NTT DATA

Designed with the power of **GenAI**

Information discovery acceleration: a moonshot realized

Lessons learned from our work enabling Virgin Media O2 to extract new value from enterprise IT documentation.

Helping telcos see **Eye to AI.**

Pushing the boundaries of GenAI use cases

NTT DATA's **Global GenAI Report**, based on input from over 2,300 senior decision-makers across 34 countries reveals the ambition of business leaders to make strides with GenAI by embedding the technology into their operations to deliver tangible value.

Our approach is to test the business benefits of new use cases while also looking at how we can push the technology further, building on 40 years of global AI innovation. That's why we called our latest work with Virgin Media O2 a GenAI "moonshot".

Like most IT teams, VMO2 had a proportion of technical knowledge embedded in diagrams in their design documents. So the starting point was asking Gen Ai to describe those images. Once this was done, VMO2 could query this knowledge alongside the text of their IT documentation. To achieve everything VMO2 were hoping for, we used this corpus of text to develop a knowledge graph of their IT landscape. Combining this knowledge graph and the text in turn allowed us to generate new design documentation using AI agents.

These cutting-edge techniques combine to create powerful new tools that can tackle complex information discovery challenges In this way, we believe we can help typical organizations to cut 20% of the time and effort they spend on architecture and design.

Extracting structure from unstructured information

The pace of change in GenAI makes the things we achieved just a year ago seem passé. New services from Amazon Web Services (AWS), such as Amazon Bedrock Studio, allow us to achieve use cases like retrieval-augmented generation (RAG) over corporate knowledge in a simple way And the powerful models which are part of Bedrock, such as Anthropic Claude, enable us to address more complex challenges.

One we're particularly excited about at NTT DATA is integrating AWS and the leading graph database from Neo4J, then asking large language models (LLMs) to extract a structure of our choosing from unstructured information.

Although this is not an entirely new paradigm, it has the power to address some key use cases. One we see often in NTT DATA's Data and AI practice is finding enterprise metadata from multiple silos in varying formats.

A new challenge

Virgin Media O2 (VMO2) set us a different challenge: to generate design documentation. While we apply GenAI extensively in later stages of the software delivery lifecycle, the design stage is a different type of problem because architecture and design are inherently less structured than code, test and deploy. This is where we swap application and cloud engineers for data and AI specialists.

Part of the U.K.'s critical national infrastructure, VMO2 is a leading mobile and fixed-line telco provider that's continually innovating with new products and services. Each of these has to be carefully designed to ensure VMO2's operations remain robust.

The challenge for NTT DATA was whether GenAI technologies could help draft new designs based on existing materials, project requirements and established best practice.

We recognized right away that a traditional RAG solution would not work (this was later proven to be true, as other teams gave it a try and failed!).

So, we started with the underlying principles of how enterprise architecture tools work. These are a real asset when it comes to generating design materials, but organizations need to invest properly in their adoption, and this is incredibly rare. These tools are supported by a flexible knowledge graph (even if the data volumes do not always justify an actual graph database underneath).

Eyes on the prize(s)

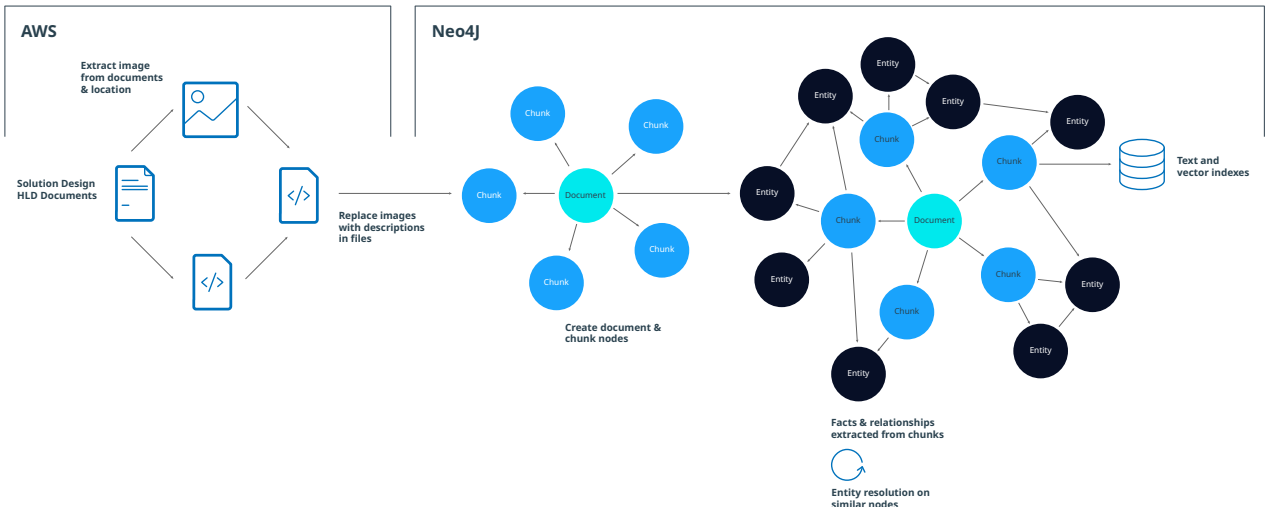
We asked ourselves if GenAI could find the nodes and edges (relationships) that made up the knowledge graph of VMO2's technology landscape. If this (very difficult) problem could be cracked, the other parts of the problem were probably solvable.

We identified other "prizes" in addition to design generation, such as a Q&A bot, so we scoped out a moonshot project aimed at proving all these areas.

We knew that we'd need multiple GenAI agents as well as multimodal approaches, because rich knowledge was contained in architectural diagrams. We selected Neo4J as the graph database both because of our long-standing partnership and because of its proven abilities to store both the graph nodes and edges as well as LLM vector embeddings.

Design details

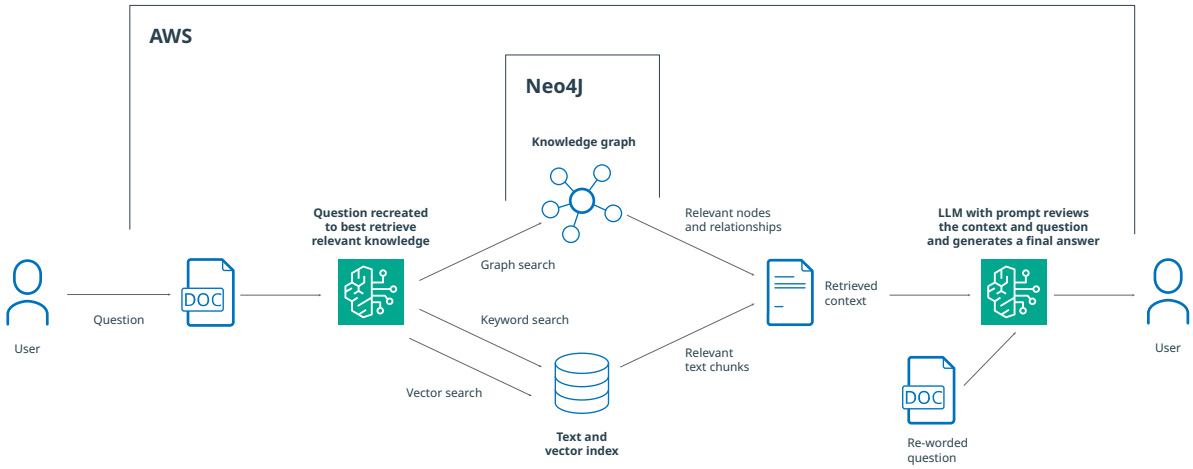
Ingestion technique



Here's how it was done:

1. Using AWS Sagemaker Studio Labs, we first pre-processed the documents, including using the Anthropic Claude Sonnet 3.5 through AWS Bedrock, to create detailed descriptions of any images or diagrams.
2. We then used LangChain to chunk the documents and the LLMGraphBuilder to extract the desired structured ontology.
3. The first nodes ingested into Neo4J are text chunks, which include the vector embeddings of the text. From these, the extracted ontology from each chunk is ingested, creating the structured nodes and relationships that underly VMO2s infrastructure.
4. We then undertook an entity resolution process to resolve any similar nodes with the same meaning to ensure the graph is as connected as possible to give higher accuracy results.
5. Finally, we created the full-text, vector and entity indexes within the graph, which enable the quick and effective retrieval of information.

Q and A retrieval technique



The solution allows for rich Q&A by extracting keywords, vectorizing the question and using the graph to find related nodes with one level of expansion. This information gives a richer set of context information for the final prompt.

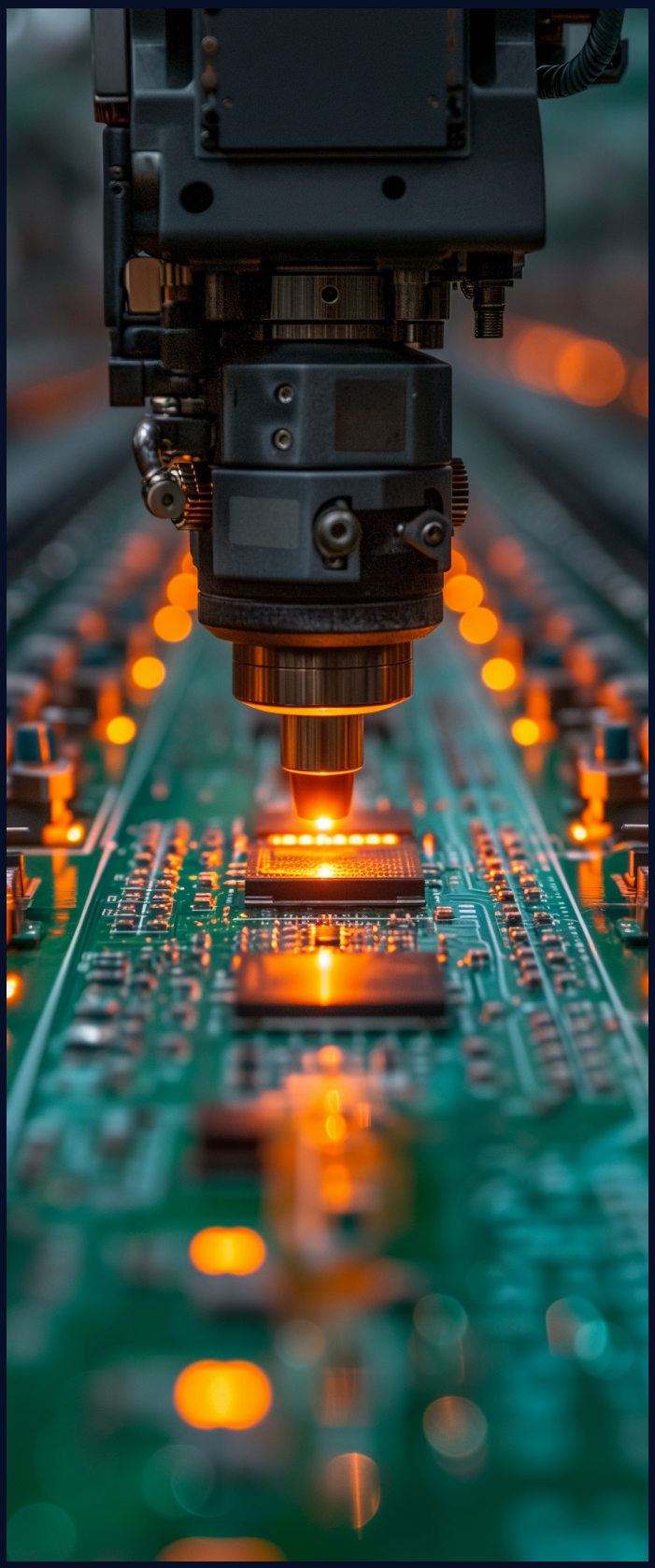
A design document is then created through a multiagent system, where LLMs are targeted to do different tasks, from drafting to reviewing to supervising the system. This process produces each part of the design document in the desired template. This simplifies a process which usually requires the most skilled engineers to perfect.

One important advantage of the design is that it becomes easy to see what the solution "knows". This means it's easy to correct small errors. It's also possible to access precise information where it's needed, particularly for the forward generation of new diagrams. Known graph relationships can be used to tease out further information, going back to the source material in a process we call re-harvesting.

In a nutshell ...

By extracting the knowledge latent in dozens of documents and diagrams into a structured representation, we can put that knowledge to work quickly – saving a great deal of time and effort in producing and refining high-quality design documents.

Execution



We described this as a moonshot project because of the combination of technologies, including agentic and multimodal approaches, that had to work for this approach to succeed with VMO2's corpus of design documentation.

Several features of the graph became immediately apparent:

- There was inconsistent terminology, which is common in many technology landscapes. Entities which are candidates for fusion (typically due to malapropism) are easily identified by lexical similarity combined with shared relationships. For example, the core HR platform had four different names across the documents we were given. Some of the names were similar to each other, and in all four cases, they shared a very similar set of relationships with other nodes (for example, the node for the core SaaS product that the HR platform used).

- Although architects' names were anonymized, key linchpin personnel became obvious. Originally, the plan had been to remove people nodes from the graph entirely, but this finding led to their being included.

Moonshot achieved in weeks!

A design document was produced and was judged to be a great start when compared with a document the solution had not seen. The answers to key questions were judged by experts to be good or excellent.

A key test was comparing RAG alone with GraphRAG.

GraphRAG far surpassed RAG in terms of quality of outputs, the result of a combination of factors.

RAG-only solutions will preselect content that is semantically relevant to the question. However, they will miss the whole context when it comes to questions relating to technical design or even metadata when the target information is distributed across many source documents.

As NTT DATA rolls out this solution, which we call KANO (Knowledge Accelerated by Network Optimization), to more clients, we will discover the range of problems where GraphRAG aces traditional RAG approaches. Kanō means “possibility” in Japanese!

The next steps for VMO2 are to quantify the savings in several areas of the design lifecycle, including knowledge upskilling, finding SMEs, generating diagrams and preparing for governance boards. Our estimate is that they will save at least 20%, based on established benchmarks.

Demonstrating the capabilities of the solution to a new client

In order to showcase KANO, we can demonstrate the same approach but applied to open source architecture documentation such as AWS reference architectures. One downside is that these sources are much better curated than internal design documents. However, the UK Post Office scandal has necessitated the publication of the Horizon system's design documents (Horizon's undisclosed faults led to the prosecution and conviction of more than 900 innocent people). The model has done a good job of rearchitecting Horizon for the cloud, which could be a good next step for the Post Office (especially considering the project spend currently exceeds £2 billion).

NTT DATA, AWS and Neo4J are grateful for the assistance and thought leadership from Sergio Rubio, Head of Architecture at VM02. Commenting on the partnership, Sergio said: "At VMO2, we're harnessing the full potential of generative AI to speed up the launch of new products and services for our customers. As a technology leader, I'm directly involved in making this ambition a reality. This proof of concept has already delivered impressive results, with real promise for transforming our approach to solution architecture. It's driving up productivity, enhancing the quality of our deliverables, and ultimately enabling faster delivery across the board. A huge part of this success comes down to the fantastic collaboration with NTT DATA, alongside Neo4J and AWS. Their team dove deep into our unique challenges and worked closely with us to push the boundaries of innovation. This partnership has been pivotal in showing us what's possible as we move forward.".

Matt Connon VP, Indirect Channels at Neo4J added "Neo4j helps address enterprise challenges requiring advanced reasoning and contextual understanding. By enabling graph-based RAG and integrating with AWS services, we empower organizations like Virgin Media O2 to accelerate solution architecture and design processes while achieving greater explainability and precision. Our partnership with NTT DATA demonstrates how generative AI and knowledge graphs together deliver transformative results, driving productivity and innovation at scale."

Never stop shooting for the moon.

Bill Wilson, Head of Data & AI Solutions
Rebecca Modiano, Data Science & AI Lead
Abhishek Thota, Gen AI Lead Architect
Colin Woods, Solution Director


NTT DATA