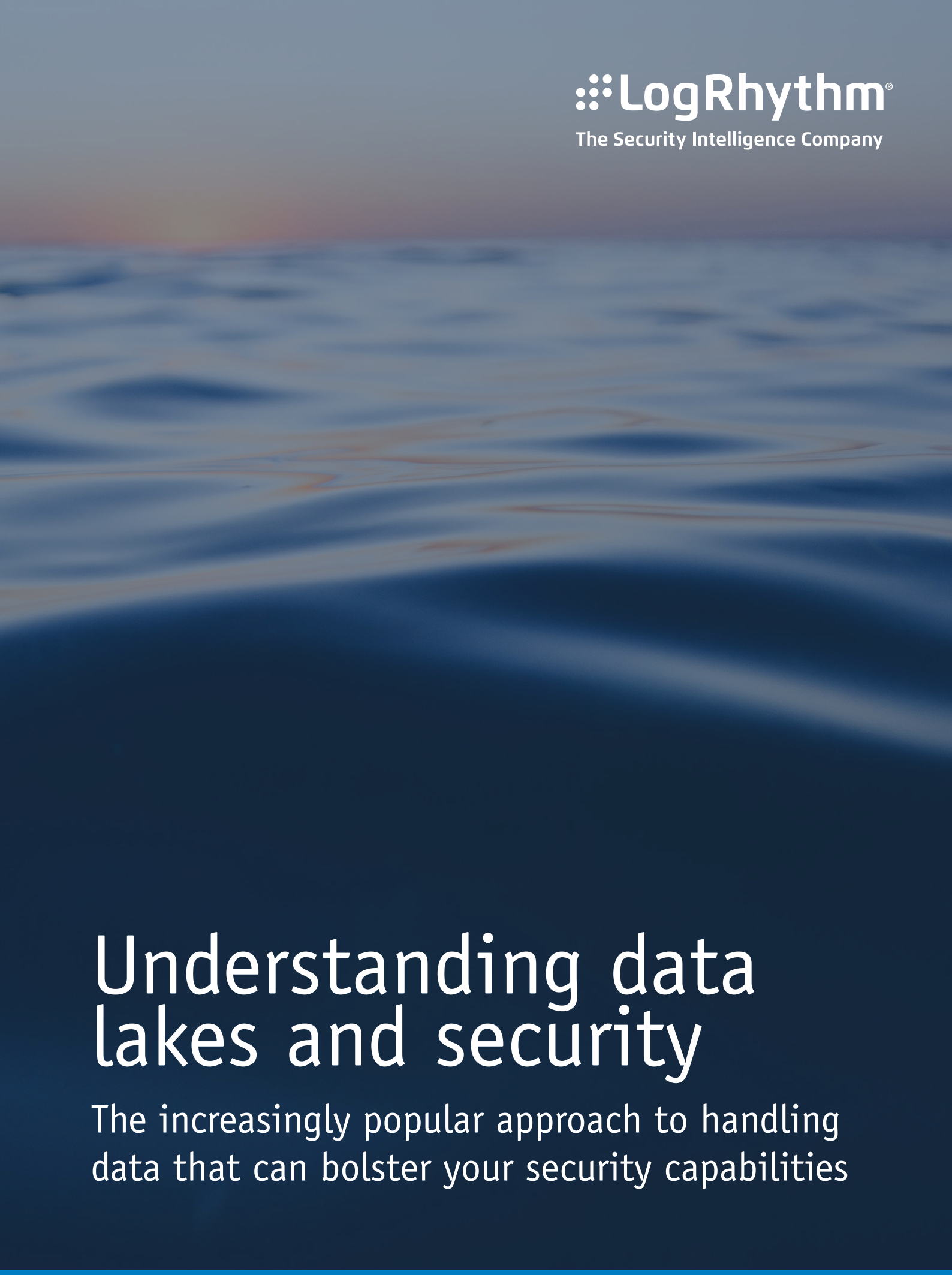


# Understanding data lakes and security

## The increasingly popular approach to handling data that can bolster your security capabilities

## The more flexible and scalable way in which data lakes store and access data is well suited to tackling today's cyberthreats

The cyberthreat landscape has never been more complex and fast moving. This means any new technology that can be used in the fight against cybercrime can bring real value to the security operations centre (SOC).

Given how organisations across numerous industries are increasingly using big data and analytics to extract the maximum value from their information, it makes sense that the same techniques could be applied to security data, which is becoming increasingly plentiful.

A relatively recent addition to the big data armoury is the concept of data lakes, in which a vast range of raw data is pulled together and stored in its original format. Thanks to a flat architecture, data lakes allow analytics tools to work across data that may not have been associated before, generating new insights for businesses and security teams.

Data lakes can also be quickly scaled to fit requirements, meaning they can accommodate a rapid influx of data from a corporate network if, for example, a cyberattack has breached perimeter defences.

These features make data lakes a powerful tool for corporate security teams, particularly with the high likelihood that corporate networks will be compromised.

Given how organisations across numerous industries are increasingly using big data and analytics to extract the maximum value from their information, it makes sense that the same techniques could be applied to security data.



01628 918330 | europe@logrhythm.com | www.logrhythm.com

## Comprehensive threat detection

The sheer variety and complexity of data being produced by compromises and active cyberattacks means the ability for data lakes to deal with a huge range of data types is also key.

Network monitoring and forensics tools are becoming increasingly important for protecting organisations as perimeter defences become more susceptible to breaches.

The ability of data lakes to enable data of all types to be interrogated by forensics and log management tools is therefore becoming more and more relevant. In turn, this means security information and event management (SIEM) systems have access to a broader range of data.

Data lakes enable analytics tools to work across data that would previously have been held in different locations. This means patterns of activity in seemingly unrelated data, but which is the result of malicious intent, can be linked together to detect threats that may have been missed before.

The scalability of data lakes means searchable data generated over a long period of time can be stored and accessed to help analysts shed light on activity that has taken place months or even years before reaching a point where it becomes a threat.

This ability to make historical data more accessible also means analysts can look back at security compromises that have arisen before, giving them guidance on how best to deal with similar issues that arise in the future.

Data lakes allow analytics tools to work across data that may not have been associated before, generating new insights for businesses and security teams.

## Wider benefits

Data lakes can also make security operations more efficient, and even bring benefits to other parts of the business. They avoid duplication of security data generated by different security products by collecting it once and making it available to all tools that need it. This means security processes can be made quicker and more efficient.

A well-structured data lake can also be accessed by other parts of an organisation, such as IT operations, meaning they can benefit from the collection of data for other purposes, such as performance and health monitoring.

Having a data lake comprised of rich log data and network sessions also enables threat hunters to more efficiently hunt for threats.

LogRhythm
The Security Intelligence Company

## Tackling data lake challenges

A couple of areas need close attention when working with data lakes. The first is the preparation of data. According to research conducted for a 2014 New York Times article, data scientists can spend up to 80 per cent of their time preparing data for analysis[1].

There are ways to make this process more efficient, however. LogRhythm's Machine Data Intelligence (MDI) Fabric, for instance, allows the LogRhythm platform to interpret data from virtually any device or technology partner. It also includes a large set of metadata fields to provide essential context, such as event criticality, impacted host and origin host.

The other major issue is the 'garbage-in, garbage-out' problem, which relates to the importing of terabytes of unstructured data into the data lake, then taking it out, making sense of it and deriving value from it.

The MDI Fabric can also help here, as it helps populate the platform's data lake, or that of a customer, with well-structured data that can be more easily made sense of.

## What LogRhythm brings

In addition to the MDI Fabric, LogRhythm supports a range of data lake needs, depending on what a customer requires.

LogRhythm's platform can act as a data lake on its own, by collecting data and analysing it using security analytics and artificial intelligence, as well as security automation and orchestration.

The platform can also act as a feed to larger security data lakes that have been created by customers to include a range of security solutions. It can provide raw logs and metadata, as well as retrieve data from a data lake for further analysis.

Ultimately, the way in which the LogRhythm platform is used depends on what a customer is trying to achieve with a security data lake. But rest assured, LogRhythm can help you take advantage of the huge benefits that data lakes can bring to the SOC.

### About LogRhythm

LogRhythm, the security intelligence company, is leading the way with NextGen SIEM to empower organisations to rapidly detect, respond to and investigate cyberthreats. LogRhythm's award-winning platform unifies leading-edge data lake technology, artificial intelligence, security analytics and security automation and orchestration in a single end-to-end solution. LogRhythm serves as the foundation for the AI-enabled security operations centre, helping customers secure their cloud, physical and virtual infrastructures for both IT and OT environments. Among other accolades, LogRhythm is positioned as a Leader in Gartner's SIEM Magic Quadrant.

**www.logrhythm.com**

1 For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html