Datameer
stay curious

Sophos increases **security** with big data analytics

SOPHOS

As the complexity of IT networks has grown, the inventiveness and sophistication of threats and attacks has grown just as quickly. Industry statistics show that over a million new malware are produced daily. The good news is, even the most skilled perpetrators leave digital footprints that can be discovered with the right analytics.

Sophos began producing antivirus and encryption products nearly 30 years ago. The company helps secure the networks used by 100 million people in 150 countries and 100,000 businesses. As IT networks grow in complexity, Sophos' mission is to keep IT security simple and reliable.

Challenge

Malware is constantly being created and brought to market. Sophos products examine billions of events per day to detect malicious files. Each day, over 300,000 new potentially malicious files are reported to SophosLabs and require analysis.

The Sophos threat research analysts need to analyze these newly suspicious files to determine if they truly are threats. An automated analysis process produces a large set of metadata for each file, generating many millions of records every day. As the volume and complexity of the data grew, their old analytic infrastructure could not keep pace.

Another challenge came in a form of cloud telemetry data consisting of billions of lookups for website and file information. A particular aspect of the analysis – correlating patterns across previous analysis – had become too complex for their SQL-based database and analytic tools to manage.

Dmitry Samosseiko, Director of Global Threat Research, states, “Our SQL solution did not scale well with the growing volume of data and it was becoming costly to maintain. It couldn’t ingest all the data, so we tried to filter and aggregate the data first, but it was tricky and difficult to decide what to filter out and what to keep. The bigger the database got, the slower the queries became. Threat data correlation became a slow and painful process.”

Sophos investigated NoSQL technologies available at the time and selected Hadoop for big data analytics needs related to telemetry and threat correlation. Unfortunately, out-of-the-box Hadoop was lacking any enterprise-ready tools for creating analytic reports, dashboards, data access controls or mechanisms to easily import or export data in and out of various storage systems. It was going to take a great deal of time and expertise to build out the ecosystem--time and resources to fill the gap.

Solution

Sophos required an analytic platform that would combine the best of everything they needed – an infrastructure which leveraged Hadoop for power and scalability, while abstracting the technical complexity so their analysts could be productive. In Datameer they found:

- A scalable analytic infrastructure built natively on Hadoop
- An Excel-like workbook interface that was very familiar to analysts
- Powerful analytic functions that could be applied to data using point-and-click operations in an easy to use user interface
- Easy data connectivity and integration that could combine the variety of data sources and formats they required

With Datameer, the Sophos' team was able to ramp up quickly without becoming Hadoop experts. Threat researchers and malware analysts use Datameer to analyze data from multiple sources, including its Threat Telemetry (reputation queries), threat feeds (urls, hashes, etc.), product feedback, and other imported datasets. In addition to comparing file samples to the millions of known malware, analysts are canvassing the threat landscape to identify new malware and stop it.

"Our researchers and analysts don't want to write code or have to rely on system developers to do their job. Datameer's point-and-click functionality removes these bottlenecks."

Dmitry Samosseiko,
Director of Global Threat Research

"We needed a solution that can handle a lot of data but also allow many people in our company to use it," says Dimitry. "Our researchers and analysts don't want to write code or have to rely on system developers to do their job. Datameer's point-and-click functionality removes these bottlenecks."

Analyzing billions of rows per day, or 2-3 TB per month, analysts built algorithms in Datameer workbooks to automatically analyze log data and detect anomalous trends. As Dimitry states, "We need to look at patterns and trends that we haven't seen before, and Datameer is really valuable to us for this ad-hoc analysis and reporting."

As a Data Scientist, Madeline Schiappa is analyzing silent detections at Sophos. Using Datameer every day to run jobs analyzing basic telemetry data on billions of logs, Madeline states, "It's not just one use case, but a lot of little ones added up together that makes Datameer valuable."

"It's not just one use case, but a lot of little ones added up together that makes Datameer valuable."

Madeline Schiappa,
Data Scientist, Sophos

Results

With Datameer, Sophos no longer needed to filter or aggregate its data, resulting in better insights and faster detection of malware and security breaches. Datameer is integral to Sophos' daily malware detection in multiple use cases:

1. **Malware research and analysis.** Malware is becoming more evasive and pervasive. Sophos analyzes the characteristics of suspicious files and report the analysis outcome.
2. **Macro trend analysis.** Sophos analysts also analyze the data for macro trends of malware movements to better understand and anticipate the direction of the threat landscape.
3. **Measuring detection performance.** Analyzing statistics on the performance of malware detection to understand which protection technology is providing us the most value.

"Datameer changed the world for us," says Dimitry. "The best value we got out of it was not just ease of use but empowering any of the researchers globally to mine the data without having to learn how to code, build special UIs, or embark on a steep learning curve."

While Sophos began with threat research and malware analyst team, product and development managers also started using Datameer. It helps them understand product version usage and ramp up.

FREE TRIAL
datameer.com/free-trial

TWITTER
@Datameer

LINKEDIN
linkedin.com/company/datameer

SAN FRANCISCO 1550 Bryant Street, Suite 490 • San Francisco, CA 94103 USA • Tel: +1 415 817 9558 • Fax: +1 415 814 1243
NEW YORK 9 East 19th Street, 5th floor • New York, NY 10003 USA • Tel: +1 646 586 5526
HALLE Datameer GmbH • Große Ulrichstraße 7 – 9 • 06108 Halle (Saale), Germany • Tel: +49 345 2795030
SINGAPORE Datameer Singapore Pte Ltd • 03-20 Galaxis 1, Fusionopolis Place • Singapore 138522, Singapore • Tel: +65 6809 1157
HONG KONG 12/F International Commerce Centre, 1 Austin Road, West Kowloon, Hong Kong • Tel: +852 2824 8646