Security Data Maturity Model

DataBee®
A COMCAST COMPANY

databee.ai

Introduction

Did you know that humanity reached the “Zettabyte Era” in 2012? This is when our collective data creation surpassed one zettabyte (equivalent to 10^{21}, or 1,000,000,000,000,000,000,000 bytes). By 2025, data creation is projected to grow to more than 180 zettabytes.[1] The sheer volume and exponential growth of “Big Data” has put it in the same category as philosophical concepts like “infinity,” going beyond the massive and into a place that’s difficult for most people to comprehend. For organizations of all stripes — large, small, public, private, or criminal — long-term viability hinges on being able to use all this data for insights into how to grow and protect the organization.

[1] “Amount of data created, consumed, and stored 2010-2020, with forecasts to 2025,” Statista, 2023

Challenges posed **by big data**

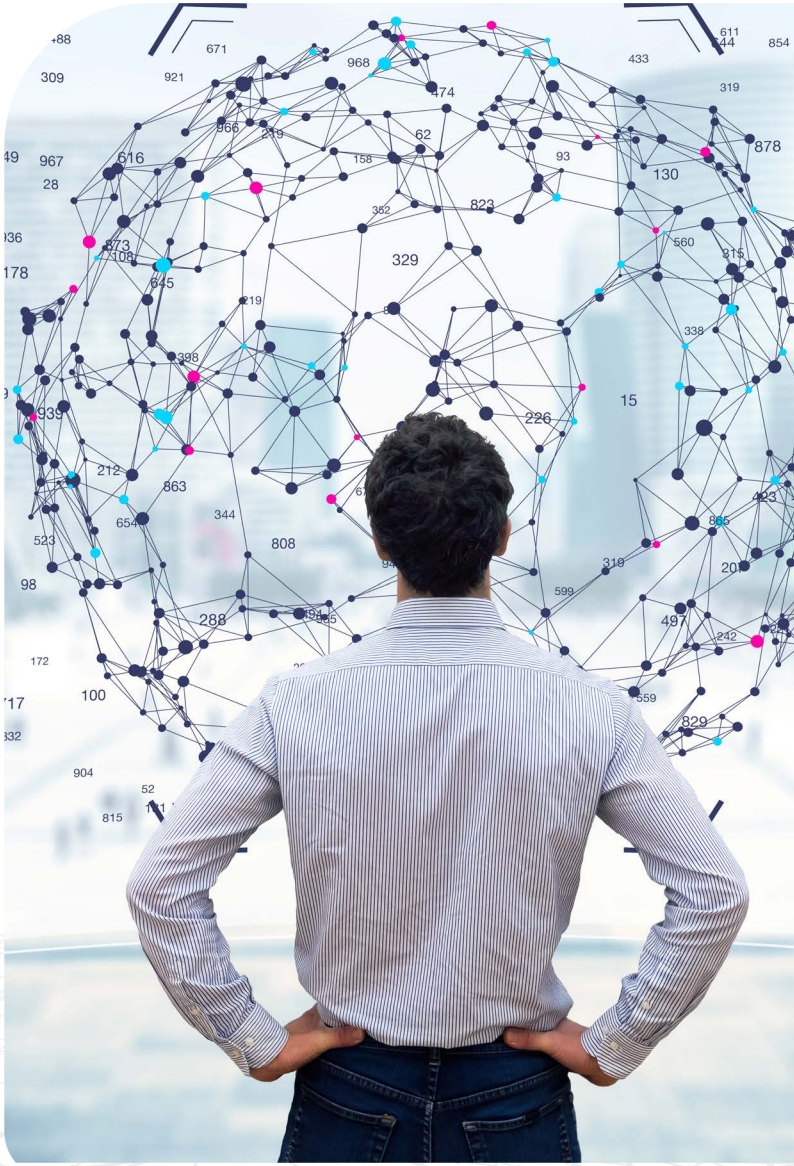
While all data consumers within an organization stand to benefit from big data insights, security teams might have the most to gain, but also face some of the greatest challenges. Why?

- First, the volume of data that security teams deal with is huge; if you think about it, the security team's purview is everyone, everything, everywhere, all the time.
- Second, when trying to analyze this data, there are too many vendor-specific security tools with unique semantics and file formats that make it difficult to gather insights quickly.

Enterprises can easily have over 100 security tools, some with overlapping solutions or features. Security teams tend to buy many tools to analyze and understand their disparate data generated from the hundreds of thousands (or millions) of sensors deployed in a large enterprise ecosystem. With their big budgets, there is increasing pressure to ensure investments deliver the protection and efficacy they promise.

Context is crucial to telling a complete story, relating data from different security controls to each other to flesh out the data narrative, and answering questions about what is truly happening within the enterprise's ecosystem.

Big data's other challenge for security leaders is that the data lives in silos. Making sense of that disparate and decentralized data takes time and effort. How do you combine all of it to gain insights and become informed?



Introducing the **Security Data Maturity Model**

Critical to assessing and answering these questions is taking stock of the enterprise's security data maturity status. Security data maturity measures and categorizes how well an organization can assess, analyze, and leverage data to help protect and secure its infrastructure. In short, developing and using a Security Data Maturity Model (SDMM) is vital for an organization to understand its security posture and journey.

To address these challenges within the broad Comcast enterprise structure, Comcast's cybersecurity team spent years building a security data fabric to better operationalize and optimize security data to help improve threat responses and cybersecurity oversight.

As part of that framework, the team also defined an SDMM, a conceptual framework to help their security teams self-identify their progress in moving toward a prescriptive approach to making data-driven decisions. The model provides strategic value to help organizations better assess their security status and risk profile. It also helps foster a critical assessment to understand better what policies, procedures, and resources must be adopted to reduce the overall risk exposure.

There are many beneficial reasons for an organization to investigate and use an SDMM, including:

Developing an overall security data strategy. The advent of big data, artificial intelligence (AI), and machine learning (ML) has spotlighted that data is a core asset, and many organizations have a data strategy. But to have an enterprise-wide global data strategy, security data must be included. Moreover, to have meaningful security data, you must add data from outside the security realm to provide deeper context.

Data can be used to make centralized, intelligent decisions. Higher levels of security data maturity lead to better organizational decision-making.

With context, an SDMM can help provide deeper organizational insights. Enterprises can make better data-driven decisions based on the data's context and by showing how the data relates to specific areas such as governance, compliance, and risk management.

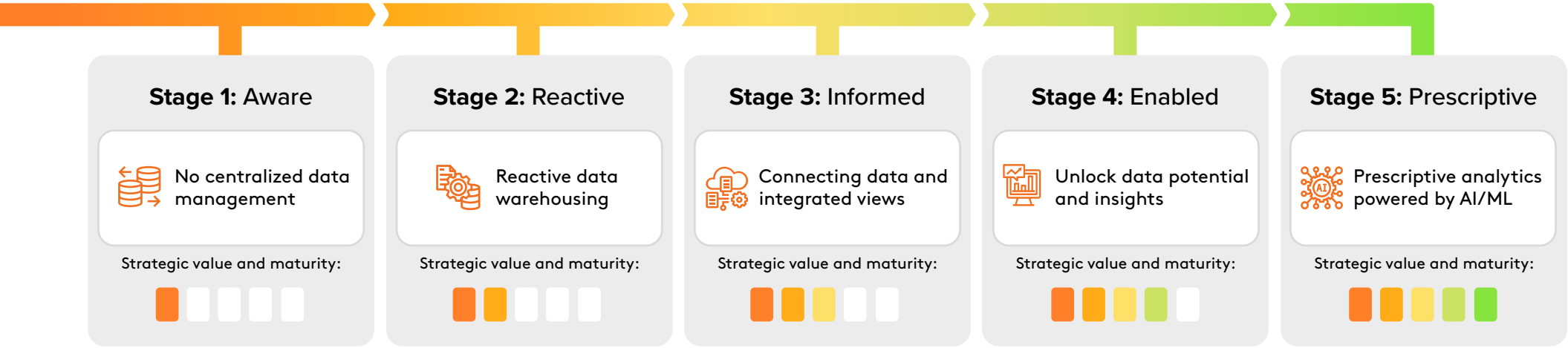
Help facilitate organizations moving from a reactive to a proactive state. SDMMs can facilitate moving an organization's cultural state from one that is reactive to one that is more proactive and prescriptive. From a security standpoint, this means getting ahead of the bad guys and putting systems, resources, and processes in place to anticipate risks and breaches before they occur.

Business goals alignment. An SDMM can help a decentralized enterprise align its security and business goals across business units.

Cost savings. An SDMM can help save the organization money by focusing on tools and resources that best deliver actionable insights. For example, an SDMM may help an organization reduce the number of ineffective or overlapping security tools.

SDMMs provide efficiency and are vital to helping democratize data across organizations, enabling large and small teams to make better data-driven decisions for security, governance, and compliance.

Comcast's SDMM comprises five stages, mapped from less mature to more mature. Greater maturity implies greater strategic value and efficacy to the organization. **The stages are depicted below and described in detail in the following sections.**



Understanding where your organization lies within the SDMM can help determine the resources and investment required to move to the next maturity level. It is tempting to look at this model and assume that if you build a security data strategy, understand your business requirements, and get the investment needed, the job is done. Unfortunately, this is easier said than done because it is difficult to extrapolate the investment level required to secure an enterprise infrastructure. The SDMM can help approximate the investment required.

As an organization increases its security data maturity, the core topic becomes less about what type of data is stored or what type of threat hunting is effective and more about benefiting from prescriptive analytics using AI and ML. The strategic value of an organization's data moves in parallel with its data maturity.

Stage 1: **Aware**

Stage 1: Aware

No centralized data management

- **No data strategy** and business or stakeholder buy-in
- Standardized reporting limited by **siloed data**
- Desire to be data-driven and define data investments

Strategic value and maturity:

Stage 2: Reactive

Reactive data warehousing

- Start of data warehousing, **but data silos persist**
- Security data remains segmented
- Significant **effort on data manipulation** and prep
- Multiple data schemas and "sources of truth"
- Limited access and data-driven insights

Strategic value and maturity:

Stage 3: Informed

Connecting data and integrated views

- Exploring ML models and higher levels of data understanding
- Data sources connected with common data schema
- Real- and near-real-time data for **360-degree data entity** view and proactive **reporting**
- Growing data producers, use cases, and personas
- Improvements on data latency, quality, and availability

Strategic value and maturity:

Stage 4: Enabled

Unlock data potential and insights

- **Multiple users benefit** from custom-built data products aligned to business logic
- Automated data infrastructure and pipeline
- ML accelerates data discovery and use cases
- Data used to **solve complex business problems** and advanced use cases (i.e., threat hunting)
- Exploring and using multi-cloud deployments

Strategic value and maturity:

Stage 5: Prescriptive

Prescriptive analytics powered by AI/ML

- **Global data strategy** that is inclusive of security data
- Continuous optimization of algorithms and models for analytics at scale
- **Augmented cybersecurity workflows** with AI/ML
- Data is parsed and normalized more rapidly
- Ability to historically re-run data during incidents to see what exploits may have been lurking
- Data teams focused on building organizational awareness and expanding:
  - AI
  - Deep learning
  - Supervised and unsupervised learning
  - Classification and regression
  - Clustering

Strategic value and maturity:

Stage 1: Aware

The **Aware** stage represents the least mature stage of the SDMM. Newly formed organizations, merged entities, and those that lack a security focus and resources are likely represented in this stage. An organization in this stage is aware that security data exists, but the data is underutilized. Furthermore, the security team is likely uninformed that the data can provide better visibility into the organization's security posture. Accordingly, the team can easily miss signs of threats or compliance violations.

By far, the most significant attribute of organizations in the Aware stage is their lack of centralized data management. Decentralized data results in a lack of consistency and focus, resulting in more time needed to find and cleanse data to make some sense of it. Decentralized data is potentially less secure as there is no consistent set of policy controlsto secure the data throughout the organization.

Other attributes of the **Aware** stage are:

- No data strategy and business or stakeholder buy-in
- Standardized reporting limited by siloed data
- Desire to be data-driven and define data investments

Stage 2: **Reactive**

Stage 1: Aware

No centralized data management

- **No data strategy** and business or stakeholder buy-in
- Standardized reporting limited by **siloed data**
- Desire to be data-driven and define data investments

Strategic value and maturity:

Stage 2: Reactive

Reactive data warehousing

- Start of data warehousing, **but data silos persist**
- Security data remains segmented
- Significant **effort on data manipulation** and prep
- Multiple data schemas and "sources of truth"
- Limited access and data-driven insights

Strategic value and maturity:

Stage 3: Informed

Connecting data and integrated views

- Exploring ML models and higher levels of data understanding
- Data sources connected with common data schema
- Real- and near-real-time data for **360-degree data entity** view and proactive **reporting**
- Growing data producers, use cases, and personas
- Improvements on data latency, quality, and availability

Strategic value and maturity:

Stage 4: Enabled

Unlock data potential and insights

- **Multiple users benefit** from custom-built data products aligned to business logic
- Automated data infrastructure and pipeline
- ML accelerates data discovery and use cases
- Data used to **solve complex business problems** and advanced use cases (i.e., threat hunting)
- Exploring and using multi-cloud deployments

Strategic value and maturity:

Stage 5: Prescriptive

Prescriptive analytics powered by AI/ML

- **Global data strategy** that is inclusive of security data
- Continuous optimization of algorithms and models for analytics at scale
- **Augmented cybersecurity workflows** with AI/ML
- Data is parsed and normalized more rapidly
- Ability to historically re-run data during incidents to see what exploits may have been lurking
- Data teams focused on building organizational awareness and expanding:
  - AI
  - Deep learning
  - Supervised and unsupervised learning
  - Classification and regression
  - Clustering

Strategic value and maturity:

Stage 2: Reactive

The **Reactive** stage is another early period in an enterprise's security maturity. Often, it is characterized as a very ad hoc approach to overall security data management and knee-jerk investments in data storage or warehousing solutions. For example, a third-party consultant may have identified a potential vulnerability in an organization's infrastructure, or an organization may have become aware of malware from a news article or a colleague. In both cases, the enterprise reacts to news or circumstances.

Accordingly, the organization needs to identify its risk factors. Addressing the risk factors may be challenging if the organization is not adequately prepared or resourced. For example, assessing the risk may require analyzing a large volume of data, which requires the data to be stored in a data warehouse or other repository, such as a data lake. All of that requires time and resources. Every single incident can be a highly demanding manual data collection activity.

Other attributes of the **Reactive** stage are:

- Start of data warehousing, but data silos persist
- Security data remains segmented
- Significant effort on data manipulation and prep
- Multiple data schemas and "sources of truth"
- Limited access and data-driven insights

Stage 3: **Informed**

Stage 1: Aware

No centralized data management

- **No data strategy** and business or stakeholder buy-in
- Standardized reporting limited by **siloed data**
- Desire to be data-driven and define data investments

Strategic value and maturity:

Stage 2: Reactive

Reactive data warehousing

- Start of data warehousing, **but data silos persist**
- Security data remains segmented
- Significant **effort on data manipulation** and prep
- Multiple data schemas and "sources of truth"
- Limited access and data-driven insights

Strategic value and maturity:

Stage 3: Informed

Connecting data and integrated views

- Exploring ML models and higher levels of data understanding
- Data sources connected with common data schema
- Real- and near-real-time data for **360-degree data entity** view and proactive **reporting**
- Growing data producers, use cases, and personas
- Improvements on data latency, quality, and availability

Strategic value and maturity:

Stage 4: Enabled

Unlock data potential and insights

- **Multiple users benefit** from custom-built data products aligned to business logic
- Automated data infrastructure and pipeline
- ML accelerates data discovery and use cases
- Data used to **solve complex business problems** and advanced use cases (i.e., threat hunting)
- Exploring and using multi-cloud deployments

Strategic value and maturity:

Stage 5: Prescriptive

Prescriptive analytics powered by AI/ML

- **Global data strategy** that is inclusive of security data
- Continuous optimization of algorithms and models for analytics at scale
- **Augmented cybersecurity workflows** with AI/ML
- Data is parsed and normalized more rapidly
- Ability to historically re-run data during incidents to see what exploits may have been lurking
- Data teams focused on building organizational awareness and expanding:
  - AI
  - Deep learning
  - Supervised and unsupervised learning
  - Classification and regression
  - Clustering

Strategic value and maturity:

Stage 3: Informed

The **Informed** stage represents an enterprise's intermediate phase of security data maturity, where disparate data and security views are first connected. It is the first time data-driven decisions are made with a defined security controls framework. At this stage, security teams can ask and answer more insightful questions regarding their infrastructure, such as:

- How can I get visibility and more protection on my endpoints?
- What does my current risk posture look like for that particular area?
- Is my signal-to-noise ratio increasing to provide better visibility?
- What areas of focus do I want to address now?
- What regulatory or compliance issues do I need to address?

Other attributes of the **Informed** stage are:

- Exploring ML models and higher levels of data understanding
- Data sources connected with common data schema
- Real- and near-real-time data for 360-degree data entity view and proactive reporting
- Growing data producers, use cases, and personas
- Improvements on data latency, quality, and availability

Stage 4: **Enabled**

Stage 1: Aware

No centralized data management

- **No data strategy** and business or stakeholder buy-in
- Standardized reporting limited by **siloed data**
- Desire to be data-driven and define data investments

Strategic value and maturity:

Stage 2: Reactive

Reactive data warehousing

- Start of data warehousing, **but data silos persist**
- Security data remains segmented
- Significant **effort on data manipulation** and prep
- Multiple data schemas and "sources of truth"
- Limited access and data-driven insights

Strategic value and maturity:

Stage 3: Informed

Connecting data and integrated views

- Exploring ML models and higher levels of data understanding
- Data sources connected with common data schema
- Real- and near-real-time data for **360-degree data entity** view and proactive **reporting**
- Growing data producers, use cases, and personas
- Improvements on data latency, quality, and availability

Strategic value and maturity:

Stage 4: Enabled

Unlock data potential and insights

- **Multiple users benefit** from custom-built data products aligned to business logic
- Automated data infrastructure and pipeline
- ML accelerates data discovery and use cases
- Data used to **solve complex business problems** and advanced use cases (i.e., threat hunting)
- Exploring and using multi-cloud deployments

Strategic value and maturity:

Stage 5: Prescriptive

Prescriptive analytics powered by AI/ML

- **Global data strategy** that is inclusive of security data
- Continuous optimization of algorithms and models for analytics at scale
- **Augmented cybersecurity workflows** with AI/ML
- Data is parsed and normalized more rapidly
- Ability to historically re-run data during incidents to see what exploits may have been lurking
- Data teams focused on building organizational awareness and expanding:
  - AI
  - Deep learning
  - Supervised and unsupervised learning
  - Classification and regression
  - Clustering

Strategic value and maturity:

Stage 4: Enabled

The **Enabled** stage unlocks the data's potential and provides insights for more users, including enabling the GRC and outside auditors. Data visibility and usage significantly increase in this stage, and data is used to allow new functions. Many of these use cases require people resources with deep expertise in analytics, security, threat hunting, AI, and ML to assess the data.

The Enabled stage is where security data is connected and automated, transforming enterprise processes and functions. For example, in the case of threat hunting, automation reduces the work, time, and effort to find and respond to threats. For business teams, better business intelligence is available.

For some organizations, progress in the SDMM stops at the Enabled stage. These organizations are held back by needing help with what to do next or understanding the resources required to achieve more maturity. They may be relatively happy with the Enabled stage and have no future plans. Or they may need more expertise and technology investment to progress to the next stage. Or they might possess the right expertise but not the right tools.

Other attributes of the **Enabled** stage are:

- Multiple users benefit from custom-built data products aligned to business logic
- Persistent data used for both security and non-security teams on an as-needed basis
- Automated data infrastructure and pipeline
- ML accelerates data discovery and use cases
- Data used to solve complex business problems and advanced use cases (i.e., threat hunting)
- Exploring and using multi-cloud deployments

The **Enabled** stage is particularly impactful because this is where the enterprise begins to use and accelerate data's potential for previously unavailable insights. Moreover, data is now a resource for enterprise-wide team members, answering previously unasked (and unanswered) questions. Coupled with AI and ML tools, analytics, security, and threat hunting, teams can address security threats before they negatively impact the organization.

Stage 5: **Prescriptive**

Stage 1: Aware

No centralized data management

- **No data strategy** and business or stakeholder buy-in
- Standardized reporting limited by **siloed data**
- Desire to be data-driven and define data investments

Strategic value and maturity:

Stage 2: Reactive

Reactive data warehousing

- Start of data warehousing, **but data silos persist**
- Security data remains segmented
- Significant **effort on data manipulation** and prep
- Multiple data schemas and "sources of truth"
- Limited access and data-driven insights

Strategic value and maturity:

Stage 3: Informed

Connecting data and integrated views

- Exploring ML models and higher levels of data understanding
- Data sources connected with common data schema
- Real- and near-real-time data for **360-degree data entity** view and proactive **reporting**
- Growing data producers, use cases, and personas
- Improvements on data latency, quality, and availability

Strategic value and maturity:

Stage 4: Enabled

Unlock data potential and insights

- **Multiple users benefit** from custom-built data products aligned to business logic
- Automated data infrastructure and pipeline
- ML accelerates data discovery and use cases
- Data used to **solve complex business problems** and advanced use cases (i.e., threat hunting)
- Exploring and using multi-cloud deployments

Strategic value and maturity:

Stage 5: Prescriptive

Prescriptive analytics powered by AI/ML

- **Global data strategy** that is inclusive of security data
- Continuous optimization of algorithms and models for analytics at scale
- **Augmented cybersecurity workflows** with AI/ML
- Data is parsed and normalized more rapidly
- Ability to historically re-run data during incidents to see what exploits may have been lurking
- Data teams focused on building organizational awareness and expanding:
  - AI
  - Deep learning
  - Supervised and unsupervised learning
  - Classification and regression
  - Clustering

Strategic value and maturity:

Stage 5: Prescriptive

In the Prescriptive stage, prescriptive analytics using AI/ML can help build models (such as chatbots) to leverage this data. The data can now guide teams to self-identify incidents before they become more significant events. Compliance and threat hunting are improved dramatically, as are threat detection and incident response. All in all, getting to the Prescriptive stage profoundly impacts managing security and compliance with positive outcomes.

This stage takes advantage of lessons learned from previous stages, and users spend time preparing for the future of data. Teams have more insights into the data lineage, analytics, and visualizations that can run, and more people access and utilize the data without losing governance and control.

Other attributes of the **Prescriptive** stage are:

- Global data strategy that is inclusive of security data
- Continuous optimization of algorithms and models for analytics at scale
- Augmented cybersecurity workflows with AI/ML
- Data is parsed and normalized more rapidly
- Ability to historically re-run data during incidents to see what exploits may have been lurking
- Data teams focused on building organizational awareness and expanding:
  - AI
  - Deep learning
  - Supervised and unsupervised learning
  - Classification and regression
  - Clustering

The **Prescriptive** stage focuses on using prescriptive analytics. AI/ML helps build models to provide additional deep data insights in this stage. Compliance, threat hunting, threat detection, and incident response are far more effective and quicker to respond. False positives are significantly reduced.

Determining your security **data maturity level**

In a perfect world, assessing your organizational risk is fast, easy, and readily available. But the reality is risk levels are constantly changing, and predictability is difficult. The threat of ransomware, breaches, lost data, and stolen financial and personal data is real — it happens every day. Whether you are just beginning to consider your risk profile or are already in a Prescriptive stage using AI and ML to enhance your defenses, **there are questions you should be asking your team:**

- Is there an appropriate and current security data strategy in place?
- Who owns data throughout the enterprise?
- Where are all the places that data lives?
- How is data utilized today?
- What stage of security data maturity is the company in?
- Who, or what team, is responsible for the organization's overall data strategy?
- What are the organization's business requirements for security, governance, and compliance?
- Is appropriate funding in place for the investments in security and data infrastructure?

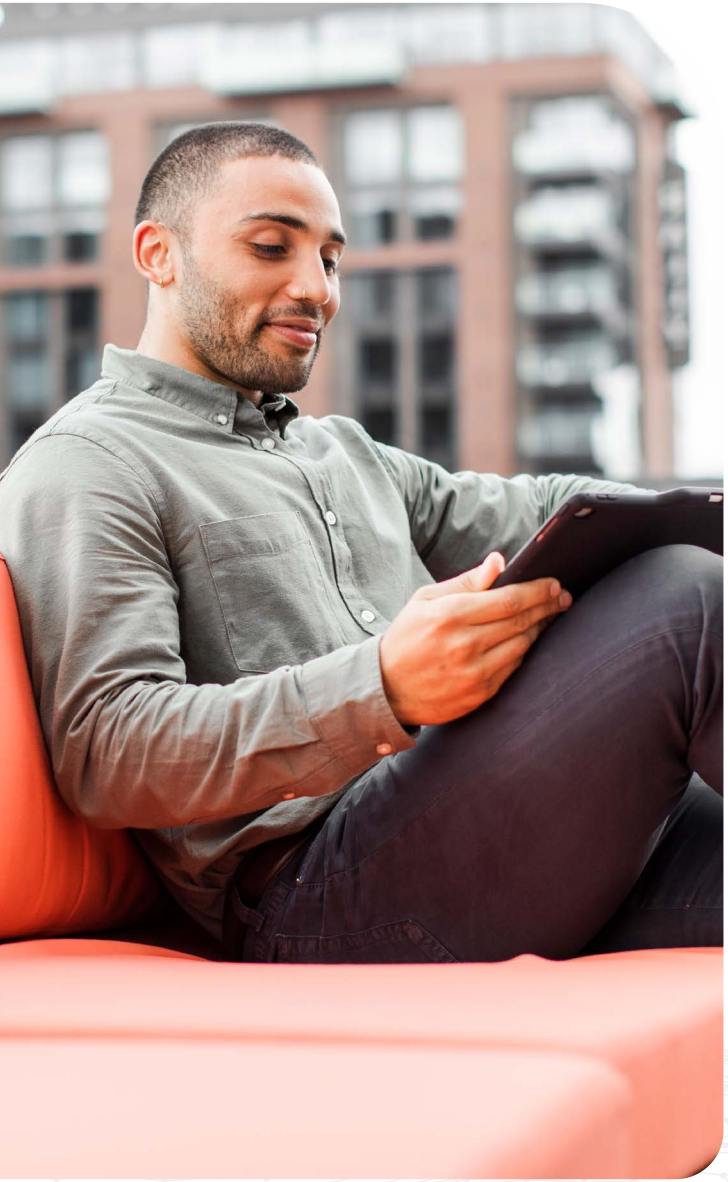
Operationalize the **SDMM with DataBee®**

DataBee®, a cloud-native security data fabric platform inspired by Comcast's cybersecurity team, can help accelerate your security data maturity journey and establish a long-term foundation for the evolution of your organization's global data strategy. DataBee operationalizes and optimizes security data, and the SDMM maps right to that journey. DataBee was created based on the internally successful security, risk, and compliance data fabric (mentioned earlier) that focuses on delivering business outcomes from data.

The DataBee data fabric platform:

- Eliminates data silos, bringing security data together by transforming data early in the pipeline and normalizing it to the Open Cybersecurity Schema Framework (OCSF).
- Makes data readily usable by multiple personas, from everyday users like data analysts to the board of directors.
- Significantly reduces the cost of data by flattening and compressing data into optimized, time-series datasets.
- Provides faster time-to-value because data is optimized.
- Combines security data with enterprise data to provide deep insights into an organization's security and compliance posture with business context.

DataBee weaves together disparate security data from various technologies and tools, correlating and enriching large historical datasets with business context into a single data layer that is standardized, usable, and searchable for analyses, monitoring, and reporting at scale. The platform spans security, risk, compliance, and privacy missions with use cases in advanced threat detection, threat hunting, continuous controls assurance, security information and event management (SIEM) decoupling, and behavioral analysis. DataBee sends enriched and optimized data into your data lake or other cloud or on-premises storage repository, where it is collated for quicker analysis and yields more accurate long-term insights. DataBee meets you where you are in your security journey, and organizations in the beginning of the Reactive stage can accelerate their journey even faster. Once data is actively warehoused and proactively stored, it can be more easily segmented and processed. And this is where the most significant data leverage and value begins.



SDMM positive **outcomes**

Based on Comcast's internal experiences, the benefits and positive outcomes of using an SDMM as a blueprint to map security maturity were far-reaching and continue to impact the organization favorably.

From a business perspective, the SDMM brings "security" into the enterprise's global data strategy. It encouraged discussions resulting in balanced investments in new security tools, cost savings through eliminating redundant tools, and the ability to adapt quickly to customer, market, business, and threat conditions.

From an overall security perspective, security teams can now respond with context and confidence and surface real, accurate signals from noise. Threat hunting teams found they could hunt faster without restrictions.

Furthermore, security teams found cost savings opportunities revealed by the SDMM.
For example, data sent to the SIEM is optimized via deduplicating data. Marrying insights further reduces the storage needed from the SIEM. In addition, storing raw and processed data in a data lake further reduces costs.

From a compliance standpoint, teams discovered they could now gain real-time compliance insights, proactively and continuously manage compliance, and rigorously validate compliance levels.

Harnessing security data **for better business outcomes**

Clichéd yet undeniable: Knowledge is power. This Security Data Maturity Model is a programmatic way for organizations to chart and understand where they are in their security journey and how they need to evolve to adapt quickly as threats change, regulations change, and business conditions change.

A successful SDMM is not just a “one and done” exercise, but a living framework that organizations can use as they seek new ways to gather and extract insights from security and business data to thrive. All those zettabytes of data will just keep growing (dare we start thinking about yottabytes?); an SDMM is the path to putting all of this amazing data to work protecting your organization through the 2020s and into the future.

Whether your organization is at the Aware phase or has already progressed to the Enabled or even Prescriptive phase of its security journey, the DataBee security, risk, and compliance data fabric platform can help you navigate your journey to adaptability and the future.

Need help determining where your organization is on its journey to security data maturity? Let’s talk. Better yet, we’d love to show you how the DataBee platform can help you pave a path to Stage 5, Prescriptive.

Request a Demo →