Streaming Defense Student Study Guide

Component Concepts, Operating Methodology, and Practical Review

---

AOT | Investigator | SDGPT Air-Gapped LLM | n8n Automation | MISP Ticketing | Superset Analytics | ClickHouse Big Data

---

Prepared for Streaming Defense students, operators, analysts, engineers, and partner learners.

Signal. Structure. Strength.

Contents

1 How to Use This Study Guide

This study guide is the student-facing companion to the Streaming Defense component training plan. It is designed to help students understand each major component, remember the operating methodology, prepare for hands-on labs, and review the competitive strengths of the platform.

Student Goal

By the end of this guide, you should be able to explain how Streaming Defense observes network activity, investigates risk, contextualizes evidence with SDGPT, automates follow-up actions, creates structured MISP records, visualizes trends in Superset, and stores high-volume data in ClickHouse.

1.1 Recommended Study Method

1. Read the component explanation first.
2. Memorize the input, method, and output of each component.
3. Use the "Key Things to Remember" box for quick review.
4. Complete the review questions at the end of each module.
5. Practice explaining the full evidence chain in plain language.

2 The Streaming Defense Operating Loop

Streaming Defense should be understood as an integrated evidence pipeline. The system is not a random set of tools. Each component has a role in moving from raw observation to action.

Stage	Component Role	Student Explanation
Observe	AOT and collectors capture network and operational signals.	The system first needs high-quality visibility into what is happening.
Classify	AOT and backend services structure protocol, flow, geography, asset, and risk fields.	Raw events become searchable and explainable records.
Investigate	Investigator applies playbooks, enrichment, thresholds, and evidence logic.	Suspicious behavior is transformed into a defensible finding.
Contextualize	SDGPT explains evidence using approved local knowledge, policies, schemas, and playbooks.	Analysts receive private, policy-aware assistance without sending data outside the environment.
Automate	n8n routes outcomes, performs enrichment, and connects systems.	Repeatable actions happen safely with validation and approval gates.
Ticket and Share	MISP stores structured incidents, indicators, tags, and intelligence.	Findings become controlled records that can support response and future detection.
Analyze	Superset creates dashboards and visualizations.	Teams can understand trends, workload, risk, and performance over time.
Persist	ClickHouse stores high-volume event history for fast analytical queries.	The system can retain and query large amounts of telemetry efficiently.

Checkpoint

A strong student can explain the platform as one evidence chain: AOT observes, Investigator reasons, SDGPT contextualizes, n8n routes, MISP records, Superset visualizes, and ClickHouse persists.

3 Module 1: Attack Operations Theater (AOT)

3.1 What AOT Is

The Attack Operations Theater is the real-time operational view of Streaming Defense. It gives analysts visibility into network activity, risk, protocol context, geography, threat stream events, and investigation outcomes.

3.2 How AOT Works

AOT receives structured network-derived telemetry and displays it in a way that supports rapid triage. It helps operators answer: What is happening now? Which events matter? Where is the traffic going? What protocols and destinations are involved? Does the activity require investigation?

3.3 Key Things to Remember

- AOT is the operational theater for real-time visibility.
- AOT emphasizes network-derived evidence, not just endpoint alerts.
- AOT displays risk, protocol, geography, asset, and threat context.
- AOT feeds downstream investigation, automation, ticketing, analytics, and storage.

3.4 Competitive Strength

AOT is strong because it provides an independent network-centered source of truth. It can observe behavior in environments where endpoint agents may be missing, disabled, unmanaged, or incomplete. This is especially important in enterprise, OT, sovereign, and air-gapped environments.

3.5 Review Questions

1. What kind of visibility does AOT provide?
2. Why is network-derived evidence valuable?
3. What are three types of context AOT can show for a network event?

4 Module 2: Investigator

4.1 What Investigator Is

Investigator is the triage and evidence-generation layer. It applies playbooks, enrichment, severity logic, confidence gates, suppression rules, and historical context to produce analyst-ready outcomes.

4.2 How Investigator Works

Investigator takes candidate events from AOT or stored data, normalizes them, enriches them, evaluates them against playbooks, and creates a finding that an analyst can explain and defend.

4.3 Evidence Discipline

Students must understand the difference between these categories:

Category	Meaning
Observed fact	Something directly seen in the telemetry, such as source IP, destination IP, bytes, packets, port, or timestamp.
Enriched context	Additional information added by the system, such as geography, ASN, threat intelligence, or asset context.
Derived interpretation	A conclusion produced by rules or playbooks, such as unusual movement, likely exfiltration, or high-risk destination.
Unsupported claim	A statement that is not proven by the available evidence and should not be presented as fact.

4.4 Key Things to Remember

- Investigator turns visibility into explainable findings.
- Playbooks make investigations repeatable.
- Evidence must be separated from inference.
- Suppression and deduplication reduce alert fatigue.

4.5 Competitive Strength

Investigator is competitively strong because it produces evidence-backed outcomes rather than opaque alerts. It helps analysts explain why something matters and what evidence supports the finding.

4.6 Review Questions

1. What is the difference between AOT visibility and Investigator evidence?
2. Why is evidence discipline important?
3. What is one example of an observed fact and one example of enriched context?

5 Module 3: Air-Gapped LLM: SDGPT

5.1 What SDGPT Is

SDGPT is the Streaming Defense air-gapped large language model capability. It helps students, analysts, operators, and engineers interpret Streaming Defense evidence using approved local context, policy documents, schemas, playbooks, SOPs, and system knowledge.

SDGPT should be understood as a private contextualization assistant. It is not a replacement for analysts, engineering judgment, policy approval, or customer authority.

5.2 Why Air-Gapped Matters

In sensitive, sovereign, regulated, or air-gapped environments, customer telemetry and operational evidence must not be sent to external cloud AI services. SDGPT supports AI-assisted analysis while keeping sensitive data inside the protected environment.

5.3 How SDGPT Should Be Used

1. Provide approved local context, such as policies, playbooks, schema references, and training documents.
2. Ask precise questions about the event, field, detection, policy, or workflow.
3. Require the answer to distinguish observed facts, enriched context, inference, and policy recommendations.
4. Use SDGPT to improve understanding, not to bypass analyst review.
5. Keep final decisions with authorized humans and approved workflows.

5.4 Example SDGPT Prompts

- Explain this Investigator outcome using only the attached exfiltration playbook.
- Identify which fields in this event are observed facts and which are enriched context.
- Summarize the MISP ticketing requirements for this detection based on the local policy document.
- Draft an analyst explanation of this event without making claims beyond the evidence.
- Tell me what additional evidence I should check before escalating this finding.

5.5 Key Things to Remember

- SDGPT is for private, local, policy-aware contextualization.
- SDGPT must be grounded in approved context.
- SDGPT should not invent missing facts.
- SDGPT advises; humans approve and act.
- SDGPT is especially valuable in air-gapped and sovereign environments.

5.6 Competitive Strength

SDGPT gives Streaming Defense a private AI knowledge advantage. Many AI assistants require cloud access or lack local policy awareness. SDGPT can be aligned to Streaming Defense documentation, customer procedures, and approved playbooks while keeping sensitive data under local control.

5.7 Review Questions

1. Why is SDGPT useful in an air-gapped environment?
2. What is the difference between contextualization and automated decision-making?
3. What should a student do if SDGPT gives an answer that is not supported by the evidence?

6 Module 4: Automation with n8n

6.1 What n8n Is

n8n is the workflow automation layer. It connects Streaming Defense outcomes to repeatable actions such as enrichment, notifications, ticket creation, approvals, reporting, and system-to-system routing.

6.2 How n8n Works

n8n workflows are built from connected nodes. A workflow may start from a webhook, schedule, manual action, or system event. It can validate fields, call APIs, query ClickHouse, create MISP events, notify teams, or wait for human approval.

6.3 Safe Automation Principles

- Validate required fields before action.
- Use manual approval for sensitive or disruptive actions.
- Log executions for audit.
- Handle missing fields and failed API calls.
- Design workflows that are reversible or safely recoverable.

6.4 Key Things to Remember

- Automation should reduce repeat work without bypassing control.
- Workflows should be auditable.
- Customer policy determines which actions can be automated.
- Approval gates are critical for sensitive actions.

6.5 Competitive Strength

n8n gives Streaming Defense flexible, self-hostable workflow automation. This is valuable in private, sovereign, and air-gapped environments where cloud SOAR tools may be difficult or impossible to use.

6.6 Review Questions

1. What are three actions n8n can automate?
2. When should a workflow require human approval?
3. Why is audit history important for automation?

7 Module 5: Ticketing and Threat Intelligence with MISP

7.1 What MISP Is

MISP is the ticketing and threat-intelligence layer. It stores structured events, attributes, tags, threat levels, distribution rules, and analyst context. In Streaming Defense, important findings can become MISP events that preserve evidence and support response.

7.2 How MISP Works

A MISP event can contain observables such as IP addresses, domains, hashes, URLs, ports, protocols, or other indicators. Tags and classifications help describe severity, handling rules, confidence, tenant, or campaign context.

7.3 Key Things to Remember

- MISP turns findings into structured intelligence records.
- MISP attributes should be meaningful and evidence-backed.
- Sharing and distribution controls matter.
- MISP can support future detection by feeding confirmed intelligence back into the system.

7.4 Competitive Strength

MISP helps Streaming Defense move beyond simple alerting. It creates structured, reusable, controlled intelligence records that can support investigation, collaboration, and long-term learning.

7.5 Review Questions

1. What is the difference between a MISP event and a MISP attribute?
2. Why should weak or unsupported indicators not be added as strong attributes?
3. How can MISP intelligence improve future detection?

8 Module 6: Analytics with Apache Superset

8.1 What Superset Is

Apache Superset is the analytics and visualization layer. It helps users create charts, dashboards, filters, and reports from trusted datasets, usually backed by ClickHouse.

8.2 How Superset Works

Superset allows students and operators to answer questions such as: Which assets generated the most high-risk activity? Which destinations are most common? How many escalations occurred this week? Which playbooks create the most outcomes? How is analyst workload trending?

8.3 Key Things to Remember

- Superset turns operational history into dashboards.
- Dashboards must be built from trusted datasets.
- Charts should answer clear questions.
- Students must avoid misleading conclusions from incomplete or stale data.

8.4 Competitive Strength

Superset gives Streaming Defense open and flexible analytics. It supports analyst, engineering, executive, and customer reporting without forcing every use case into a proprietary BI tool.

8.5 Review Questions

1. What is one analyst dashboard and one executive dashboard you could build in Superset?
2. Why must dashboard data be validated?
3. How does Superset complement AOT?

9 Module 7: Big Data with ClickHouse

9.1 What ClickHouse Is

ClickHouse is the high-volume analytical database layer. It stores structured event history and supports fast queries over large amounts of telemetry.

9.2 How ClickHouse Works

ClickHouse uses column-oriented storage. This is useful for security analytics because many queries only need a subset of fields, such as timestamp, source IP, destination IP, bytes, severity, protocol, or tenant.

9.3 Key Things to Remember

- ClickHouse is optimized for analytical queries over large datasets.
- Good schemas make investigations and dashboards easier.
- Batch inserts are important for high-volume telemetry.
- Retention policies control how long data is stored.
- ClickHouse supports AOT, Investigator, Superset, automation, and reporting.

9.4 Competitive Strength

ClickHouse gives Streaming Defense fast and cost-effective analytics over high-volume event data. This helps preserve richer evidence and makes historical investigation practical.

9.5 Review Questions

1. Why is columnar storage useful for analytics?
2. What are five fields commonly used in security queries?
3. Why are retention policies important?

10 Integrated Capstone Scenario

10.1 Scenario

AOT shows a large outbound movement from an internal source to an external destination. The event includes bytes, packets, protocol, geography, ASN, timestamp, and severity context.

Investigator evaluates the event using a movement or exfiltration playbook. It enriches the finding with threat intelligence and historical context from ClickHouse. SDGPT helps the analyst interpret the outcome against approved playbooks and policy guidance. If the analyst approves escalation, n8n routes the outcome into MISP and triggers the required notifications. Superset later shows the event as part of weekly operational reporting.

10.2 Student Task

Explain the full evidence chain in your own words:

1. What did AOT observe?
2. What did Investigator decide, and what evidence supported it?

3. How did SDGPT help contextualize the event?
4. What did n8n automate?
5. What went into MISP?
6. What did ClickHouse preserve?
7. What could Superset show later?

11 Quick Review Tables

11.1 Component Summary

Component	Primary Role	Remember This
AOT	Real-time operational visibility.	Shows what is happening now.
Investigator	Evidence-backed triage and playbooks.	Explains why an event matters.
SDGPT	Air-gapped contextualization and policy knowledge.	Helps interpret evidence privately and safely.
n8n	Workflow automation.	Routes approved outcomes through controlled workflows.
MISP	Ticketing and threat intelligence.	Stores structured incident and indicator records.
Superset	Dashboards and visualization.	Shows trends, posture, and performance over time.
ClickHouse	Big-data analytical persistence.	Stores and queries high-volume telemetry.

11.2 Competitive Strength Summary

Component	Competitive Strength
AOT	Network-centered visibility independent of endpoint-only or log-only coverage.
Investigator	Repeatable, explainable, evidence-backed findings.
SDGPT	Private, policy-aware AI assistance inside air-gapped environments.
n8n	Flexible, self-hosted automation aligned to customer workflows.
MISP	Structured intelligence records with tags, attributes, and distribution control.
Superset	Open analytics and visualization over trusted security data.
ClickHouse	Fast analytical queries over large-scale telemetry.

12 Final Review Questions

1. Describe the full Streaming Defense evidence chain from observation to dashboard.

2. Why is AOT not simply a dashboard?
3. What makes an Investigator outcome defensible?
4. How does SDGPT support policy knowledge without replacing analyst judgment?
5. Why is n8n useful for safe automation?
6. What should be included in a MISP event created from an investigation outcome?
7. How does Superset help executives and analysts differently?
8. Why is ClickHouse important for high-volume network telemetry?
9. What is the risk of confusing enriched context with observed fact?
10. What is the competitive advantage of combining these components into one operating model?

13 Glossary

Term	Meaning
AOT	Attack Operations Theater; the real-time operational visibility layer.
Investigator	The playbook-driven evidence and triage layer.
SDGPT	Streaming Defense air-gapped LLM used for private contextualization and policy knowledge.
n8n	Workflow automation engine used to route and automate approved actions.
MISP	Threat intelligence and ticketing system for structured events and indicators.
Superset	Analytics and dashboard platform for visualizing trusted datasets.
ClickHouse	Column-oriented analytical database used for high-volume telemetry.
Observed fact	A data point directly present in telemetry.
Enriched context	Added information such as geography, ASN, asset, or threat intelligence.
Inference	A conclusion derived from facts, enrichment, rules, or playbooks.
Air-gapped	Isolated from external networks or cloud services for security and sovereignty.