Cybersecurity: Building the Next Generation of Threat Intelligence with MongoDB

October 2016

Table of Contents

Introduction

Cybersecurity has always been a top priority for IT leaders, and now, as organizations embark on digital transformation initiatives, so it is increasingly becoming a priority for the boardroom as well. And for good reason. Industry analysts predict cybercrime will cost the global economy $6 trillion annually by 2021. To respond, organizations are expected to spend more than $1 trillion on cybersecurity from 2017 to 2021. But with the onslaught of new threat classes and threat actors, are those investments being directed towards the right solutions?

With the ubiquity of mobile access from smartphones and tablets, coupled with the emergence of Internet of Things (IoT) connectivity for billions of devices, and the rise of cloud computing, the surface area for cyberattack has radically increased in just half a decade. Threats are becoming more sophisticated with the emergence of social engineering, Advanced Persistent Threats (APTs), ransomware, and fraud committed through digital identity theft. At the same time, threat actors are multiplying – today our systems are vulnerable from armies of hackers-for-hire, well-organized criminal gangs and state-sponsored initiatives.

Beyond protecting sensitive data and infrastructure, executives are aware that building a strong cybersecurity posture is essential for competitive advantage. In the digital age, organizations must create and maintain trust with its customers, whether those be consumers, citizens or others enterprises. Companies that fail to maintain trust with their customers will ultimately fail in the market.

Core perimeter security defences such as firewalls and encryption remain key, but most innovation in cybersecurity is coming today from "big data" analytics and machine learning, protecting systems in real time. These approaches are being used together to unlock the more sophisticated behavioral insight and threat intelligence necessary to combat new classes of threats and actors, which in turn demand investment in new technologies and skills sets.

In this whitepaper, we will identify the requirements of data management in modern cybersecurity initiatives. We will also explore how MongoDB is being used by Fortune 500 companies and governments alike to counter emerging threats, while reducing operational efforts and time between detection and containment.

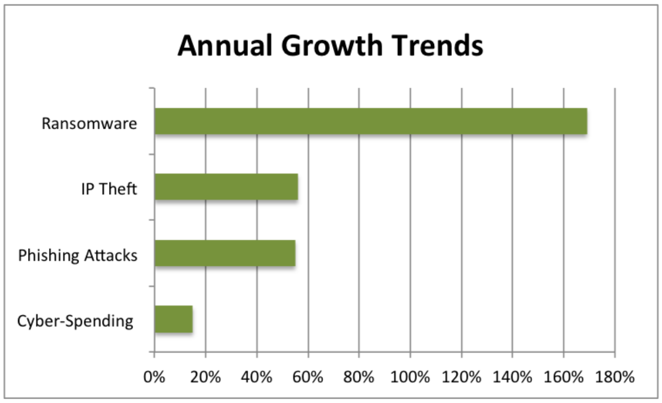

Figure 1: Growth in spending to counter new threat classes and actors

New Capabilities Required to Combat The Changing Face of Cyber Attacks

Most security surveys agree that data theft or corruption by internal users remains an organization's greatest cybersecurity risk – research from Accenture indicates 69% of respondents have been subject to such an attack. However the threat actors of increasing concern are now external to the enterprise. In fact, the Verizon Data Breach Investigations Report cites over 80% of breaches come from attackers outside of the enterprise. These actors are typically well-organized criminal gangs or state-sponsored teams pursuing objectives that include fraud, corporate espionage, and the disruption of critical infrastructure.

The modes of attack are changing, and security teams are challenged to respond. Symantec reports a 55% growth in phishing attacks, while Intel estimates ransomware incidents grew an incredible 169% in just 12 months. PWC uncovered intellectual property theft increased 56% in 2015, while key industrial infrastructure and military operations are being disrupted. Especially concerning are statistics from Cisco stating the industry average time to detect a compromised system now stands at between 100 and 200 days.

Enterprises have traditionally relied on the same established technologies, such as firewalls and encryption, as the primary defense against cyberthreats. While this technology has been effective at detecting attacks, it is less well suited to combat long lasting threats where an intruder has already breached the security perimeter and is within the organization's firewall. APTs rely on a "slow and low" approach where Indicators of Compromise (IOC) are much less visible. The attacker can sit within internal systems for weeks or more, continuously monitoring and extracting sensitive data from specific targets, and compromising systems for future data theft and extortion.

To counter these more sophisticated attacks, organizations are increasingly exploring new approaches to cybersecurity. Rather than just relying on serially scanning potential attack vectors, organizations are seeking to implement systems that enable continuous monitoring and data collection from their infrastructure. Behavioral analytics and machine learning can be applied in real-time to this data to create intelligent insights that enable not just detection and response to threats, but can actually predict them before systems are breached.

Such a data-driven approach to cybersecurity requires a fundamental re-evaluation of the data management technologies employed as the foundation for defense systems.

What Changes are Needed in Cybersecurity Data Management?

Data-driven cybersecurity relies on continuous monitoring and data collection from all internal systems, users and endpoints. This data must be centralized in a single repository to allow complete enterprise visibility. The data must be processed in real-time, and threat models served to cybersecurity applications with millisecond latency. These requirements place demands that cannot be met by the traditional relational databases and enterprise data warehouses used in current generations of cybersecurity solutions:

- **Exponential increases in data volume**. It's not unusual for tens to hundreds of gigabytes of new event data to be added every day. Trying to keep pace with this data growth by deploying ever larger "scale-up" servers is neither economic nor practical. Instead, data needs to be distributed across fleets of commodity nodes, often across data centers for data locality and disaster recovery, with new database instances transparently added in a “scale-out” design as data volumes grow.

- **“Messy”, rapidly changing data**. The database must be able to handle data from multiple sources, including applications, identity management systems, network packet capture, user sessions, asset catalogs, operating systems, firewalls, proxy servers, filesystems, web server logs, clickstreams, and many more components. The data is presented in multiple formats and structures, which can change as systems are updated or renewed. Trying to compress multi-structured data into the rigid and static row-and-column format of relational databases imposes significant delays to application development, evolution and extension of the data model as new threats emerge.
- **Intelligent insights, delivered in real time**. To identify threats, correlating data from multiple endpoints is essential in providing a complete view and context into system activity. Analytics and machine learning must be able to run across vast swathes of data stored in a centralized repository. Traditional technologies are unable to store and process these data volumes at scale. Hadoop-based platforms are unable to serve the risk models generated from this analysis, or ad-hoc investigative queries, with the low latency demanded by real-time cybersecurity systems.

Meeting Cybersecurity Challenges with a Modern Database

MongoDB is the most popular and widely used non-relational database available today. With its unique Nexus Architecture, MongoDB is at the center of many digital transformation initiatives across a range of organizations including ADP, AstraZeneca, Barclays, Bosch, Cisco, Forbes, KPMG, Lockheed Martin, Metlife, the UK Government’s Digital Service, UPS, Verizon, and many more. MongoDB is the only database that harnesses the innovations of NoSQL – data model flexibility, always-on global deployments, and scalability – while maintaining the foundation of rich query capabilities and enterprise-grade security that have made relational databases an essential technology for cybersecurity over the past three decades.

Data Model Flexibility

MongoDB's document data model makes it easy for developers to store and combine data of any structure within the database, without giving up sophisticated governance rules to enforce data quality. The schema can be dynamically modified without application or database downtime. If, for example, we want to start to store geospatial data associated with a specific event class, the application simply writes the updated object to the database, without costly schema modifications or redesign.

MongoDB documents are typically modeled to store data for a given entity – such as an event or user – within a single document, rather than spread across multiple relational tables. Document access can be completed in a single round trip to MongoDB, rather than having to JOIN separate tables spread across the database. With data localized into a single document, application performance is often much higher when using MongoDB, which can be the decisive factor in predicting a cybersecurity risk or detecting a breach.

Elastic Scalability with Always-on Availability

MongoDB provides horizontal scale-out for databases on low cost, commodity hardware using a technique called sharding, which is transparent to applications. Sharding distributes data across multiple physical partitions called shards. Sharding allows MongoDB deployments to address the hardware limitations of a single server, such as bottlenecks in RAM or disk I/O, without adding complexity to the application. MongoDB automatically balances the security data in the cluster as the data grows or the size of the cluster increases or decreases.

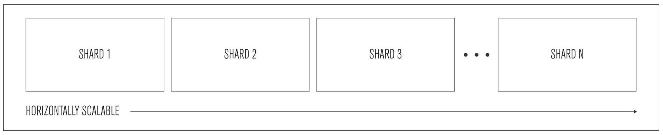

Figure 2: MongoDB scales out as security data grows

MongoDB maintains multiple replicas of the data to maintain database availability. Replica failures are self-healing, and so threat intelligence applications remain unaffected by outages or planned maintenance. Replicas can be distributed across regions for disaster recovery and data locality necessary to support the protection of global infrastructure.

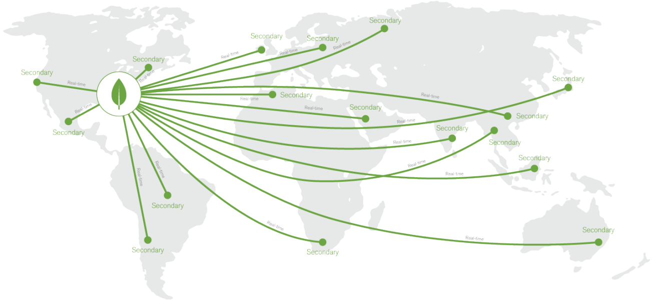

Figure 3: Global distribution of cybersecurity applications

Advanced, Real-Time Analytics

The MongoDB query language and rich secondary indexes enables developers to build cybersecurity applications that can query and analyze the data in multiple ways. Data can be accessed by single keys, ranges, text search, graph, and geospatial queries through to complex aggregations and MapReduce jobs, returning responses in milliseconds. Data can be dynamically enriched with elements such as user identity, location and last*access*time to add context to events, providing behavioral insights and actionable threat intelligence. Complex queries are executed natively in the database without having to use additional analytics frameworks or tools, and avoiding the latency that comes from ETL processes that are necessary to move data between operational and analytical systems in legacy enterprise architectures.

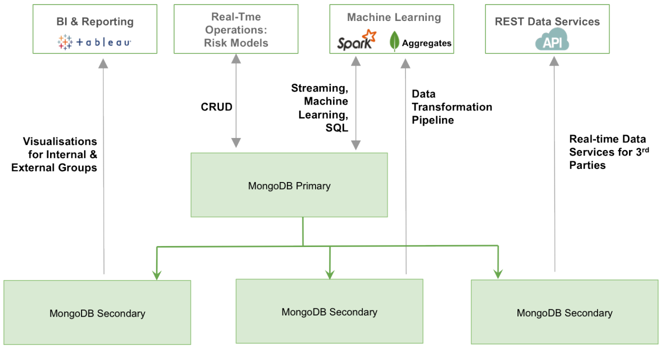
BI & Reporting
tableau
Real-Tme Operations: Risk Models
Machine Learning
Spark
Aggregates
REST Data Services
API
CRUD
Streaming, Machine Learning, SQL
Data Transformation Pipeline
Visualisations for Internal & External Groups
Real-time Data Services for 3rd Parties
MongoDB Primary
MongoDB Secondary
MongoDB Secondary
MongoDB Secondary

Figure 4: Single cybersecurity platform converging real-time operational and analytic workloads

MongoDB replica sets can be provisioned with dedicated analytics nodes. This allows the security team to simultaneously run exploratory queries and reporting against live data, without impacting nodes serving risk models to operational applications, again avoiding lengthy ETL cycles.

Native, idiomatic drivers in over a dozen languages, including Python and Scala allow data scientists to build sophisticated machine learning models using data from MongoDB. The certified MongoDB Connector for Apache Spark exposes all of Spark's libraries, enabling MongoDB data to be materialized for analysis with SQL, streaming, machine learning, and graph APIs.

Enterprise-Grade Data Protection

While based on a modern, distributed database design, MongoDB does not compromise on essential enterprise-grade capabilities demanded by mission-critical applications. This is one reason why over 50% of the Fortune 100 are MongoDB customers.

MongoDB Enterprise Advanced is the production-certified, secure, and supported version of MongoDB, offering:

- **Advanced Security**. Robust access controls via LDAP, Active Directory, Kerberos, x.509 PKI certificates and role-based access control to ensure a separation of privileges across apps and users. Data anonymization can be enforced by read-only views and field level redaction. Data in-flight and at-rest can be encrypted to FIPS 140-2 standards, and an auditing framework for forensic analysis is provided. You can learn more from the MongoDB Security Reference Architecture Guide.
- **Point-in-time Recovery**. Continuous backup and consistent snapshots of distributed clusters allow seamless data recovery in the event of corruption caused by malicious internal users or ransomware.
- **Automated Provisioning and Upgrades**. Operations teams can deploy and upgrade distributed MongoDB clusters from a powerful GUI or programmatic API. With Symantec estimating 75% of websites are exposed to security vulnerabilities due to unpatched systems, MongoDB Ops Manager can deploy patches in seconds, without application downtime.

Cybersecurity with MongoDB in Action

MongoDB has been adopted in cybersecurity solutions built by ISVs, Systems Integrators and global enterprises.

McAfee Global Threat Intelligence

Based on activity from millions of sensors world-wide and an extensive research team, McAfee Labs publishes timely, relevant threat activity via McAfee Global Threat Intelligence (GTI). This always-on, cloud-based threat intelligence service enables accurate protection against known and fast-emerging threats by providing threat determination and contextual reputation metrics. McAfee GTI instantly protects against emerging threats to reduce operational efforts and time between detection and containment.

When it became clear that its existing databases couldn't keep pace with GTI's demands, McAfee turned to MongoDB to achieve the scale, performance and flexibility required for big data analysis.

In order to provide up to date, comprehensive threat information, McAfee needs to quickly process terabytes of different data types (such as IP address or domain) into meaningful contexts: e.g. Is this web site good or bad? What other sites have been interacting with it? The success of the cloud-based system also depends on a bidirectional data flow: GTI gathers data from millions of client sensors and provides real-time intelligence back to these end products, at a rate of 100s of billions of queries per month.

As the authoritative source for McAfee threat information, MongoDB enables big data analytics and supports the real-time flow of cyberthreat data between GTI's cloud-based system and end client products. MongoDB was selected for multiple reasons: Flexible data model, supporting the native ingest of multi-structured and rapidly changing JSON data, quickly accessed by geospatial indexes. Distributed, scale-out architecture, allowing McAfee to accommodate data volumes doubling every 24 months, and replicate data to users around the world for low-latency access. ACID compliance and strong data consistency ensures incomplete or stale data is never acted upon by the threat intelligence systems.

You can learn more about MongoDB within McAfee GTI from our case study.

Lockheed Martin Cybersecurity Database

Cyber intelligence analysts at Lockheed Martin understood that storing, searching, and analyzing cybersecurity data such as network logs, file metadata, or security tool output would present unique challenges. To maintain agility with the different organizational threats such as APTs, hacktivists and cybercriminals the team had to gather metadata from a multitude of sources with diverse output formats and verbosities.

These demands posed many cybersecurity and big data challenges such as data size, normalization, query centralization, indexing, real time performance sensitivities, and the handling of temporal datasets. To address these challenges Lockheed Martin turned to MongoDB to deliver high throughput, low latency queries and complex analytics across a diversity of datasets totaling hundreds of terabytes of data stored in nodes distributed around the globe. Additionally, they found MongoDB could be configured, operated and maintained with limited database administration knowledge by associate engineers, rather than highly skilled DBAs.

You can learn more about MongoDB at Lockheed Martin from the MongoDB World presentation.

Ogilvy & Mather Delivers Security Compliance with MongoDB

Ogilvy & Mather (O&M) is one of the largest marketing communications companies on the planet with 500 offices spread across 120+ countries. MongoDB is being used for O&M's core auditing application, capturing authentication and authorization activities of all users as they access its systems. MongoDB replaced an existing relational database that was not able to keep pace with the growth in data volumes or low latency analytics capabilities demanded by users.

From events written to MongoDB, analysts are able to build an audit trail of system access, which is used by the support, compliance and security teams. From this data, O&M's teams have fine-grained visibility into who did what and when, what privileges each user has, and how each account is configured. The teams can enforce security policies such as password resets; resolve application access issues; monitor application usage by user, business unit and region; and much more. The application is write-heavy, with MongoDB ingesting tens of gigabytes of data every day, from tens of thousands of users distributed around the globe.

MongoDB is used for both data ingest, and in generating real time analytics. The company uses the MongoDB aggregation pipeline to roll up key metrics, such as snapshots of how many users are active on the system at any one time. The database is deployed in an active/active multi-data center configuration spanning two continents, thus providing disaster recovery, and local writes with read anywhere access.

By using MongoDB Enterprise Advanced, O&M gets access to 24x7 proactive technical support, Ops Manager, and the MongoDB Connector for BI for advanced data visualizations.

You can learn more about MongoDB at Ogilvy & Mather from our case study.

We Can Help

We are the MongoDB experts. Over 2,000 organizations rely on our commercial products, including startups and more than a half of the Fortune 100. We offer software and services to make your life easier:

MongoDB Enterprise Advanced is the best way to run MongoDB in your data center. It's a finely-tuned package of advanced software, support, certifications, and other services designed for the way you do business.

MongoDB Atlas is a database as a service for MongoDB, letting you focus on apps instead of ops. With MongoDB Atlas, you only pay for what you use with a convenient hourly billing model. With the click of a button, you can scale up and down when you need to, with no downtime, full security, and high performance.

MongoDB Cloud Manager is a cloud-based tool that helps you manage MongoDB on your own infrastructure. With automated provisioning, fine-grained monitoring, and continuous backups, you get a full management suite that reduces operational overhead, while maintaining full control over your databases.

MongoDB Professional helps you manage your deployment and keep it running smoothly. It includes support from MongoDB engineers, as well as access to MongoDB Cloud Manager.

Development Support helps you get up and running quickly. It gives you a complete package of software and services for the early stages of your project.

MongoDB Consulting packages get you to production faster, help you tune performance in production, help you scale, and free you up to focus on your next release.

MongoDB Training helps you become a MongoDB expert, from design to operating mission-critical systems at scale. Whether you're a developer, DBA, or architect, we can make you better at MongoDB.

Resources

For more information, please visit mongodb.com or contact us at sales@mongodb.com.

Case Studies (mongodb.com/customers)
Presentations (mongodb.com/presentations)
Free Online Training (university.mongodb.com)
Webinars and Events (mongodb.com/events)
Documentation (docs.mongodb.com)
MongoDB Enterprise Download (mongodb.com/download)
MongoDB Atlas database as a service for MongoDB (mongodb.com/cloud)

mongoDB

New York • Palo Alto • Washington, D.C. • London • Dublin • Barcelona • Sydney • Tel Aviv
US 866-237-8815 • INTL +1-650-440-4474 • info@mongodb.com