Chronicle Data Lake structure - Reference Guide

This document describes the Chronicle Data Lake, the data contained within it, the structure of the data lake, and some of its characteristics.

Summary

Chronicle provides an interface to search data ingested into it, to build and run detection content, and to provide dashboarding for your security telemetry data. As well as this, normalized events and entities, aggregated statistics about the data in Chronicle, and some other useful information is exported to a BigQuery dataset. This is used to power the embedded Looker dashboards within Chronicle, and can be made available externally to enable you to use the dataset and BigQuery interface to report on and query your data.

Data overview

The below table provides an overview of the tables included in the Chronicle Data Lake, followed by a more detailed overview on the sources of the data in the table, and how it can be used.

Table Name	**Details**
udm_events	Contains normalized events from Chronicle in UDM format
udm_events_aggregates	Contains aggregated data summarized by hour of normalized USER_LOGIN events from Chronicle in UDM format
ingestion_stats	Statistics related to ingestion of data into Chronicle
ioc_matches	IOC matches found in normalized events
rule_detections	Detections based on rules defined in Chronicle
udm_enum_value_to_name_mapping	Mapping of numerical values in udm_events to their string representations
entity_enum_value_to_name_mapping	Mapping of numerical values in entity_graph to their string representations

entity_graph	Contains normalized entities from Chronicle in UDM format
job_metadata	Internal table used for tracking the export of data to BigQuery, this has no value for customers

UDM Events

During the data ingestion process, Chronicle parses raw log data into the Unified Data Model (UDM) to normalize the data, in order to make the data easy to search and run detections against it. Normalized event data is then exported each hour to the Chronicle Data Lake. This data can be exported or queried using BigQuery to allow you to join this data with other data sets, or run custom search or detection algorithms on it based on the BigQuery SQL-like syntax, this data can also be used for reporting or visualization.

UDM Events Aggregates

This table aggregates USER_LOGIN events, grouping by source and target fields and allowing for the reporting of

Ingestion Stats

This table contains details of log types ingested into Chronicle, sliced into hourly time windows. It shows the number of events of a given type ingested, and parsing success rates across that time period. It can be used to report on and monitor the ingestion of data into Chronicle, helping you to identify logs are being received, and are normalizing correctly.

IOC Matches

Chronicle retrospectively matches new and updated IOCs (IP addresses and domains) against all of the data ingested into the system on a continuous basis. This table contains an export of any matches found so that you can use it for reporting or visualization purposes.

Rule Detections

As Chronicle detection content identifies events of interest, the details of these detections are typically extracted via the Detection API and fed into a SOAR or ITSM system for security analysts to work on. These rule detections are also exported to BigQuery so that you can use this data for reporting on these detections.

UDM Enum Value to Name Mapping

Where enumerators are used within the UDM Events table, the enumerated value will be recorded as a number. The enum value to name mapping table can be used to convert these numbers back into a string value for use in reporting or visualizations.

Entity Enum Value to Name Mapping

Where enumerators are used within the Entity Graph table, the enumerated value will be recorded as a number. The enum value to name mapping table can be used to convert these numbers back into a string value for use in reporting or visualizations.

Entity Graph

During the data ingestion process, Chronicle parses raw log data into the Unified Data Model (UDM) to normalize the data, in order to make the data easy to search and run detections against it. Normalized entity data is then exported each hour to the Chronicle Data Lake. This data can be exported or queried using BigQuery to allow you to join this data with other data sets, or run custom searches on it based on the BigQuery SQL-like syntax, this data can also be used for reporting or visualization.

Job Metadata

As stated above, this is an internal table used to track the export of data to BigQuery, and for troubleshooting by Google support personnel in the case of failures. This table has no value for customers beyond this.

Table structure

udm_events

Field name	Type	Mode
metadata	RECORD	NULLABLE
additional*	RECORD	NULLABLE
principal	RECORD	NULLABLE
src	RECORD	NULLABLE
target	RECORD	NULLABLE

intermediary	RECORD	REPEATED
observer	RECORD	NULLABLE
about	RECORD	REPEATED
security_result	RECORD	REPEATED
network	RECORD	NULLABLE
extensions	RECORD	NULLABLE

* this contains repeated key/value pairs

udm_events_aggregates

Field name	Type	Mode
event_hour	TIMESTAMP	NULLABLE
principal_ip	STRING	NULLABLE
principal_hostname	STRING	NULLABLE
principal_userid	STRING	NULLABLE
principal_location	RECORD	NULLABLE
target_ip	STRING	NULLABLE
target_hostname	STRING	NULLABLE
target_userid	STRING	NULLABLE
target_application	STRING	NULLABLE
target_location	RECORD	NULLABLE
action	INTEGER	NULLABLE
event_count	INTEGER	NULLABLE

udm_events_aggregates.principal_location

Field name	Type	Mode

city	STRING	NULLABLE
state	STRING	NULLABLE
country_or_region	STRING	NULLABLE
name	STRING	NULLABLE
desk_name	STRING	NULLABLE
floor_name	STRING	NULLABLE
region_latitude	FLOAT	NULLABLE
region_longitude	FLOAT	NULLABLE

udm_events_aggregates.target_location

Field name	**Type**	**Mode**
city	STRING	NULLABLE
state	STRING	NULLABLE
country_or_region	STRING	NULLABLE
name	STRING	NULLABLE
desk_name	STRING	NULLABLE
floor_name	STRING	NULLABLE
region_latitude	FLOAT	NULLABLE
region_longitude	FLOAT	NULLABLE

ingestion_stats

Field name	**Type**	**Mode**
time_bucket	STRING	NULLABLE
log_type	STRING	NULLABLE
entry_number	INTEGER	NULLABLE

size_bytes	INTEGER	NULLABLE
source	STRING	NULLABLE
timestamp_sec	INTEGER	NULLABLE
normalized_event_count	INTEGER	NULLABLE
parsing_error_count	INTEGER	NULLABLE
validation_error_count	INTEGER	NULLABLE
enrichment_error_count	INTEGER	NULLABLE
total_error_count	INTEGER	NULLABLE
normalization_ratio	FLOAT	NULLABLE
parsing_error_ratio	FLOAT	NULLABLE
validation_error_ratio	FLOAT	NULLABLE
enrichment_error_ratio	FLOAT	NULLABLE
total_error_ratio	FLOAT	NULLABLE

ioc_matches

Field name	**Type**	**Mode**
ioc_value	STRING	NULLABLE
ioc_type	STRING	NULLABLE
feed_log_type	STRING	NULLABLE
is_global	BOOLEAN	NULLABLE
day_bucket_seconds	INTEGER	NULLABLE
category	STRING	NULLABLE
confidence_score	STRING	NULLABLE
feed_name	STRING	NULLABLE
severity	STRING	NULLABLE

ioc_ingest_time	RECORD	NULLABLE
asset	RECORD	NULLABLE
location	RECORD	NULLABLE

ioc_matches.ioc_ingest_time

Field name	Type	Mode
seconds	INTEGER	NULLABLE
nanos	INTEGER	NULLABLE

ioc_matches.asset

Field name	Type	Mode
hostname	STRING	NULLABLE
asset_ip_address	STRING	NULLABLE
mac	STRING	NULLABLE
product_id	STRING	NULLABLE
namespace	STRING	NULLABLE
is_any_namespace	BOOLEAN	NULLABLE

ioc_matches.location

Field name	Type	Mode
city	STRING	NULLABLE
state	STRING	NULLABLE
country_or_region	STRING	NULLABLE
name	STRING	NULLABLE
desk_name	STRING	NULLABLE
floor_name	STRING	NULLABLE

region_latitude	FLOAT	NULLABLE
region_longitude	FLOAT	NULLABLE

rule_detections

Field name	**Type**	**Mode**
rule_id	STRING	NULLABLE
rule_name	STRING	NULLABLE
rule_text	STRING	NULLABLE
version_timestamp	RECORD	NULLABLE
detection	RECORD	NULLABLE
severity	STRING	NULLABLE

rule_detections.version_timestamp

Field name	**Type**	**Mode**
hostname	STRING	NULLABLE
asset_ip_address	STRING	NULLABLE
mac	STRING	NULLABLE
product_id	STRING	NULLABLE
namespace	STRING	NULLABLE
is_any_namespace	BOOLEAN	NULLABLE

rule_detections.detection

Field name	**Type**	**Mode**
hostname	STRING	NULLABLE
asset_ip_address	STRING	NULLABLE
mac	STRING	NULLABLE

product_id	STRING	NULLABLE
namespace	STRING	NULLABLE
is_any_namespace	BOOLEAN	NULLABLE

udm_enum_value_to_name_mapping

Field name	Type	Mode
field_path	STRING	NULLABLE
enum_value	INTEGER	NULLABLE
enum_name	STRING	NULLABLE

entity_enum_value_to_name_mapping

Field name	Type	Mode
field_path	STRING	NULLABLE
enum_value	INTEGER	NULLABLE
enum_name	STRING	NULLABLE

entity_graph

Field name	Type	Mode
metadata	RECORD	NULLABLE
entity	RECORD	NULLABLE
relations	RECORD	REPEATED
additional*	RECORD	NULLABLE

* this contains repeated key/value pairs

job_metadata

Field name	Type	Mode

customer_id	BYTES	NULLABLE
event_time_bucket	INTEGER	NULLABLE
export_path	STRING	NULLABLE
data_type	STRING	NULLABLE
partition_name	STRING	NULLABLE
num_events	INTEGER	NULLABLE

Data retention

Data is retained in BigQuery for 180 days.