💻 Lecture 28 – Splunk

🧠 What is Splunk?

Splunk is a powerful platform used for **collecting, analyzing, and visualizing machine-generated data (logs)** from various systems in real-time.

Originally, Splunk was only a **data analysis tool**, but now it has evolved into a full **SIEM (Security Information and Event Management)** system.

⚙️ Purpose of Splunk

Splunk is mainly used to:

- 🔍 **Monitor** system and application logs.
- 🚨 **Detect** security threats and anomalies.
- 📊 **Analyze** large volumes of data quickly.
- 📝 **Generate reports** and dashboards for real-time monitoring.
- 🧠 **Correlate events** from multiple systems (firewalls, servers, etc.) to detect attacks or failures.

🧩 Why Splunk is Needed

👉 You can manually check logs from **Event Viewer** or command-line tools for a few systems.
But when you have **hundreds of devices (100–200+)**, manual checking becomes impossible.

That's where **Splunk** helps — it **collects logs from all systems automatically**, indexes them, and makes them **searchable and analyzable** in one place.

📚 What are Logs?

Logs are automatically recorded data entries that store information about system events, user actions, network traffic, and errors.
Splunk reads, processes, and analyzes these logs.

📜 Common Types of Logs Splunk Can Analyze:

1. **System Logs:**
    - OS activities, browser history, application usage, updates.
2. **Application Logs:**
    - Records from software like MS Office, web servers, antivirus, etc.
3. **Security Logs:**
    - Authentication logs, firewall alerts, intrusion detection logs.
4. **Network Logs:**
    - Data from routers, switches, firewalls, VPNs, or proxy servers.
5. **Database Logs:**
    - Queries, transactions, errors, and performance events.
6. **Server Logs:**
    - Access logs, performance, uptime, and resource usage.
7. **Active Directory Logs:**
    - User login/logout, password changes, and privilege modifications.

🧭 How Splunk Works (Architecture Overview)

Splunk's architecture is divided into three major components:

🛰️ 1. Forwarder

- Installed on the **source machine** (like a server, PC, or router).
- Its job is to **collect log data** and **send** it to the Splunk **indexer**.
- It acts like an agent that **forwards real-time data streams**.

🧩 Example:
Forwarder on a web server sends Apache access logs to the main Splunk server.

🗄️ 2. Indexer

- The **core component** of Splunk.
- It **receives data** from forwarders, **indexes** it (organizes into searchable format), and **stores it**.
- Helps in **fast searching and correlation** of events.

💡 Example:
If logs are collected from 10 routers, the indexer organizes and labels them by source, time, and type.

🖥️ 3. Search Head (or GUI)

- The **interface** where users search and visualize data.
- Provides tools like dashboards, alerts, and reports.
- You can write **queries** using Splunk's Search Processing Language (SPL).

🧱 Indexing Based on Source (Source-based Indexing)

- Splunk indexes data based on **source type** — like system logs, application logs, firewall logs, etc.
- This makes searching and filtering data faster and more accurate.

🧩 Example:
You can create indexes such as:

index=firewall_logs

index=webserver_logs

index=windows_security

Then you can easily run queries like:

index=webserver_logs error OR warning

🧩 How Splunk Helps in Cybersecurity (SIEM Role)

Splunk as a **SIEM tool** performs:

- 🛡️ **Real-time threat detection**
- 🧠 **Correlation of multiple events** (to detect suspicious patterns)
- 🚨 **Alert generation**
- 📈 **Incident investigation and timeline creation**
- 📜 **Compliance reporting** (for SOC audits, ISO, GDPR, etc.)

Example:

If an employee logs in from two different countries within 10 minutes, Splunk can automatically generate a **security alert** 🚨.

🧰 Installation Overview (Splunk Enterprise)

🧩 1. Download Splunk Enterprise

- Go to https://www.splunk.com
- Navigate to **Products → Splunk Enterprise**
- Choose your operating system (Windows/Linux/macOS).

🧩 2. Installation Steps

For Windows:

1. Run the installer .msi file.
2. Choose **installation directory** (default: C:\Program Files\Splunk).
3. Set up **admin username and password**.
4. Splunk service starts automatically on port **8000**.
    - Access GUI: **http://localhost:8000**

For Linux:

wget -O splunk.tgz 'https://download.splunk.com/...'

tar -xvzf splunk.tgz

cd splunk/bin

./splunk start --accept-license

🧩 3. Install Forwarder

- Download **Splunk Universal Forwarder** from the same website.
- Install it on endpoint systems.
- Configure to forward logs to the main Splunk indexer's IP and port (default: 9997).

☁️ Integration with Office 365 and Intune

Splunk can integrate with **Office 365**, **Azure**, and **Intune** to monitor:

- 📨 Email security and login attempts (O365)
- 💻 Device compliance and app usage (Intune)
- ☁️ Cloud activity and access logs (Azure AD)

🔗 Integration Steps:

1. Create an **Azure App Registration** (for API access).
2. Enable **Microsoft Graph API** permissions.
3. Use **Splunk Add-on for Microsoft Cloud Services**.
4. Configure your credentials and choose log sources (Exchange, Teams, SharePoint, Intune).
5. Logs will start appearing in Splunk dashboards.

📜 Splunk Resume Points (for Job Readiness)

If you're building your **resume** for cybersecurity/SOC jobs, include:

- ✅ Installed and configured **Splunk Enterprise & Forwarder**.
- ✅ Created and managed **custom dashboards and alerts**.
- ✅ Analyzed **firewall, system, and application logs**.

- ✅ Performed **incident investigation using Splunk SPL queries**.
- ✅ Integrated **Office 365 and Intune logs** for centralized monitoring.
- ✅ Worked on **log correlation, visualization, and reporting**.

💡 **Example line for resume:**

"Configured and maintained Splunk Enterprise environment for centralized log monitoring and incident analysis, integrated with O365 and Intune for real-time visibility."

🧩 **Summary of Key Components**

Component	Function	Example
Forwarder	Collects and sends log data	Installed on endpoint
Indexer	Stores and organizes data	Splunk main server
Search Head	Provides GUI for analysis	Web dashboard
Logs	Raw data from systems	Firewall, OS, app
Source-based Indexing	Sorts logs by origin	Web, DB, Network

🧠 **Why Splunk is Used**

- Handles **massive log data** (terabytes easily).
- Provides **real-time analytics** for quick response.
- Simplifies **incident response**.
- Used in **SOC, NOC, DevOps, and Cloud monitoring**.
- Compatible with **thousands of devices and APIs**.

⚡ **Conclusion**

Splunk is a modern SIEM and data analysis platform that converts **machine data into security intelligence** 🧠.

It helps analysts quickly identify issues, detect threats, and make data-driven security decisions.