Using Knowledge Graphs and LLMs for Deep Entity Exploration – EDW 2024

Using Knowledge Graphs and LLMs for Deep Entity Exploration Wednesday, March 27, 2024

Gartner and Forrester emphasize the importance of constructing knowledge graphs to connect data silos. By doing so, companies can achieve a comprehensive enterprise data fabric solution that enables deeper analytics but also optimizes AI investments. Usually, too much emphasis is placed on structured data but the reality is that in many enterprises there is even more information and knowledge hidden in unstructured data. So the ultimate goal is to provide nontechnical end users with the additional ability to query across the business knowledge contained in unstructured data, business correspondence, financial files, and contracts.

Recent advancements, like LLMs and vector-enabled knowledge graphs, permit a blend of natural language and structured queries to retrieve data from documents putting the goal one step closer to delivering queries that span your data fabric.

During our demonstration, we will compare and contrast three approaches data architects should consider in developing a knowledge graph-based approach to connecting valuable enterprise and industry data. For demonstration purposes we will use a collection of legal documents that pertain to the financial industry, in our case we take the entire collection of FINRA rules that are publicly available on the web.

1. Standard LLM Interaction: We demonstrate best practices for querying a legal contract via an LLM website. Are the answers sufficient and do they provide the depth and references necessary for users? Could a better answer have been in another document?

2. LLM combined with web search: Combining web search with an LLM greatly improves answer quality for more complex questions and in some cases, rule references are provided. But are the references specific enough to point back into our local documents?

3, Contract Knowledge Graph as the Source of Truth for LLMs: By storing the example FINRA contracts along with vector embeddings in a knowledge graph, we yield accurate answers directly linked to specific rule passages in the documents that provide evidence for the answers. In addition, we show the deep entity connections exposed in the graph as a result of all the cross references.

This presentation will show users how to efficiently link siloed knowledge and query across documents with natural language techniques for richer insights on entities of interest.

PitchFest 2023: Unlocking the Vision of the Financial Data Transparency Act

Franz Inc. has been selected to participate in the Data Foundation's PitchFest, Unlocking the Vision of the Financial Data Transparency Act.

Franz will be presenting "A Visual Time Machine for Influence Networks in a National Legal Entity Knowledge Graph"

Once implemented, the FDTA will improve efficiency and

government operations, reduce compliance burden, and improve data accessibility, searchability, and comparability to the benefit of regulators, investors, companies, and society. The FDTA encourages financial regulators to harmonize data collections and standards on information already collected – moving to machine-readable forms, reduce filers' compliance burden, increase analytical capabilities, and enhance transparency and accountability. The new law amends securities and banking laws to make the information reported to financial regulators including the Securities and Exchange Commission, Consumer Financial Protection Bureau, and Financial Deposit Insurance Commision, among others, electronically searchable and to further enable the development of regulatory technologies and artificial intelligence applications.

Recording of the presentation:

AllegroGraph Named "2023 Best Knowledge Graph" by KMWorld Readers' Choice

Franz Inc., is proud to announce it has been named the "Best Knowledge Graph" in the 2023 KMWorld Readers' Choice Award voting.

According to KMWorld, Technologies such as knowledge graphs, cloud computing and storage, data mesh and data fabric, chatbots, natural language processing, machine learning, and, most recently, generative AI (GenAI) have come to the forefront in our attempts to manage the myriad formats and knowledge silos rampant within organizations. Business practices are changing fast, and so are knowledge management offerings. To put the spotlight on the innovative and dependable products and services that KMWorld readers depend on, the publication presents the KMWorld Readers' Choice Award winners. After all, who best to know what products serve them best as they wrestle with so many changes happening so quickly?

In the November 2023 issue, KMWorld magazine announces the winners of the 2023 KMWorld Readers' Choice Awards. The categories for competition were wide-ranging. In all, there were 13 areas in which products and technologies could be nominated and ultimately voted upon. They include business process management, cognitive computing and AI, customer service and support, e-discovery, knowledge graphs, text analytics, and NLP.

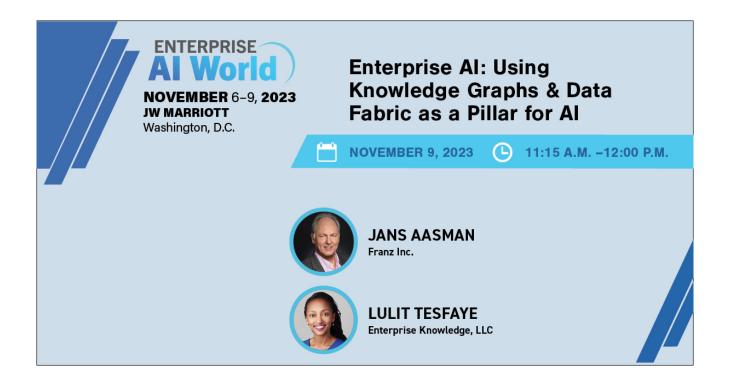
With the diverse array of knowledge management products, services, and technologies to consider, and the stakes getting higher for information-driven success, it can be challenging to make the right choices. There are many ways to learn more about what is available, including white papers, research reports, and webinars, as well as consulting with experts and peers. We hope the KMWorld Readers' Choice Awards list provides an additional resource to help make the job of identifying solutions to investigate easier.

Enterprise AI World – Using

Knowledge Graphs & Data Fabric as a Pillar for AI

Hype or no hype, organizations have seen a significant spike in capabilities around advanced knowledge engineering and AI abilities thanks to advancements in higher computing and abundance of open source solutions such as OpenAI (ChatGPT), and other large language models (LLMs). Data fabrics are emerging as the most effective means of integrating knowledge throughout the enterprise, and experts agree these fabrics are the future of enterprise analytics and AI. However, many organizations continue to face challenges in realizing the promise of AI. For many organizations, one of the top reasons why AI projects get stalled, despite leadership support, is the lack of clear strategies for sourcing the right knowledge and data that AI requires, ultimately resulting in their inability to explain how AI arrived at a certain decision. real-world examples, our speakers share Using their experiences, ideas, and applications of knowledge graphs and semantics in AI, as well as the benefits they offer to organizations seeking to leverage the power of enterprise AI. Join this fast-paced session for lots of tips and ideas.

Dr. Jans Aasman will be presenting at Enterprise AI World.



AI 100 Top Company — Franz Inc.

Franz Inc. is proud to announce it has been named an "AI 100 Top Company."

AllegroGraph provides organizations with essential Knowledge Graph solutions, including Large Language Models (LLMs), Graph Neural Networks, Graph Virtualization, GraphQL, Apache Spark graph analytics, and Kafka streaming graph pipelines. These capabilities exemplify AllegroGraph's leadership in empowering data analytics professionals to derive business value out of Knowledge Graphs.

"Today, AI has the potential to impact almost every part of an organization's structure and operations, including their customer-facing presence," remarked Tom Hogan Jr., publisher of KMWorld. "We see AI reaching into marketing, customer service, legal, finance, human resources, compliance, fleet maintenance, manufacturing, sales, and many other business units."

"Franz Inc. is continually innovating and we are honored to receive this acknowledgement for our efforts to deliver leading AI solutions in Data management," said Dr. Jans Aasman, CEO, Franz Inc. "Organizations across a range of industries are realizing the critical role that Knowledge Graphs play in creating rich, yet flexible AI-driven applications. AllegroGraph with its patented FedShard™ technology uniquely provides companies with the foundational environment for delivering Graph based AI solutions with the ability to continually enrich and contextualize the understanding of data."

Enterprise Data World Conference 2023

Why an Event-Native Mindset is Now Essential for Data Architecture

Dr. Jans Aasman - Tuesday, September 19, 2023

Enterprise Data World Conference

Gartner has noted that it is not enough for businesses today to be "ready to change;" instead, companies need to be "ready to act" in real-time and understand the context of the action. Simply being prepared to respond to change is not sufficient in today's fast-paced and constantly evolving business environment. Organizations need to be able to anticipate and proactively respond to changes in order to stay competitive.

But how can a business create a data architecture that supports "ready to act" applications and systems? Gartner suggests embracing an Event-Native Mindset, and with an Event-Driven Architecture, organizations can create an enterprise nervous system that delivers continuous intelligence and keeps the business always ready.

Applying an Event-Native Mindset to Data Modeling

Consider for a moment that everything that happens within a business environment is an event, and every event impacts an entity or is carried out by an entity. An entity in this context is a core business concept like a customer, patient, or product. Everything a patient does – getting diagnosed, visiting a specialist, being discharged, or receiving a prescription – is an event. Anything that happens to a business's customer, from making purchases, returns, or calling for support, is an event. When products are created, tested, and updated, these activities are also events.

By adopting an Event-Native Mindset in data modeling, organizations can develop more agile and responsive data architectures that are better suited to handle complex and rapidly changing data environments. This approach can help organizations to more quickly identify and respond to changes in data patterns and to more effectively leverage the insights and value that can be derived from event-driven data models.

During this presentation, we will discuss the necessary steps to building your Event-Driven Knowledge Architecture to enable event prediction with machine learning, so your organization can understand the next-best actions and recommendations and prepare for the future. We will cover how Event-based data models also minimize the complexities of processing real-time data, which is why many streaming data platforms and services

DBTA 100 Companies that Matter Most in Data

Franz Inc. is proud to announce it has been named to Database Trends and Applications "100 Companies That Matter Most in Data." .

AllegroGraph provides organizations with essential Knowledge Graph solutions, including Large Language Models (LLMs), Graph Neural Networks, Graph Virtualization, GraphQL, Apache Spark graph analytics, and Kafka streaming graph pipelines. These capabilities exemplify AllegroGraph's leadership in empowering data analytics professionals to derive business value out of Knowledge Graphs.

"The data landscape continues to increase in size and complexity and is more distributed than ever before," said Tom Hogan, Group Publisher, Database Trends and Applications. "Spanning the wide range of established legacy technologies from MultiValue to cutting-edge breakthroughs such as LLMs, the DBTA 100 is a list of hardware, software, and service providers working to enable their customers' data-driven future."

"Franz Inc. is continually innovating and we are honored to receive this acknowledgement for our efforts to deliver leading edge solutions in Data managment," said Dr. Jans Aasman, CEO, Franz Inc. "Organizations across a range of industries are realizing the critical role that Knowledge Graphs play in creating rich, yet flexible Enterprise Data Fabrics and AI-driven applications. AllegroGraph with FedShard uniquely provides companies with the foundational environment for delivering Graph based AI solutions with the ability to continually enrich and contextualize the understanding of data."

AllegroGraph Semantic Layer for Databricks (Delta Lake)

This AllegroGraph tutorial is available on our Github Examples page.

Databricks is a popular choice for hosting lakehouses – a new architecture that unifies data storage, analytics, and AI on one platform. On the other hand, as an enterprise knowledge graph platform, AllegroGraph provides quick semantic layer integration with Databricks transparently through our advanced VKG (virtual knowledge graph) interface.

In this tutorial, we will show you how to load RDF triples directly from your Delta Tables that are hosted in Databricks, and we assume the readers have prior experience with AllegroGraph and our agtool facility.

For users starting with the open-source Delta Lake but not hosted on Databricks, this tutorial may still apply, as long as your platform exposes a JDBC connection and enables SQL as (one of) its query interface.

Requirements

Obviously, you will need to have a running cluster or a SQL warehouse in your Databricks workspace as well as an AllegroGraph server. This tutorial uses a cluster to demonstrate.

Note that Databricks provides trial clusters and one can start from here. If all is successfully set up, the cluster's dashboard should look similar to this:

Đ						
	Franzinc's Demo Cluster 🖗					More ••• Terminate Edit
	Configuration Notebooks (0) Libraries Event log	Spark UI D	river logs N	Metrics Apps	Spark cluster UI - M	aster 🗸
چ)	Policy 🚱					UI <u>JSON</u>
()	Unrestricted					
	Multi node Single node				Summary	/
	Access mode 🕢 Single user access 🕢				2-8 Workers	61-244 GB Memory 8-32 Cores
£	Single user Tianyu Gu (gty@franz.com)				1 Driver	30.5 GB Memory, 4 Cores
≝					Runtime	12.2.x-scala2.12
	Performance				Photon i3.	xlarge 6–18 DBU/h
	Databricks Runtime Version					
	12.2 LTS (includes Apache Spark 3.3.2, Scala 2.12)					
	✓ Use Photon Acceleration					
	Worker type 🚱	Min workers	Max workers	Current		
	i3.xlarge 30.5 GB Memory, 4 Cores	2	8	2		
	Driver type					
	i3.xlarge 30.5 GB Memory, 4 Cores					
	Enable autoscaling Ø					
	Enable autoscaling local storage 🛛					
	✓ Terminate after 120 ♀ minutes of inactivity Ø					
	Instance profile 🕑					
	-					
	None					
	Tags 🚱					
	No custom tags					
	> Automatically added tags					
	Advanced options					

Create a table and load a sample dataset

We use a sample dataset called *people10m* for this tutorial. As documented by Databricks, we can load it into a table by executing this SQL query:

CREATE TABLE default.people10m OPTIONS (PATH 'dbfs:/databricks-datasets/learning-sparkv2/people/people-10m.delta')

After being successfully loaded, you can find the table in the Data Explorer:

Data Explorer		-	+ Add O Franzinc's De 91.5 GB, 12 Cores
\Box hive_metastore \rightarrow \ominus default	default >		
Filter tables	default.people10m		
∰ people10m	⊞Table ▲Delta ≗Not s	et 👻 211MiB, 8 files 🕒 Add co	mment Create
	Columns Sample Data	Details Permissions History	
	Filter columns		
		A	
	Column	Туре	
	id	int	<u>(+)</u>
	firstName	string	\odot
	middleName	string	(\mathbf{f})
	lastName	string	(\mathbf{f})
	gender	string	÷
	birthDate	timestamp	÷
	ssn	string	÷
	salary	int	Ð

as well as a few sample data rows:

Data E	xplorer						+ Add 🛛 📿 Fran	zlnc's De 91.5 GB, 1	2 Cores
defaul defa	t→ ult.people	10m							
⊞ Tal		은 Not set 💉		 Add comm History 	ent			: 0	Create
id	firs	tName n	niddleName	lastName	gender	birthDate		ssn	salary
3766	824 His	ako li	sabella	Malitrott	F	1961-02-12T05:00:00	.000+0000	938-80-1874	5886
3766	825 Dai	sy N	Merissa	Fibben	F	1998-05-19T04:00:00	.000+0000	971-14-3755	6622
3766	826 Car	en E	Blossom	Henner	F	1962-08-06T04:00:00	.000+0000	954-19-8973	5437
3766	827 Dar	leen G	Gertie	Goodinson	F	1980-03-12T05:00:00	.000+0000	981-65-5269	6995
3766	828 Kyle	e L	_u	Habben	F	1974-02-15T04:00:00	.000+0000	936-95-3240	5668
3766	829 Mel	ia k	Kristy	Bonhill	F	1970-09-13T04:00:00	.000+0000	960-91-9232	7399
3766	830 Yev	ette F	Faye	Bebbell	F	1972-09-07T04:00:00	.000+0000	987-72-3701	9288
3766	831 Del	pha K	Kenisha	Gillison	F	1979-06-25T04:00:00	.000+0000	962-66-5404	5120
3766	832 Mik	aela J	Jenifer	Hallan	F	1973-05-23T04:00:00	.000+0000	911-38-3114	9888
3766	833 Cin	di F	Renita	Cousin	F	1979-03-19T05:00:00	.000+0000	666-50-3216	6364
3766	834 Ros	ana S	Sari	Griston	F	1956-10-25T04:00:00	.000+0000	998-70-9023	1925
3766	835 Lee	na V	/anita	Cajkler	F	1971-08-15T04:00:00	.000+0000	954-81-8367	7476
3766	836 Ele	onore S	Stephany	Goldhill	F	1987-12-15T05:00:00	.000+0000	978-64-1528	6794
3766	837 Olg	a J	Jona	Klos	F	1978-05-26T04:00:00	.000+0000	961-27-8570	5248
3766	838 Cha	indra S	Sherrill	Kelsey	F	1962-11-18T05:00:00	.000+0000	910-53-2697	6592
3766	839 Var	ia L	_ashon	Clixby	F	1975-10-31T05:00:00	.000+0000	985-43-1411	7375
3766	840 Dag	ımar N	Misti	Poole	F	1962-05-30T04:00:00	.000+0000	916-94-8252	8517
3766	841 Mai	ta J	Joel	Copelli	F	1953-01-23T05:00:00	.000+0000	929-52-3074	8234
3766	842 Dar	nae D	Darline	Mithun	F	1993-08-07T04:00:00	.000+0000	957-79-3027	6246
3766	843 Fre	dda S	Shawn	Denness	F	1956-06-03T04:00:00	.000+0000	946-20-3144	9299

Υς,

Prepare Databricks JDBC Connection

Now we need to prepare the Databricks JDBC connection details. You may follow these steps to retrieve the JDBC URL, which may look similar to:

1 2 3 4 5 > 20/page ∨

jdbc:databricks://dbc-0bf1f204-2226.cloud.databricks.com:443/d efault;transportMode=http;ssl=1;httpPath=sql/protocolv1/o/3267 754737859861/0405-070225tumf7a9c;AuthMech=3;UID=token;PWD=<personal-access-token>

A personal access token is needed, see here for how to generate one.

Last but not least, we will need to download the Databricks JDBC driver from here. This tutorial uses version 2.6.32. Both the URL and the driver are needed by AllegroGraph's virtual knowledge graph interface, as we will see later.

vload – Load RDF triples from Databricks

The vload facility of agtool is able to load data from relational databases as RDF triples. For a tutorial for vload itself, please refer to this page.

To configure vload, we need 2 files:

demo.properties

This file contains information about the Databricks JDBC connection details as what we have shown in the previous section:

jdbc.url=<your-JDBC-url>
jdbc.driver=com.databricks.client.jdbc.Driver

Note that the downloaded Databricks JDBC driver also needs to be properly installed. See more details here.

demo.mapping.obda

This file defines the rules of how to map the columns from the people10m table between our expected RDF triples. As the target and source sections indicate, we will map id, firstName, lastName, gender, and salary into RDF triples by executing a SQL query.

```
[PrefixDeclaration]
: http://example.org/
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs: http://www.w3.org/2000/01/rdf-schema#
```

owl:http://www.w3.org/2002/07/owl#xsd:http://www.w3.org/2001/XMLSchema#obda:https://w3id.org/obda/vocabulary#

[MappingDeclaration] @collection [[

```
mappingId people10m
target :{id} a :Person ; rdfs:label "{firstName}
{lastName}" ; :gender "{gender}"; :salary "{salary}"^^xsd:int
.
source SELECT * FROM
`hive metastore`.`default`.`people10m` LIMIT 1000
```

]]

By using this mapping, a row of such data:

id	firstName	middleName	lastName	gender	birthDate	ssn	salary
3766824	Hisako	Isabella	Malitrott	F	1961-02-12T05:00:00.000+0000	938-80-1874	58863
will be mapped to these RDF triples (in Turtle syntax):							

```
@prefix : <http://example.org/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
```

:3766824 a	:Person ;
rdfs:label	"Hisako Malitrott" ;
:gender	"F" ;
:salary	"58863"^^xsd:int .

For more details on creating mappings, please refer to this page.

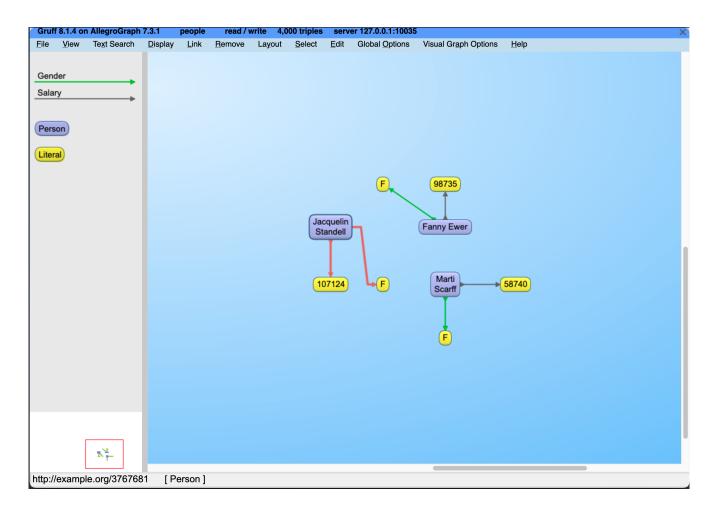
Finally, we can start vloading by running this command:

agtool vload --ontop-home /path/to/ontop --properties /path/to/your/demo.properties --mapping /path/to/your/demo.mapping.obda people

2023-04-12T19:04:13| Creating a temporary workspace

2023-04-12T19:04:13 | Temporary workspace successfully created: "/tmp/agtool-vload-1dfd731d-4862-e844-fde6-0242164d5260/" 2023-04-12T19:04:13 | Mapping file is given, skip bootstrapping 2023-04-12T19:04:13| Starting materialization 2023-04-12T19:04:16 Materialization - OBTAINED FROM SPARK JDBC DRIVER: hive metastore, default 2023-04-12T19:04:18 | Materialization - 19:04:18.398 |-INF0 in i.u.i.o.a.r.impl.QuestQueryProcessor - Ontop has completed the setup and it is ready for guery answering! 2023-04-12T19:04:30 Materialization -WARNING: sun.reflect.Reflection.getCallerClass is not supported. This will impact performance. 2023-04-12T19:04:31 | Materialization - NR of TRIPLES: 1000 2023-04-12T19:04:31 Materialization - Elapsed time to materialize: 13218 {ms} 2023-04-12T19:04:34 | Materialization - NR of TRIPLES: 1000 2023-04-12T19:04:34 | Materialization - Elapsed time to materialize: 2786 {ms} 2023-04-12T19:04:35 | Materialization - NR of TRIPLES: 1000 2023-04-12T19:04:35 | Materialization - Elapsed time to materialize: 1339 {ms} 2023-04-12T19:04:36 | Materialization - NR of TRIPLES: 1000 2023-04-12T19:04:36 Materialization - Elapsed time to materialize: 1210 {ms} 2023-04-12T19:05:36| Materialization successfully exited 2023-04-12T19:05:36| Start loading triples 2023-04-12T19:05:37 | Load finished 4 sources in 78ms (0.08 4,000, Average Rate: Triples added: seconds). 51,282 tps.

It will load the RDF triples into the people repository. You may display a few sample instances through Gruff:



Now let's try to query all the information of 10 top-paid people:

I F

134145 |

I

```
agtool query --output-format table people - <<EOF
PREFIX : <http://example.org/>
SELECT ?person ?name ?gender ?salary {
  ?person a :Person ;
        rdfs:label ?name ;
        :gender ?gender ;
        :salary ?salary .
}
ORDER BY DESC(?salary)
LIMIT 10
E0F
                                  | gender | salary |
  person
           | name
  :3767538 | Shameka Mitcham
                                             135931
                                   F
```

:3767690 | Adelia Salters

Ι	:3767101	Eldora Welbeck	F	134099
	:3767137	Rosalie Challenger	F	129091
	:3767409	Hassie Sides	F	127972
	:3767659	Bridget Inwood	F	126424
	:3767771	Lovie Dorn	F	124903
	:3767631	Latoya Stogill	F	120098
	:3766922	Dot Murkus	F	119509
	:3767736	Ima Adnam	F	119195

Query information:

time : output: 0.001829, overall: 0.045899, parse: 0.000000, plan: 0.020477, query: 0.005075, system: 0.000072, total: 0.027381, user: 0.042787

memory : consCells: 5829080, majorPageFaults: 0, maximumChunk: 5200000, maximumMap: 10131448, minorPageFaults: 2787 other : generation: 2. info: "bindings-set". rowCount:

other : generation: 2, info: "bindings-set", rowCount: 10

Summary

This tutorial has shown AllegroGraph's capability of creating a Semantic Layer for the Databricks lakehouse platform.

Adding a semantic layer, via AllegroGraph, ascribes business meaning to data so end users can better understand their data and associated metadata. A semantic layer provides a number of advantages in terms of Enterprise-wide data management. Users can define business concepts and connections which add meaning to their desired use-case. Some specific advantages of a semantic layer include: improved data integration, enhanced data accessibility, improved data governance, enhanced data quality, and enhanced data security.

Franz Inc. Named to KMWorld's - 100 Companies That Matter in Knowledge Management

Companies have been bought, sold, and merged, and some have disappeared altogether. Others have moved from startups to established leaders. Still more have shown great resilience, changing with the times and remaining as companies that continue to matter in KM.

Technologies that affect knowledge sharing have changed as well, some dramatically. The increasing power of augmented and artificial intelligence, machine learning, natural language processing (NLP), semantic layering, vectorization, knowledge graphs, cloud computing and storage, chatbots, text analytics, and a host of others has revolutionized many aspects of KM.

Some have moved from bright, shiny, and brand-new to mundane, accepted, and pretty much table stakes in the technology game. Newer tools inspire us with their innovative approaches to KM. The impact of generative AI tools, such as ChatGPT, with the ability to draw on an enormous knowledgebase to produce relevant responses to human prompts and have human-like conversations, has jolted not only the KM world but also a plethora of other professions and industries as well.

Putting together the list of 100 companies that matter in KM causes us to look at organizations with pioneering solutions and notable modifications to existing products, and those that are just plain interesting. We applaud innovation, agility, and a focus on the customer. We are excited about the future.

Franz is proud to be named to KMWorld's "100 Companies That Matter in Knowledge Management." Which follows closely on the heals of AllegroGraph being recently named "Best Knowledge Graph" by KMWorld Readers' Choice Award aware voting.

Overlaying of a Knowledge Graph onto a Lakehouse architecture

Dr. Jans Aasman presented to the Estes Park Group.

The Estes Park Group is a monthly online presentation and discussion forum on knowledge graph and data-centric architecture (DCA) trends. SA President Dave McComb first brought the Group together in person in 2017 for a weekend retreat in Estes Park, Colorado, thus the name. This is designed for open discussion among enthusiasts and practitioners, not a vendor sales pitch.