



**REDEFINING  
BIG DATA ANALYTICS  
WITH AI**



# Introduction

---

Over the next 5-15 years, AI's impact on the world will ramp up dramatically. Some say we may literally be plugging our brains into the internet. Many of these changes will be positive, and allow us to concentrate on the things that matter most while the machines do all the rest. As for Big Data and analytics, AI is going to help us generate more useful insights, predict things before they happen, and definitely save us time (and time is something that technology won't be able to bring back—at least over the next decade or two...).

But first, let's roll back the clock and talk about the technologies that the industry is working on today that are going to make Big Data analytics with AI possible.

## Big Data Analytics and AI Today

---

Despite massive investments in Big Data technologies, querying Big Data sources is still a technical challenge. We've broken down the technologies that exist today into four, high-level examples, including their drawbacks:

### **Expensive High-Performance Data Systems**

Multiple vendors offer large, expensive proprietary computing systems which attempt to tackle the problem of latency when working with larger and larger datasets. The cost of such systems is extremely prohibitive to smaller organizations or anything other than the most financially important projects.

### **MapReduce & Parallel Processing**

With the introduction of Hadoop and the evolution of the adjacent technologies, Big Data analytics became more accessible. These systems employ a method of breaking up Big Data requests across multiple parallel computing systems using a programming model called MapReduce. This method divides up the request into many smaller parts using a cluster of computing units and then brings those back together into a final answer. However, even these clusters of commodity computing units can add considerable expense and complexity to an implementation.

### **Query Approximation Systems**

One novel method developed in the last few years uses statistical sampling of the data to simulate the likely outcomes of a query without actually running one, sacrificing precision for speed. The system samples the underlying data and bases the resulting answer on the sample rather than the full set of underlying data. The answers are not exactly the same as the results that you would have received if you had queried all of the data. In most cases, though, they're close enough for decision making.

### **Data Summaries**

Data Summaries pre-calculate various attributes of the data (e.g., min, max) and store this information for fast retrieval. Of course, usage of data summaries is limited to those points that were pre-calculated and stored in the summary, giving them limited applicability to modern interactive reporting, where users can choose from any number of permutations of data selection.

These existing solutions fail to offer a high level of interactivity with Big Data due to one basic design concept—they all require continued access to the underlying data.

# What Does the Future Hold?

---

More and more data, that's for sure. In fact, IDC's paper, [Data Age 2025](#), predicts the world will be creating:

**163**  
**ZETTABYTES**  
**(ONE TRILLION GIGABYTES)**  
**OF DATA A YEAR BY 2025**

With so much data being produced, we are going to have to decide what to store and what to discard. We can't just keep building larger repositories to store this data; we will never keep up, and let's face it, not all of that data is relevant.

The same IDC report mentions that 60% of this stored data will be managed by businesses. This can be seen in embedded systems, the Internet of Things, and mobility, all enterprise initiatives.

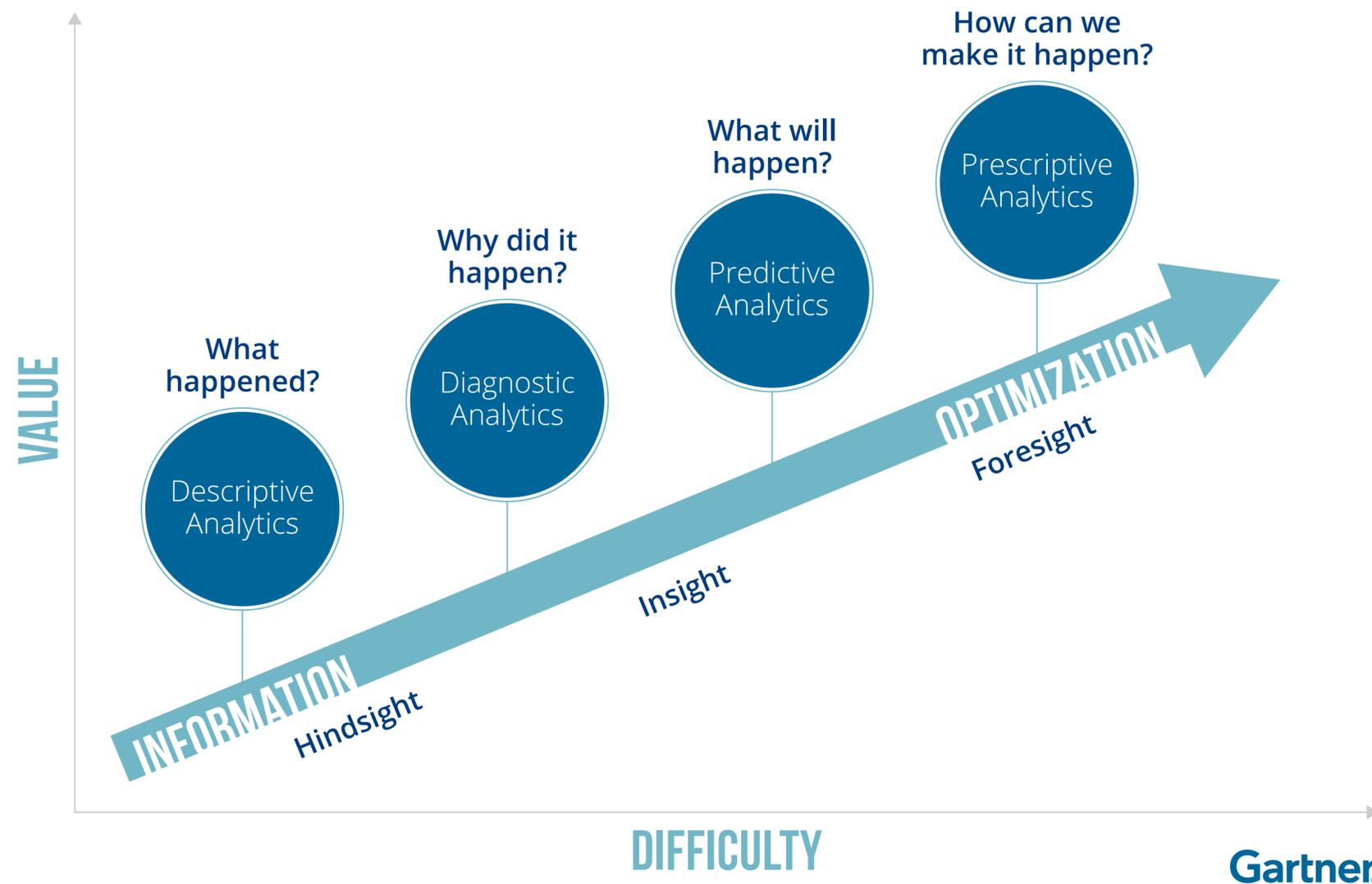
## **Enter AI and machine learning.**

Businesses will have to use AI and machine learning to extract the true business value from the data they collect every minute. This will generate a better human experience and more personal value.

This is the chance for enterprises to step up to the data plate and start hitting home runs with data collection, utilization, and management.

Definition: Advanced Analytics

# Analytics Value Escalator



Advanced analytics refers to *future-oriented* analysis that will:

- Discover deeper insights
- Predict future outcomes
- Generate recommendations to improve business practices
- Drive changes

On-demand webinar: "What Advanced Analytics will Look Like"

# What Needs to Change

---

It's time to start thinking differently about the way we deal with Big Data analysis.

First, we are going to have to make Big Data interactive. In other words, take this huge mass of data and start working with it to get all the information we need. And fast. You don't want to wait weeks, or even days, to get answers. And with all that data piling up, we need to find a way to sift through it quickly and efficiently.

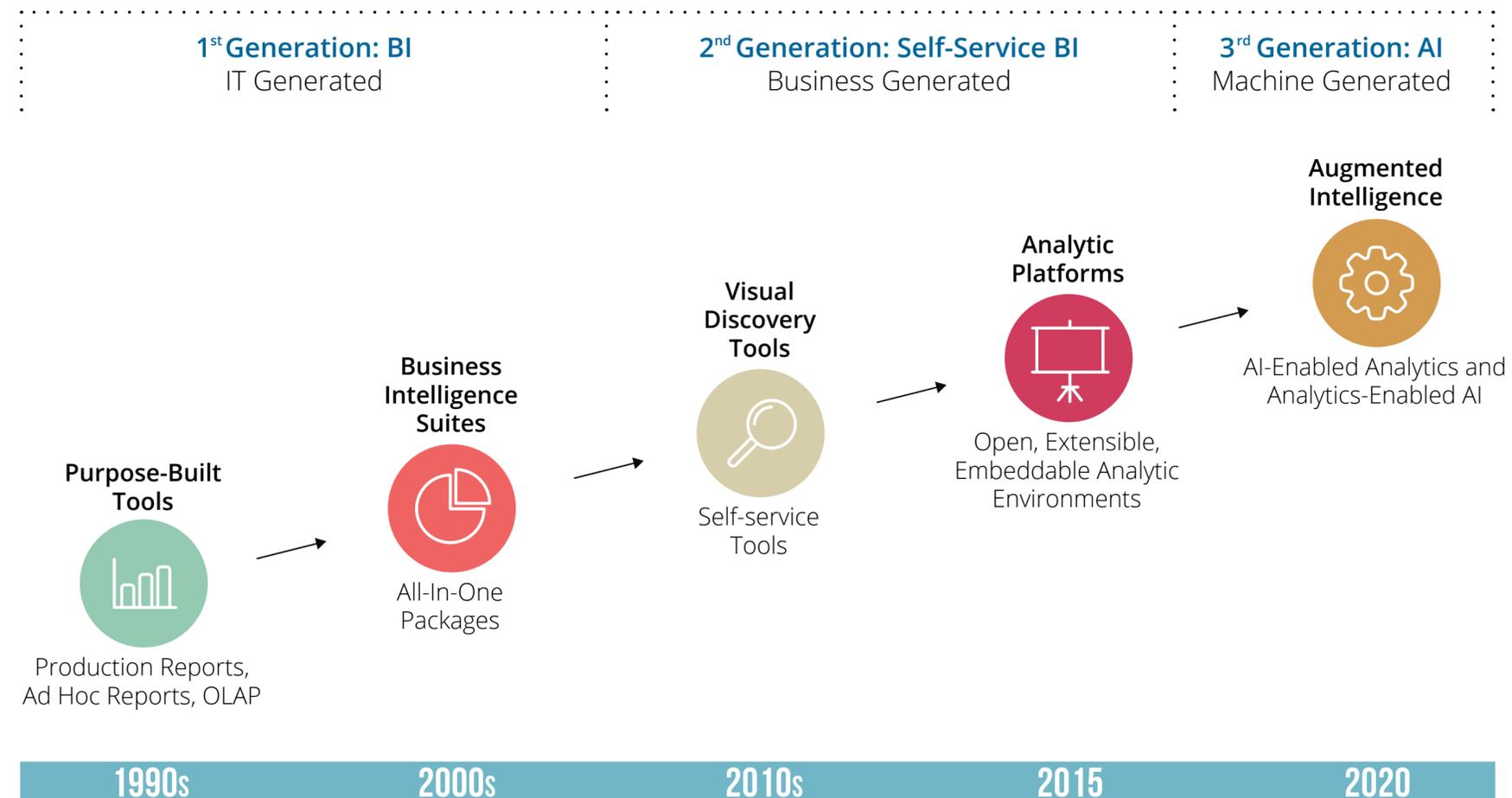
Then, we are going to need to get this data to the edge. What does that mean? That means we need to put that data where it matters, on the smart devices, closer to where the data is created. One of the biggest wastes of time and money is moving the data back and forth. For instance, data from your health device needs to be sent to a storage area, synced with other data, stored, processed, analyzed, and then actions are taken. If we can work with the data at the source, then we are already saving ourselves the storage, processing power, and bandwidth used in the complicated process today.

Data privacy needs to be kept top-of-mind. With more and more information out there, there is a need to rethink how we show and store that private information. Take [GDPR](#) for instance: here are strict new rules that require businesses to protect the personal data and privacy of EU citizens.

# AI is Using Advanced Technologies Today

There are many advances already in AI technologies. And like all the other trends before us, we can see these as true steps in the right direction to doing Big Data analytics efficiently. As [Wayne Eckerson](#) puts it, we've entered the third stage of analytics with machine-generated intelligence.

Figure 1: The Evolution of Business Intelligence



Source: Wayne Eckerson, "The Impact of AI on Analytics: Machine-Generated Intelligence"

Let's take a look at two of the technologies that are driving AI adoption today:

### NLP

Natural Language Processing (NLP), otherwise known as computational linguistics, is the combination of machine learning, AI, and linguistics that allows us to talk to machines as if they were human.

Search giants like Google and Bing use natural language understanding to allow users to enter their searches in whatever language they are comfortable in. By breaking down search requests into simpler terms and understanding linguistic and structural cues, AI can return better and more relevant results.

In business intelligence and advanced analytics, natural language understanding allows for easier searches and more actionable results. Instead of having to create complex algorithms to find data in warehouses, individual users can quickly find what they need in a familiar way, and without the assistance of costly specialists. This is one technology that is advancing in the direction of making it easier and easier for you to troll (or talk) through all that stored Big Data. The technology is also seen in chatbots used by companies like Facebook on their messenger platform.

### Bots

Bots are going to play a significant role in the future of analytics and Big Data, as they can generate analytics responses quickly. Bots can ask relevant questions proactively to further add to the knowledge base, which can then be used to deliver important insights. Bots can also extract data from various data systems in order to analyze and generate key insights.

Putting both of these AI technologies together will enable employees to "chat" with the bot and obtain relevant information regarding key metrics. A conversational interface using NLP will assist with relevant information for faster and more efficient decision making. This means a bot can answer questions whenever and wherever you need information, searching vast amounts of data to come up with a quick answer.



Infographic: "The Relationship between Business Professionals and KPIs"

# Why We Need to Get to the Edge

---

With so much data being generated on the internet, cloud computing may become overloaded sooner than we think—especially with our need for speed and our expectations to receive information quickly and effortlessly.

Edge computing makes faster data processing possible. This can mean the difference between life and death in some emergency use cases, but in this paper, we will refer to business cases for BI and analysis (making businesses more efficient and effective with their decision making).

When our traditional cloud infrastructure starts to fall short, edge computing will come in and provide the solution, and in some cases, may even make it more efficient.

According to [Markets and Markets research](#), edge computing market size is expected to grow from \$1.47B in 2017 to \$6.72B by 2022.

Edge computing allows data to be processed closer to where it's created (i.e., sensors, machine engines, healthcare devices, wearables, and smartphones). This reduces the need to transfer data back and forth between the cloud.

For example, a manufacturing company may put sensors in their machinery that can immediately provide the status of its engines. In this scenario, sensor data does not need to travel to a data center (on the cloud or in a data system) to see whether something is impacting operations or not.

Take our recent example of GDPR above that seeks to protect individuals' personal identifiable information from data abuse. With edge devices, data is collected at the local level, on the sensors themselves. And because this data can be acted upon right on the sensor, there is no need to transfer the data to the cloud. Sensitive information stays on the sensor, avoiding the risk of being hacked or leaked over the world wide web.

# What Needs to be Done to Get There

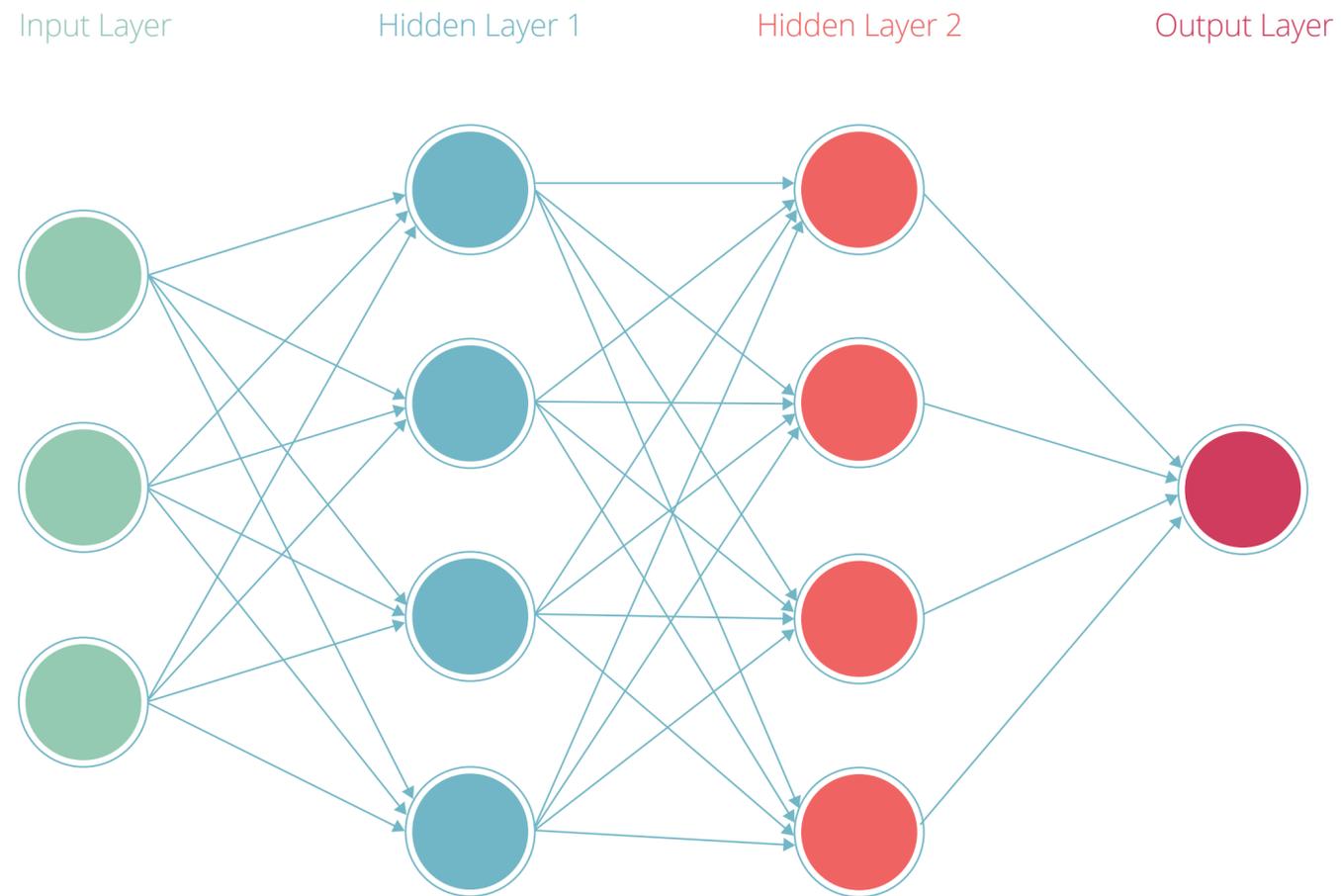
---

Querying these huge datasets with the near-instantaneous response times needed for interactive analysis is a huge technological hurdle. Other analytic solutions have focused on finding ways to process Big Data queries faster by just putting in more expensive computational "horsepower" or via architectural manipulation.

But what if 100% accuracy isn't needed for every decision?

Supposing that the accuracy level was high enough, we could enable a rapid and scalable query processing engine and an efficient data analytics platform. Let's go back a few years to 2015.

Most say that the year 2015 was a breakthrough year for AI. Computer started to “open their eyes,” and today, [Deep Neural Networks](#) (DNNs) are part of many modern AI applications. These DNNs are currently running many applications from self-driving cars and detecting cancer to complex games. The superior performance of DNNs comes from its ability to extract high-level features from raw sensory data after using statistical learning over a large amount of data to obtain an effective result.



*Multi-layer fully-connected neural nets that consist of an input layer, multiple hidden layers, and an output layer. Every node in one layer is connected to every other node in the next layer. We make the network deeper by increasing the number of hidden layers.*

Source: Medium, "Applied Deep Learning - Part 1: Artificial Neural Networks"

The advantage is clear. Instead of the lengthy hit-or-miss approach of creating a custom program to solve each individual problem, the algorithm simply needs to learn, via a process called “training,” to handle each new problem.

For Big Data analytics, we will be able to analyze larger and larger datasets, and still get quick answers that we can count on. This is what’s going to make the difference in the future, drastically reducing the cost and complexity of Big Data analytics projects.

## Summary

We can expect many changes in the way we handle Big Data, edge computing, and analysis over the next decade as devices become smarter, technologies become smaller, and the needs and expectations of the consumers continue to grow.

If we continue to think outside of the traditional storage of data, we can decouple ourselves from the underlying data, and free ourselves up to analysis at the edge. And once we figure out how to analyze all this Big Data with AI and push it to the edge, we will all become junior data scientists, without even lifting a finger.

# Sisense Redefines Big Analytics with Sisense Hunch™

Sisense Hunch™ is a radically different approach to Big Data analytics. In fact, Sisense Hunch is not just a new capability; it is an entirely new class of analytics that helps organizations solve data and analytics challenges that were previously insoluble.

Instead of investing huge amounts of effort and technology in an attempt to process Big Data queries faster, Sisense Hunch condenses immense amounts of data into lightweight neural networks for quick and easy analysis that can be processed at the edge.

With Sisense Hunch, organizations can leverage insights from petabytes of data; processing hundreds of thousands of queries with sub-second response time using the latest AI technologies without the need for access to the underlying Big Data.

See how we're helping our customers tackle their Big Data challenges.

[Watch a demo here](#)



## SPEED

Sisense Hunch delivers sub-second response times ensuring that both individuals and automated processes can get the answers they need immediately. Coupled with traditional methods, users only need to query the original dataset when precision is needed.



## SECURITY

Since Sisense Hunch doesn't retain any knowledge about the lowest level of detail in your data, there is zero risk that queries could return sensitive data that is prohibited by your policy or regulatory compliance requirements.



## STORAGE

Go ahead and drain your data swamp. In cases where the row-level detail is not needed or can not be stored (such as with GDPR), Sisense Hunch can completely replace big data repositories with a nearly accurate solution that will satisfy most analytic needs.



## SAVINGS

Once in place, Sisense Hunch requires no access to the underlying data, putting the power of your data anywhere you need it all while eliminating the need for expensive storage, processing power, and bandwidth.