Forcepoint

Datasheet

Forcepoint Behavioral Analytics

Platform architecture overview: From data to behavior insights–functional architecture

The Forcepoint Behavioral Analytics platform is a distributed, fault-tolerant, full-stack application that enables deep visibility into your critical data streams for identifying enterprise insider risks. The purpose of this document is to provide an understanding of the major components within Behavior Analytics, how they contribute to the overall solution, and specific technical benefits of these components.

As described below, and illustrated on the following page, raw data from a multitude of sources flows into the Ingest Architecture, and finally into the Application layer, where enriched Forcepoint Behavioral Analytics events are correlated, analyzed, and presented to analysts for investigative review.

Architecture layers

→ **Layer I: Data ›** Forcepoint collects raw data from a wide variety of enterprise data feeds, including communications, physical access, endpoint, and network activity. Forcepoint DLP and Forcepoint Insider Threat are recommended—but not required—data sources.

→ **Layer II: Ingest ›** In the Ingest layer, raw data feeds are transformed and prepared for analysis. Leveraging a flexible data collection platform (e.g., TCP listener, FTP download), Forcepoint gathers raw data, transforms to event format, and passes data through its ingest pipeline and analytics engine.

→ **Layer III: Application ›** The Application layer provides massively scalable data storage and querying capabilities, runtime behavior analytics, an analyst and administrator interface, as well as an outbound API.

Forcepoint analytics framework

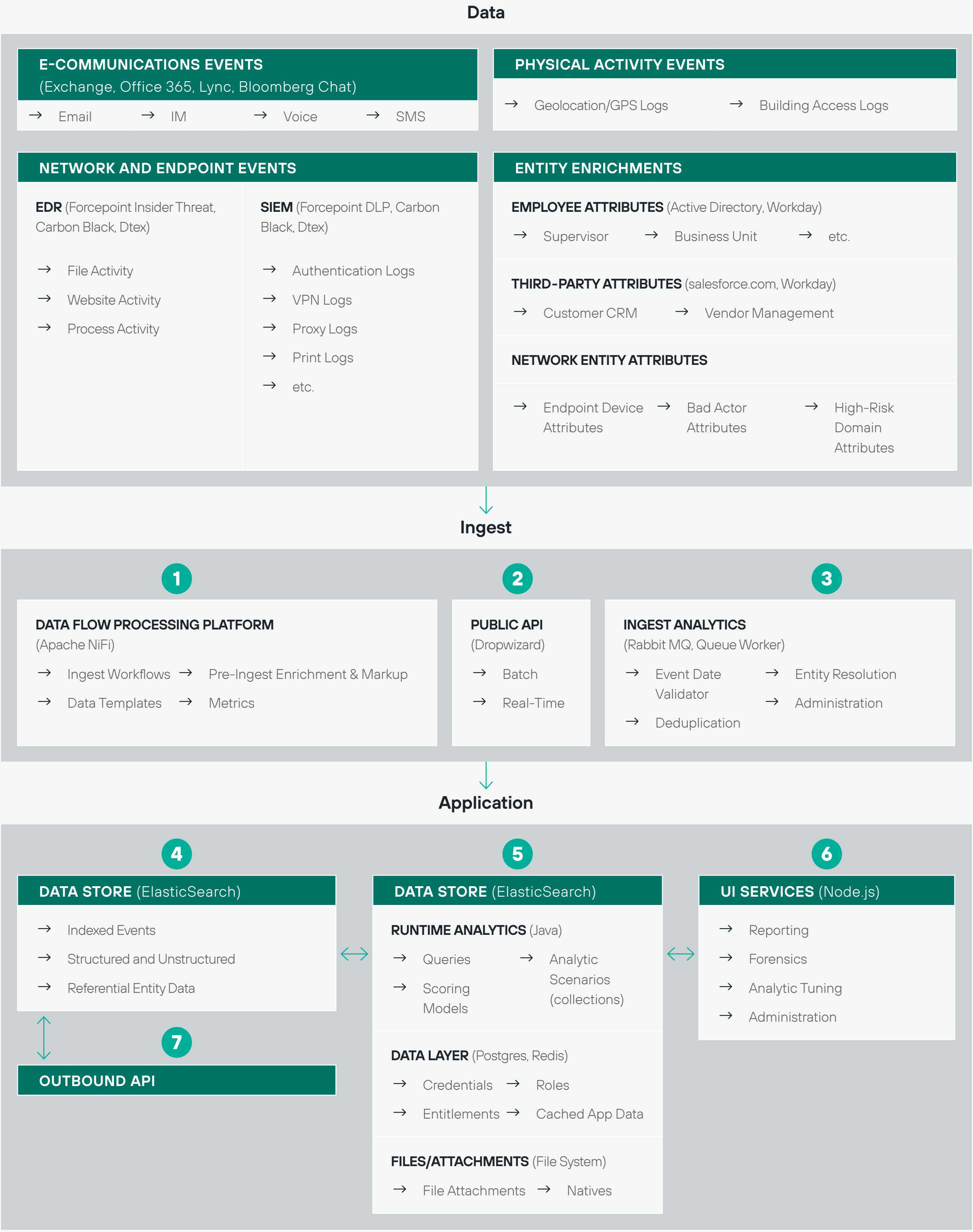
Forcepoint's analytic approach is founded on three core elements that establish: what can and should be captured; how to capture activity and enrichment information; and which analytic models to run to identify insider risk. Each element is briefly described below.

→ **Audit guidance ›** Recommends the ideal logging and collection settings given a customer's network and security ecosystem.

→ **Information model ›** Ensures the consistency of data mapping across very different data sources.

→ **Baseline analytic models ›** Provide out-of-the box analytic configurations of features, models, and scenarios.

Architecture diagram



Data

E-COMMUNICATIONS EVENTS
(Exchange, Office 365, Lync, Bloomberg Chat)

- → Email
- → IM
- → Voice
- → SMS

PHYSICAL ACTIVITY EVENTS

- → Geolocation/GPS Logs
- → Building Access Logs

NETWORK AND ENDPOINT EVENTS

EDR (Forcepoint Insider Threat, Carbon Black, Dtex)

- → File Activity
- → Website Activity
- → Process Activity

SIEM (Forcepoint DLP, Carbon Black, Dtex)

- → Authentication Logs
- → VPN Logs
- → Proxy Logs
- → Print Logs
- → etc.

ENTITY ENRICHMENTS

EMPLOYEE ATTRIBUTES (Active Directory, Workday)

- → Supervisor
- → Business Unit
- → etc.

THIRD-PARTY ATTRIBUTES (salesforce.com, Workday)

- → Customer CRM
- → Vendor Management

NETWORK ENTITY ATTRIBUTES

- → Endpoint Device Attributes
- → Bad Actor Attributes
- → High-Risk Domain Attributes

Ingest

1

DATA FLOW PROCESSING PLATFORM
(Apache NiFi)

- → Ingest Workflows
- → Pre-Ingest Enrichment & Markup
- → Data Templates
- → Metrics

2

PUBLIC API
(Dropwizard)

- → Batch
- → Real-Time

3

INGEST ANALYTICS
(Rabbit MQ, Queue Worker)

- → Event Date Validator
- → Deduplication
- → Entity Resolution
- → Administration

Application

4

DATA STORE (ElasticSearch)

- → Indexed Events
- → Structured and Unstructured
- → Referential Entity Data

7

OUTBOUND API

5

DATA STORE (ElasticSearch)

RUNTIME ANALYTICS (Java)

- → Queries
- → Scoring Models
- → Analytic Scenarios (collections)

DATA LAYER (Postgres, Redis)

- → Credentials
- → Roles
- → Entitlements
- → Cached App Data

FILES/ATTACHMENTS (File System)

- → File Attachments
- → Natives

6

UI SERVICES (Node.js)

- → Reporting
- → Forensics
- → Analytic Tuning
- → Administration

Layer I: Data processing platform

Forcepoint Behavioral Analytics relies on feeds from an organization's existing sensors, logs, and network security fabric to deliver its insider threat behavioral analytics. To consume raw data, Forcepoint's data processing platform leverages a range of collections mechanisms which can include listening for incoming TCP or UDP streams (e.g., syslog), API queries (e.g., Splunk API), or batch data pulls using FTP, file share access, etc.

Layer II: Ingest

The first set of components in the platform handles data ingest and enrichment. Once the types of data to be analyzed are determined, as prescribed by the Audit Guidance, those data sources are mapped to the Information Model in the pre-ingest dataflow processing platform. They are then ingested via the Public API and are finally piped through a series of enrichment and analytic processes. Each of these components and their benefits are discussed below.

1 Pre-ingest data flow processing platform › Forcepoint primarily uses the Apache NiFi framework for processing data prior to ingest into the platform. Originally developed by the National Security Agency, Apache NiFi moved to open source in 2014. The key concepts of the NiFi framework—data provenance, transformation, loose coupling, high concurrency, metrics—align closely with Forcepoint's ingest objectives, and the productized Forcepoint solution provides customers with a set of workflows, templates, and processors that are standardized, hardened and resilient, for datasource consolidation and publishing to the Forcepoint Public API. The data flow processing platform also enables customers to develop and implement environment-specific enrichment processors for marking up events prior to ingest. This allows data feeds to be customized without the added cost of Forcepoint's professional services. In summary, the pre-ingest data flow processing platform provides additional extensibility and flexibility without introducing added cost or complexity.

2 Streaming ingest Public API › The Forcepoint Behavioral Analytics Public API is a RESTful API used for ingesting event and entity information into the application, either in real-time or via bulk upload to the API. The API contains numerous convenience endpoints that align with the Information Model mappings (also used in the pre-ingest data flow processing platform), and also offers a standardized set of metrics that measure request latencies. Convenience and visibility into ingest are two primary benefits offered by the Public API. (Note: Certain data sources can be placed directly into the ingest pipeline discussed below, but Forcepoint strongly recommends use of the Public API.)

3 Ingest pipeline (validation › enrichment › analytics) › Once event and entity information is ingested via the Public API, it is placed on the Message Queue and into the Queue Worker for further processing and enrichment. Each of the processors in the Queue Worker ultimately provides benefits downstream for analytics and forensic investigation. Let's look at each:

→ **Event date validator ›** Enables Forcepoint to ingest only events that are relevant to the configured analytic time window

→ **Deduplication ›** Allows for removal of duplicate events so they do not create noise and unnecessary work for users in the application

→ **Entity resolution ›** Provides resolution of identifiers on an event to a specific entity, so that a single entity can easily be tied to multiple identifiers and modes of activity

→ **Disclaimer detection ›** Removes analytically uninteresting text, namely disclaimers, from communications events so that they do not generate noise in the system

→ **Labeling ›** Facilitates labeling of events based on a defined set of policies

→ **Feature scoring ›** Represents the first building block in our analytic process, by scoring every event based on the configured set of Baseline Analytic Models loaded into the system

Layer III: Forcepoint application

Once events and entities are ingested into the system, they are then stored for use in the application. This section outlines the functions and benefits of each component in the application itself.

4 Data store › Forcepoint Behavioral Analytics uses ElasticSearch (ES) as the primary data store for event and entity information. ElasticSearch is proven at scale and provides significant end-user benefits for text search, analytics, and aggregation that other database technologies simply cannot provide.

5 Master Data Service (MDS) › Forcepoint's proprietary Master Data Service provides much of the application's analytic capability and also correlates data from within the ElasticSearch data store to other supporting technologies (e.g., Postgres and Redis are used to store relational and transactional data). One advantage of that separation of data stores is the ability to scale them independently of each other, allowing for more deployment configuration controls based on users monitored and data ingested, and ultimately lower operational expense for our customers.

- **Runtime analytics ›** These analytics enable the execution of ad hoc, real-time analyst queries, entity-centric risk score calculations, and scenario-based user behavior analysis (e.g., data exfiltration, privileged user abuse, and flight risk scenario-based analytic rollups).
- **Data layer ›** User credentials, entitlements, and roles are stored in the Master Data Service. Additionally, cached application data is maintained here, speeding query response times.
- **Files/attachments ›** For those event feeds which include files and attachments, the Master Data Service indexes the contents of those files for attachments, and optionally persists the original attachments for easy end-user access and drill-down from the user interface. This vastly improves analyst productivity by allowing an integrated forensic deep dive, directly from a scored event.

6 **UI services ›** Forcepoint's user interface layer is a web application, whereas the client is a browser. A server running node.js HTTPS is used for all browser/server interactions, Redis for caching application data, as well as user sessions.

7 **Outbound API ›** Lastly, Forcepoint provides an Outbound API service that allows external applications to retrieve processed events and their associated analytic metadata (i.e., Feature and Models). This enables customers to leverage Forcepoint's insights in other systems (i.e., security orchestration or workflow).

Key architecture takeaways

In summary, the Forcepoint Behavioral Analytics ingest and application stack uses a variety of well-understood and highly respected open source technologies combined with proprietary solutions in order to offer the following technical differentiators and benefits for our customers:

- Modern, scalable architecture that grows horizontally as data volumes grow
- Flexible entity and event data modelling that is "baked into" the ingest pipeline, which can consume, resolve, and analyze both structured and unstructured data and allows for rapid integration of new feeds
- Pipelined data processing that sequences content classification, event scoring, risk modeling, and behavioral profiling
- Highly configurable real-time analytics that allow Forcepoint Behavioral Analytics to tackle a variety of use cases, from rogue trading to market manipulation to data exfiltration and corporate espionage
- Real-time query engine allows ad hoc, analyst-driven investigation and behavioral analytics that do not create noise and unnecessary work for users in the application

forcepoint.com/contact

[FP-Datasheet_UEBA Platform Architecture Overview_EN] 21Dec2020