



The Security Division of EMC

RSA Solution Brief

# RSA enVision LogSmart™ Internet Protocol Database



The global enterprise puts tough demands on a security information & event management (SIEM) system. It must capture all event data – not just most of it – from worldwide sites, large and small. It must give alerts, not only for immediate and obvious attacks, but also for the stealthy variety which take place over time, anywhere on the network.

The ideal SIEM system must keep all the captured data readily, securely and quickly available through an easy-to-use centralized interface for flexible historical, trend and audit reporting – even reports not yet anticipated. But the system must not touch the data in a manner that would invalidate it as evidence.

It must be fast, so as not to lose data or delay reporting; but it must also be cost effective. It should not create a data volume explosion and thereby greater storage purchases. And with all that, it should be simple, standard, open and not place further burdens on the IT staff.

This brief describes the RSA enVision LogSmart Internet Protocol Database (IPDB) architecture. It details the advantages of the distributed and agent-less aspects of the system, as well as how the Logsmart IPDB architecture makes the system much smaller, faster, and more complete than one that relies on a relational database to store message data.

RSA's third generation LogSmart IPDB architecture is based on an innovative yet conceptually simple approach.

The RSA enVision LogSmart™ IPDB turns several traditional relational database management system (RDBMS) notions on their heads. First, the Logsmart IPDB architecture is designed to work efficiently with unstructured data natively without any pre-processing or data normalization. In the case of event logs, the architecture uses the raw event logs themselves to form a database with no further overhead required.

Next, the Logsmart IPDB architecture is highly write-optimized. Even though the number of reads will far outweigh the number of writes, by optimizing the writes any subsequent reads will become far more efficient and save overall I/O load on the host system. Thirdly, the architecture utilizes a write-once-read-many approach to the data itself. This assures that once data is committed to the database, it can never be altered. Fourth, unlike most data schemas used in RDBMS-based solutions, the Logsmart IPDB architecture is designed specifically to store internet protocol-based information and it always stores each source element (IP address, MAC address, hostname, etc) in a separate container. IPDB never co-mingles multiple elements together into a common row or table like a RDBMS would; this enables fine-grained access control. Finally, since the Logsmart IPDB

architecture is designed to store unstructured data natively, it only parses the contained information on the way out of the database, when requested, instead of parsing 100% of the data on the way in, saving precious machine resources and preserving the native structure of the incoming data.

Add strong security, authentication and compression techniques, and the Logsmart IPDB architecture more than measures up to the demands of the global enterprise.

---

## Challenges

---

Any SIEM system faces challenges in the enterprise network – log message volume and rate, retention for reporting, analysis, data safety, system compatibility, scalability, resource use and performance. In designing three generations of systems to deal with these challenges, RSA has evolved an approach that handles them all successfully.

### Log Message Volume and Rate

Log messages are generated for each event in every network resident device, including those dedicated to computing and applications, networking, storage and security. The volume of log messages varies with time of day, often peaking at mid-morning and again once or twice in the afternoon. The SIEM system must be able to handle peak loads of incoming log messages without loss. The system must catch every message and also typically poll up-to thousands of devices three or four times per minute to collect event logs. It must understand each message it receives and handle and report on them all in a centralized manner, whether it actually uses a centralized or a distributed architecture.



Syslog and SNMP messages from UNIX-based systems, firewalls, routers and switches use push technology and UDP. This lightweight protocol uses few resources, but it has the disadvantage that the sender does not find out whether the message reaches its destination and therefore does not retransmit on failure. Thus the SIEM system must listen with perfect accuracy and catch every UDP message. Otherwise the message could be lost permanently.

In contrast, Windows®-based devices, like application databases and web servers, write event logs to the local disk. The SIEM system must authenticate itself and collect the event log with pull technology using TCP/IP. This approach detects failure and retransmits, but it has higher overhead than does UDP.

The Logsmart IPDB architecture handles both the push technology of UDP and the pull technology of TCP/IP. It listens with such high accuracy for messages that it captures over 99.9% percent of them, even at peak transmission volumes. LogSmart IPDB technology uses a distributed architecture that, among other benefits, allows local message collection to continue normally during WAN outages.

The Logsmart IPDB architecture does not employ agents on devices. Agents are designed for low CPU usage to minimize their impact on running devices, with the result that their transmission rates can be too low to send all log messages in a timely way. Agents filter data and modify it to a standard set of fields, thus “touching” the data and discarding some of it. Agents also are prone to being turned off by system administrators who are trying to correct problems, resulting in unknown “blind spots” across the network. The agent-less LogSmart IPDB architecture avoids both issues.

### Retention for Reporting

The message data is itself metadata and can be used directly as an object-oriented database if it can be well defined and described and if retrieval for reporting can be done efficiently.

Among the greatest challenges for a typical SIEM system is parsing every message that comes in and retaining all desired information at speeds that can keep up with peak loads. Parsing each message on input is not only slow, but also discards information and retains only a lowest-common-denominator

subset of message contents. It is quite possible, however, that the unsaved information might be required later for audit reports.

The third-generation Logsmart IPDB architecture design does not parse log messages on input, but retains all of them in original form close to their origins and retrieves and parses them as needed only on output for reporting. Taking this approach, the system can easily handle high data input rates and does not arbitrarily discard information to fit a limited RDBMS schema. Furthermore, the message payload is untouched— an important factor in preserving the integrity of evidence.

### Analysis

A SIEM system needs to provide two types of analysis:

- Real-time analysis to detect attacks and provide alerts
- Historical analysis for trend and audit reporting

For accurate real-time alerting, the system must capture *all* log messages. It then must retain all data that might possibly be needed for later historical analysis. But a further type of security monitoring is also needed that too often is ignored. Sophisticated attackers do not (except in movies) hack into a corporate network all at once, but often mount what is called a low-and-slow attack by probing for one seldom-used port, then waiting days or weeks, probing another, and so on until an opening is found. Such attacks are unseen by the majority of SIEM systems, which are tuned only for the dramatic onslaught of day-one virus or worm attacks. The Logsmart IPDB architecture, however, automatically sets sensitivity to high on seeing more than three attempts on any two or more ports, and is thereby able to warn of stealthy attacks conducted over time as well as of the rapid ones.

### Data Safety

An important aspect of storing information about an enterprise network is keeping that information secure. As log message descriptions are prepared for storage in the IPDB, they pass through three steps that guard them permanently against tampering: authentication, lossless compression and encryption.



### Compatibility

Putting in a SIEM system should not require training an already strained IT staff in yet another new set of proprietary protocols, data formats, etc. The Logsmart IPDB architecture is an open, non-proprietary system that makes use of well understood mechanisms such as the Windows® file system and can be integrated on any common network. It understands output from over one hundred types of network devices and also handles messages from unknown devices without data loss.

### Scalability

Scalability refers to the ability of a system to grow smoothly in size with the growth of the enterprise it supports. Logsmart IPDB distributed architecture easily supports this growth and is designed to include data capture from small remote sites with no loss of reporting efficiency or completeness. The design supports collection, retention and reporting of log message data. It also allows for future development of capabilities to store and report on lower-level types of data.

### Resource Use

The phrase “resource use” often conjures up images of endless disk-drive purchases, but enterprise resources actually come in several forms: storage and related network hardware – and also IT time, skill and effort to install, configure and troubleshoot myriad devices and their interactions. The Logsmart IPDB architecture minimizes resource use in storage and hardware – as well as undocumented and often unrecognized wear and tear on IT team members from the labor involved in integrating and maintaining new systems (especially those using agents) .

The Logsmart IPDB architecture provides storage savings due to a low data explosion (DE) factor, when compared with RDBMS-based SIEM systems. Bringing 1K bytes of raw data into a traditional relational database results in 12K to 15K of new storage – caused by construction of tables and other overhead – despite data loss in the process. In contrast, the Logsmart IPDB architecture requires less storage than the incoming message data: its DE factor is 0.29, including all IPDB and analysis overhead as well as the compressed data itself. Further, it requires much less time and effort to install and maintain an IPDB system than it does for an RDBMS based system.

### Performance

SIEM systems that must work around the issues of a large RDBMS are time-restricted in how they write data. The Logsmart IPDB architecture is optimized for a write-once read-many approach, with the philosophy that writing correctly the first time will enable efficient multiple reads later. The speed with which the IPDB handles data results partly from this optimization: far fewer processing steps are involved in a write than are used in the typical RDBMS.

---

## Architecture Overview

---

The LogSmart IPDB architecture is shown in the diagram. Data flows from network devices to local data collectors, where the raw data (packaged and secured) resides permanently. Metadata (information about the data’s location) is derived and stored on the server at the management level for use in locating the data. Finally, queries initiated at the application server level cause the data management level to do efficient retrieval from the local level. Historical and trend analysis also takes place on the application server .

The key to performance in this layered architecture is the idea of parsing data on the way out (to be analyzed) rather than on the way in (to initial storage). The IPDB architecture permits this approach and supports small, efficient storage together with rapid retrieval of complete data. Data flow through the layers of this architecture is described in more detail through three main functions: capturing, analyzing and managing the data.

---

## Capturing the Data

---

Four main infrastructure pillars work together to support enterprise IT: security, network, infrastructure and applications. Enterprise networks may contain thousands of devices, and each of them writes details describing each transaction to event logs. This enormous body of continuously generated information becomes the basis for measuring actual network security as well as regulatory compliance.



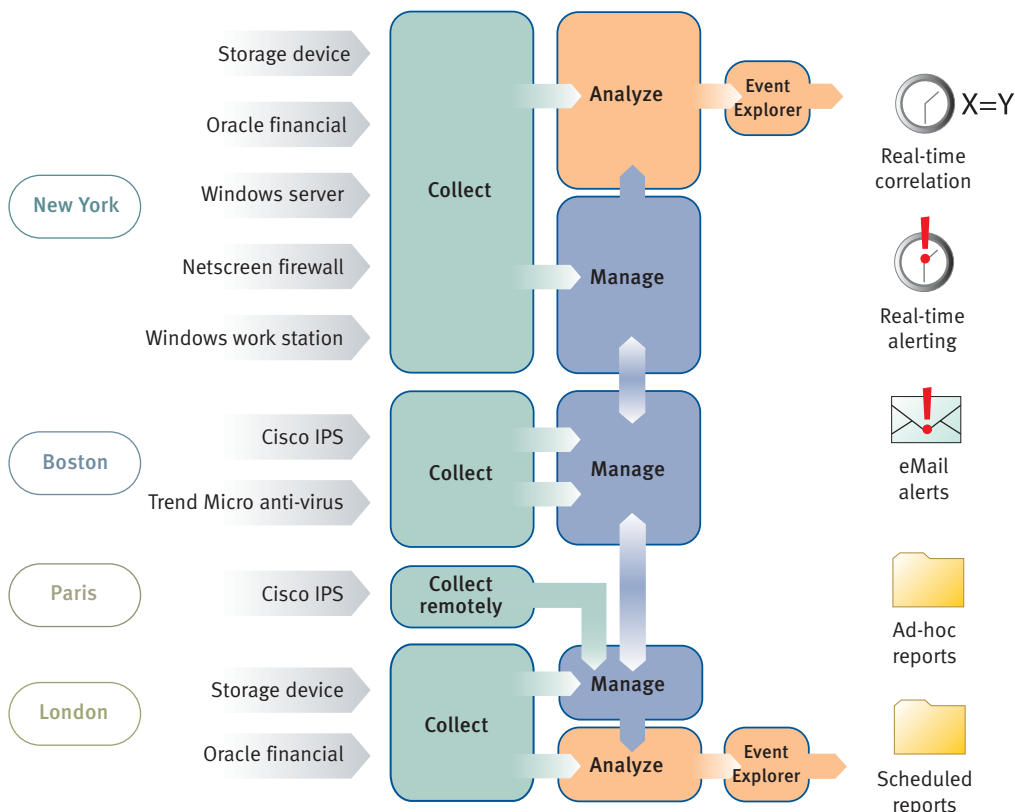
Collecting this data is complicated by the fact that it is recorded and transmitted in numerous formats and protocols, from the traditional UNIX syslog and simple network management protocol (SNMP) traps through open database connectivity (ODBC) and proprietary interfaces such as Check Point® Software’s OPSEC API, SecurePOP, SecureXML, and so on.

As CPU performance continues to increase, the number of events generated only continues to escalate. Log data traffic in a global enterprise can be staggering – but a distributed LogSmart IPDB architecture deployment using local and remote collection devices and distributed data management can collect and process from 1000 events per second (EPS) to over 300,000 EPS from tens of thousands of devices – or over 8 billion events per 24 hour day. The RSA Logsmart IPDB architecture supports well over one hundred network device types and their protocols.

### Local Capture

Many SIEM systems employ agents installed on the various network devices to capture events and create reports. However, agents have several inherent problems. They may have to be installed device by device or they may be advertised as self-installing. But self-installation does not always succeed, and an installation failure may not be detected immediately. Once installed, agents require maintenance and they may also serve as troublemakers on the device. IT team members may turn them off while troubleshooting other problems and may then leave them off, creating blind spots on the network.

Further, agents have a parasitic effect. Usually an agent is designed to use no more than 5 percent of a device’s CPU time; but if seven or eight agents are present, e.g., for backup, virus protection, policy management, etc., the impact becomes significant. The Logsmart IPDB architecture does not use agents. It



RSA enVision LogSmart IPDB Distributed Architecture



captures data on collector devices on the local network by listening and capturing UDP messages at a success rate much better than 99 percent, while also polling other devices, like those operating on Windows, that do not transmit on UDP. Polling rates are configurable but are typically three to four times per minute for real-time connections.

### Building the Internet Protocol Database

Since each log data message identifies the IP address of the device that originated it, LogSmart IPDB technology maps this IP address to its originating device and thus determines the format of the incoming message directly. A log message does not contain the data associated with the event: it may, for example, report that an email was sent, but it does not contain the email's text. Thus the log message is itself metadata that describes the event. A local LogSmart IPDB collector can contain all these events, sorted by originating IP address. This idea is the basic concept underlying the LogSmart Internet Protocol Database.

Saved messages must be protected and they must not cause data explosion and storage problems. The collector contains a process that, for every IP address on its local network, generates a "nugget" file, a temporary file with a lifetime of 60 seconds, into which all messages from that IP address are written. Every 60 seconds, the packager process on the collector reads the nugget file, generates metadata describing it briefly and then authenticates, compresses and encrypts the nugget and appends it to a package file. The packaging process appends to this file for one hour, then closes it and begins a new package file. These files persist indefinitely in the local data storage directory on the local collector. (Remote collectors are slightly different.) In this process, the incoming data is compressed to 29 percent of its original size, alleviating data explosion.

RSA utilizes a "write-once, read-many" philosophy in writing the data. Data writing in the Logsmart IPDB architecture is optimized to two steps, whereas writing data to the typical RDBMS takes six or seven steps.

### Data Management

Each component of the data management layer in the diagram is a data server, or D-SRV. Each D-SRV knows the contents of its local collectors. Using peer-to-peer technology, every D-SRV updates its knowledge of the contents of all others once per minute. Recall that when the local collector packages a nugget file, it first generates metadata describing the file. This metadata is transmitted in real time over the local network to the D-SRV, so that location of the described message data is quickly known throughout the data management layer.

### Remote Collectors

Remote collection devices are used in small offices with few devices. Unlike local collectors, they do not permanently store package files locally. Real-time alerts are communicated immediately to a designated D-SRV; routinely packaged data is held in a temporary IPDB for secure FTP transmission in bulk to the D-SRV at off-peak hours. A D-SRV that handles a remote collector maintains the transferred IPDB locally to itself rather than at the remote collector. Throughout the distributed database, there is neither data loss nor interruption in collection if a WAN connection goes down. Data and data management information are updated when the connection is restored.

---

### Analyzing the Data

Serious events that could indicate system problems, such as an intrusion, virus outbreak or device failure are evaluated and transmitted immediately from both local and remote collectors to the data management layer, and thence to the analysis layer. This layer in turn generates an alert within 60 seconds via any of the standard output actions. Possible actions include screen output to the Web interface, SNMP trap, SMTP (e-mail) message, SNPP (pager) message, instant message to three major public IM infrastructures (AOL, Yahoo, or MSN) or a user-defined script. SNMP trap and SMTP message contents are completely configurable for proper integration with help desk or trouble ticket software already deployed within the enterprise.

Historical and trend analysis are initiated by an administrator through the interface to the analysis server (A-SRV). The A-SRV in turn queries its closest D-SRV to find the data relevant to the query. From the



user's perspective, the Logsmart IPDB architecture appears to be centralized: because all D-SRVs have all location information, they operate virtually as one in locating and retrieving the needed data from IPDBs all over the network. Because the LogSmart Internet Protocol Database architecture is optimized for repeated reading, query scan rates exceed 650 million events per second and query data extraction rates exceed 100,000 events per second. Temporary tables are built in a small RDBMS on the A-SRV and the messages are completely parsed appropriately for the query, without omitting any data that might be needed. The tables are then discarded, so A-SRV storage does not grow.

---

## Managing the Data

The RSA enVision platform uses a hardened Windows operating system which, among other advantages, allows the use of the standard and familiar Windows file system and associated management tools. Backup in this environment is simply file system backup, so it does not suffer from the type of data loss problems that characterize backing up a RDBMS. Defragmentation, anti-virus and other tools also are standard. There are no proprietary utilities to learn or troubleshoot. The RSA enVision platform supports the major vendors of backup solutions and provides automated backup processes.

Impact on the enterprise network is minimal: there is no proprietary database or protocol, and the Logsmart IPDB architecture can interoperate in any kind of network. The architecture can use any protocol or storage technology. The LogSmart IPDB management interface supports role-based administration and controls. Management and updating of all components of the system is centralized and can be done by an authorized administrator.

---

## Summary

The demands of the global enterprise on a SIEM system are great: too great, in fact, for a system based on the traditional RDBMS. The RSA enVision platform, with its scalable distributed architecture based on the LogSmart Internet Protocol Database, meets all these demands and does so at a total cost of ownership lower than that of many systems that do not perform as well.

When examining SIEM systems, administrators should consider the following:

- What is included in the DE numbers? Do they include all overhead, or just the data in the RDBMS?
- Does the system continue to collect data during backup, defragmentation and other management scenarios?
- Does it collect and measure events per second (EPS) or events per day?
- Are agents used? How many EPS can they handle? What happens to dropped messages?
- What percentage of UDP messages are captured?
- Does the system utilize proprietary protocols or formats?
- Does the system capture and retain log data from unrecognized network devices?
- How many steps are involved in writing data to storage?
- What data is stored? – all fields of the data captured or just a standard subset?
- Does the architecture permit extension to capture different types of data in the future?
- Does the system see “low and slow” attacks? How are they judged?
- How many network device types are recognized?
- How does the system handle small remote installations?
- How is the database secured?



## RSA is your trusted partner

RSA, the Security Division of EMC, is the premier provider of security solutions for business acceleration, helping the world's leading organizations succeed by solving their most complex and sensitive security challenges. RSA's information-centric approach to security guards the integrity and confidentiality of information throughout its lifecycle – no matter where it moves, who accesses it or how it is used.

RSA offers industry-leading solutions in identity assurance & access control, data loss prevention & encryption, compliance & security information management and fraud protection. These solutions bring trust to millions of user identities, the transactions that they perform, and the data that is generated. For more information, please visit [www.RSA.com](http://www.RSA.com) and [www.EMC.com](http://www.EMC.com).

RSA, LogSmart and *All the Data* are either registered trademarks or trademarks of RSA Security Inc. in the United States and/or other countries. EMC is a registered trademark of EMC. Windows is a registered trademark of Microsoft Corporation in the U.S. and/or other countries. All other products or services mentioned are trademarks of their respective companies.



RSA Security Inc.  
RSA Security Ireland Limited  
[www.rsa.com](http://www.rsa.com)