TECHNICAL BRIEF

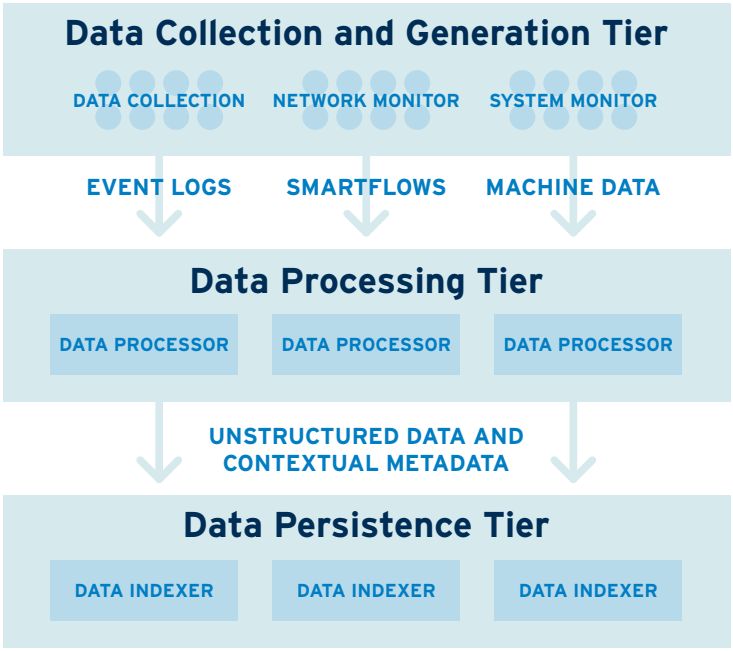
Data Processing & Indexing Tiers

LogRhythm
The Security Intelligence Company

Enterprise IT infrastructures are under attack. Despite having access to a wealth of forensic evidence, most organizations don't have the tools to make effective use of the data. As a result, the organization often takes weeks or months to detect and respond to threats, when this should only take days, hours, or even seconds.

Forensic data can enable quick and focused discovery of true threats. Yet due to the volume and complexity of this data, organizations struggle to capture meaningful information. To use forensic data effectively, it must be processed for machine-automated analysis and indexed for efficient search-based discovery. And this data processing and storage needs to be supported by an architecture that is both highly scalable and cost-effective.

LogRhythm solves big data management challenges with an architecture that supports the responsiveness, scalability, and performance needs of the largest and most sophisticated IT and security operations environments. This architecture is built on data processing and data indexing layers that can process and persist petabytes of machine data at high rates across a wide variety of data sources. These layers support machine-automated security analytics and high-performance search on massive data volumes collected from across the entire environment. This tiered architecture also reduces capital expenditures and operating costs by enabling independent scaling of processing and indexing layers.



LogRhythm Machine Data Processing and Persistence Architecture

Data Processing Tier

LogRhythm Data Processors receive unstructured machine data from the collection tier and process it into a contextualized format that supports consistent analytics across petabytes of heterogeneous machine data.

Data Processors normalize and categorize disparate data from many different products and vendors, creating a consistent Machine Data Intelligence (MDI) Fabric for both machine-based and search-based analytics. The MDI Fabric includes a large set of metadata fields that distill essential context, including factors such as event criticality, impacted host, and origin host.

Collected data are securely forwarded to the Data Processor, where it is parsed and normalized into contextualized metadata that is sent both to the AI Engine for machine-automated security analytics and to the Data Indexer to enable contextual and unstructured search. A copy of every raw message is also forwarded to an archive. If necessary, archived messages can be recalled, re-processed, and added back into the indexing tier for further forensic analysis.

Machine Data Intelligence Fabric

LogRhythm MDI Fabric arms the Security Intelligence and Analytics Platform with a consistent understanding of processed log and machine data across all applications and devices. This enables downstream security analytics, forensic searches, dashboards, reports, and compliance automation modules.

The MDI Fabric is powered by a library of processing rules that parse and normalize machine data from over 750 sources, generating metadata that is specifically structured to enable security analytics. LogRhythm enriches this metadata with vital contextual information, such as event classification, risk-based prioritization, and geo-location. LogRhythm TrueTime™ normalizes timestamps by eliminating discrepancies from time zones, internal clocks, and time of collection and reception, enabling sequence-based security analytics algorithms.

Secure Transport, Audit Ready

The collection tier ensures secure collection and delivery of raw data, establishing a digital chain of custody. Every archived file is digitally signed in order to provide tamper-proof validation for auditing and legal purposes. Archive files are organized by date and are self-describing, making it easy to move individual files to secondary storage.

Processing Scalability
Data Processors are horizontally scalable, with each new node incrementally increasing the deployment's capacity. Horizontally scaled architectures also enable load-balancing, improving performance and delivering active/active high availability.

Data Indexing Tier

Data Indexers receive processed messages from the Data Processing tier and insert them into a highly scalable persistence tier built on Elasticsearch. This back-end delivers a distributed, highly scalable, multitenant search engine with schema-free documents. Using Elasticsearch enables distributed, full-text, unstructured search, as well as access to the contextualized, structured data outputted from the MDI Fabric. Elasticsearch also provides native clustering, enabling enterprise-grade data availability and extremely responsive search.

Unstructured Search
Unstructured Search allows analysts to rapidly search unprocessed machine data using unstructured search syntax. LogRhythm achieves high-performance ad hoc search across vast security and machine data sets with its Elasticsearch back-end, giving analysts quick access to the data they need to investigate and qualify security incidents quickly.

Contextual Search
Contextual Search is enabled by the MDI Fabric. Contextual search improves forensic accuracy, effectiveness and speed by providing immediate access to all processed activity with full event context, regardless of the data source. Advanced visualizations, on-the-fly filtering and data pivoting support rapid decision making.

Precision Search
Precision Search allows unstructured queries against data pre-filtered by contextual fields (e.g., filter down to logs with a destination IP of "X" and then return results from this subset that contain "Y" string anywhere within the raw log). Such searches quickly and easily cut down massive data sets, helping analysts more efficiently review data for evidence of high-risk activities.

Indexing Scalability & Fault Tolerance
Elasticsearch is optimized for active/active clustering, which maximizes performance and scalability and delivers native fault tolerance. The Data Indexing tier is independently scaled through the addition of nodes. Adding Data Indexers also improves search performance and increases data retention lifespans. Replicating data across multiple nodes ensures that machine data is safely stored and available for use at all times. All data is replicated to multiple nodes, and the loss of a single node in a cluster leaves no data at risk and has a minimal impact on performance.

Architected for the Enterprise

The separation of LogRhythm's collection, processing and persistence layers enables its Security Intelligence and Analytics Platform to deliver integrated security analysis and incident management at enterprise scale. By fully integrating enterprise machine data management with next-generation SIEM, network forensics, and endpoint monitoring, LogRhythm helps security teams stay ahead of complex threats. LogRhythm also brings an intuitive modern user Security Intelligence and Analytics Platform an integrated security-focused incident orchestration workflow under a single, cohesive Security Intelligence and Analytics Platform.