

# Redefining security analytics with Chronicle Backstory



# Table of contents

Understanding the salient needs and gaps in modern security analytics .....	3
--	---

Threat and complexity

How to make sense of it all?

New types of tools needed

Scale

Speed

Simplicity

Cost

Coverage

Chronicle Backstory: a different approach to security analytics .....	8
--	---

a. Data ingestion

b. Data analysis

What is the threat?

What is it doing?

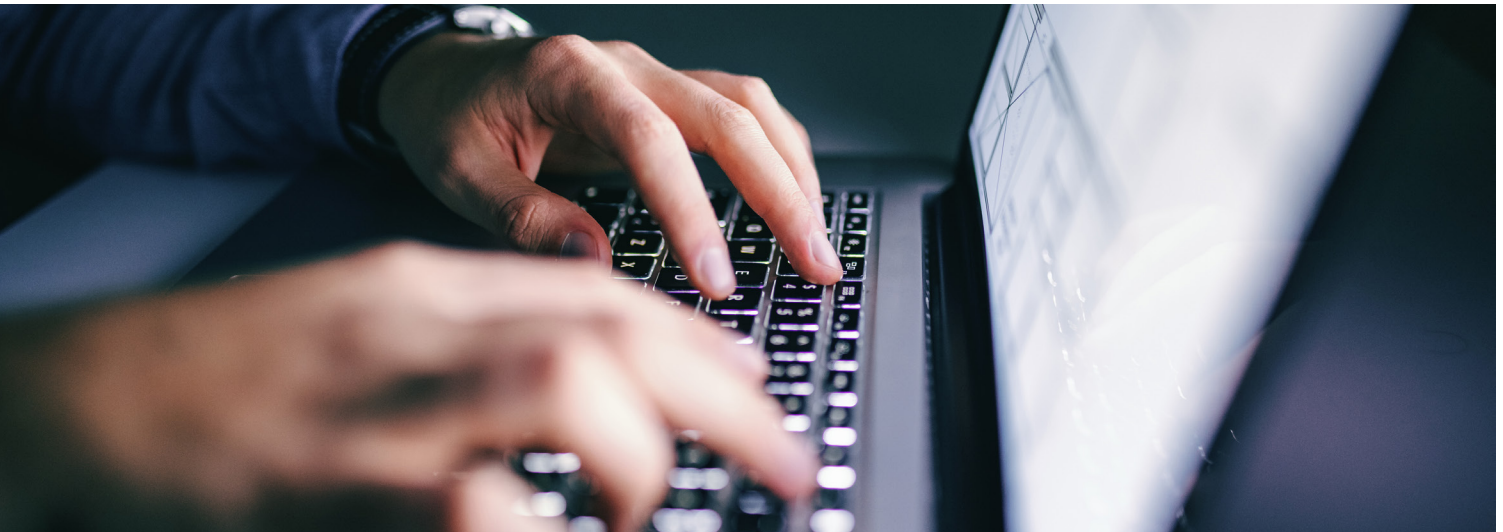
Does it matter?

How to respond?

c. Security and compliance

Summary of capabilities and benefits

Chronicle Backstory is a global security telemetry platform for investigation, hunting and detection of threats within your enterprise network. Backstory makes security analytics instant, easy, and cost-effective.



# Understanding the salient needs and gaps in modern security analytics

## Threats and complexity

Today's security professionals face not only an ever-expanding list of threats, old and new, but also an excruciating choice of security approaches and tools. Nearly 2000 security vendors are trying to sell to large enterprises and small businesses. Most organizations have already invested heavily in cybersecurity solutions. From firewalls to proxies and endpoint protection to DLP, the average firm has more than 40 types of security technologies in place, each generating telemetry that grows every year.

With employees, contractors, partners, and customers all accessing online business processes that were once hidden in a back office, corporate network traffic has also increased significantly. There is more noise than ever before, hiding more attacks than ever before – with greater impacts than ever before.



## How to make sense of it all?

Security analytics promises to help analysts make sense of this data, to find useful signals in the noise before it's too late. For most organizations, however, an effective security analytics solution is an expensive and complex exercise in systems integration, with heavy IT operations support required simply to keep the system up and running as it grows. As CIOs migrate corporate IT to the cloud, CISOs roll out advanced threat protection such as EDR and network traffic analyzers to protect the pieces that remain under their control.

In theory, a SIEM or centralized log management product consolidates and correlates all of this information, but in practice, SIEMs buckle under large data volumes. Moreover, high-volume telemetry from EDR systems is rarely fed into a SIEM. If high volume telemetry is ingested, it's typically only retained for a few weeks, if at all.

## New types of tools needed

While it sounds counter-intuitive that we need another type of a tool, the world has changed dramatically and many existing security tools did not evolve fast enough to maintain relevance.

Today, organizations still operate legacy systems, have vast on-premise IT presence, but also a large cloud presence, often across multiple cloud providers. The types of security telemetry they collect expands, and the volumes grow.

### New requirements:

- Scale
- Speed
- Simplicity
- Cost
- Coverage

Backstory, a global security telemetry platform designed to help those same analysts understand threats that exist in their own corporate networks. With Backstory, analysts can hunt for threats and investigate petabytes their data to get answers in milliseconds. And all that without paying for data volumes.



## Scale

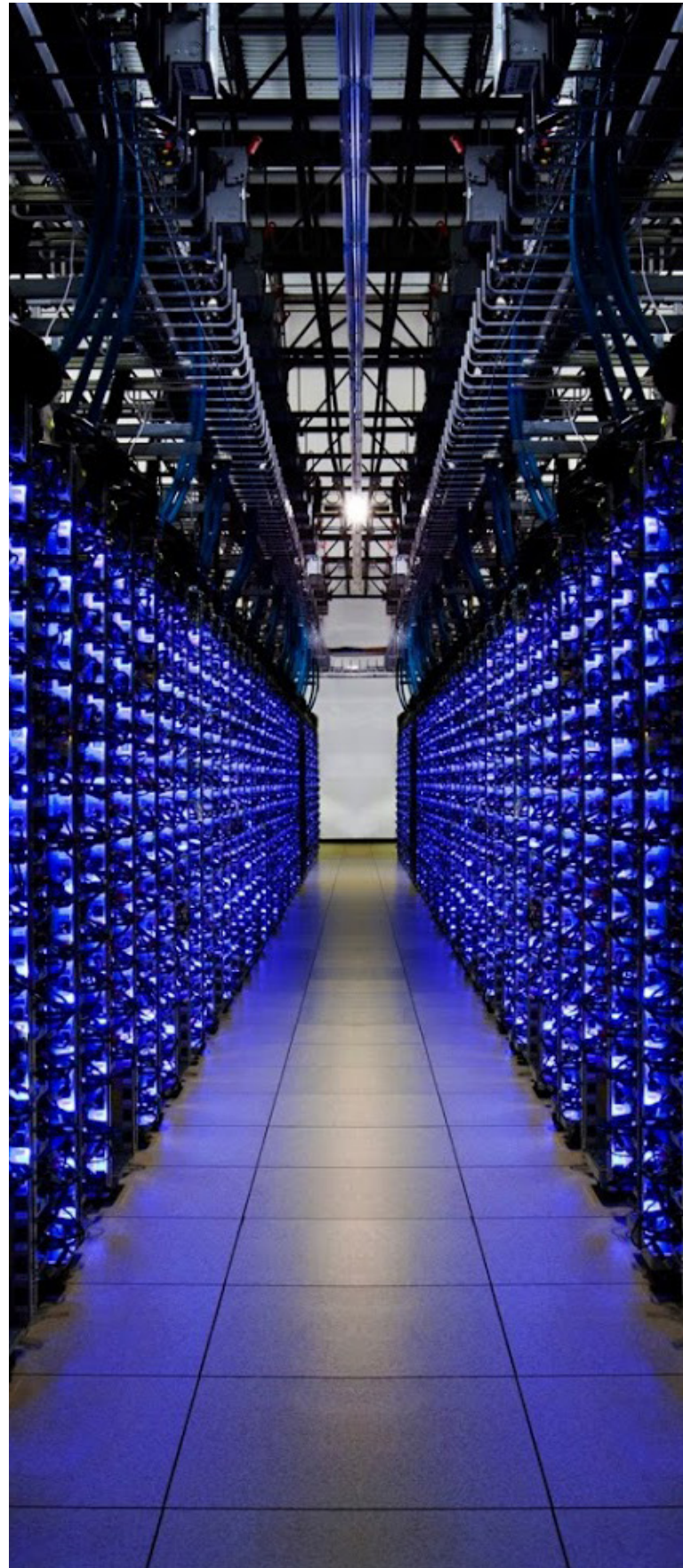
Data volumes have increased dramatically over the last decade. Using a popular SIEM metric, events per second (EPS), a large environment 15 years ago may have had “thousands” of EPS. Today, it is not uncommon to see environments with hundreds of thousands and more events per second, pushing into high daily terabyte telemetry volumes.

Admittedly, increased focus on security monitoring is one of the primary reasons. We just have more detection tools. Some of the newer tools, such as EDR, are also more verbose.

In addition, more digitization and just more IT infrastructure has pushed the log volumes way up. Both cloud and on-premise logs and a growing range of security telemetry types from traditional windows logs to cloud VPC flow logs.

Now, apart from higher volumes of data, we also have longer retention periods of data, and not merely for compliance. Breaches may be discovered 200-300 days after the initial compromise and this calls for investigative data to be retained.

As a result there are many orders of magnitude greater security telemetry that need to be ingested and analyzed. This either breaks the legacy tools or breaks their economic model. In the latter case, the challenge is as severe. Gartner back in 2009 said that “security is becoming a big data analytics problem” but today customers are the ones paying for it. Economic scaling is as important since nobody wants to buy \$10m worth of hardware to run their security data search tools.



## Speed

How hard is it to search the entire internet in under a second? Well, it has not been hard since the day Google launched back in 1999. Then why do people tolerate waiting for minutes if not hours while searching their security telemetry?

Today, organizations deal with too many alerts and confirming alerts is anything but fast for many. As organizations increasingly shift focus to threat detection and response, there's one issue that seems to get worse over time: alert triage. Prioritizing security alerts has been a critical function for security teams since the early 1990s — so why does it remain such a challenge decades later? Today we have more logs to search and more alerts to confirm.

Worse, many of the alerts are false positives, and the recent advances in machine learning for threat detection has, sadly, contributed to the problem. False alarms and ambiguous alerts all need investigating - and fast.

Finally, there is also a real need for performance around incident investigations which are too complex and slow today. Search EDR data from 50,000 machines?

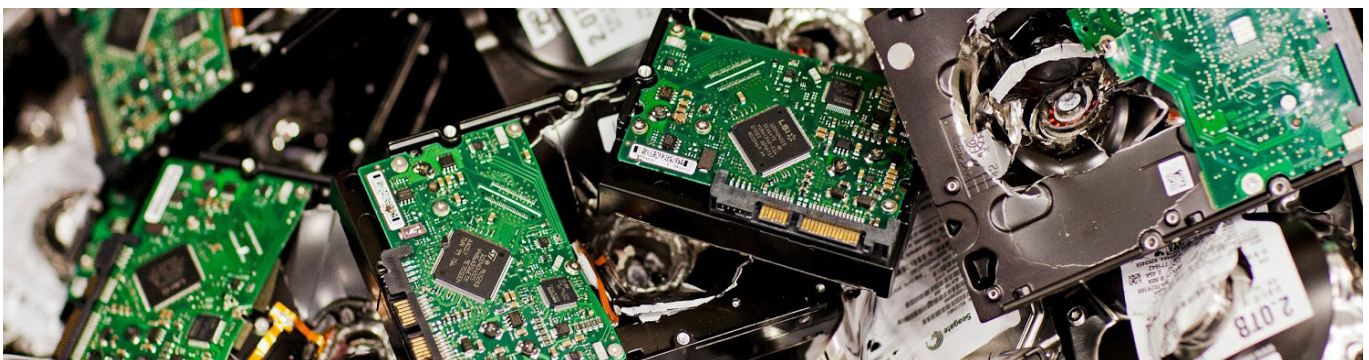
## Simplicity

It is often said that SMBs need simple tools for everything in security. But guess who else does? Large enterprises! Extreme staff shortages, and in fact worse talent shortages in security has led to many security analysts and even threat hunters being hired without deep experience. They all need simple tools to detect, confirm and investigate without stress - and without a 3 day class on a new search language.

Furthermore, using threat intelligence can be simple - but it almost never is. Threat intelligence feeds are supposed to add more context, but are often too noisy or redundant and cause more static than they eliminate. Over time, the indicators in those embedded feeds and signals may change. For example, a domain first seen a year earlier and judged as "good" in the threat feeds may begin hosting malware; the indicator turns "bad" in the feed.

Thus there is a need to automatically and instantly re-calculate any customer activity to that now-bad indicator and alert analysts about all machines that have ever communicated with this domain. Note that the focus on simple and automatic here.

Similarly, having the threat data, security telemetry enriched and connected makes alert triage much simpler. Given such a system, alert triage should not require deep knowledge of threats and the environment.







## Cost

Does a tool scale? This is a useful question to answer and security tools should cover the data needs of an organization. However, does the tool scale without an exponential increase in hardware cost and commensurate increase in complexity? Or, if deployed in public cloud, do they scale without you being stuck with a 7 digit cloud provider bill for collected and stored data, that keeps growing and growing? In fact, even some open source tools with a license cost of free has led to huge cloud bills if used for security telemetry collection without adequate planning.

Security analytics tool should not be priced based on collected data value. This license models create a disincentive to send all the data, and causes the customer to make decisions not based on security

Cost is ultimately is not only about low price, but about economic scalability and predictability - can the tool grow with you without breaking the bank. Predictable pricing is often more important than low pricing for some scenarios.

## Coverage

It is very clear that a modern security analytics tool should use a cloud backend. But it cannot only support assets deployed in the cloud since this is simply not the reality at most organizations today. Hence the ideal solution would be deployed in the cloud, but be able to analyze the telemetry from both cloud and on-premise sources, modern and legacy tools, systems and application for a wide range of security use cases.

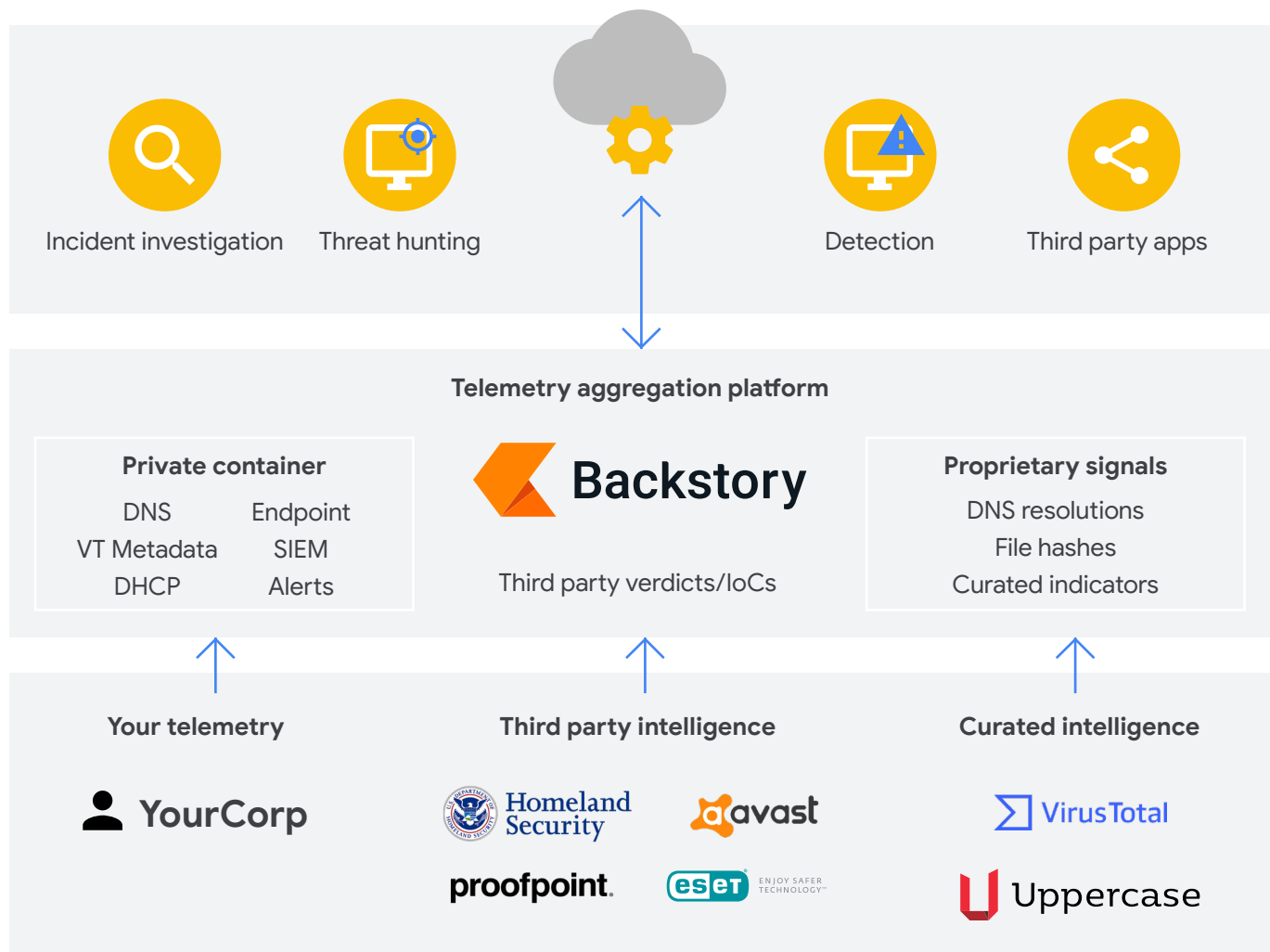
Now, it is worth clarifying that a cloud native toolset has plenty of advantages over something built for the on-premise world and then "lifted and shifted" to public IaaS. Auto-scaling, elasticity and deployment automation of cloud native tools cannot be achieved with legacy software deployed in the cloud.

# Chronicle Backstory: a different approach to security analytics

Backstory is a cloud service, built as a specialized layer on top of core Google infrastructure, designed so that enterprises can privately retain, analyze and search the massive amounts of security and network telemetry they generate today. Backstory normalizes, indexes, correlates, and analyzes the data – against itself and against third party and curated threat signals – to provide instant analysis and context regarding any risky activity.

We can drill down from this description into some of the platform's key functions:

## Backstory platform architecture





## a Data ingestion

Backstory can ingest a variety of telemetry types, through a variety of methods.

- The most common is the Backstory Forwarder, a lightweight software component, deployed in the customer's network, that supports syslog, packet capture, and existing log management / SIEM tools. The Forwarder can be installed on Windows platforms and also as a container on Linux platforms.
- Backstory also offers an ingestion API that enables logs to be sent directly to the Backstory platform, eliminating the need for additional hardware or software in customer environments. MSSPs and technology partners can leverage the Backstory ingestion APIs to forward raw logs as well as structured logs that adhere to the Backstory normalized format, directly to the Backstory data pipeline. The Backstory Ingestion API is a RESTful API with a JSON payload and API keys are used to authenticate calls.
- Additionally, Backstory can also pull telemetry from other cloud services. For example, some EDR solutions push endpoint logs to an Amazon S3 bucket, and Backstory can be configured to ingest logs directly from that location. In contrast, Carbon Black's EDR uses an event forwarder to ingest telemetry directly to Backstory. Simply put, there are many ways for customers to upload their telemetry.
- Backstory also integrates with 3rd party cloud APIs to facilitate ingestion of logs. This includes sources like Office 365 and Azure AD.

Regardless of the ingestion path, a key architectural goal during ingestion is high throughput and this is partly achieved by first writing ingested logs in the format received to disk and processing (normalizing, indexing etc.) them thereafter. This approach also has the advantage of raw log access for correction of any parsing or other errors at the processing stage.

A critical advantage of a cloud-native security analytics is that logs can be collected and then parsed (or re-parsed once improved parsers become available) in the cloud. Those who operated traditional SIEM tools have long lamented about the collector updates and changes. With Backstory, the raw logs are collected and retained and then can be "magically" parsed, normalized and enriched as needed.

Finally, the Backstory security analytics platform supports more than logs. EDR data, network traffic captures (such as those from Zeek and other capture tools) can be collected and retained - at no extra cost to a client beyond the initial investment.

## b Data analysis

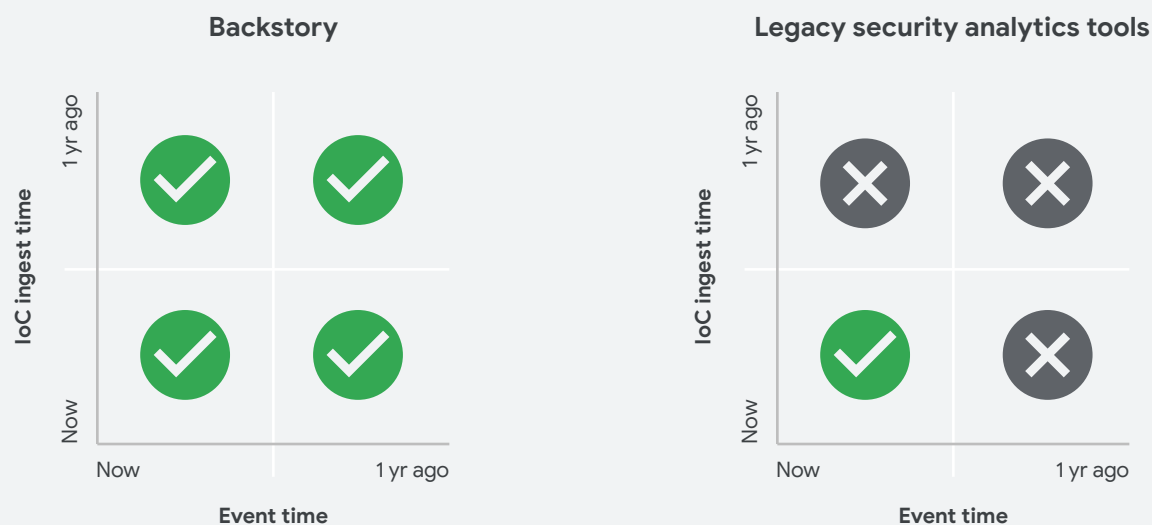
The analytical capabilities of Backstory are delivered to security professionals as a simple, browser-based application. Many of these capabilities are also accessible programmatically via read APIs and can be triggered from other security tools. At its core, the purpose of Backstory is to give analysts a way, when they see a potential threat, to determine **what it is, what it's doing, whether it matters, and how best to respond**. Next, let's understand Backstory's analytical capabilities using a real life example:

### What is the threat?

Backstory combines the scale of its core Google infrastructure backend with unique data enrichment to surface all IoC matches automatically and continuously.

For example, if a threat feed just informed Backstory about a new APT network domain, the Backstory Enterprise Insights dashboard will instantly surface every hostname that accessed that domain going back a full year, regardless of the data volume.

#### Automated, continuous, retroactive IoC matching



Similarly, Backstory continuously enriches the incoming event stream by correlating IPs to hostnames so that analysts have full information; instantly; and without the need to write complex queries.



DOMAIN	CATEGORIES	SOURCES	ASSETS	CONFIDENCE	SEVERITY	IOC INGEST TIME	FIRST SEEN	LAST SEEN
intuitstatistics.info	APT-28//SOFACY	OSINT Watchlist	john-ledue-laptop	n/a	n/a	17 hours ago	10 months ago	10 months ago
artem.sataev.com	VirusTotal YARA Rule Match	VirusTotal YARA Feed	winsrv01-a	High	High	5 months ago	a year ago	22 minutes ago
sipapals.servehalf.life.com	APT-Activity	US DHS AIS	todd-fields-pc	n/a	n/a	a year ago	9 months ago	22 minutes ago

With other security analytics solutions, data is rarely ever kept in a hot state for a full year due to storage and license costs. Even in cases where customers are willing to bear higher storage costs for fast access, retroactive IoC matches are manually initiated by an analyst and take an inordinate amount of time due to infrastructure (compute) costs. Furthermore, these searches may simply yield IP address matches which in turn have to be manually correlated with DHCP data to uncover the hostname that leased that particular IP address at that time. In short, even a common and seemingly simple security operations task of determining hostnames that have reached out to a malicious domain consumes significant time, infrastructure, and expertise (proprietary search language proficiency).

## What is it doing?

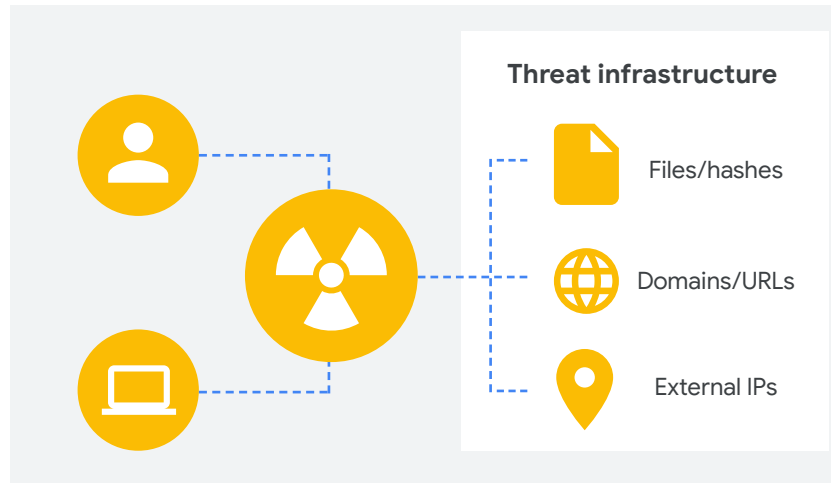
Following the above example, let's drill down into the IoC match for the US DHS (Dept. of Homeland Security) threat feed which is one of several threat intelligence sources embedded into the Backstory platform. Once an asset match is found for a known malicious domain (in this case Todd Fields' PC), analysts generally want to understand what was happening on that asset around the time of access to look for corroborating signals that confirm whether the outreach led to compromise.

### Common questions an analyst would want to answer include:

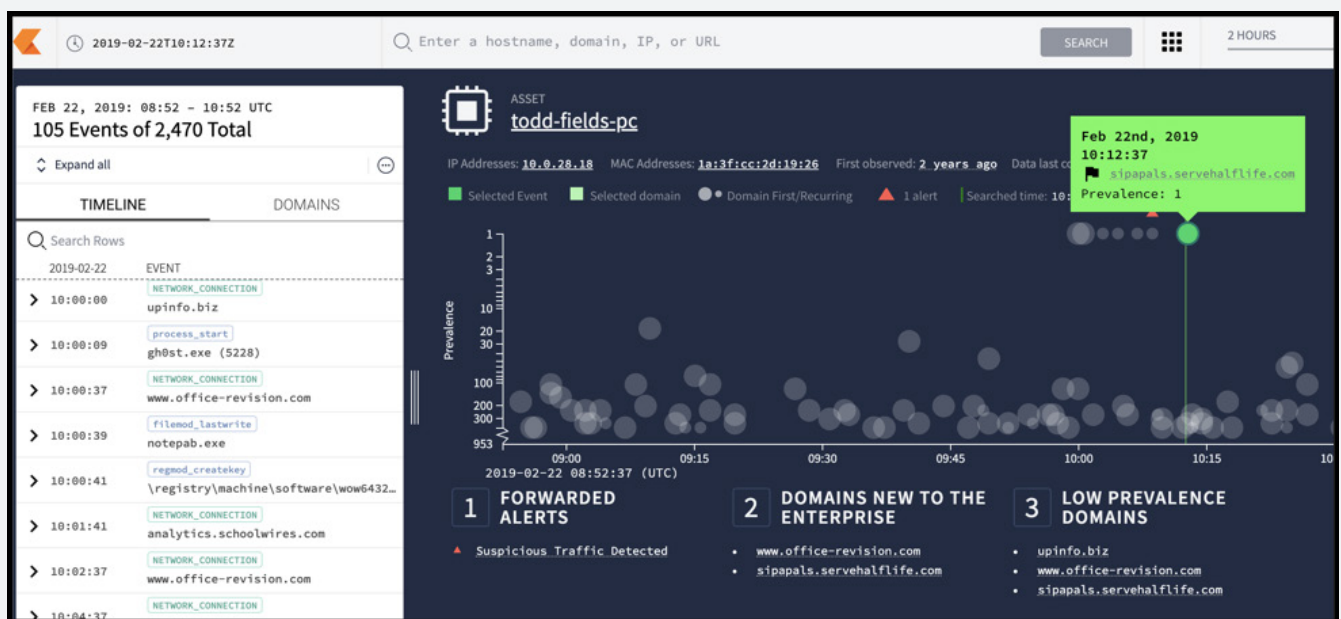
- What other domains did this asset access around the same time? Were any of those other domains malicious, rare or entirely new to my enterprise?
- What was the pattern or frequency of access to the domain in question as well as other rare domains. For example, was the outreach representative of beaconing behavior?
- Were any alerts fired around the same time for the given asset from Firewalls, CASB, EDR or other security solutions?
- Do endpoint events reveal activity that suggests the malicious domain access was followed by a successful compromise?



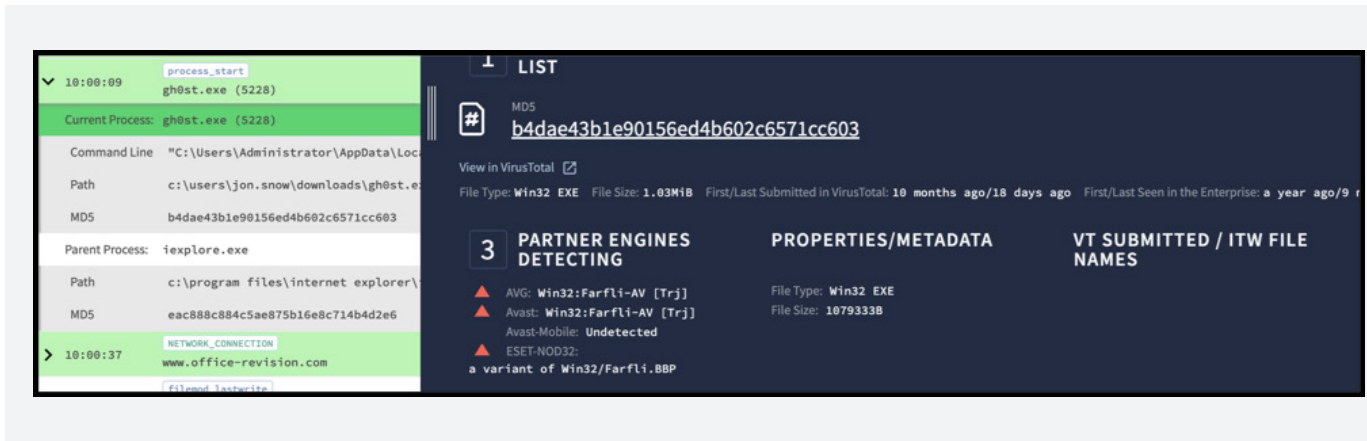
Modern threats lie at the intersection of targeted assets, users and threat campaign infrastructure (domains, URLs hosted on specific external IP addresses and malicious files deployed from there). In Backstory, curated views for internal assets and users as well as threat campaign infrastructure (domains, URLs, file hashes, external IPs) operate on a Unified Data Model which enables analysts to investigate threats quickly, intuitively, and without having to rely on a proprietary query language.



In this example, we drill down into the Asset View for the endpoint that accessed the APT domain and can quickly see a beaconing pattern to rare (low prevalence) domains graphically as well as a suspicious process in the timeline panel (ghost.exe) that is launched immediately after. Subsequently, a file (notepad.exe) is written to disk and a new registry run key is created to gain persistence across reboots. Backstory uses stateful URLs so at any time an analyst can capture the exact set of filtering conditions and pass it on to the next analyst for continued triage across shifts or escalation tiers. Similarly, the timeline of events can be easily filtered on data model dimensions and attributes and the relevant set of events can be easily exported.



With traditional security analytics and logging platforms, painting this summary picture across network and endpoint events for a single asset can take days of complex queries that have to be created and run in sequence after which results have to be combined.

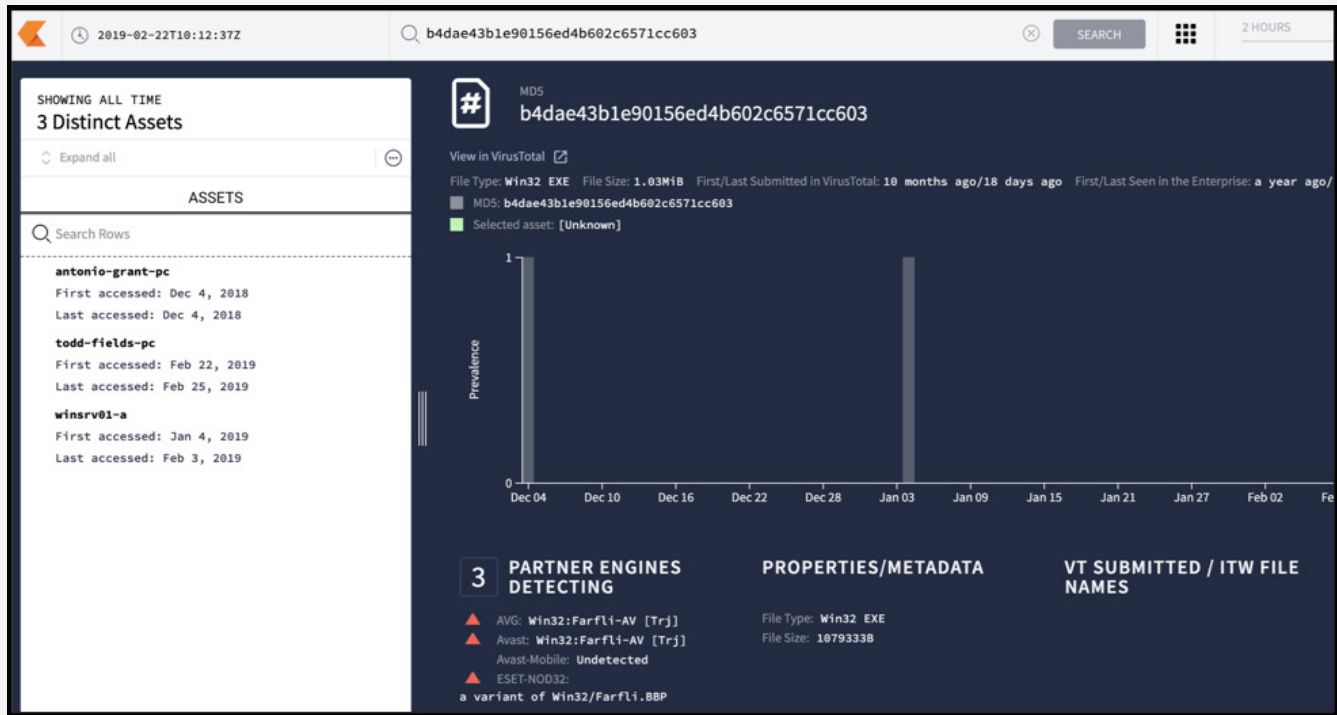


## Does it matter?

Security telemetry alone rarely provides the full picture needed to hunt, investigate or detect threats. Context is critical to giving analysts the ability to prioritize real threats and dismiss false positives. In this example, we have an endpoint (todd-fields-pc) that did reach out to a known malicious domain and is likely compromised by a malicious file (ghose.exe) that was downloaded.

Each curated view provides relevant context and insights to aid the investigation or hunt. The Asset View shows insights about domains accessed by an asset that are rarely seen in the enterprise (low prevalence); domains that are new to the enterprise; alerts from other security tools; known vulnerabilities for the given asset and more. Selecting the malicious file in the timeline panel surfaces the hash and known verdicts from Avast's 400 million consumer AV endpoint agents, and ESET's AV results – as well as embedded VirusTotal metadata. By clicking into the hash itself we pivot out from the Asset View and into a Hash View which tells us that two other assets also have a file with the same fingerprint, indicating a potentially wider compromise.





To evade detection, threat actors constantly change the domains, URLs, files and to a lesser extent - the IP addresses used in their campaigns. VirusTotal is perhaps the largest repository of malware samples with the unique ability to explore artifacts based on their relationships.

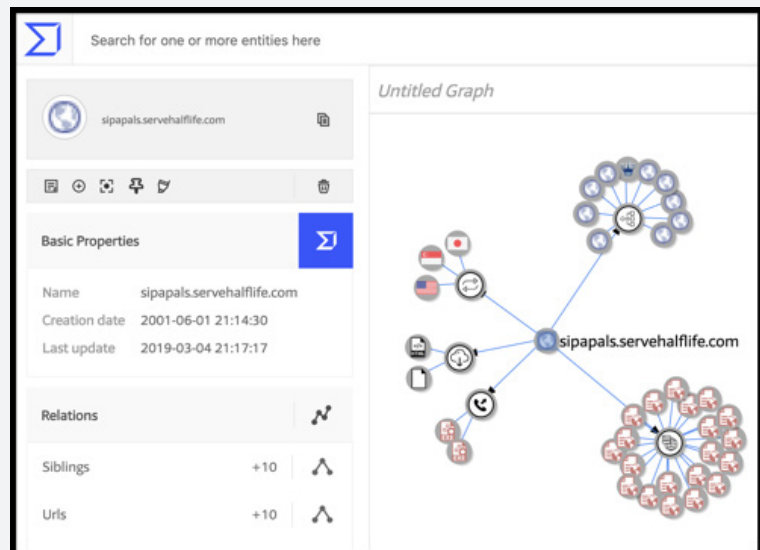
This is why VirusTotal is integrated into Backstory not just to provide summary context in the Domain, URL, External IP and Hash Views but also to enable analysts to pivot into VirusTotal (contextually) and uncover related IoCs that may be part of the same campaign. For example, an analyst investigating the APT domain in this scenario within Backstory's Domain View, can click into "see details in VirusTotal" in the VT Insights panel and start exploring related artifacts. Sure enough, there are several other sub-domains, URLs, and files that are likely tied to this campaign. The analyst can now start looking for the existence of variants of the initial threat that has been detected.



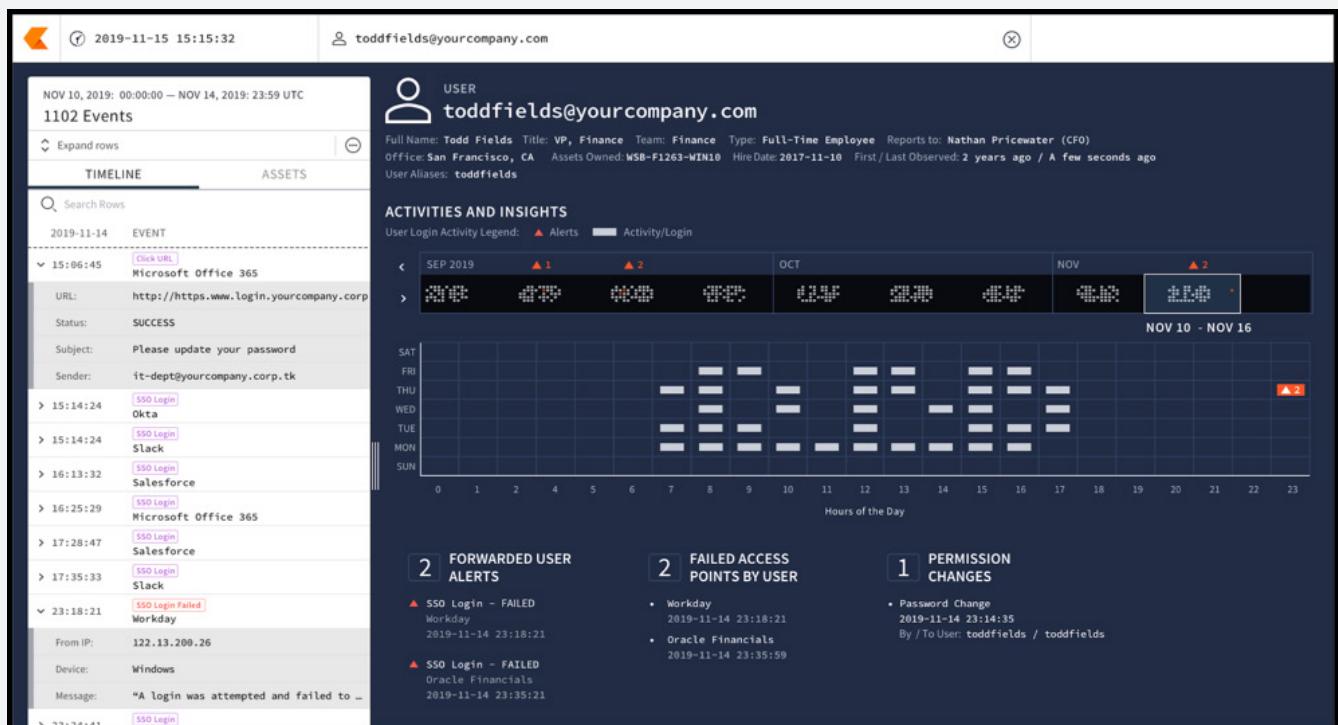


Analysts can also pivot to the Backstory User View which provides context on who the user is (from directory sources like Active Directory) and whether their behavior is anomalous. User View operates on distinct data sources and events such as Windows (login events), Office 365, and Azure AD. In this example, we see that Todd Fields is a VP of Finance and that there are a few user centric security alerts associated with his account. The graph tells us that while Todd normally only logs into his account between 7am and 5pm, after the compromise of his laptop there has been anomalous account activity past 11pm.

The combination of his role, recent account related alerts from 3rd party SaaS applications (which may also come from a CASB such as Netskope), and anomalous access by time of day suggest Todd's account may have been compromised. The likelihood is only higher because we know Todd's endpoint has been compromised by an APT that successfully downloaded a malicious payload to his laptop



To summarize, in a matter of seconds, the analyst has seen a threat, understood its behavior, gained context and judged the severity, and generated a list of all affected machines that need remediation. None of this required a single query to be written, and all of this can be performed with a single console.



## How to respond?

Large organizations with mature SOCs have documented playbooks for investigation and remediation of threats for some time now. SOAR or Orchestration technologies that automate these playbooks are now growing in adoption. Across the board, the sheer difficulty of hiring trained security professionals has increased reliance on outsourced or managed security services. Backstory's Read APIs expose some of the analytical capabilities described in the scenario above and enable enterprises and MSSPs alike to integrate Backstory findings into their security playbooks for automation and into other technologies such as ticketing systems or dashboarding tools.

The Backstory Read or Search API uses the OAuth 2.0 protocol for authentication and authorization. As a simple example, the ListAssets method returns assets that have accessed a specified artifact (a domain for example) within a specified time period, including the first and last time those assets accessed the artifact.

An example of the sample request and response for the ListAssets method follows:

## Sample request

```
https://backstory.googleapis.com/v1/artifact/listassets?start_time=2019-10-15T00:00:00Z&end_time=2019-10-17T00:00:00Z&artifact.domain_name=www.google.com&page_size=1
```

## Sample response

```
{assets: [{asset: {hostname: "rick"},
  firstSeenArtifactInfo: {artifactIndicator: {domainName: "www.google.com"},
    seenTime: "2018-09-14T20:10:27.157476Z"},
  lastSeenArtifactInfo: {artifactIndicator: {domainName: "www.google.com"},
    seenTime: "2019-10-24T22:04:04.327829Z"}},
  {asset: {hostname: "morty"},
  firstSeenArtifactInfo: {artifactIndicator: {domainName: "www.google.com"},
    seenTime: "2019-06-17T21:22:44.812738Z"},
  lastSeenArtifactInfo: {artifactIndicator: {domainName: "www.google.com"},
    seenTime: "2019-10-24T20:40:54.846676Z"}}}]}
```

## Security and compliance

As a specialized, private layer built over core Google infrastructure, Backstory inherits compute and storage capabilities as well the security design and capabilities of that infrastructure (Backstory's "Core Infrastructure").

The underlying design of our Core Infrastructure is described in more detail in a [Google whitepaper](#).

Chronicle Backstory  
certifications and attestations:

- SOC 2 Type 2 and SOC 3
- ISO 27001
- HIPAA BAA



## Summary of capabilities and benefits

Feature	Description	Benefit
Continuous enrichment	Automated IP to host correlation	<ul style="list-style-type: none"> <li>• Faster time to investigate</li> <li>• Greater analyst productivity</li> </ul>
Context and insights (threat / IoC, vulnerability, asset, user, file/process)	<ul style="list-style-type: none"> <li>• Embedded threat intelligence sources (Proofpoint, DHS AIS, OSInt, Avast, ESET)</li> <li>• Customer provided threat intelligence sources</li> <li>• Asset, Vulnerability, and User context</li> <li>• Derived Insights</li> </ul>	<ul style="list-style-type: none"> <li>• Faster time to investigate</li> <li>• Greater analyst productivity</li> </ul>
Read APIs	High performance APIs that expose Backstory functionality to downstream enterprise and MSSP SOC playbook stages and tools (ticketing, orchestration, dashboarding)	<ul style="list-style-type: none"> <li>• Automation of SOC playbooks</li> <li>• Integration with MSSP portals</li> <li>• Faster time to remediation</li> </ul>
Ingest APIs and Unified Data Model	High throughput APIs that enable sending data directly to the Backstory data pipeline without the need for a Forwarder	<ul style="list-style-type: none"> <li>• Faster time to value</li> <li>• Zero deployment footprint</li> </ul>
Raw Log Scan	<ul style="list-style-type: none"> <li>• Access to all unparsed fields</li> <li>• Search any raw security telemetry</li> </ul>	<ul style="list-style-type: none"> <li>• Faster onboarding of all security telemetry</li> </ul>
Security / Compliance	<ul style="list-style-type: none"> <li>• Adherence to Google Cloud common controls</li> <li>• SOC 2 and SOC 3</li> <li>• ISO 27001</li> <li>• HIPAA BAA</li> </ul>	<ul style="list-style-type: none"> <li>• Documented, stringent controls to protect your data at every layer</li> <li>• Key attestations and certifications</li> </ul>