IDC ANALYZE THE FUTURE

IDC TECHNOLOGY SPOTLIGHT

Sponsored by: Oracle

In the modern reality of hybrid, on-premises, and cloud IT environments, data security is more important than ever. Knowing how your systems are configured, what your users are doing, and what sensitive data is in your systems is critical to protecting the data.

Data Intelligence: Safely Protecting Cloud and On-Premises Databases

February 2021

Written by: Stewart Bond, Research Director, Data Integration and Intelligence Software

Introduction

We live in a world focused on the growing value of data, especially personal data. That focus is heightened by the increasing number and severity of data breaches accompanied by an ever-expanding list of data privacy regulations with serious consequences for noncompliance, such as the European Union General Data Protection Regulation (EU GDPR) and the California Consumer Privacy Act (CCPA). Add in a cybersecurity skills shortage that forces most organizations to choose where to invest their efforts, and you have a good picture of today's information security perspective and the associated challenges.

AT A GLANCE

WHAT'S IMPORTANT

» Data should be the focus of organizations — and a significant amount of organizational data lives in databases.

KEY TAKEAWAY

» If you must choose where to invest in security, the database is the logical place.

Numerous events such as the SolarWinds breach demonstrate that the effectiveness of the network perimeter has long since disappeared and securing data at the edge through network access controls is increasingly ineffective. The growing pace of data breaches in the past few years has proven costly for many corporations and consumers.

In addition, the global pandemic is accelerating cloud migration while increasing the pace of digital transformation and exponentially increasing the volume of data that is being generated, analyzed, and leveraged for digital experiences. According to a December 2020 IDC survey of 406 database professionals in North America, 63% of organizations are actively migrating to the cloud and another 29% are planning to start migration within three years, leaving only 8% still considering migrating. IDC's *Worldwide Global DataSphere Forecast, 2020–2024* estimated that the amount of new data created will grow from 45 zettabytes (ZB) in 2019 to 143ZB in 2024. Published at the beginning of the pandemic, the forecast predicted a "bump" in data growth due of the pandemic and stated that the amount of data created over the next three years will exceed the amount of data created over the past 30 years.

Here's the problem: The most common database security breaches revolve around compromised user credentials or rogue users. Software alone cannot solve issues of people knowingly or unknowingly losing or leaking credentials, nor can it stop a rogue user from doing damage. However, software can assess the security risks of databases, profile user accounts, discover where sensitive data is in a database and even mask that data, and audit database activity. Having actionable intelligence about database users, database configuration, database content, and database activity is a key factor to enabling database security, data governance, and data protection.

Data Intelligence for Security

Metadata, the data about the data, is the intelligence that is foundational to many data-oriented disciplines. At a technical level, metadata tells us how data is defined using dictionaries and how data is organized in schemas. In business analytics, metadata provides business glossaries to help people relate data to business processes and outcomes, and metadata also tells analysts where data came from in lineage diagrams. In data governance, metadata can provide intelligence about where sensitive data is located within databases, where those databases are located, and who is accessing data, when, and how often. Data security benefits from the metadata captured and used across all these disciplines because the more an organization knows about its data, its users, and database configurations, the easier it will be to secure and protect data. Specific examples of how intelligence about data and database configurations is being applied to data security disciplines are as follows:

- Database security configuration drift detection involves periodically capturing database configurations into metadata and then monitoring for changes from the previously stored configurations. Human error resulting in changes to database configurations is often a major factor in data breaches, such as privileges being granted by mistake or changes to a setting that increase a system's vulnerability to attacks. As the old question goes — if a tree falls in the forest and no one is around, does it make a sound? Likewise, if a configuration change is made that affects security but nobody sees it, does it cause damage? Unfortunately, by the time everyone is made aware of the change, the damage may already be done. Automated assessment and drift detection are key to finding and closing security gaps. Verizon's *Data Breach Investigations Report* highlights the need for configuration drift detection and provides specific insights into how human error has contributed to confirmed data breaches.

- Metadata about individual user accounts can be captured, analyzed individually and in aggregate, and monitored for changes. Details about each user account, such as when it was created, last updated, and assigned roles and privileges as well as when and where the user last logged in, can be used to assign a risk level to individual users. At an aggregate level across multiple users, outliers can be identified that may also point to compromised credentials and/or rogue users.

- If an organization doesn't know where the sensitive data resides within its data estate, how can it know what data needs to be protected in order to avoid a breach and achieve and maintain regulatory compliance? All data cannot be treated equally because most organizations live with constrained resources. Knowing where to invest can be determined by understanding the level of risk and the value in a system. Understanding the amount and types of sensitive data within a system contributes to the risk assessment and resulting risk mitigation decisions. Data schemas as well as related data dictionaries and business glossaries can go a long way toward knowing where sensitive data is just by how it is defined. Unfortunately, not all sensitive data is always documented as being sensitive, and people can unknowingly put sensitive data into unprotected places. Additionally, as new regulations are introduced, the definition of sensitive can change, and unprotected data may require new protection based on new definitions or mandates. Sensitive data discovery needs to go beyond technical and business definitions of data by looking at the data itself, and this intelligence needs to be updated regularly to know which policies are needed along with the investment required to protect that sensitive data.

- Securing production data is paramount to protecting organizations from breaches and failed audits, but data used in development and test environments must also be considered so that it doesn't become a soft target. Organizations frequently use exact replicas of production data to properly simulate testing, resulting in the introduction of unnecessary risk. Permanently masking data in development and test environments can mitigate these risks without compromising the validity of data for the purpose of testing. Data masking in this instance isn't just displaying data with special characters overwriting the actual values; instead it leverages intelligence about the data and database to replace factual data with fictional data while maintaining schematic and referential data integrity rules.
- The most dynamic and probably the most effective line of data defense is database activity auditing. Analysis of data and metadata captured in access and query audit logs exposes who is doing what in a database and can identify anomalies such as compromised credentials, rogue users, and unauthorized access. In the event of a breach, activity auditing helps identify what data was accessed and by whom. This not only provides data to help identify the perpetrator but also can be leveraged to mitigate the impact and to comply with breach notification requirements from regulators.

Trends

Today, business assets in the IT environment, including data and databases, need to be secured. Products and services are emerging in the data security software market that address the need to secure data where it resides and to implement access controls at the data level. These products know what data needs to be secured and can expose how secure the data actually is through the analysis of data content and the metadata that describes it. This knowledge is essential to applying controls that trust no one and validate that everyone who is attempting to access the data is doing so appropriately.

Products and services are emerging in the data security software market that address the need to secure data where it resides and to implement access controls at the data level.

While it may seem more complicated to implement security on individual data assets, this new approach is democratizing security and reducing data security risks. The number of roles that work directly with data daily to complete tasks, make decisions, and affect business outcomes is growing. Putting an integrated set of tools that can discover potential data security threats into the hands of the people who work with data is more effective than having an army of database administrators monitoring each individual database using many different tools. The results of a December 2020 IDC survey of people involved in DataOps indicated that meeting security requirements is the most challenging requirement to overcome. This demonstrates the need for security to be part of a continuous data and application delivery process that can involve many different roles, as well as multiple steps in the pipeline, from development and test to production and even disaster recovery.

DataOps is not just DevOps for data applications because it extends continuous testing to the data itself, applying techniques to monitor and measure inputs, outputs, and changes to data and comparing those measurements to acceptable practices. This is not unlike the emerging trend of observability to continuously measure and monitor for drift and shift of data, applications, security, and operations. Extending the concept to focus on databases as well as their configurations, data schemas, and data values is natural because it provides opportunities to be proactive in database management. It is still relatively early days for this feature as descriptive analytics of metadata is the most common approach to identifying drift and shift. In time, more advanced analytics and machine learning will be applied, which will help drive down the risk of data breaches that can result from continuous change.

Considering Oracle Data Safe

Oracle Data Safe is a unified security control center for Oracle Database that leverages metadata and metrics to perform security assessments of database configurations and users, discover sensitive data within databases, mask data in development and test database instances, and monitor and audit database activity. Data Safe was first announced at Oracle Open World in 2019 as an Oracle Cloud–only service for Oracle Autonomous Database and the Oracle Database Cloud Service. Not long after the release, demand surfaced for additional deployment types and support has since been announced for Oracle Database (including Oracle's Enterprise Edition and Standard Edition) running on premises and even on third-party public clouds such as Amazon Web Services (AWS) and Microsoft Azure.

Data Safe's security assessment capability identifies security risk levels, looking at security parameters, security controls in use, and user roles and privileges, based on known best practices and generally accepted standards. Assessments provide a baseline against which future assessments are compared to identify configuration drift or shift. Data Safe users can establish new baselines as risks are identified and either accepted or resolved. The security assessment offers recommendations for remediation and compliance with regulations such as GDPR, and compliance standards such as DISA STIGs and CIS Benchmark findings are categorized and prioritized to assist with remediation. Data Safe's security assessment also detects configuration drift, capturing periodic configuration snapshots and identifying databases where the current state does not match the assigned security baseline.

The user assessment feature identifies users that present critical, high, medium, and low security risks to the database. Privileges granted to the highest-risk users are scored and reviewed, and activity for these users is captured, including static and dynamic profile information such as last log-in, IP address, password age, and database activity data. Users with elevated privileges or who have weak password policies could be soft targets for compromise that require additional auditing or immediate remediation.

User activity auditing and reporting tracks database user activity and raises alerts when risky or anomalous events are identified. Predefined audit policies include support for Defense Information Systems Agency (DISA) Security Technical Implementation Guides (STIGs) and Center for Internet Security (CIS) Benchmark recommended policies, or users can set their own custom policies. This feature helps identify potential risks as database activity is occurring, and retention of the audit log provides evidence for compliance reporting and forensic analysis. An extensive reporting capability, in addition to a simple filter-and-query capability, makes it easy to pick out relevant audit data.

Data Safe can also identify sensitive data stored in databases. Data Safe ships with over 125 predefined sensitive data types and provides organizations with the ability to add their own types as well. Knowing where and how much sensitive data is in the database system helps guide risk assessment and protection actions. Sensitive data discovery is also a prerequisite to another Data Safe feature, masking data for development and test database instances. Data Safe's masking capability is the next-generation follow-on to earlier masking capabilities from Oracle, with automated masking format assignment making it easy to mask a database without having to know regular expressions.

Oracle Data Safe was developed for the non-DBA user, providing intuitive dashboards that display the overall security health of multiple Oracle databases, including graphical reporting of information collected from security assessments, user assessments, and sensitive data discovery. Data Safe provides remediation recommendations and cross references those recommendations with applicable sections of the CIS Benchmark, EU GDPR, and DISA STIGs. It is also important to note that in Oracle Cloud instances, Oracle Data Safe can be used by customers to validate their configuration, identify

what risk is contained within the database, and detect inappropriate user activity. If the cloud provider administrators attempt to read or make any changes to the customer databases, Data Safe can quickly detect these changes and notify the database consumer.

Challenges

Oracle Data Safe offers a metadata- and metrics-driven approach to securing data in a database, protecting not only the database but also the people, places, and artifacts that are represented in the data. However, Data Safe works only with the Oracle Database, and Oracle has no immediate plans to expand coverage to include other databases in the market. This limits Oracle's market for Data Safe, but it is a benefit for Oracle database customers because a tool made explicitly for one database, by the same vendor of that database, will likely function better than a generic tool that covers multiple database products.

Data Safe currently has no ability to provide alert notifications outside of the Data Safe console, nor does it provide RESTful APIs to allow for automated configuration and analysis of new databases. These are both features that Oracle intends to address in future releases.

Conclusion

Accessing sensitive data is the goal of most, if not all, cyberattacks. Hackers target databases because most mission-critical and sensitive data is stored in them. The most common database security breaches revolve around compromised user credentials or rogue users. Software is limited in its ability to prevent people from knowingly or unknowingly leaking or losing credentials, and certainty cannot stop rogue users from doing damage. Oracle Data Safe helps apply security to data where it resides: in the database. Leveraging intelligence available from the metadata associated with a database and its configuration, Oracle Data Safe can assess security risks, discover and mask sensitive data, monitor database activity for anomalies, and audit database activity for compliance. To the extent that Oracle can reinforce the value of Data Safe and Oracle Database as a holistic solution for Oracle customers and continue to improve the product by adding more advanced security intelligence features, more organizations will adopt Data Safe.

About the Analyst

Stewart Bond, *Research Director, Data Integration and Intelligence Software*

Stewart Bond is Research Director of IDC's Data Integration and Intelligence Software service. Mr. Bond's core research coverage includes watching emerging trends that are shaping and changing data movement, ingestion, transformation, mastering, cleansing, and consumption in the era of digital transformation.

MESSAGE FROM THE SPONSOR

If you would like to learn more about Data Safe, we invite you to take the ***Data Safe quick tour***. To dig even deeper and get hands-on experience with Data Safe, try the ***Oracle Data Safe Fundamentals Workshop*** on Oracle LiveLabs.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick St., Blg. B
Needham, MA 02494 USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com