

How NASA is using knowledge graphs to find talent



Mark Samuels
@mark_samuels

July 24, 2021 2:20 PM

f t in

Transform 2022

Hear from senior executives at some of the world's leading enterprises about their experience with applied Data & AI and the strategies they've adopted for success.

[Register Here](#)

ADVERTISEMENT



Full Stack Web Developer

Learn employable skills while at home. Only \$399/month.

[Open](#)

Image Credit: Jay Yano // Getty Images

We are excited to bring Transform 2022 back in-person July 19 and virtually July 20 – 28. Join AI and data leaders for insightful talks and exciting networking opportunities. [Register today!](#)

ADVERTISEMENT



Simplified Deployments with OpenVINO™ Server and TensorFlow Serving

Ad by Intel OpenVINO™ toolkit

See More

One of NASA's biggest challenges is identifying where data science skills reside within the organization. Not only is data science a new discipline – it's also a fast-evolving one. Knowledge for each role is constantly shifting due to technological and business demands.

That's where David Meza, acting branch chief of people analytics and senior data scientist at NASA, believes [graph technology](#) can help. His team is building a talent mapping database using [Neo4j](#) technology to build a knowledge graph to show the relationships between people, skills, and projects.



All it takes is ONE hit.

Meza and his team are currently working on the implementation phase of the project. They eventually plan to formalize the end user application and create an interface to help people in NASA search for talent and job opportunities. Meza told VentureBeat more about the project.

VentureBeat: What's the broad aim of this data led project?

David Meza: It's about taking a look at how we can identify the skills, knowledge and abilities, tasks, and technology within an occupation or a work role. How do we translate that to an employee? How do we connect it to their training? And how do we connect that back to projects and programs? All of that work is a relationship issue that can be connected via certain elements that associate all of them together – and that's where the graph comes in.

VentureBeat: Why did you decide to go with Neo4j rather than develop internally?

Meza: I think there was really nothing out there that provided what we were looking for, so that's part of it. The other part of the process is that we have specific information that we're looking for. It's not very general. And so we needed to build something that was more geared towards our concepts, our thoughts, and our needs for very specific things that we do at NASA around spaceflights, operations, and things like that.

VentureBeat: What's the timeline for the introduction of Neo4j?

Meza: We're still in the implementation phase. The first six to eight months was about research and development and making sure we had the right access to the data. Like any other project, that's probably our most difficult task – making sure we have the right access, the right information and thinking about how everything is related. While we were looking at that, we also worked in parallel on other issues: what's the model going to look like, what algorithms are we going to use, and how are we going to train these models? We've got the [data in the graph system](#) now and we're starting to produce a beta phase of an application. This summer through the end of the year, we're looking towards formalizing that application to make it more of an interface that an end user can use.

VentureBeat: What's been the technical process behind the implementation of Neo4j?

Meza: The first part was trying to think about what's going to be our occupational taxonomy. We looked at: "How do we identify an occupation? What is the DNA of an occupation?" And similarly, we looked at that from an employee perspective, from a training perspective, and from a program or project perspective. So simply put, we broke everything down into three different categories for each occupation: a piece of knowledge, a skill, and a task.

VentureBeat: How are you using those

ADVERTISEMENT



ADVERTISEMENT

categories to build a data model?

Meza: If you can start identifying people that have great knowledge in natural language processing, for example, and the skills they need to do a task, then from an occupation standpoint you can say that specific workers need particular skills and abilities. Fortunately, there's a database from the Department of Labor called O*NET, which has details on hundreds of occupations and their elements. Those elements consist of knowledge, skills, abilities, tasks, workforce characteristics, licensing, and education. So that was the basis for our Neo4j graph database. We then did the same thing with training. Within training, you're going to learn a piece of knowledge; to learn that piece of knowledge, you're going to get a skill; and to get that skill, you're going to do exercises or tasks to get proficient in those skills. And it's similar for programs: we can connect back to what knowledge, skills, and tasks a person needs for each project.



VentureBeat: How will you train the model over time?

Meza: We've started looking at NASA-specific competencies and work roles to assign those to employees. Our next phase is to have employees validate and verify that the associated case — around knowledge, skills, abilities, tasks, and technologies — that what we infer based on the model is either correct or incorrect. Then, we'll use that feedback to train the model so it can do a little bit better. That's what we're hoping to do over the next few months.

VentureBeat: What will this approach mean for identifying talent at NASA?

Meza: I think it will give the employees an opportunity to see what's out there that may interest them further their career. If they want to do a career change, for example, they can see where they are in that process. But I also think it will help us align our people better across our organization, and we will help track and maybe predict where we might be losing skills, where we maybe need to modify skills based on the shifting of our programs and the shifting of our mission due to administration changes. So I think it'll make us a little bit more agile and it will be easier to move our workforce.

VentureBeat: Do you have any other best practice lessons for implementing Neo4j?

Meza: I guess the biggest lesson that I've learned over this time is to identify as many data sources that can help you provide some of the information. Start small — you don't need to know everything right away. When I look at knowledge graphs and graph databases, the beauty is that you can add and remove information fairly easily compared to a relational database system, where you have to know the schema upfront. Within a graph database or knowledge graph, you can easily add information as you get it without messing up your schema or your data model. Adding more information just enhances your model. So start small, but think big in terms of what you're trying to do. Look at how you can develop relationships, and try to identify even latent relationships across your graphs based on the information you have about those data sources.



VentureBeat's mission is to be a digital town square for technical decision-makers to gain knowledge about transformative enterprise technology and transact. [Learn more about membership.](#)

Author

Mark Samuels [FOLLOW](#)

Topics

AI [FOLLOW](#) apps [FOLLOW](#) Big Data [FOLLOW](#) Business [FOLLOW](#)
Data [FOLLOW](#) Enterprise [FOLLOW](#)

Conversation [FOLLOW](#)

[LOG IN](#) | [SIGN UP](#)

Start the conversation

newest

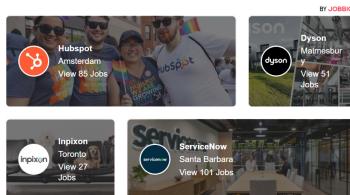
Start the conversation

ACTIVE CONVERSATIONS

Battlefield 2042 finally gets Season 1: Zero Hour on June 9 People are entering relationships with AI...but pop... 2

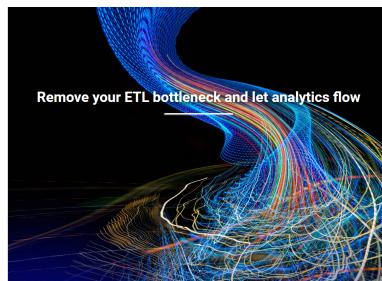
Powered by viafutura

Find your dream job on VentureBeat Careers



Remove your ETL bottleneck and let analytics flow

Microsoft Azure



Remove your ETL bottleneck and let analytics flow

Take business intelligence to the next level by converging analytics and accelerating time to insight

There is no doubt that the digital transformation age is here. Thanks to trends accelerated by the pandemic, every company is now a digital company. The question is which ones are riding this transformational wave to achieve new levels of success and which are lagging behind.

According to TechNative, [70% of companies presently aren't hitting their business objectives](#), and less than a third believe they have a cogent and well-articulated data strategy to get there.

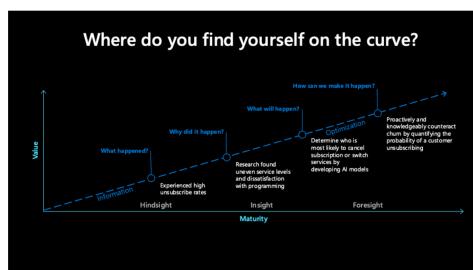
Unless you can use data to generate business intelligence and make mission-critical decisions, you're stuck at the entrance to digital transformation.

It's notoriously difficult to incorporate data into a business' decision-making processes. At the core, the challenge is moving from a descriptive and diagnostic paradigm to a predictive and prescriptive one. This is the difference between storing data – and even having access to it – and being able to use data to fuel your decision-making now, instead of after the fact.

In other words, having data is great, but unless you can use it to generate business intelligence and make mission-critical decisions from it, you're stuck at the entrance to digital transformation. Those gates will eventually close. To avoid being shut out, you need the right analytics platform and component tools to generate and act on business intelligence in near real-time.



To take the next step towards predictive and prescriptive analytics, you have to understand business intelligence on a curve. You can think of the flow of information on a line from hindsight to insight to foresight. Hindsight tells you *what happened*; insight tells you why something happened; and foresight tells you *what will happen* or what you can do to make something happen.



To gain the full breadth of real-time insight, you have to master all three. And you should begin the process by identifying the questions you have and then looking for the right data and analytics that will answer those questions in the way you need.

Business intelligence delivers insights all along this curve, including descriptive analysis that tells you *what happened*. The type of data you need for descriptive analysis includes things like sales numbers, website traffic, and customer feedback. Then, when you add an analytics layer, you're able to use diagnostic analysis to understand *why* those things happened. Understanding the "why" reflects maturity in your business intelligence.

The next step up on the curve is predictive analytics to tell you *what's going to happen* in the future, and then the final step is prescriptive analytics that help you know *what to do* to make that future happen.

Both require machine learning and artificial intelligence in combination with historical data to answer questions about things like demand forecasts, equipment failure rates, and consumer behavior.

As you glean insights closer to real time, you can begin to deliver recommendations to customers just as quickly.





It's easier to understand the need for various types of analytics and insights in context. Say you're a large intelligent manufacturing enterprise that makes hundreds of products, like medical and electrical equipment, under a variety of brands. There are billions of dollars in revenue at stake.

You need real-time insights into how well your IoT devices and other equipment are functioning.

The primary business challenges you need to solve for include ensuring minimum downtime for your machinery, forecasting the raw materials you need, and forecasting customer demand for the products you make.

To meet those challenges, you need real-time insights into how well your IoT devices and other equipment are functioning. If their performance is declining, or they're malfunctioning, that's going to impact not just hardware lifecycle management but also your overall operational efficiency.

Those insights drive predictive maintenance. Connected devices produce data. When you collect, store, and analyze that data, you can drill down to any core problems the machines are having. Then, by applying machine learning, you can have notifications about needed maintenance sent to your team, so they can jump in quickly and fix the problems.

By proactively addressing any issues, you can ensure production quality and eliminate costly delays like unplanned downtime.

By proactively addressing any issues, you can ensure production quality and eliminate costly delays like unplanned downtime. You'll also be increasing the life of your equipment, which saves costs in the long run.

From the supply chain side of the operation, you have to be able to act on data quickly. If operational data like orders, shipments, and transactions isn't in a usable, accessible format rapidly, it's not of much analytical use. Supply chains are typically linear, in the sense that each step in the process happens serially, and that often results in totally reactive responses to disruptions.

But in a networked supply chain in which all parts are connected through a digital platform, issues across any point – be it logistics and transportation, manufacturing, assembly, or product consumers – can be flagged and ameliorated at any time. That lets you understand the complexity of customers' demands and be savvy with planning for resources, leading to less waste, reining in inventory and delivery costs, and still ensuring customer retention.

Still, a digital platform is only useful if clean data can be analyzed to provide insights. For many organizations, there may be a bottleneck in the ETL (extract, transform, load) process that hampers speed. There must be a solution in place to solve that pain point.

The next step up on the curve is predictive analytics to tell you what's going to happen in the future, and then the final step is prescriptive analytics that help you know what to do to make that future happen.

Both require machine learning and artificial intelligence in combination with historical data to answer questions about things like demand forecasts, equipment failure rates, and consumer behavior.

As you glean insights closer to real time, you can begin to deliver recommendations to customers just as quickly.



Managing a large retail organization has become a greater challenge than ever because of how revenue comes from both brick-and-mortar locations and online sales. Marketing has accordingly become more eclectic, comprising not just traditional ad opportunities like TV spots, billboard, and display ads, but also a savvy and engaging social media strategy and email marketing campaign.

In addition to the logistical complication of managing inventory and forecasting demand for in-person and online sales, you have to think hard about how to create personalized experiences for your customer and find opportunities for cross-selling and upselling. You need supply chain optimization and real-time personalization.

A slow ETL process makes adjusting to these insights quickly all but impossible.

To personalize those experiences, you need a 360-degree view of your customers, linking multiple accounts from a variety of systems to understand the customer persona. From there you can be more targeted in your marketing and serve up recommendations that will appeal to your buyers. From those actions, you can apply machine learning to uncover patterns to predict end users' behaviors and monetize them.

For example, perhaps your customers are doing their browsing online but are coming into the store that same day to make the actual purchase. That's a fairly nuanced pattern of behavior that will affect the way you think about the individual customer experience.

In order to optimize the supply chain for an intelligent retail operation, you need to increase efficiency, reduce costs, and ensure the best performance. Start with a complete, unified view of your business that lets you turn operational data like point-of-sale transactions, purchase, and social media sentiment analysis into actionable insights. That will allow you to be smart about capital investments and improve business operations in areas like managing in-store inventory.

As with the prior example, a slow ETL process makes adjusting to these insights quickly all but impossible.

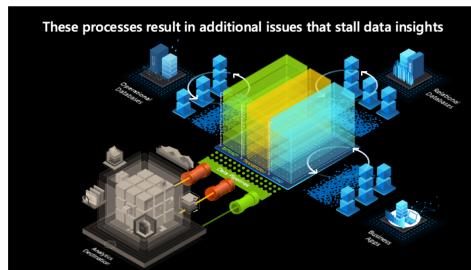


Practically speaking, then, how do you remove those roadblocks and achieve those outcomes?

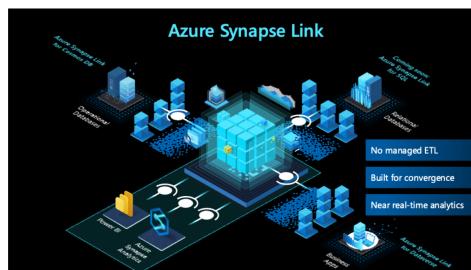
To start, you need a modernized data environment that pulls information from multiple sources and joins systems together to establish a single source of truth. Then, you need processes to prepare that data to be analytics-ready, as well as the database, applications, and analytics to glean actionable data insights.

You have to remove barriers between operational databases and analytics databases, and they need to sync with business applications.

To prepare for analysis, you have to remove barriers between operational databases and analytics databases, and they need to sync with business applications. And you need the pipelines to get the analytics to their destination.



Standing in the way of all of that is an ungainly ETL process. Microsoft's Azure Synapse Link for Azure Cosmos DB, a cloud-native hybrid transactional and analytical processing (HTAP) capability, is the key that unlocks all of it. It obviates the need for the ETL process by creating a seamless integration between the databases and Azure Synapse Analytics.



Microsoft offers products and services for every part of the business intelligence systems, including Cosmos DB, Azure Synapse Analytics, Synapse Spark, and Power BI. Azure Synapse Link stands in the middle of all of it.

There's no need for data integration pipelines, and latency is <90s to deliver insights quickly without impacting performance. Operational data is stored in a transactional store and auto syncs with an analytical store. Through Azure Synapse Link, it hooks into Azure Synapse Analytics – databases like Apache Spark and SQL – and then via machine learning, big data analytics, and BI dashboards, you get all the insights you need.

Setup is simple. Turn on Azure Synapse Link for new Azure Cosmos DB containers and create a Synapse workspace and connect it with Cosmos DB to grab near real-time data analysis.

Among the many component parts of a modernized business intelligence platform, Azure Synapse Link plays a crucial central role. It removes the ETL bottleneck, allowing all parts of the system to flow, and in that way it becomes a catalyst for achieving near real-time business insights.

Learn more about the benefits and use cases of Azure Synapse Link for Azure Cosmos DB.

[Learn more here](#)

Transform 2022

Join AI and data leaders for insightful talks and exciting networking opportunities in-person July 19 and virtually July 20-28.

[Register Now](#)

Join forces with VentureBeat at our upcoming AI & data events

[Sponsor VB Events](#)



f t in

[VB Lab](#)

[Newsletters](#)

[Events](#)

[Special Issue](#)

[Product Comparisons](#)

[Jobs](#)

[About](#)

[Contact](#)

[Careers](#)

[Privacy Policy](#)

[Terms of Service](#)

[Do Not Sell My Personal Information](#)

© 2022 VentureBeat. All rights reserved.

