

**Deloitte.**

# An octopus lives in my data lake

The evolution of more intelligent  
data access, processing, and insights





# About the Deloitte AI Institute

The Deloitte AI Institute helps organizations connect all the different dimensions of the robust, highly dynamic and rapidly evolving AI ecosystem. The AI Institute leads conversations on applied AI innovation across industries, with cutting-edge insights, to promote human-machine collaboration in the “Age of With.” Deloitte AI Institute aims to promote a dialogue and development of artificial intelligence, stimulate innovation, and examine challenges to AI implementation and ways to address them. The AI Institute collaborates with an ecosystem composed of academic research groups, startups, entrepreneurs, innovators, mature AI product leaders, and AI visionaries, to explore key areas of artificial intelligence including risks, policies, ethics, future of work and talent, and applied AI use cases. Combined with Deloitte’s deep knowledge and experience in artificial intelligence applications, the Institute helps make sense of this complex ecosystem, and as a result, delivers impactful perspectives to help organizations succeed by making informed AI decisions.

No matter what stage of the AI journey you’re in; whether you’re a board member or a C-suite leader driving strategy for your organization, or a hands on data scientist bringing an AI strategy to life, the Deloitte AI institute can help you learn more about how enterprises across the world are leveraging AI for a competitive advantage. Visit us at the Deloitte AI Institute for a full body of our work, subscribe to our podcasts and newsletter, and join us at our meetups and live events. Let’s explore the future of AI together.

[www.deloitte.com/us/AIInstitute](http://www.deloitte.com/us/AIInstitute)












# How data warehouses evolved

Octopuses are just one example of the way nature has everything organized beautifully. Fish swim in streams and oceans. Animals graze on land, apples grow on trees, carrots under the ground. We know where to find specific pieces of nature.

If you ask kids today, they think everything from nature also comes out of the refrigerator or freezer. It's warehoused there. When it comes to data, that's exactly what we have done for ages by creating huge data warehouses. We collect data from a variety of sources and put it into a central storage unit, so we can access what we want, when we want it. We keep bringing in fresh produce (data), but often we've bought things we don't need and we're not removing the old stuff fast enough, if at all. **It's not sustainable.**




A server room with blue lighting and binary code overlaid on the floor. The server racks are visible on the right side, and the floor is covered in a pattern of binary code. The lighting is a mix of blue and purple, creating a futuristic atmosphere.

The rationale for this traditional practice was that data stored locally by units throughout the enterprise was unknown or couldn't be processed as is. Hence, it needed to be pulled into one central location or warehouse. Next, rules were created around managing the data and making sure it was all consistent. Technology had to be upgraded to ensure there was enough capacity to handle all the data and accompanying rules. Governance was then established so everyone who should be accessing the data could get to it.

This practice has led to huge demand and investments in storage, software, computing power, networking, etc. over the years. Yet insights from the data underwhelm the investments that have been made. In fact, it's the old 80/20 Rule.

Eighty percent of the effort has gone toward creating these huge data warehouses, and only 20% has been devoted to analyzing insights from the data and generating value from it.





Results emerged at a trickle because most of the focus remained on simply managing the data. It's not a strategic asset if people can't effectively find and pull the information from it in the moment they need it.

With massive amounts of data being created every day and advances in computing power, better ways to store and manage data are emerging to create value in the moment. People get the answers they need to act upon in the moment when it matters, not three weeks later when a report is finally compiled. It's time the data-storage equation is flipped so the majority of time and energy is used for processing the data and using it to drive results. To do that, data access should be intelligent. We should leave the data where it originated and only pull from it what we need, when we need it.

Imagine how fresh the data would be if it was just pulled out of the ocean or plucked from the tree that same day, that same hour.



# The octopus reach

So, instead of a central warehouse, imagine data continues to lie in its native form, at the source where it is generated—a data lake. There are many such data lakes, and there is an octopus that extends its tentacles into each of these lakes, pulling information from there as needed.

That information is pumped up to the main octopus brain and processed there for simple visualization and eventual insight generation.





## What's different from typical data warehousing? **Quite a bit.**

First of all, because the data is not transferred to a central repository, it's left in its natural state with all of its characteristics, as opposed to being forced to fit into a central warehouse where some of the characteristics could get lost or left behind.



Second, the data is modular and free flowing. Parts of it can be extracted dynamically by a central brain in the cloud as needed. The rest can be left behind.



Finally, some of the data can be processed right there in its own lake, at the edge, in a more focused and timely manner. As the main brain in the cloud and the mini brain on the ground share information, they both get smarter, bringing additional insights to light.



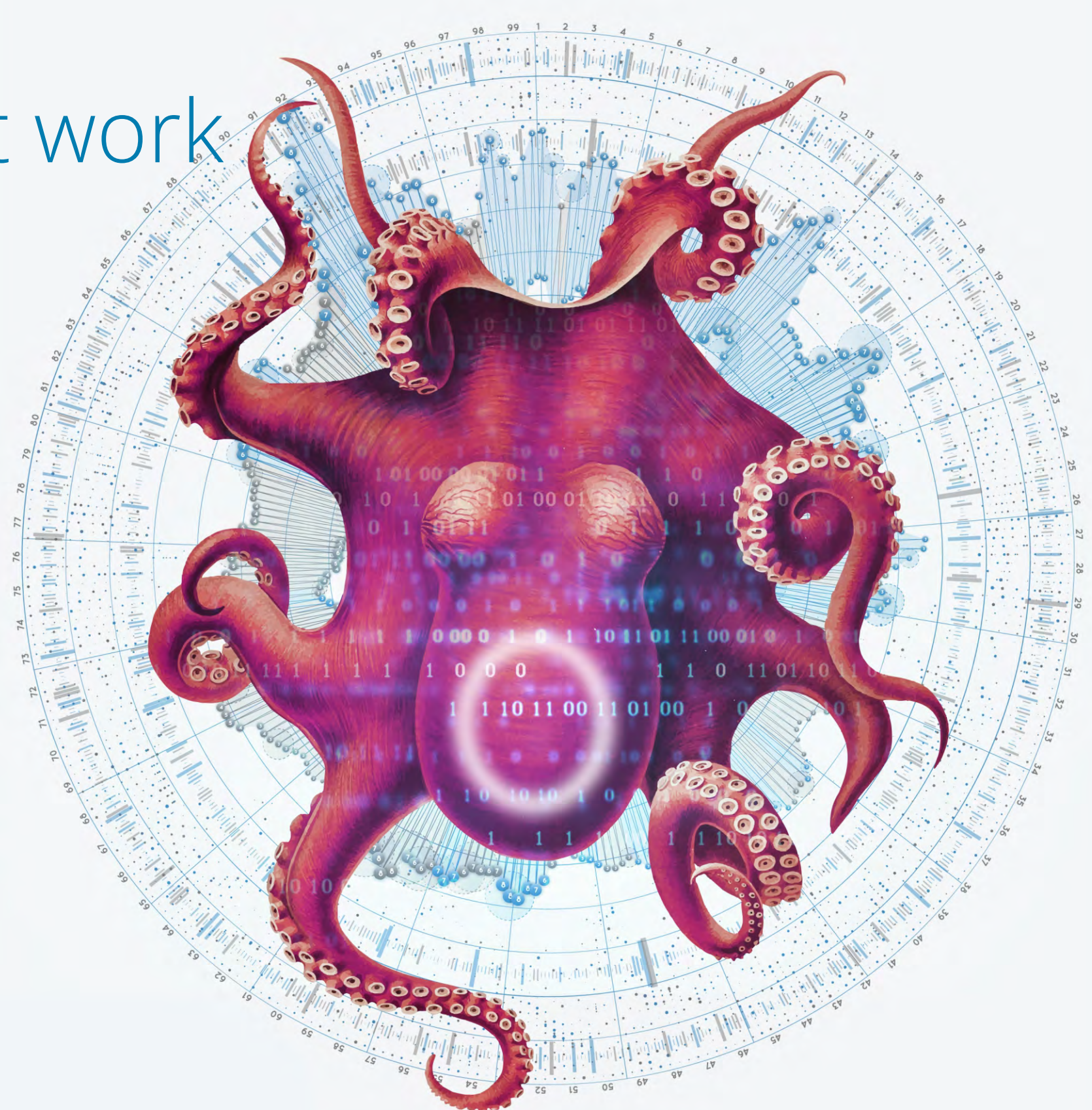
**Here's how  
it works...**



# The octopus brain at work in the cloud

Let's say the octopus is looking for a very specific, rare type of salmon. The octopus brain can extract the data from all over the world, not just from the data in its own domain, to figure out what the salmon looks like and where it's located. It can then use algorithms, or machine learning, to make sense of the data and provide a clear result. How?

The main octopus brain sorts out the three to five specific variables that are most significant to finding that rare type of salmon from the 20 or 200 variables that might be impactful. For example, it quickly determines where similar salmon were sold and disregards tuna sales. As it's looking for information about the salmon, the central brain figures out and grabs just what relates to the salmon and leaves behind information that isn't relevant. As this happens repeatedly, the octopus brain becomes very adept at learning from its experience. It has a better chance of finding the salmon again—faster than it did before.





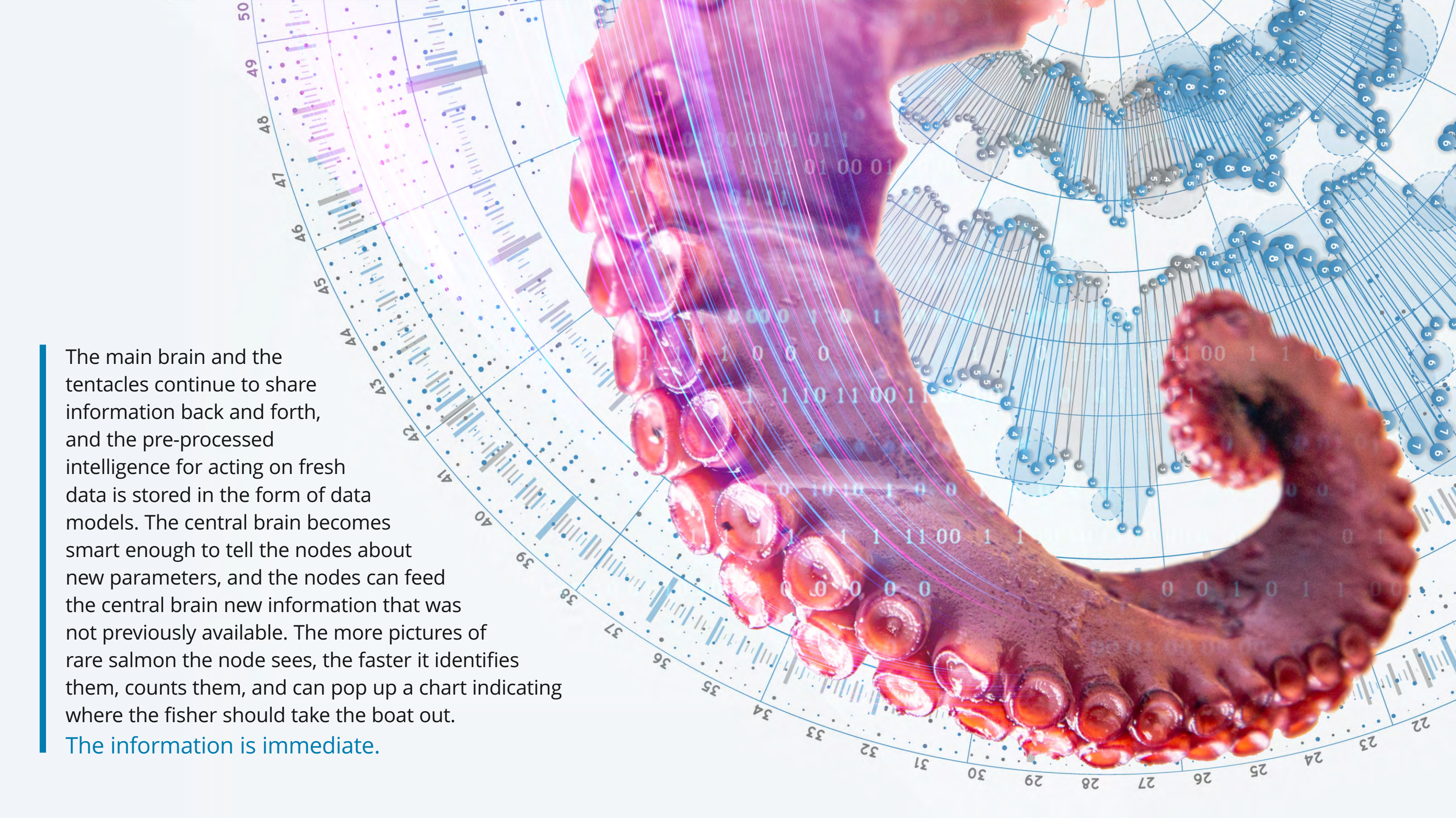
# The tentacle brain at work at the edge

This type of learning is transferred to the mini brain node in the tentacles, and eventually the irrelevant information is not even transferred to the main octopus brain because the tentacle nodes become smart. The data isn't just analyzed in the main brain, but at the node, before the computed results are sent to the main brain. So when the main brain asks for rare salmon, the mini brain at the edge doesn't bother pulling data about tuna. It knows what the big brain is looking for. Edge analytics and computing happens locally where the data originates.

This type of intelligent computing isn't easy. Global companies are just starting to solve for it, but it's fast on the horizon. Imagine this.







The main brain and the tentacles continue to share information back and forth, and the pre-processed intelligence for acting on fresh data is stored in the form of data models. The central brain becomes smart enough to tell the nodes about new parameters, and the nodes can feed the central brain new information that was not previously available. The more pictures of rare salmon the node sees, the faster it identifies them, counts them, and can pop up a chart indicating where the fisher should take the boat out.

**The information is immediate.**



# The result is smart business

In a business context, the right models and visualization are applied to the right kind of ingested data and answers to business questions are delivered at speeds we haven't achieved before and in a high-touch way. Think about watching Netflix as one individual in a family of viewers. The suggestions for what to watch are different for each family member based on previous viewing history. It's personalized.

Or, think of ads for credit card rewards that cater to the individual based on purchase history and their own digital footprints. In both cases, such high-touch personalization has been developed with machine learning and artificial intelligence (AI). Intelligence grows over time with experience, and this intelligence speeds up the processing and interpreting of fresh data. What's starting to evolve is where it should happen.

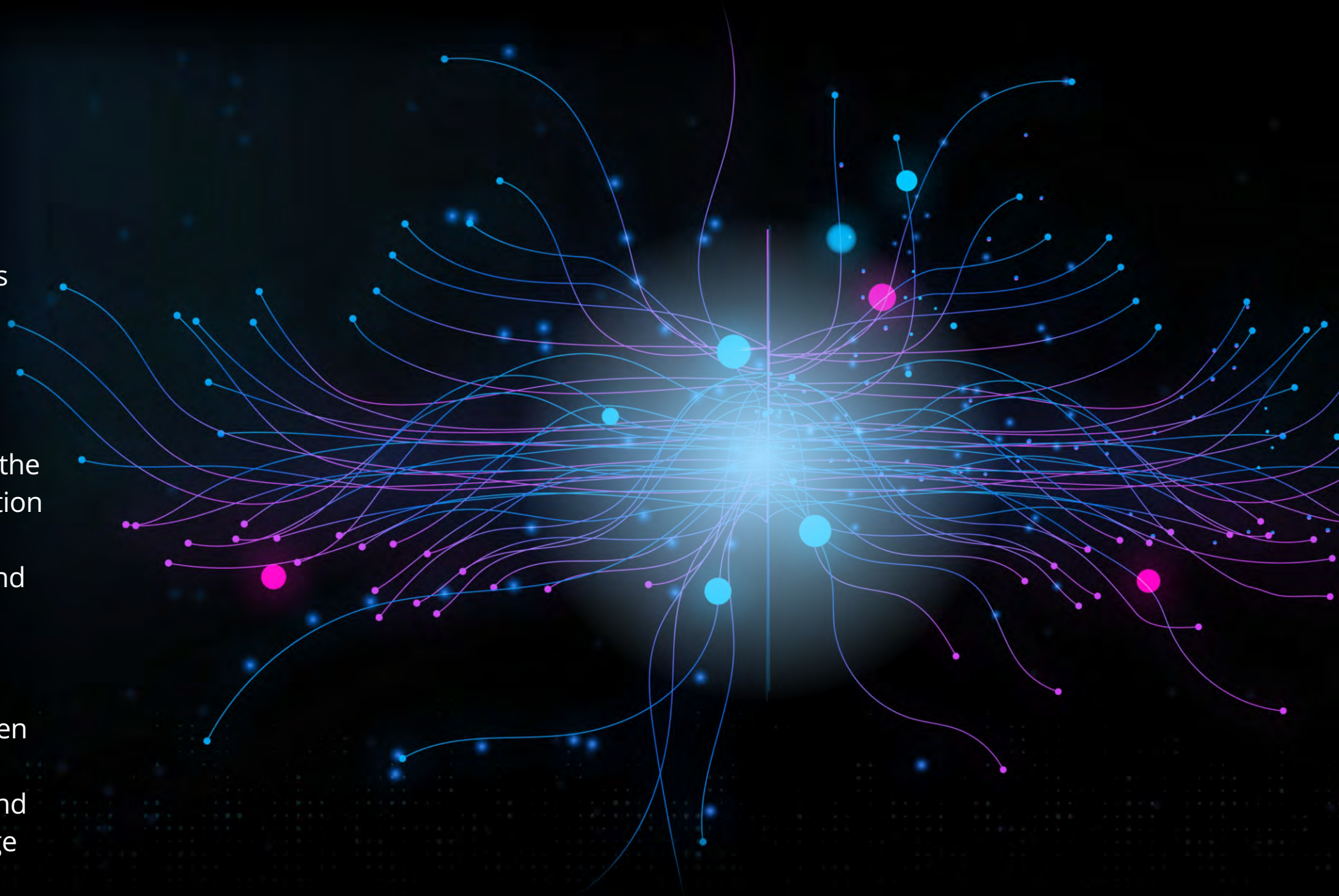




# Where will the octopus surface first?

The mini brain edge-processing technology is already being used in a variety of industries and applications. It will likely surface first for emergency situations or in really painful business cases where a service might fail. It may be best suited toward situations where the node doesn't have time to send the information to the big brain for processing before it's too late—as in the case of self-driving vehicles and wearable health care devices.

When you ask your mobile phone to play a song when you're out of range in a car, it often fails. Eventually that will change. It's because the processing is taking place in the cloud, and not at the device itself. It's a difficult challenge technologically, but let's take it a step beyond that and into the future.





What's being further developed is the octopus' central brain training the tentacles to create the right waves in the lakes—to take action on their own when necessary to prevent a problem from occurring. The mini brain at the edge receives signals from a device collecting data (a personal wearable or dozens of IoT devices across a stadium) that signals a negative event is pending, and the mini brain takes immediate action. In the near future, insulin could be released from a wearable device to a diabetic patient if needed. Or an autonomous vehicle stops if it senses an obstruction in its path, which you've likely already experienced in newer models.

The evolution from data warehouses to data lakes is ongoing. New advancements and business value, even lifesaving discoveries, are being made because two things merged: computing power and massive amounts of data. And we can all be thankful for the intelligent octopuses in the cloud and at the edge that are not only making it possible, but bringing high-touch value to the table.





## Learn more

**Rohit Tandon**

Consulting Managing Director, Deloitte Consulting LLP  
rotandon@deloitte.com

# Deloitte.

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

## About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see [www.deloitte.com/about](http://www.deloitte.com/about) to learn more about our global network of member firms.

Copyright © 2022 Deloitte Development LLC. All rights reserved.

<sup>1</sup> Finances Online, "53 important statistics about how much data is created every day," accessed February 21, 2022.

<sup>2</sup> Lisa Poncet, "Suckers for learning: Why octopuses are so intelligent," The Conversation, July 6, 2021.