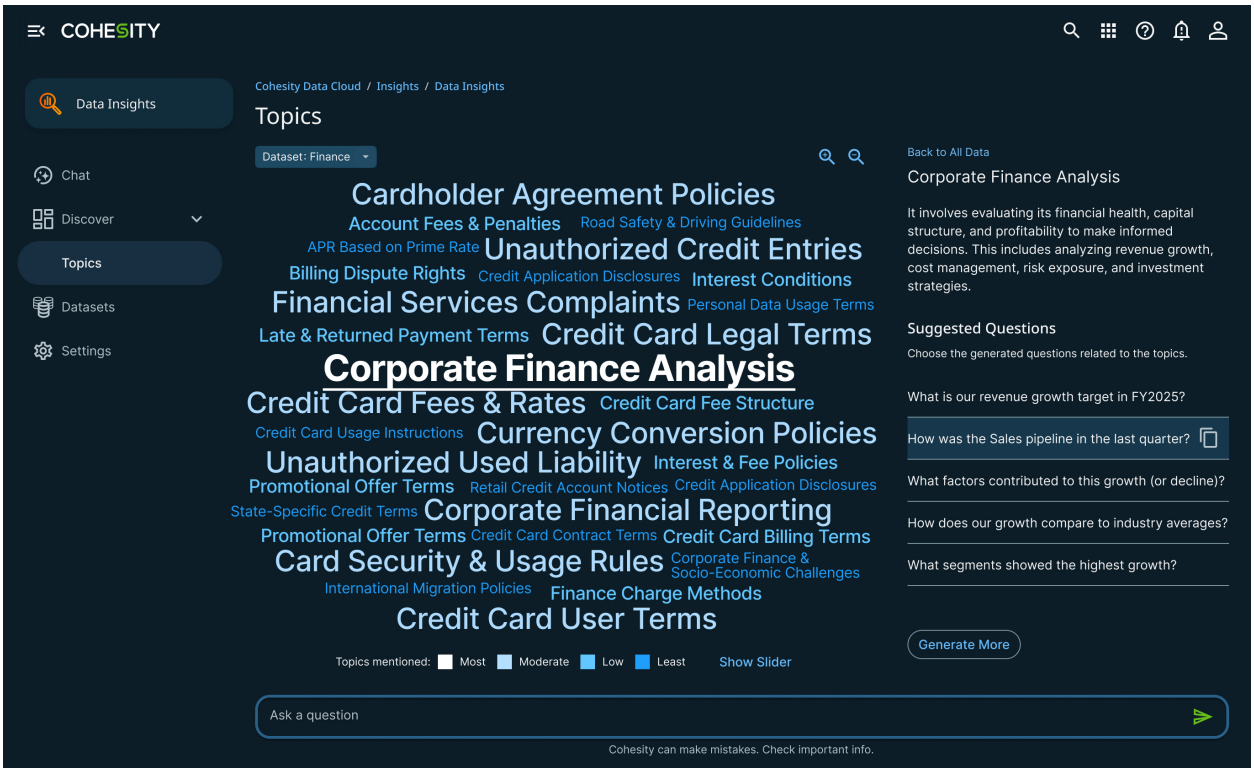


Fig . 7 : Visually explore what's in your data set and get query suggestions automatically

COHESITY

- Simply back up your M365 mailboxes, OneDrives, and/or NAS storage arrays using the Cohesity Data Cloud.
- Select your backup as the data source in Cohesity Gaia. The data is then indexed and vectorized. (See #1 in the illustration.)
- Once the indexing is complete, use Cohesity Gaia's data explorer feature to visually examine what's in your data source and automatically get query suggestions based on what Gaia discovered indexing your data.

Cohesity Gaia can conduct searches on your data and deliver results. Those results can be augmented to provide summaries or other insights from your data.

To create a query, simply ask a question in the Cohesity Gaia chat interface, using natural language. You don't need to create complicated search equations to use Cohesity Gaia.

As Cohesity Gaia delivers the results, continue to ask additional questions to augment the findings and get deeper insight.

Once the data has been indexed and you understand what's your data set, you may begin asking natural language questions in the Cohesity Gaia interface. (See #2 in the illustration.)

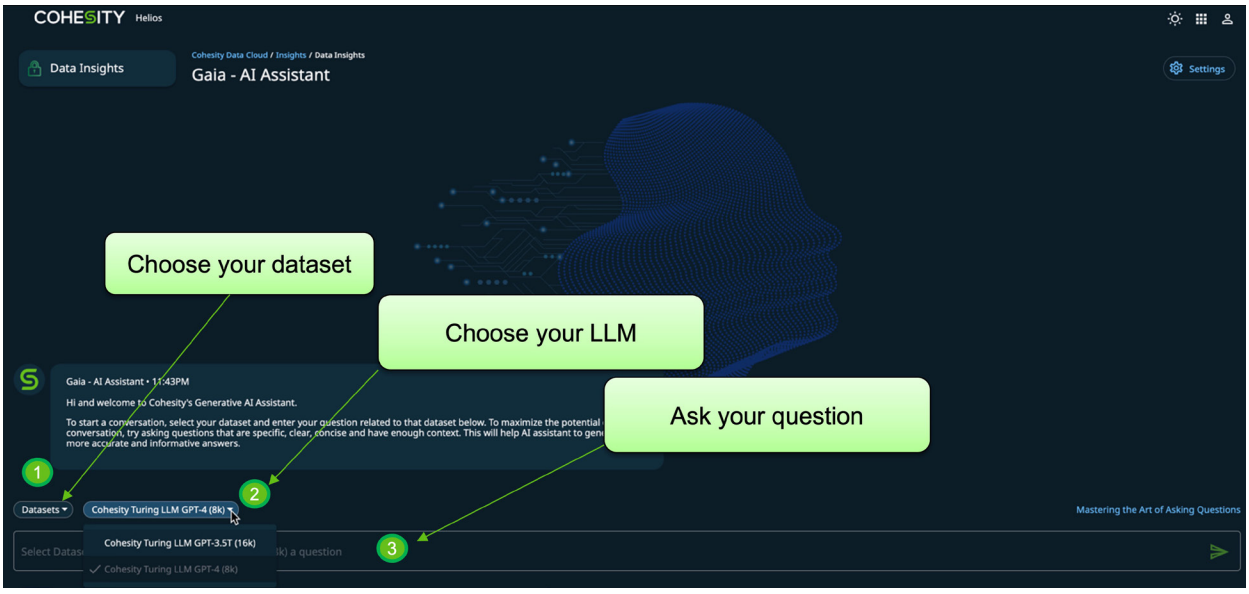


Fig. 8 The simple and easy-to-use Cohesity Gaia interface

COHESITY

Conclusion

The AI revolution has begun. We at Cohesity look forward to being your partner in helping you unlock deep insights, gain new operational efficiencies, and enjoy an ease-of-use not seen before in IT solutions. Check our website regularly.

We'll introduce new compatibilities with different data sources, environments, and clouds—plus new features, functionalities, and additional use cases—for Cohesity Gaia over time.

COHESITY

www.cohesity.com

2000049-002-EN 10-2024