Agenda

- Abstract
- Introduction to Reconnaissance
- Objective of the Project
- Selected Website – Research Details
- Tools Used for Reconnaissance
- Data Collection & Technology Stack
- Impact Analysis
- Recommendations & Mitigation Techniques
- Conclusion
- References

Abstract

- Reconnaissance (Information Gathering) is the first and most crucial phase of cybersecurity assessment. The objective of this project is to study and demonstrate various reconnaissance tools and methods used to collect publicly available information about a target website without directly attacking it.
- In this project, a real website was selected and analyzed using passive and basic active reconnaissance tools such as WHOIS, Nmap, WhatWeb, Wappalyzer, DNS tools, and Google Dorking. The final outcome of this project is a structured understanding of how attackers and security professionals gather information, analyze the technology stack, and assess potential exposure points.
- This report highlights the importance of reconnaissance in cybersecurity, its impact, and recommended mitigation techniques to reduce information leakage.

Introduction to Reconnaissance

- Reconnaissance is the first phase of cybersecurity assessment.
- It involves gathering publicly available information about a target system, network, or website to identify possible attack vectors.
- Used by ethical hackers to identify weaknesses.
- Also known as Information Gathering, is the first phase of the Cyber Kill Chain.
- Helps understand the attack surface of a website or network.
- Reconnaissance can be:
- Passive Reconnaissance – No direct interaction with the target (safe & legal)
- Active Reconnaissance – Direct interaction with the target (e.g., port scanning)

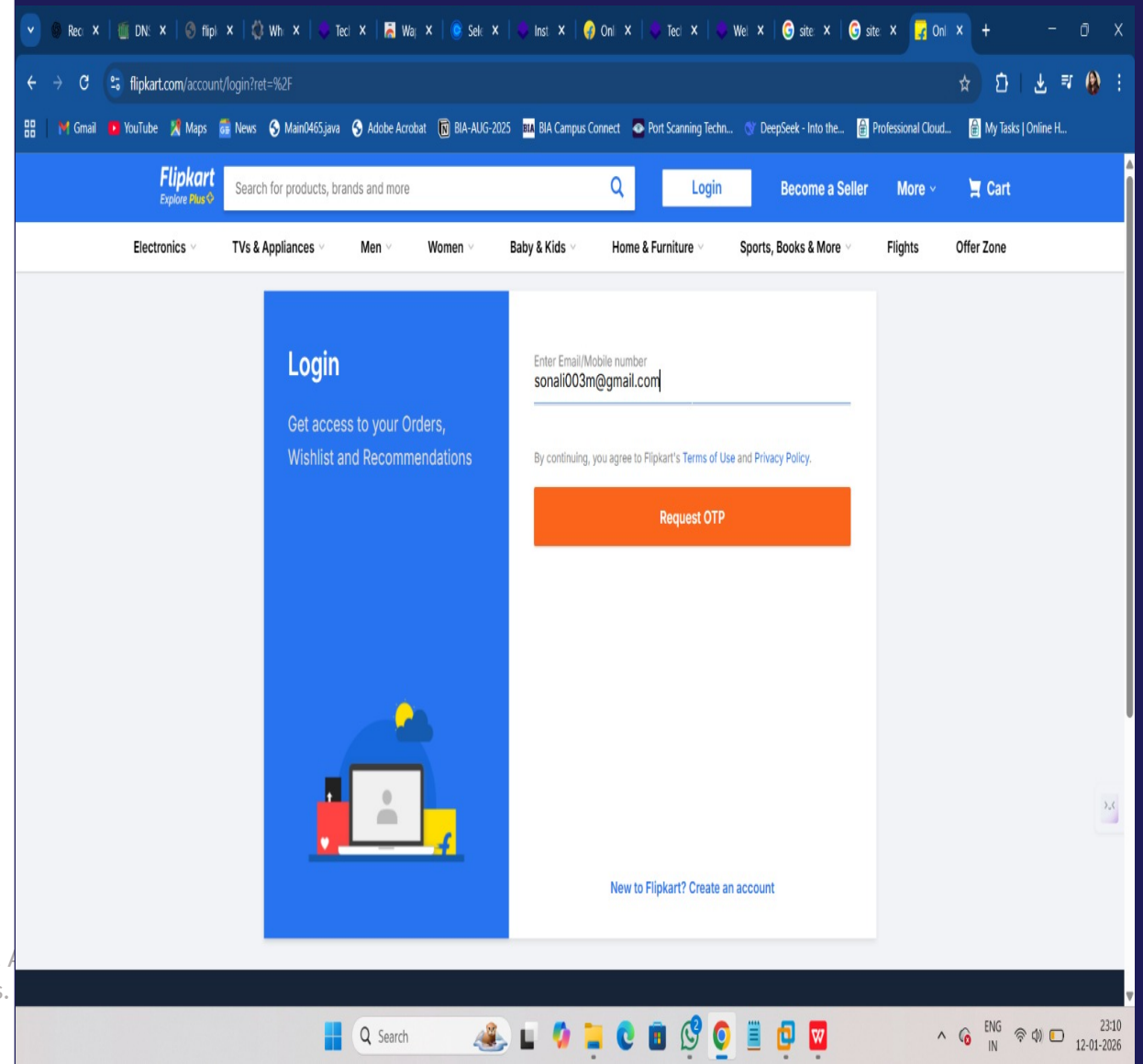
Objective of the Project

- To understand the role of reconnaissance as the first phase of cybersecurity assessment
- To perform ethical information gathering on a real-world website (Flipkart.com)
- To identify publicly exposed information such as domain details, DNS records, IP addresses, and technologies
- To analyze the technology stack used by Flipkart.com without exploiting vulnerabilities
- To differentiate between passive and active reconnaissance techniques
- To understand how attackers can use gathered information for further attacks
- To highlight the importance of minimizing information leakage
- To gain hands-on experience with reconnaissance tools used in cybersecurity

Note: The following website is used only for educational and passive analysis purposes.

Selected Website – Research Details

Parameter		Details
Website Name	:	Flipkart
URL	:	https://www.flipkart.com
Category / Type	:	E-Commerce
Business Type	:	Online Retail Platform
Popularity / Ranking	:	One of India's top e-commerce websites
Users	:	Millions of daily active users
Technology Nature	:	Large-scale cloud-based platform



Tools Used for Reconnaissance

- WHOIS Lookup
- DNS Enumeration
- Wappalyzer
- WhatWeb
- Nmap
- Google Dorking
- IP Address & Hosting Information
- OSINT via Shodan (Passive)
- THEHARVESTER (OSINT & Email Reconnaissance)
- Robots.txt File Analysis
- HTTP Header Analysis

1. WHOIS Lookup

Purpose:

To collect domain ownership and registration details.

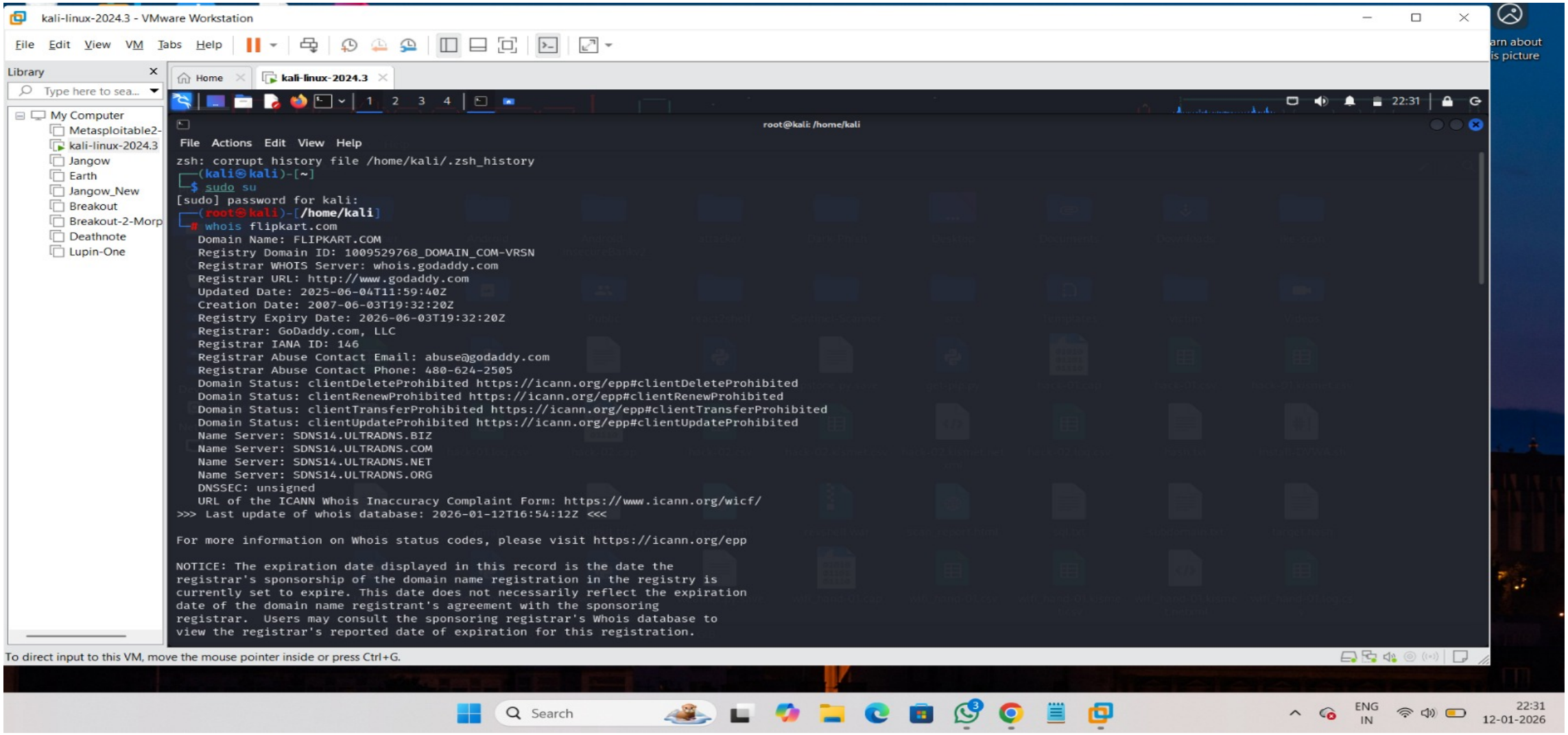
Information Collected:

- Domain owner
- Registrar and Organization details
- Creation & expiry date
- Name servers

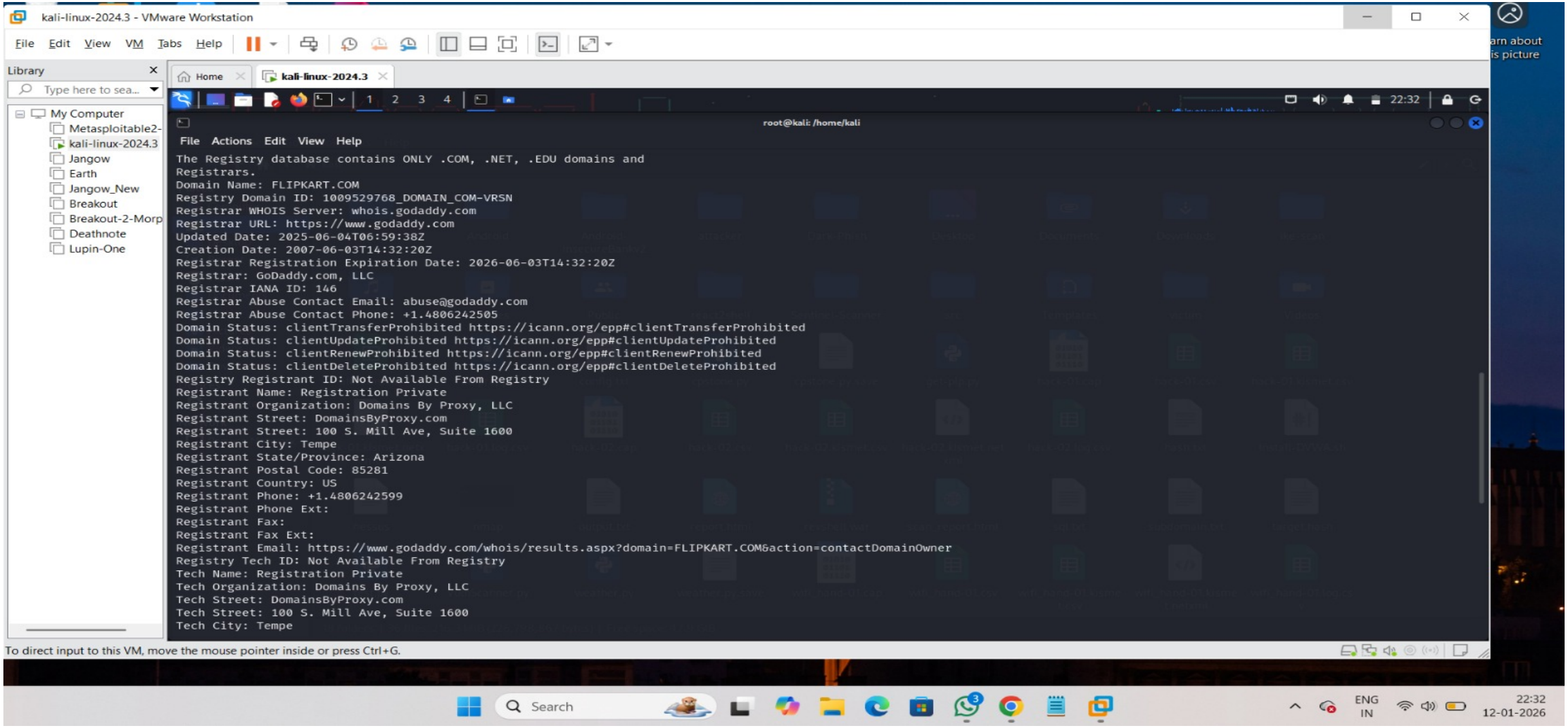
Outcome:

- Domain ownership information identified.
- WHOIS privacy protection enabled.

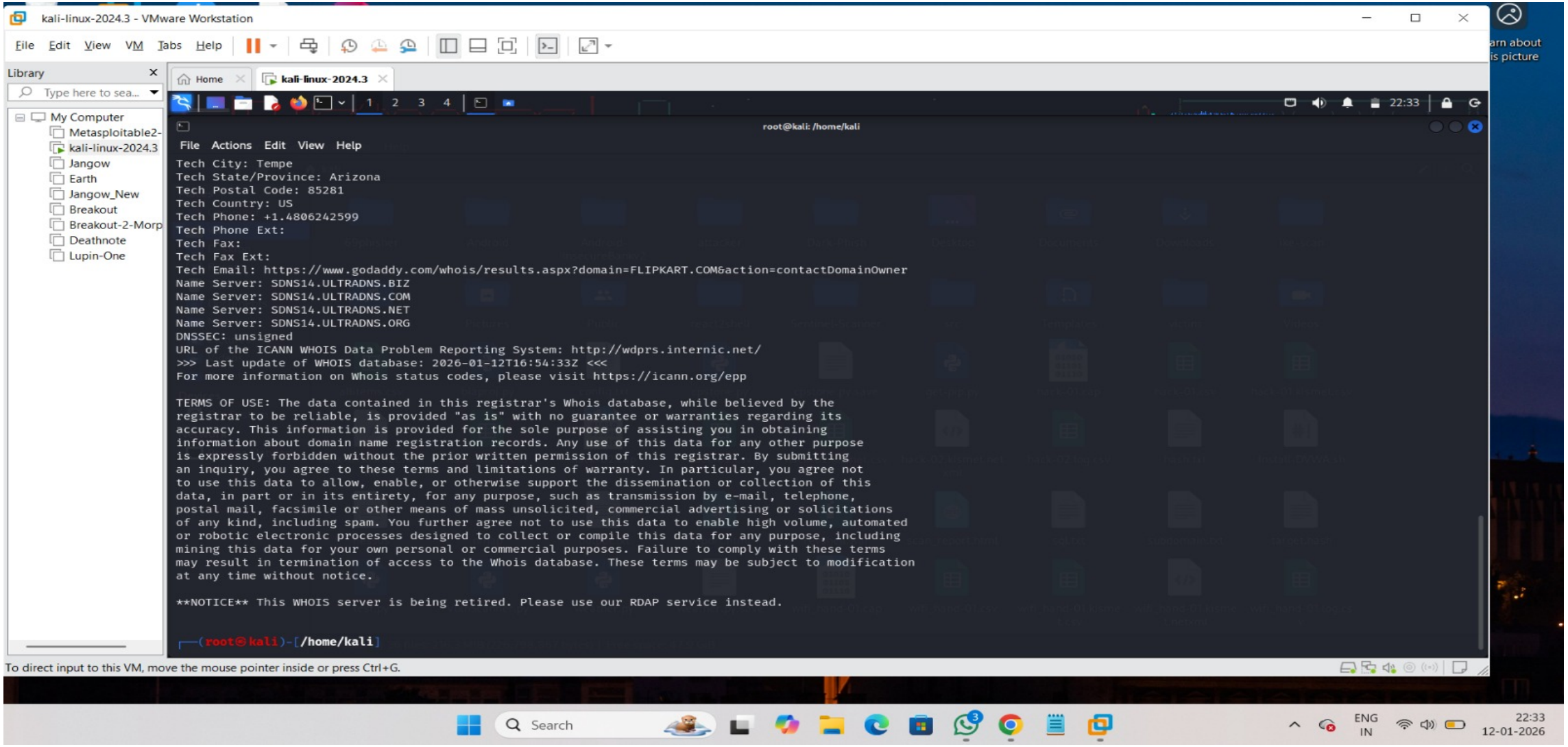
WHOIS Lookup - 1



WHOIS Lookup - 2



WHOIS Lookup - 3



2. DNS Enumeration

Tool Used:

- DNSDumpster
- nslookup

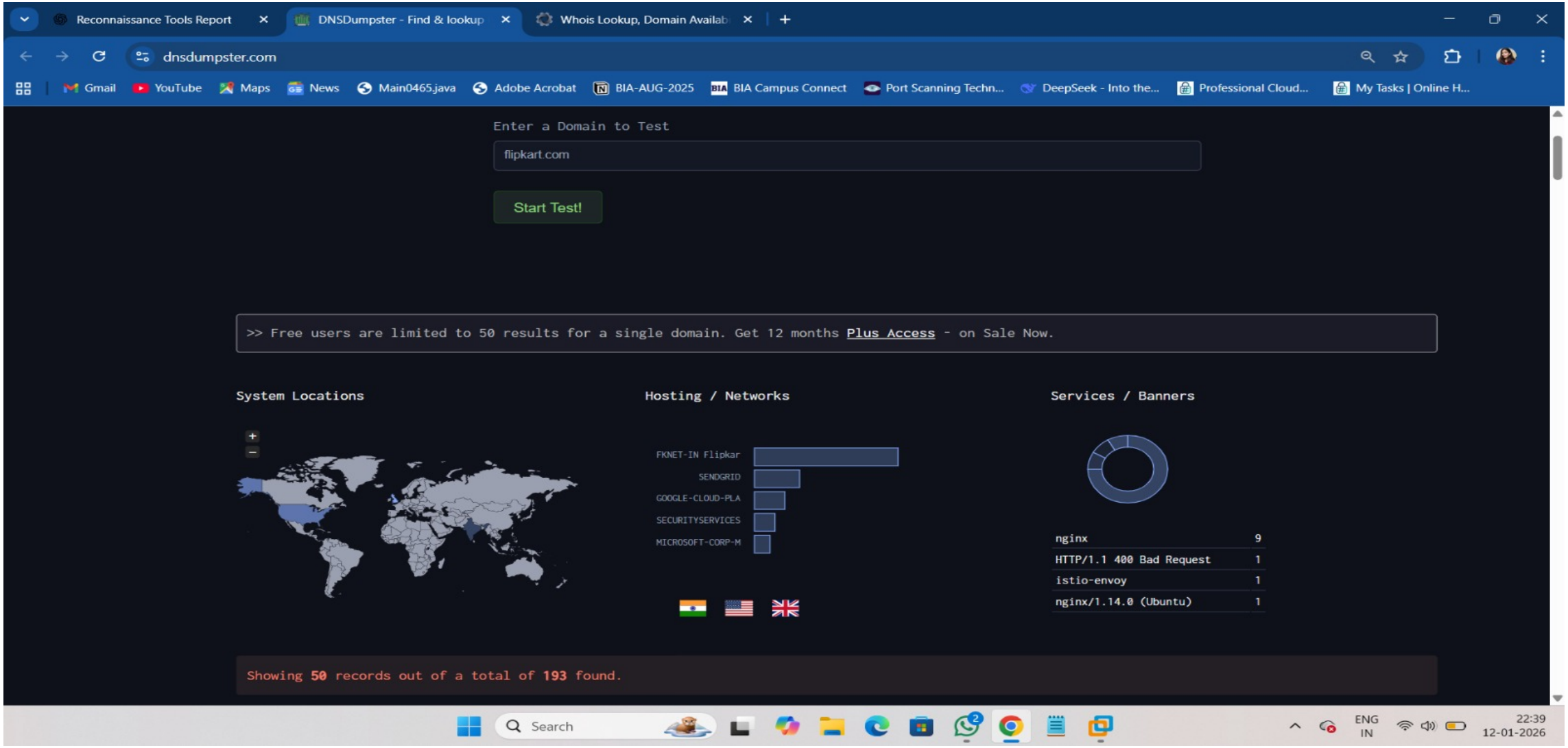
Functionality:

- Identifies DNS records (A, MX, NS)
- Finds subdomains

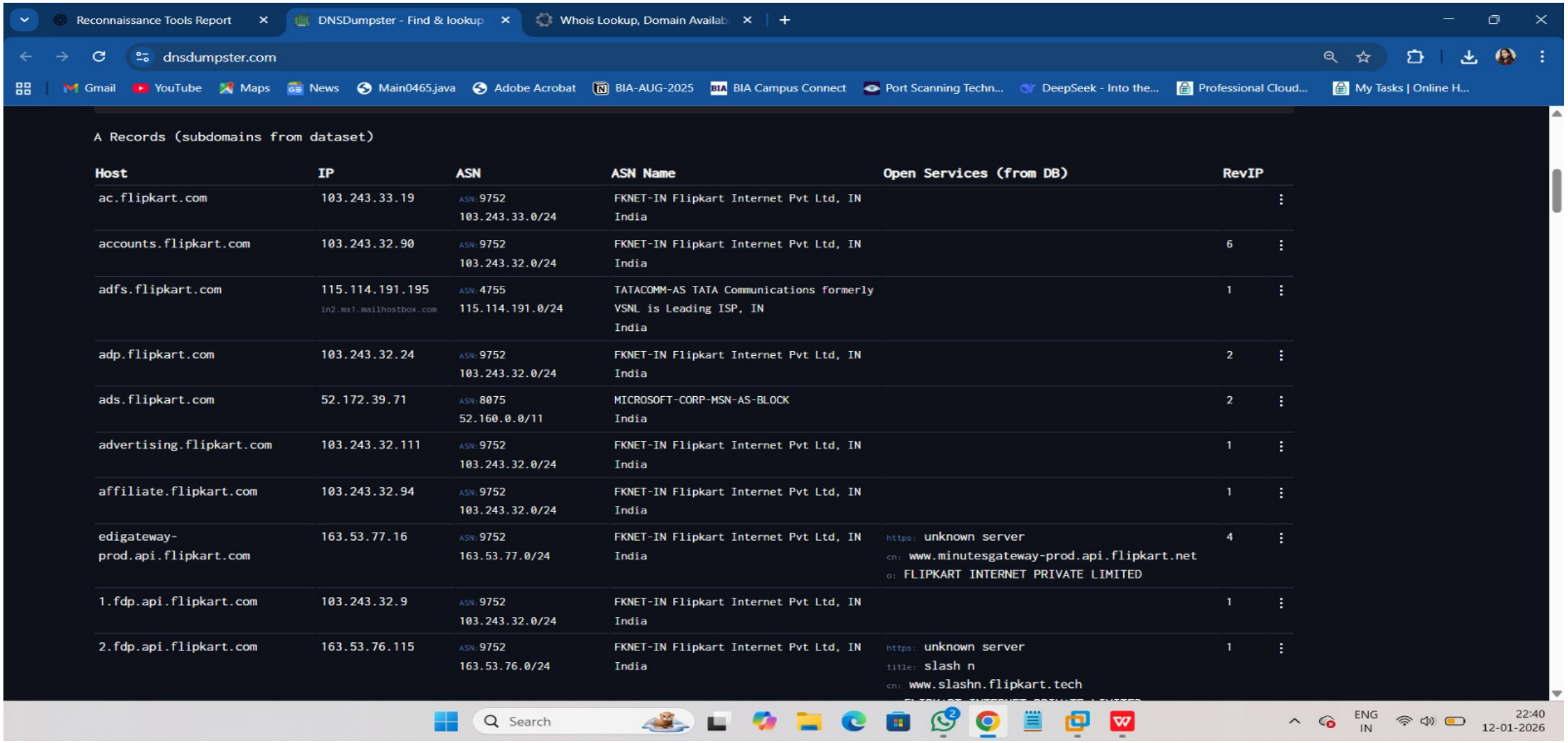
Outcome:

- Multiple subdomains detected.
- Cloud-based DNS infrastructure observed.

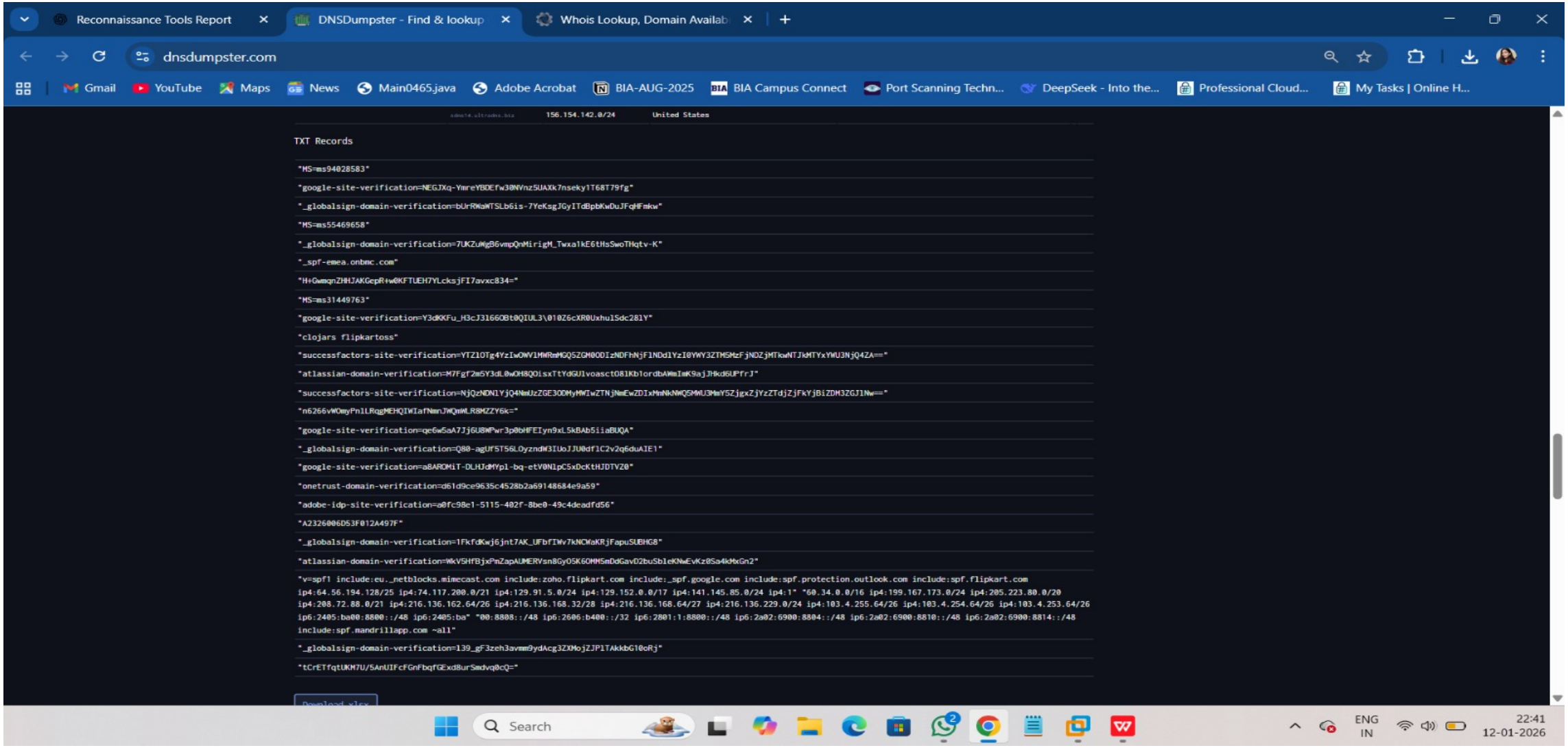
DNSDumpster - 1



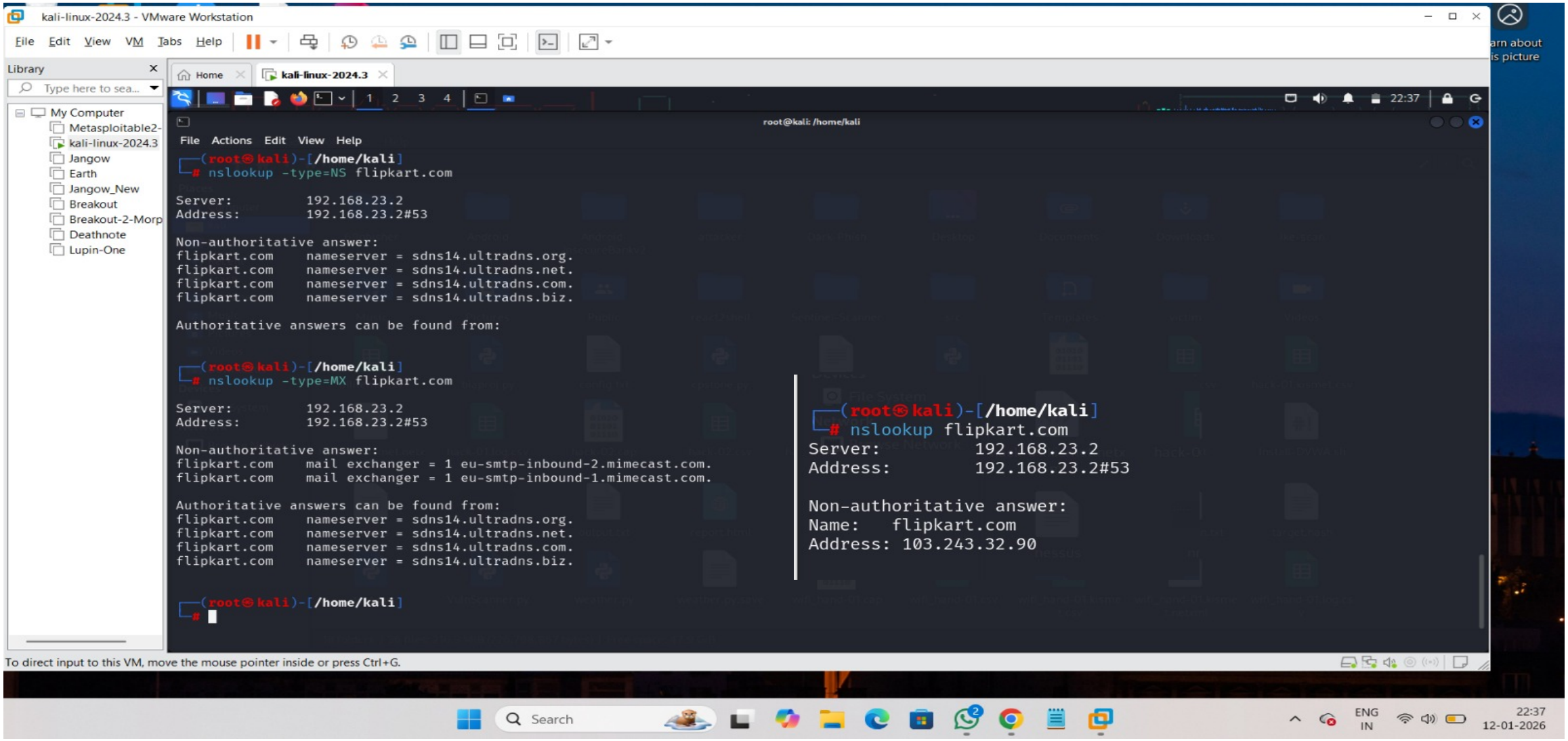
DNSDumpster - 2



DNSDumpster - 3



nslookup - 4



3. Technology Stack Detection – Wappalyzer

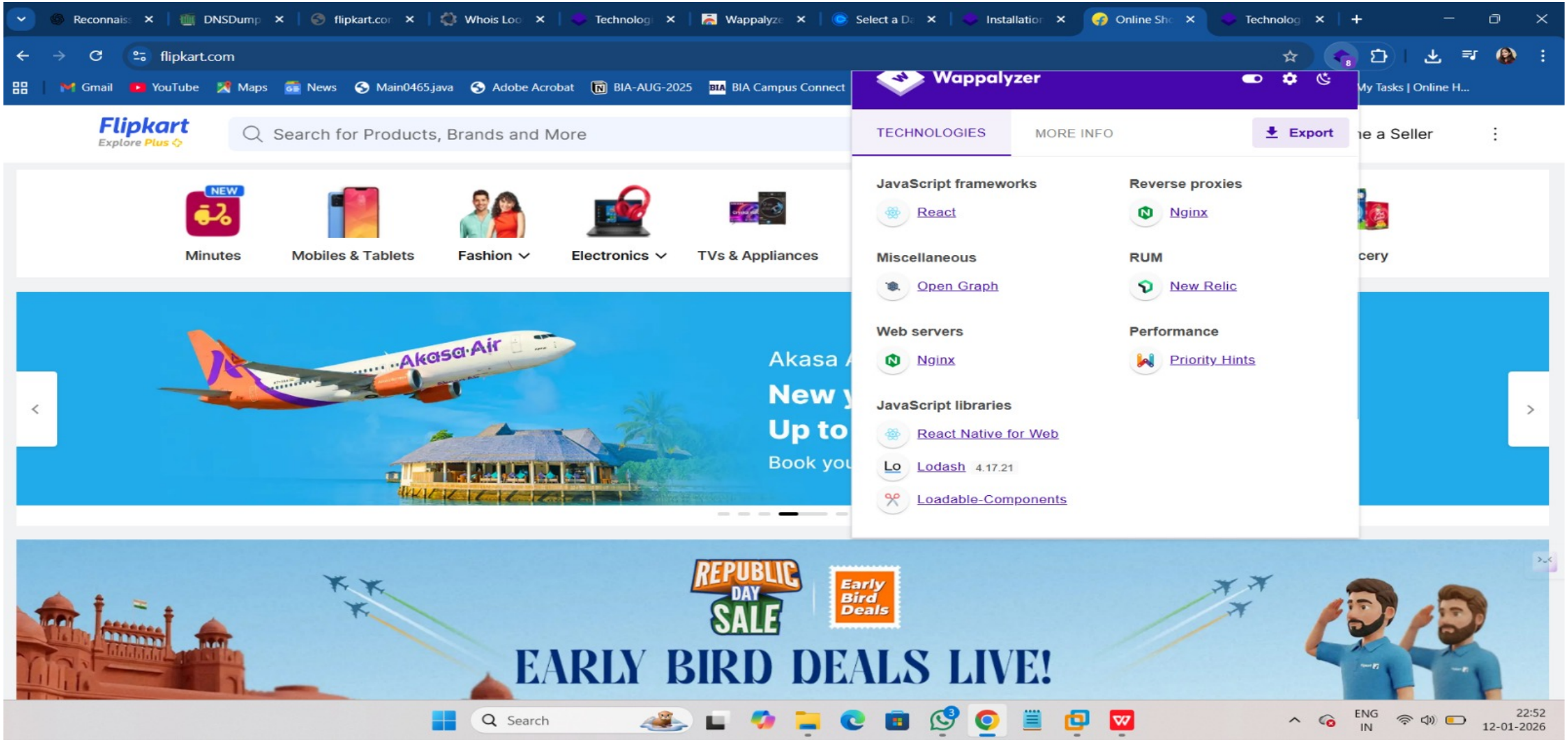
Purpose:

- Identify frontend, backend, server, and security technologies.

Detected Components:

- Web frameworks
- CDN services
- Analytics tools
- Security mechanisms

Technology Stack Detection – Wappalyzer - 1



4. Website Fingerprinting – WhatWeb

Purpose:

- Fingerprint website technologies

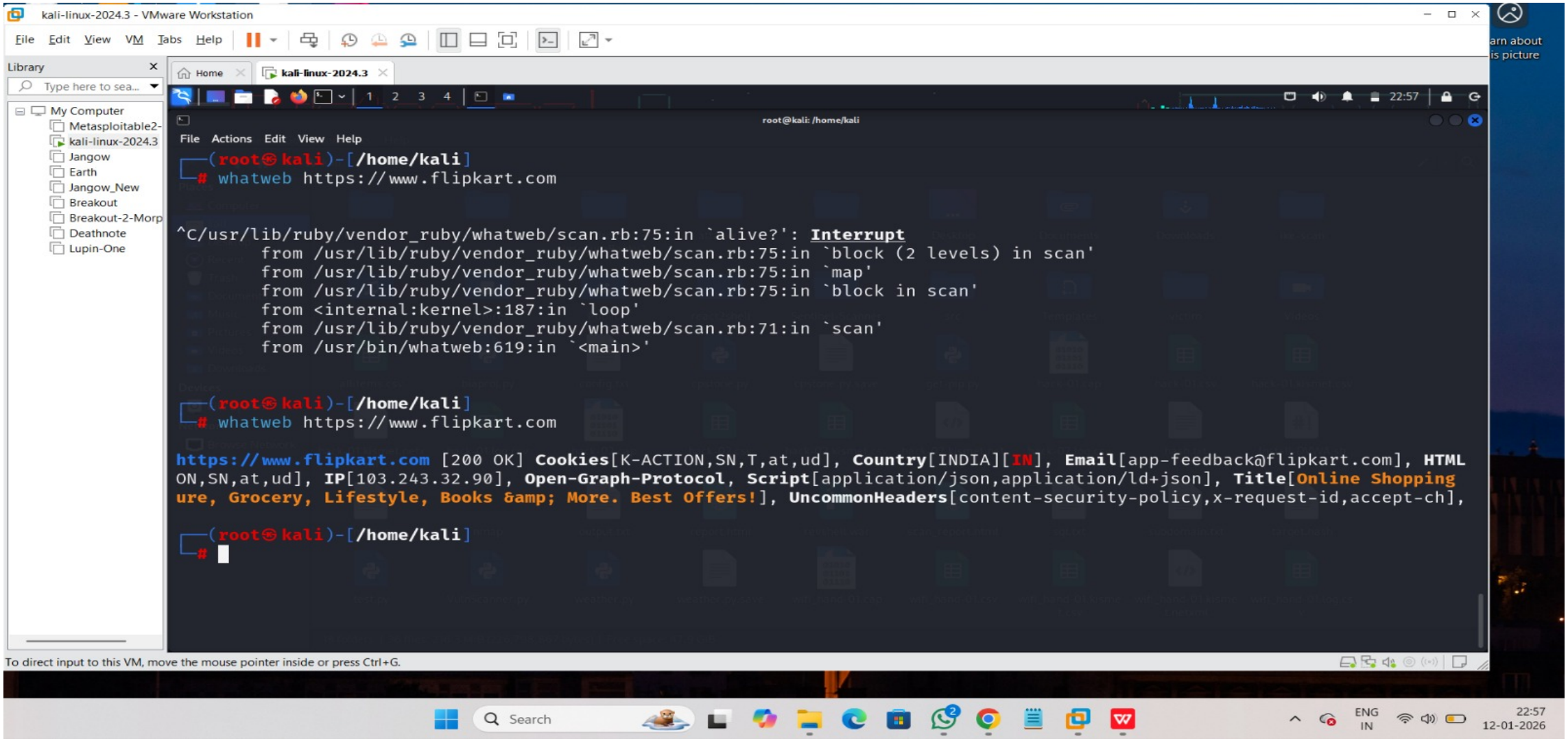
Command Used:

```
whatweb https://www.flipkart.com
```

Outcome:

- Server information
- Framework detection
- Security headers identified

Website Fingerprinting – WhatWeb - 1



5. Network Scanning – Nmap

Purpose:

- Identify open ports and services.

Command Used:

```
nmap -F flipkart.com
nmap -sV flipkart.com
```

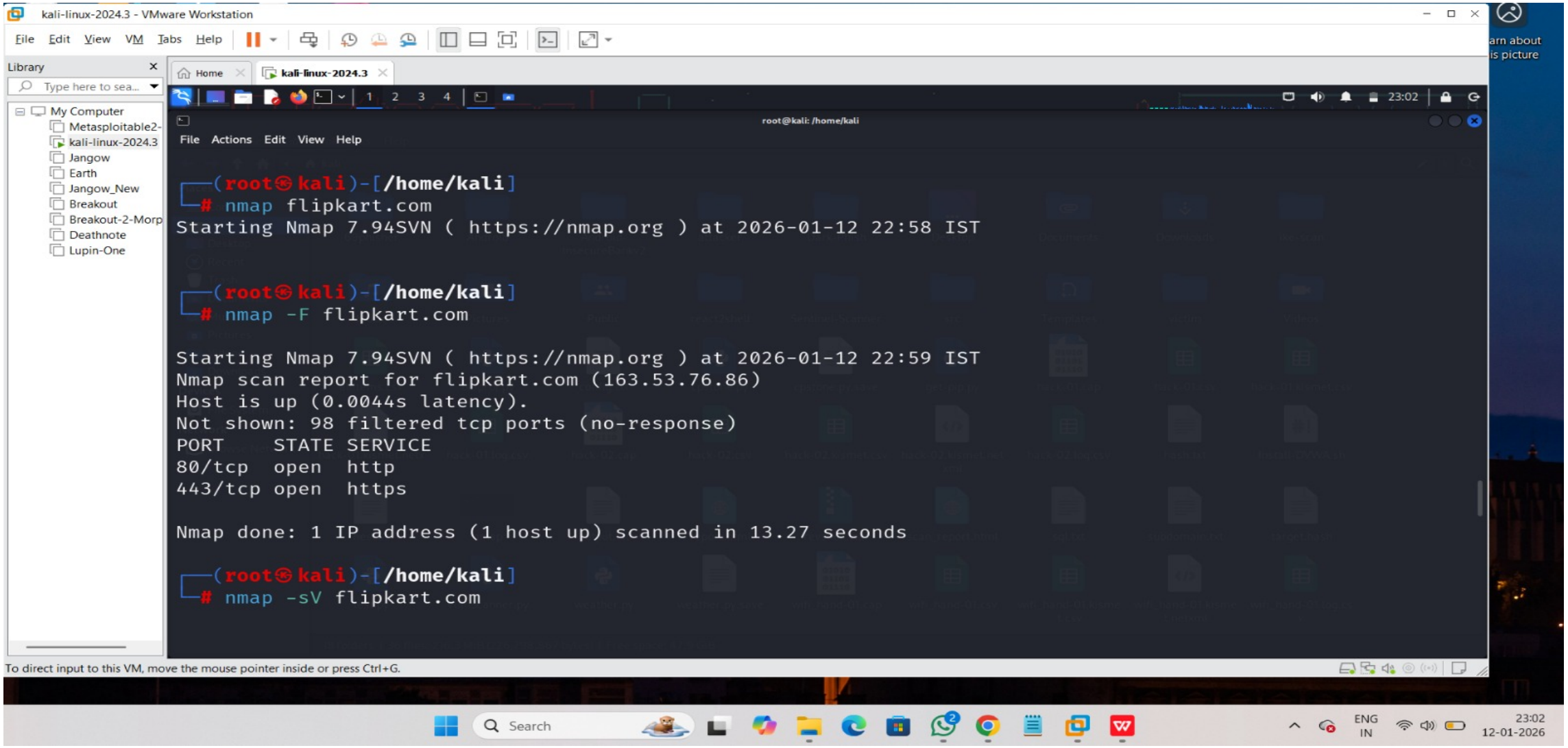
Functionality:

- Port scanning
- Service enumeration

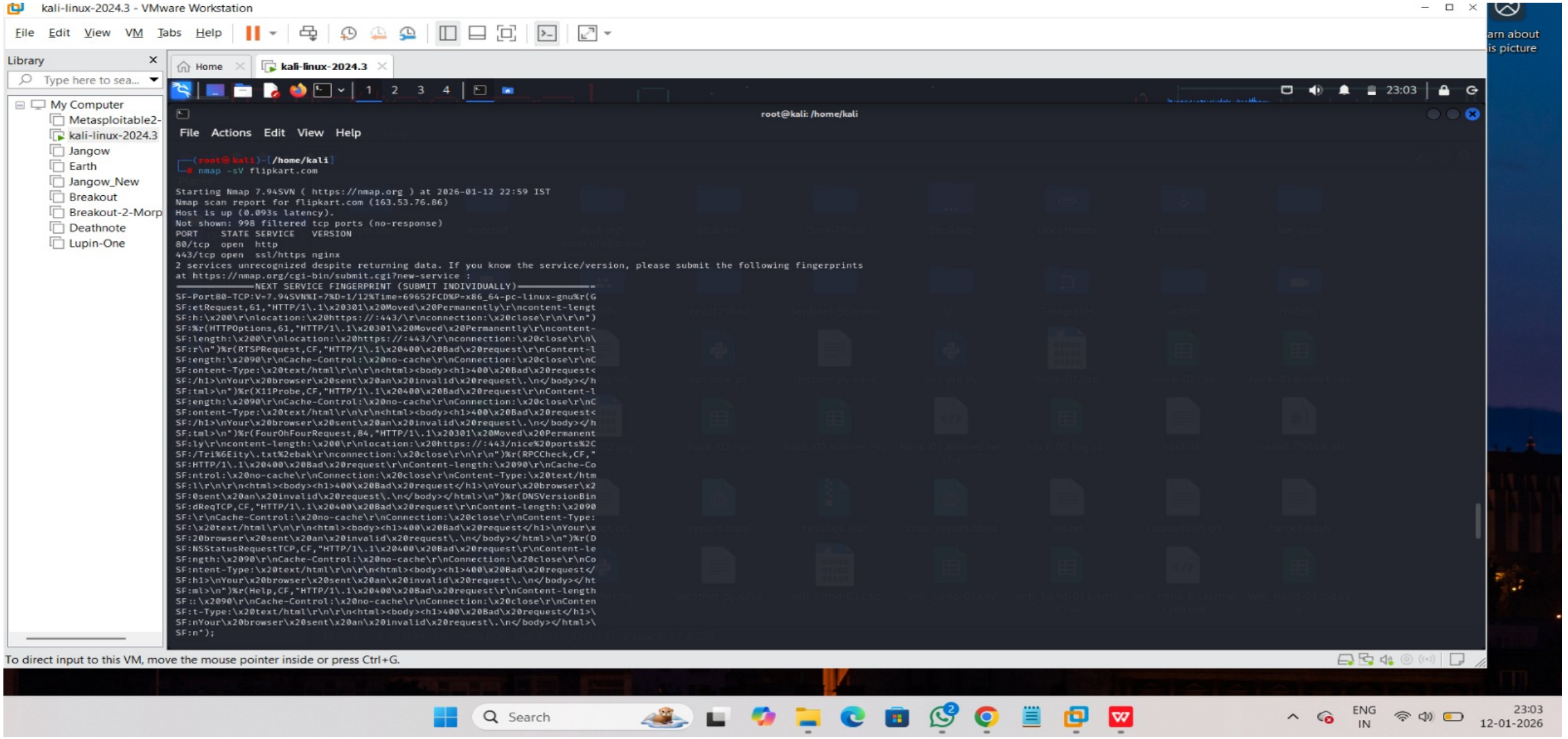
Outcome:

- Limited open ports.
- Firewall and security filtering detected.

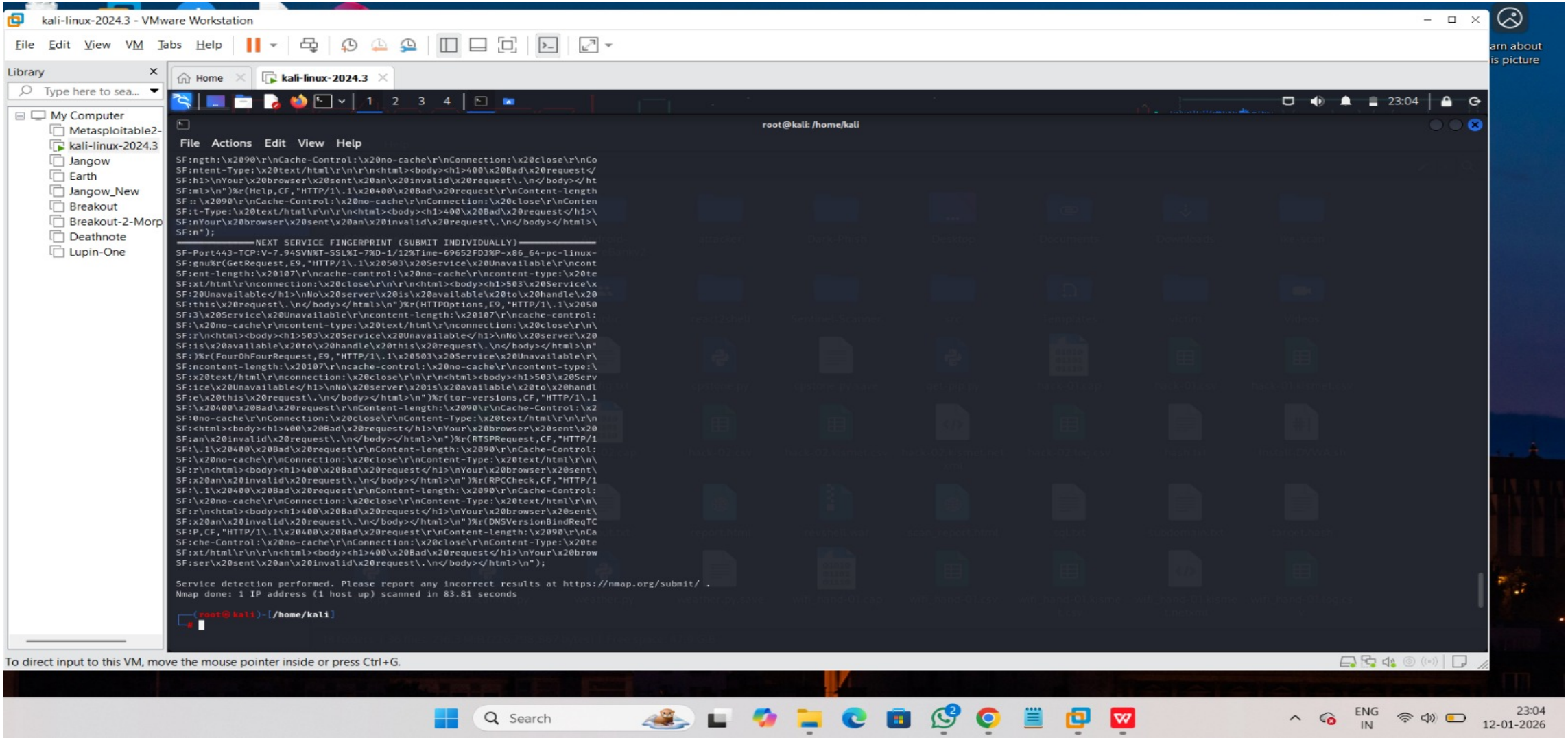
Network Scanning – Nmap - 1



Network Scanning – Nmap - 2



Network Scanning – Nmap - 3



6. Google Dorking

Purpose:

- Extract publicly exposed information

Examples:

site:flipkart.com

site:flipkart.com filetype:pdf

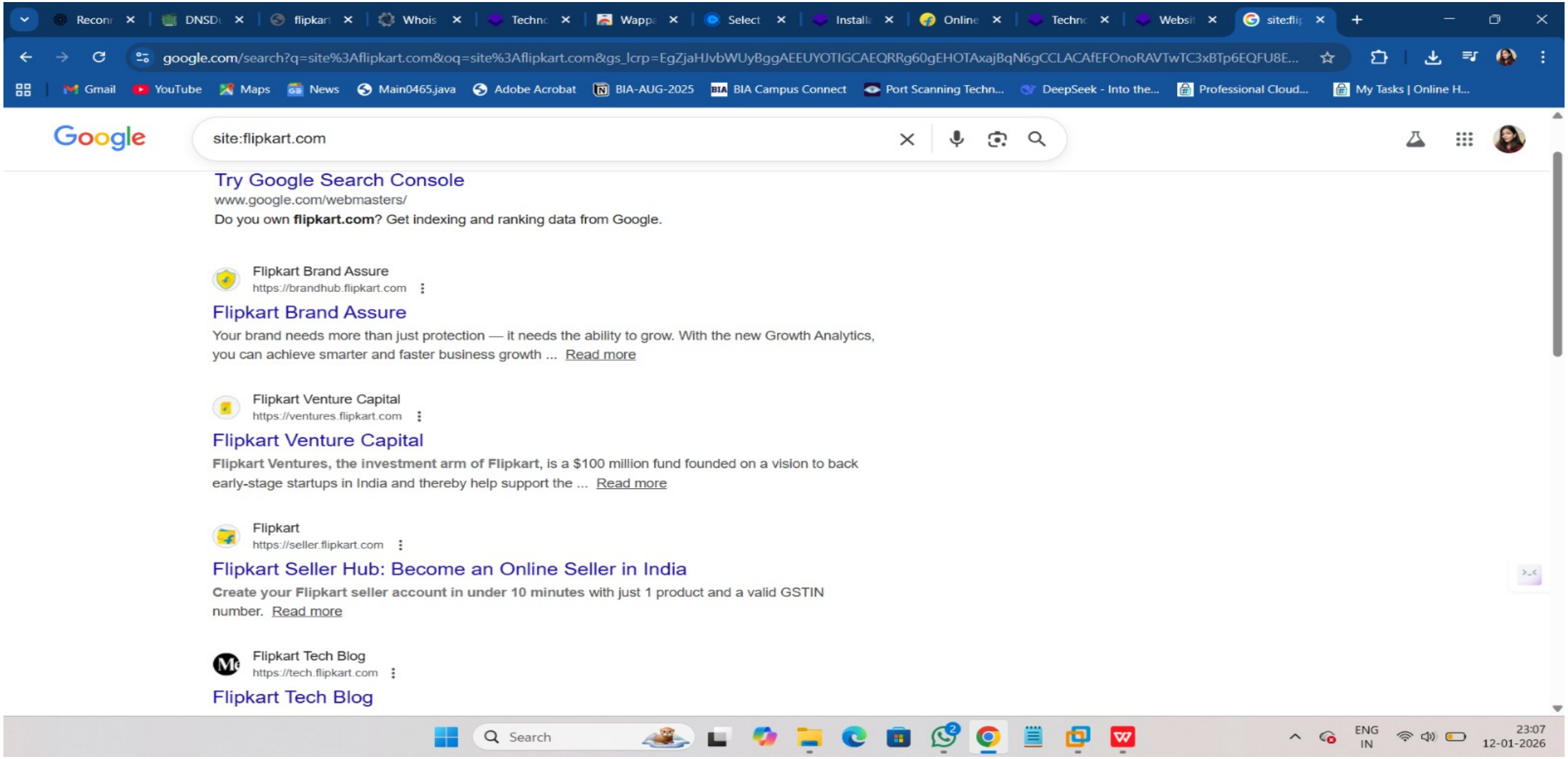
site:flipkart.com login

site:flipkart.com "@flipkart.com"

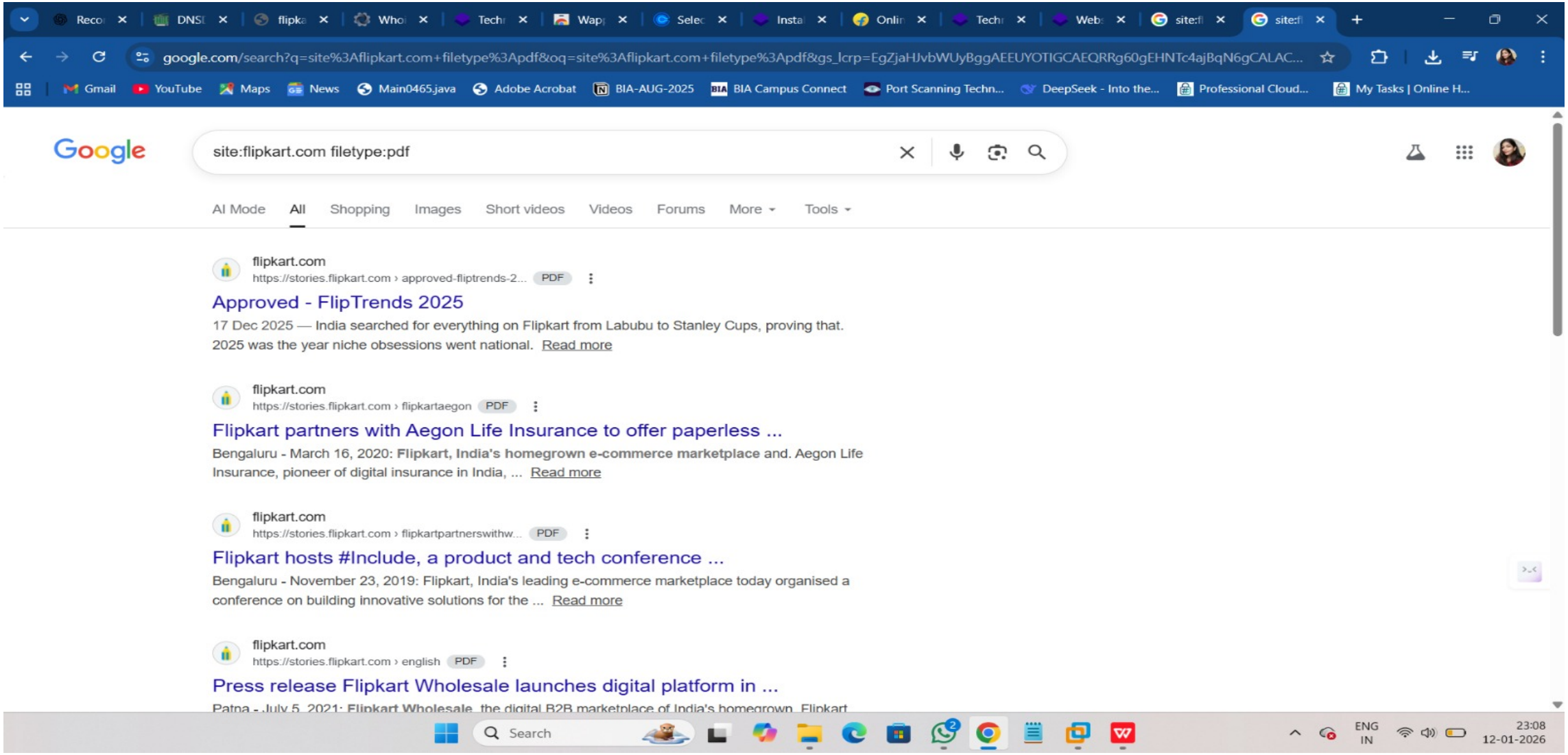
Outcome:

- Public documents identified.
- Support pages.
- No sensitive data exposure found.

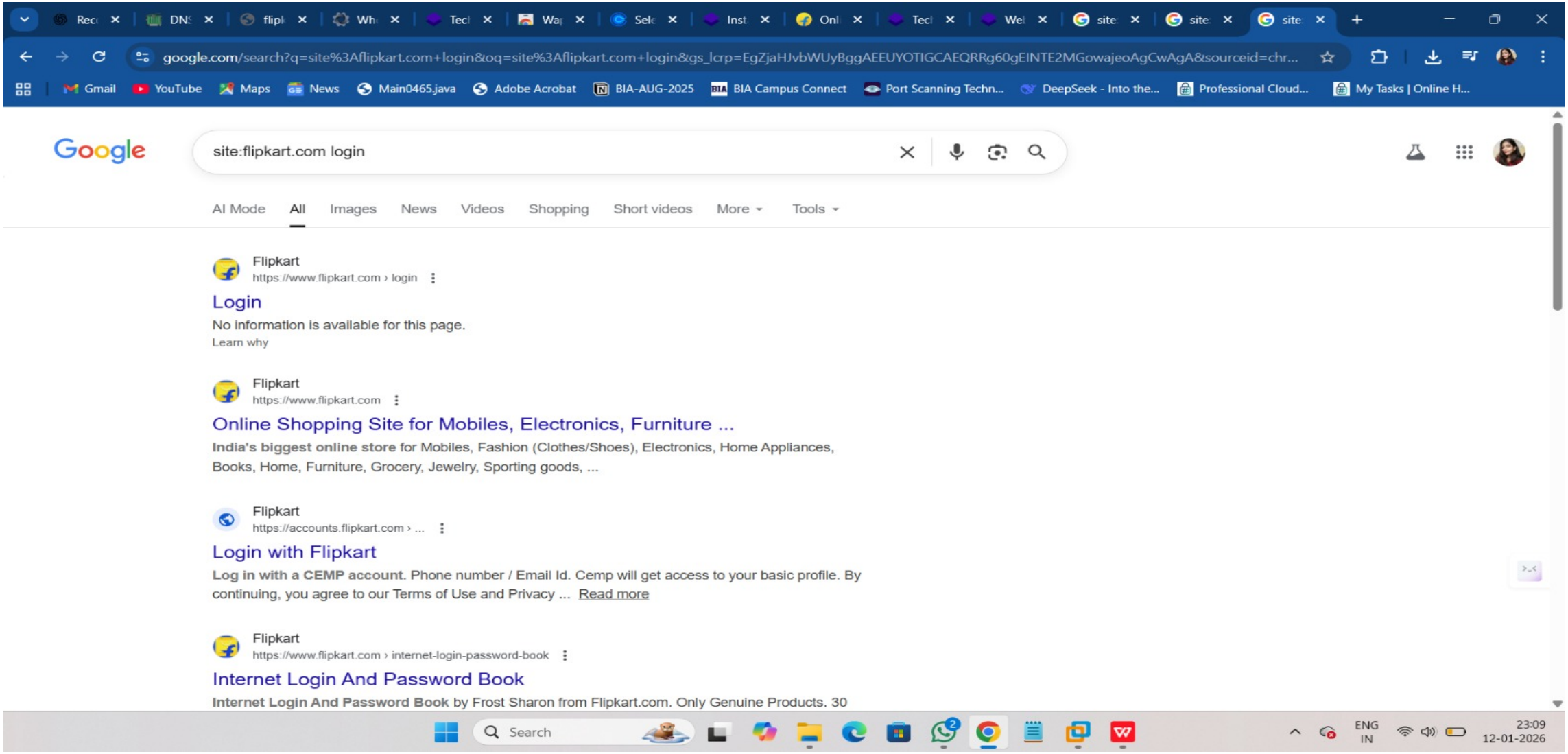
Google Dorking - 1



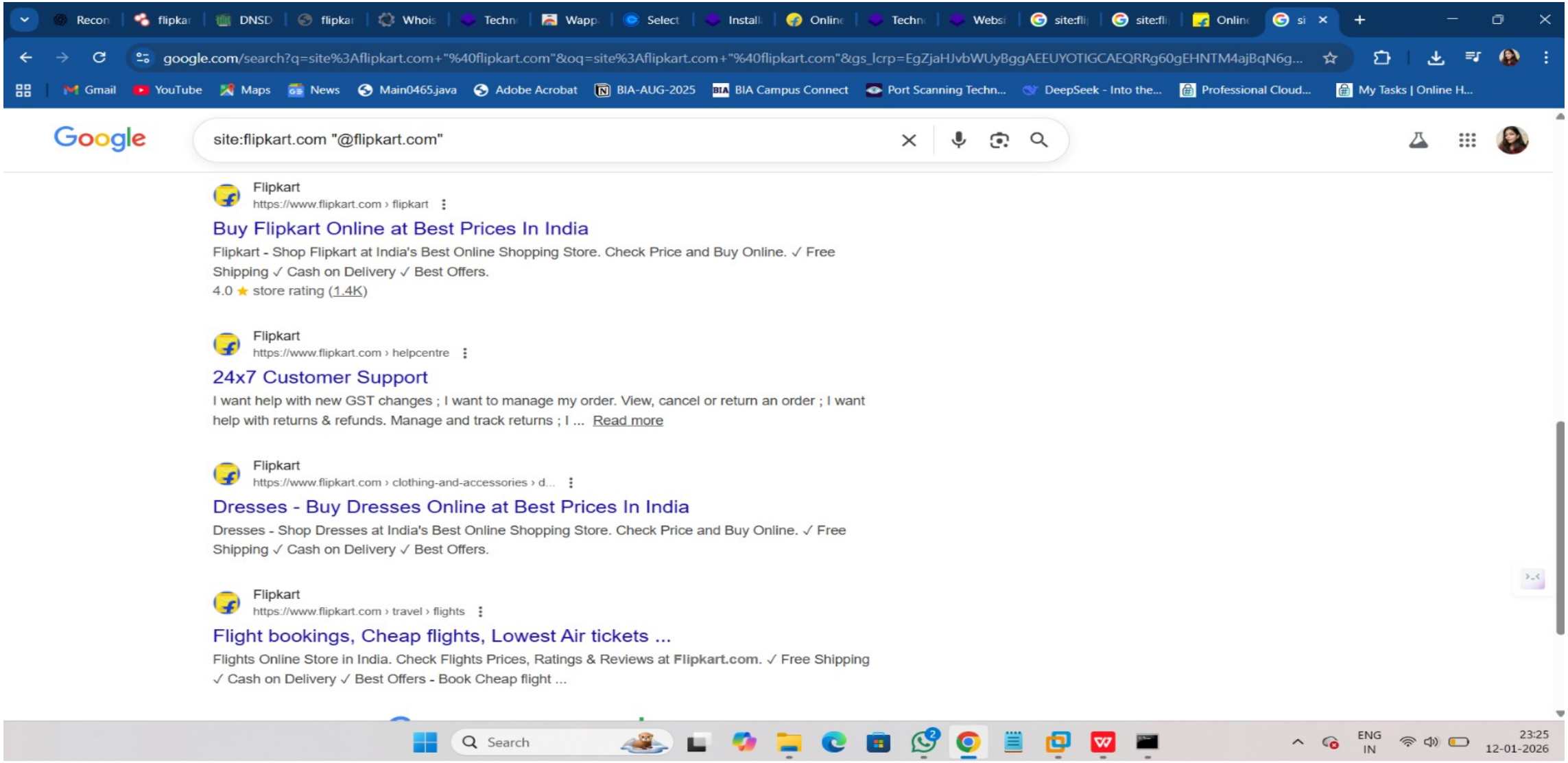
Google Dorking - 2



Google Dorking - 3



Google Dorking - 4



7. IP Address & Hosting Information

Purpose:

- Check if the host is reachable
- Identify approximate IP address

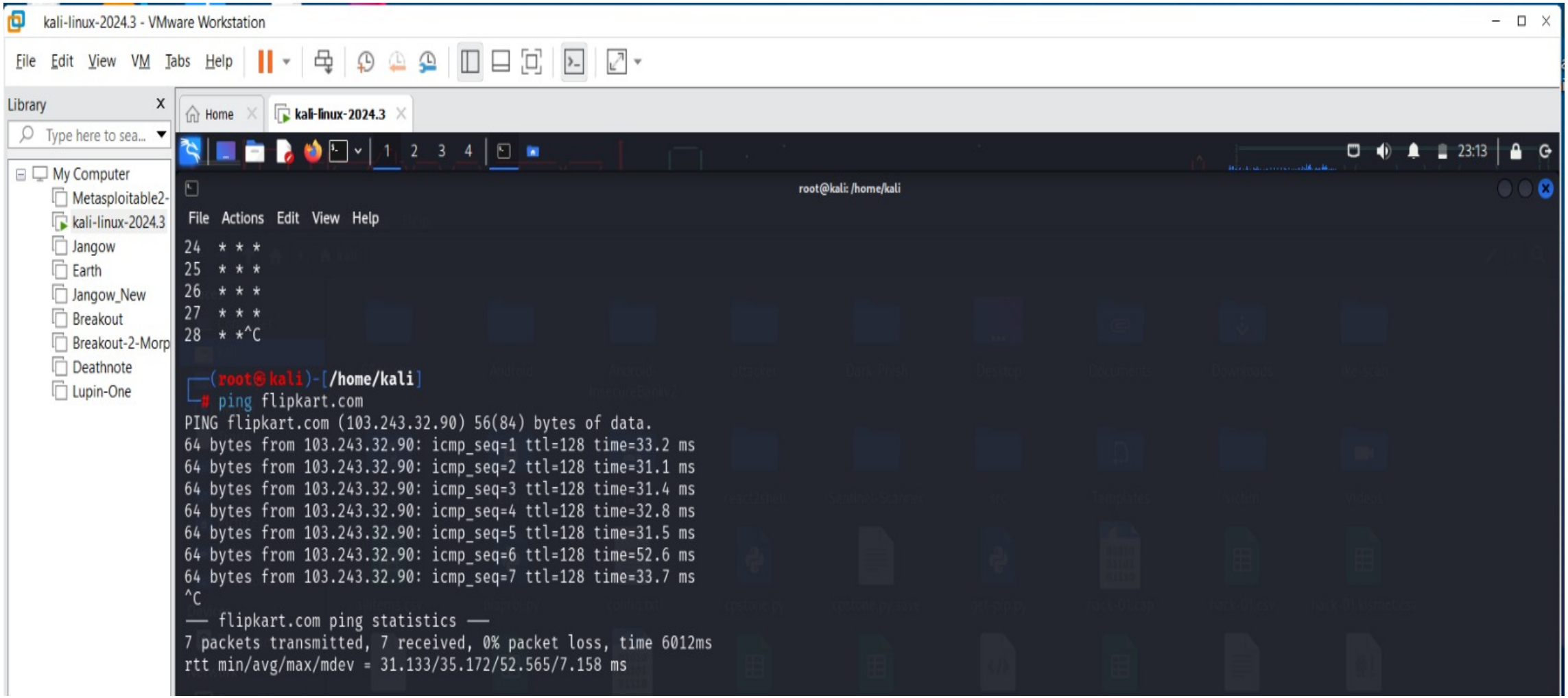
Command:

```
ping flipkart.com
```

Outcome:

- Shows response time
- Confirms CDN or load balancer usage

IP Address & Hosting Information - 1



8. OSINT via Shodan (Passive)

Purpose:

- Identify exposed services

Method:

- Visit: https://www.shodan.io

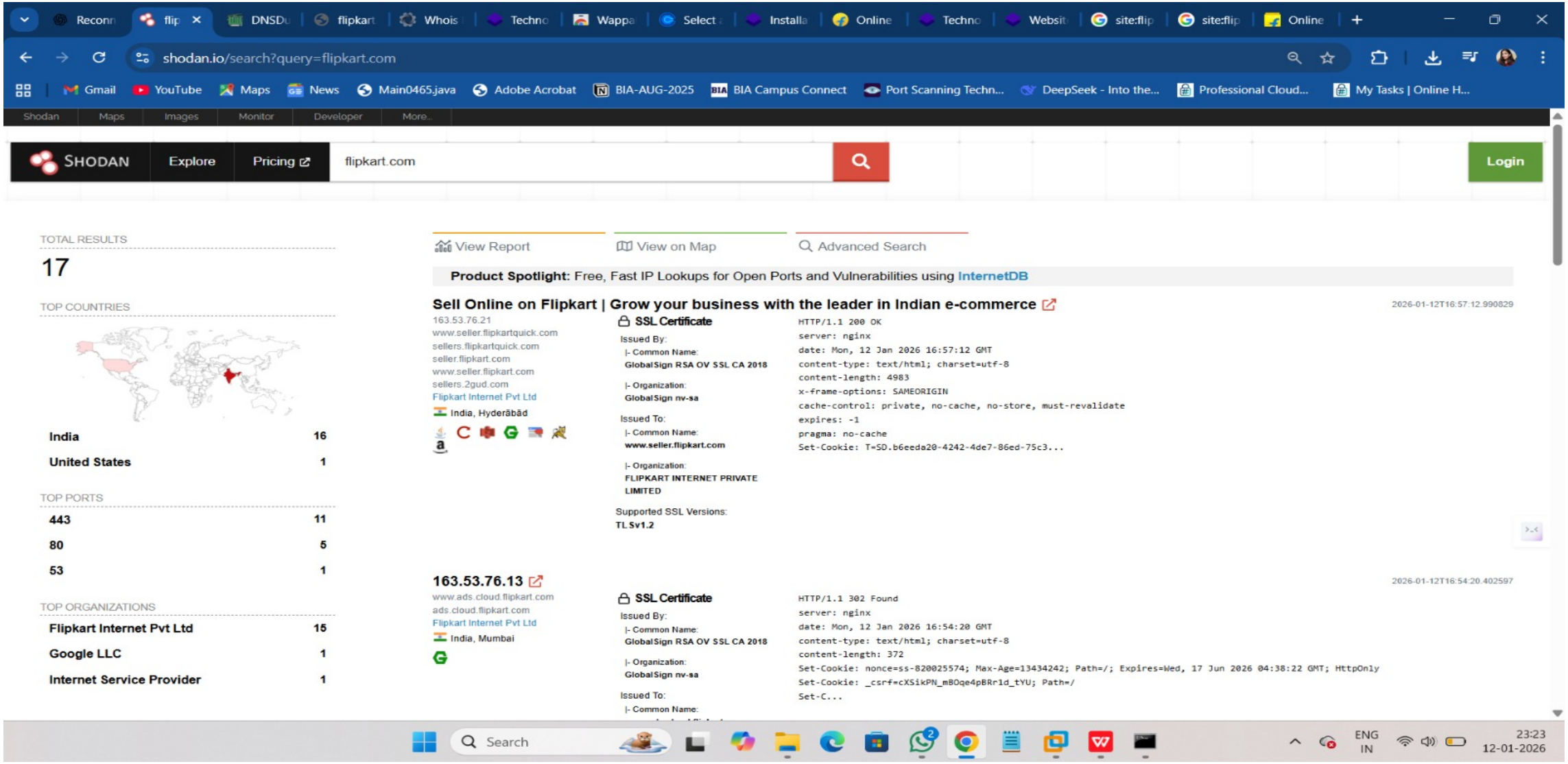
Search:

flipkart.com

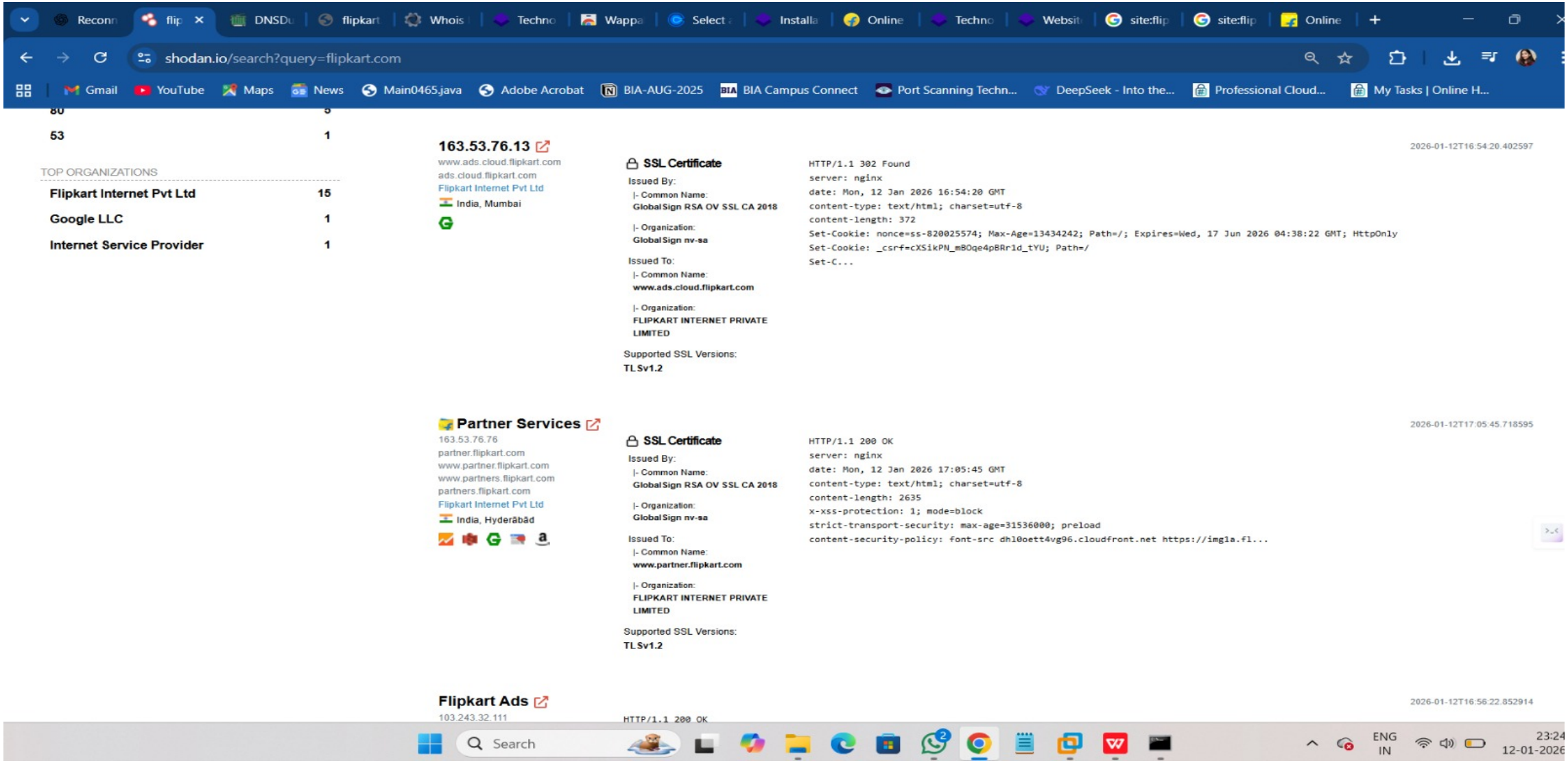
Outcome:

- Cloud infrastructure visibility

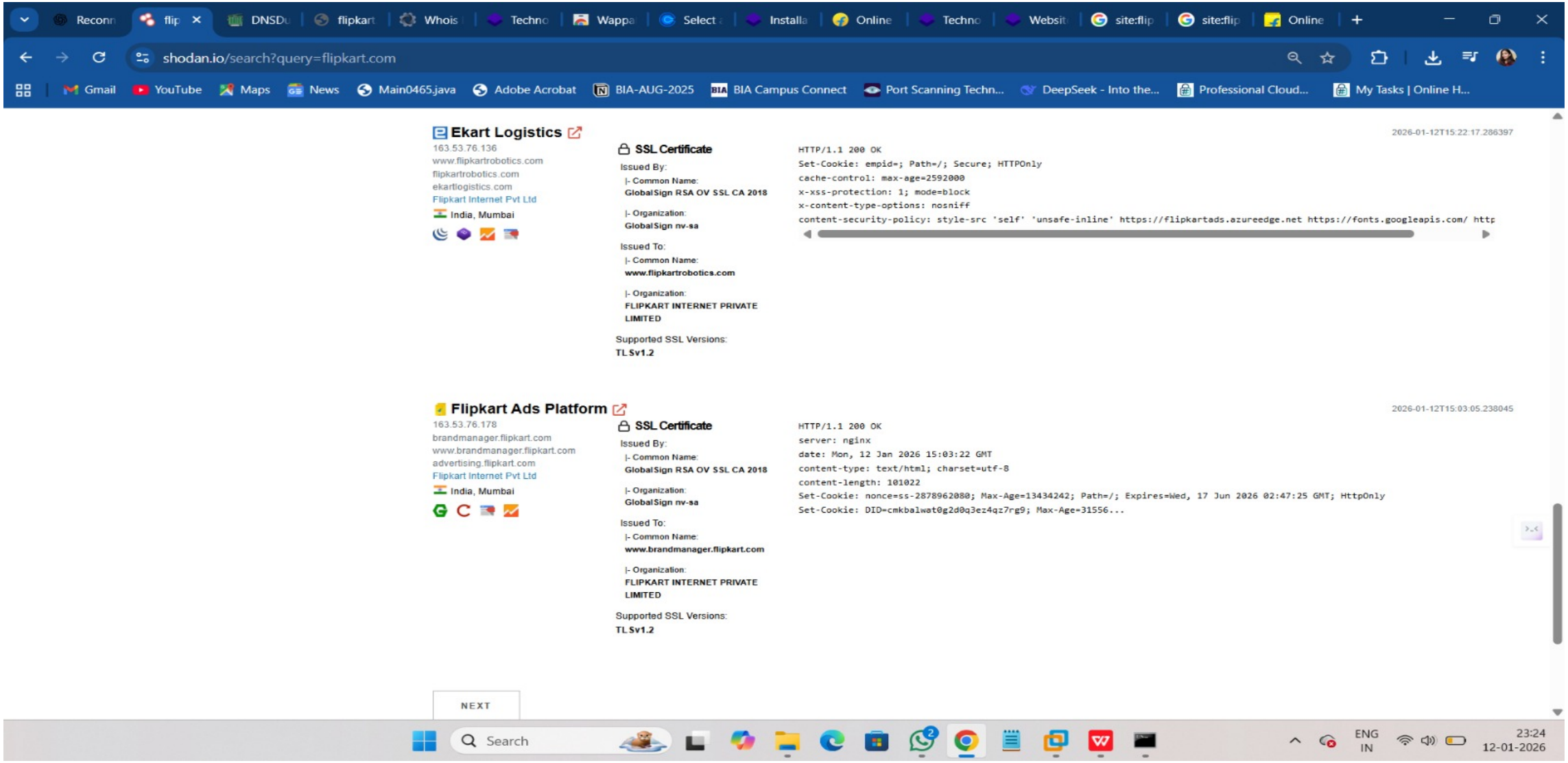
OSINT via Shodan (Passive) - 1



OSINT via Shodan (Passive) - 2



OSINT via Shodan (Passive) - 3



9. THEHARVESTER (OSINT & Email Reconnaissance)

Purpose:

- Expands OSINT coverage

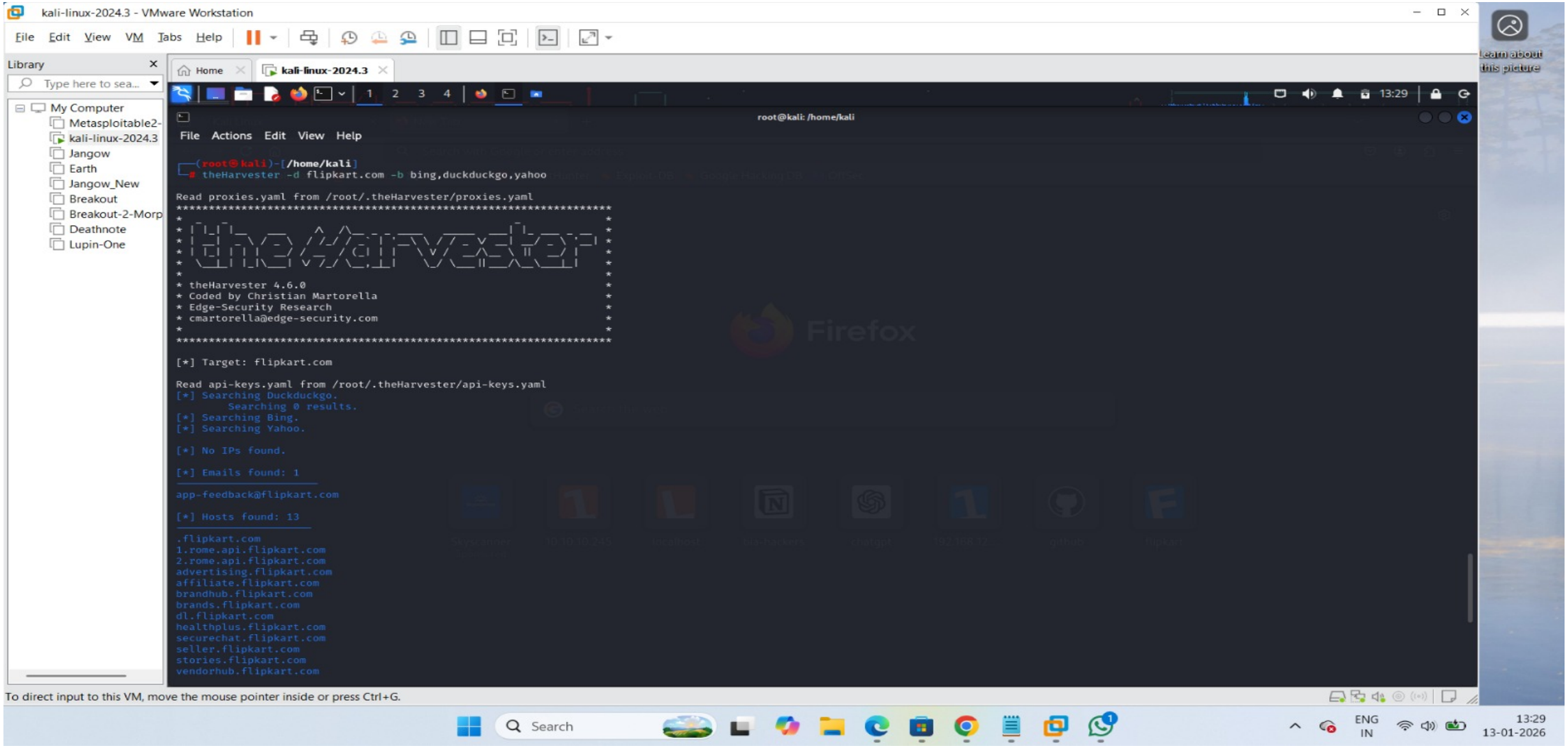
Command:

```
theHarvester -d flipkart.com -b bing,duckduckgo,yahoo
```

Outcome:

- More subdomains
- Additional hosts

THEHARVESTER (OSINT & Email Reconnaissance) - 1



10. Robots.txt File Analysis

Purpose:

- To identify directories and URLs restricted from search engine indexing
- Helps understand which sections of the website are intentionally hidden from crawlers

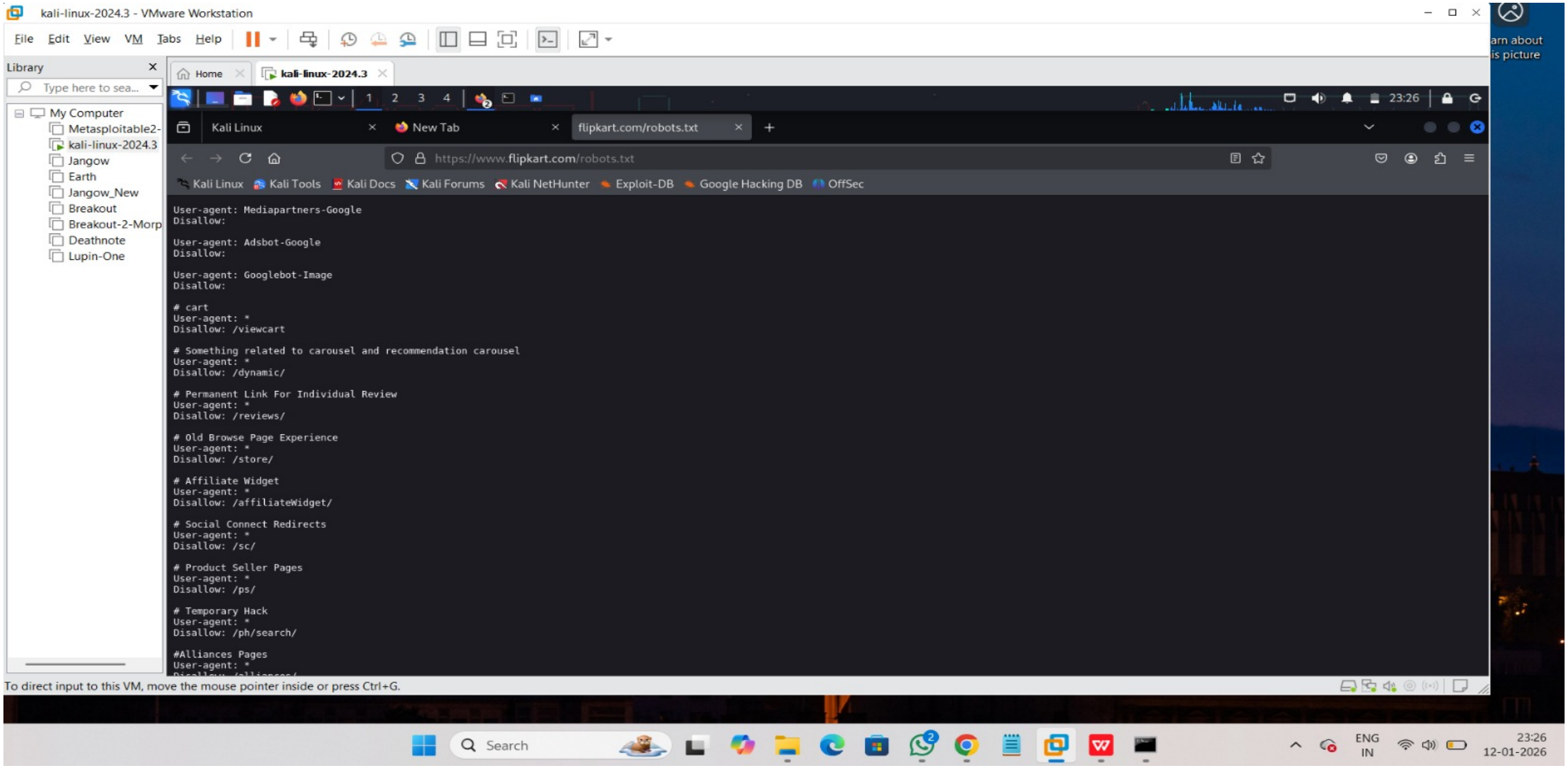
Command:

```
curl https://www.flipkart.com/robots.txt
```

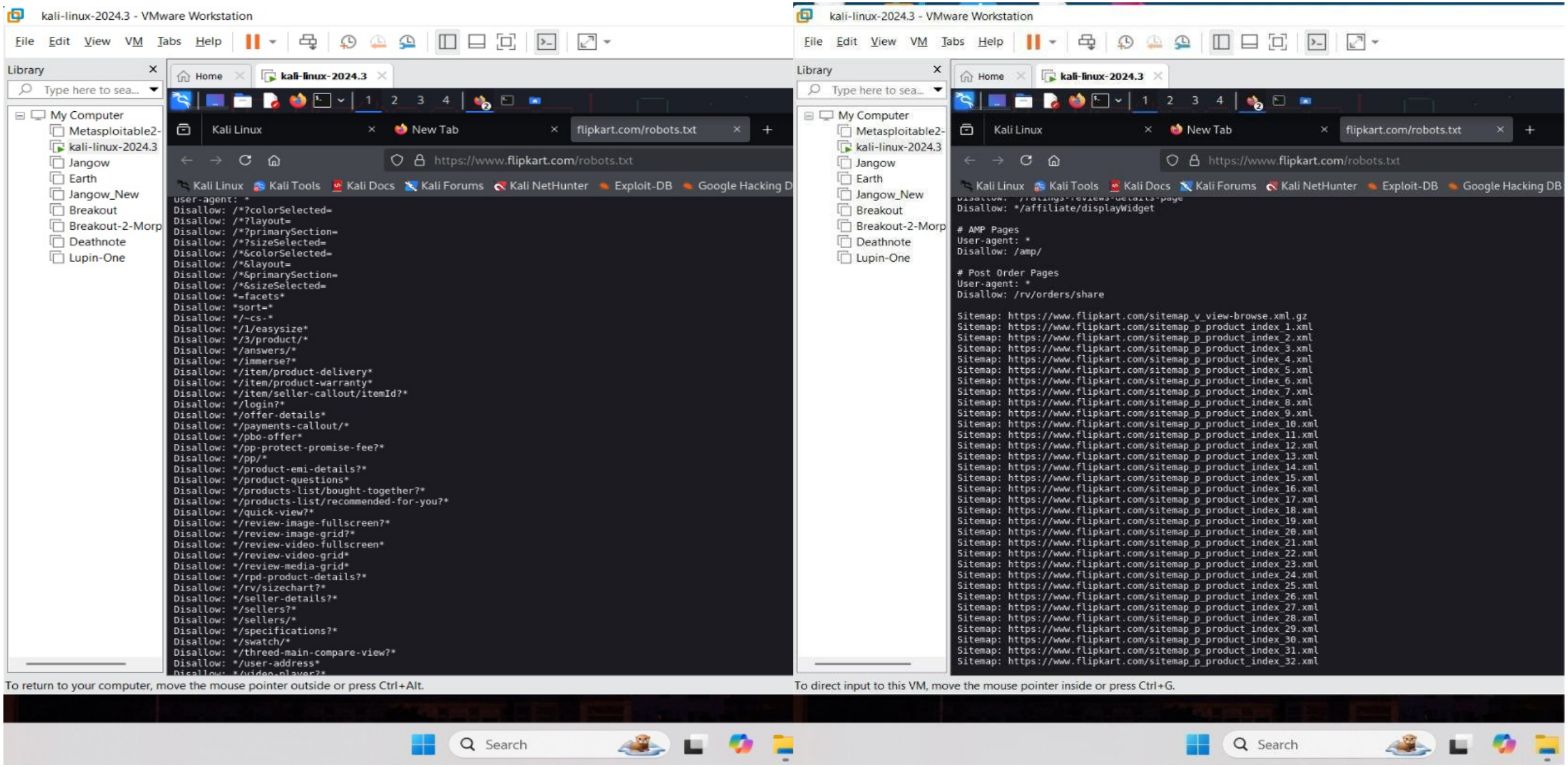
Outcome:

- Displays a list of disallowed paths for search engines
- Indicates areas that are not indexed publicly
- Useful for understanding information exposure control

Robots.txt File Analysis - 1



Robots.txt File Analysis - 2



11. HTTP Header Analysis

Purpose:

- View HTTP response headers

Command:

```
curl -I https://www.flipkart.com
```

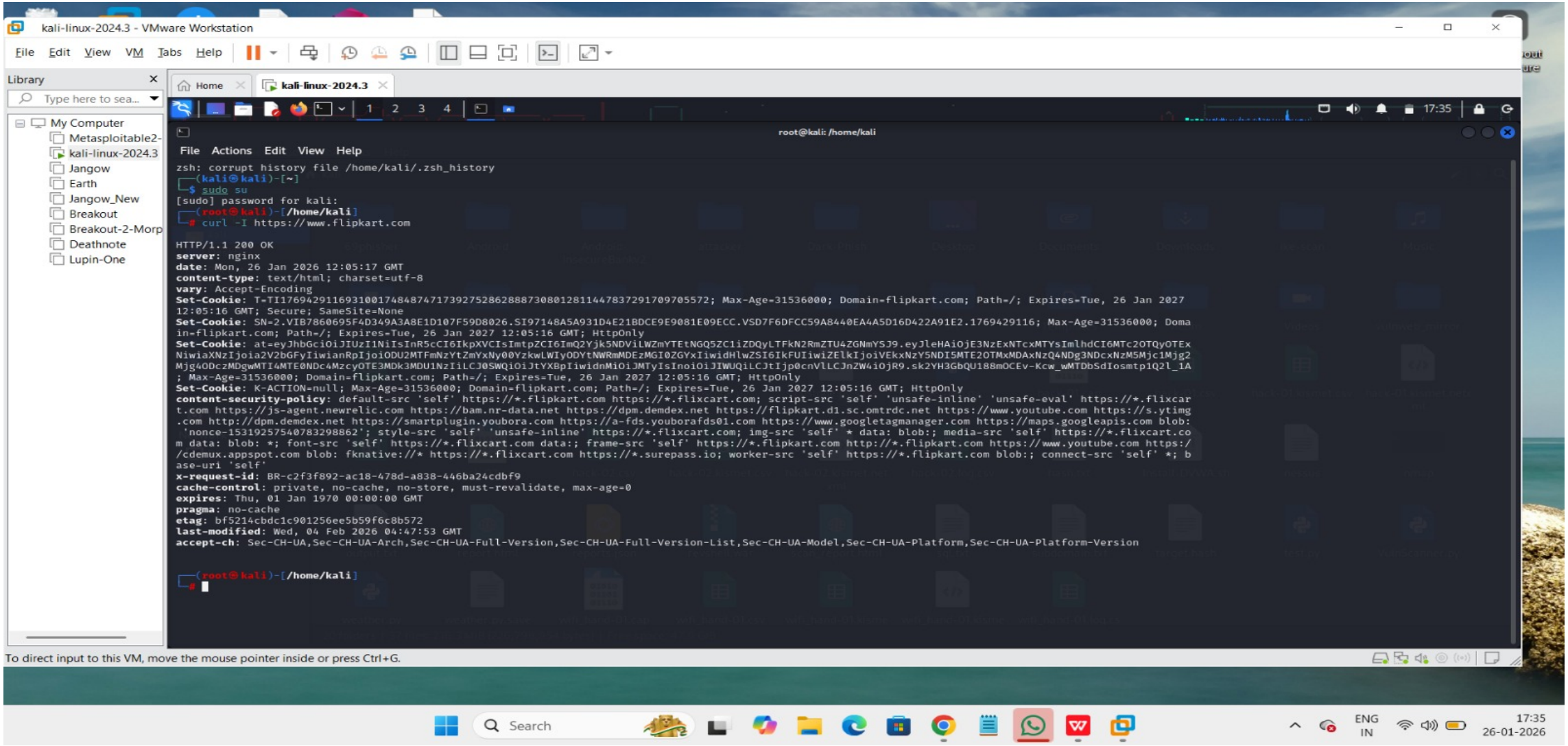
Information Collected:

- Server type
- Security headers
- Cookies

Outcome:

- Confirms HTTPS
- Security policies detected

HTTP Header Analysis - 1



Data Collection & Technology Stack

Data Collection Method:

- Passive reconnaissance using WHOIS, DNS queries, OSINT tools, and technology fingerprinting
- (No intrusive scanning performed)

Information Collected:

- Domain and registrar details
- DNS records and IP addresses
- Public subdomains
- HTTP headers and SSL certificate details

Technology Stack:

Layer	Technology
Frontend	HTML5, JavaScript
Backend	Java-based services
Server	Cloud-based servers
CDN	Akamai / Cloud CDN
Security	HTTPS, Security Headers
Analytics	Google Analytics

Impact Analysis

Positive Impact:

- Helps organizations understand exposure
- Improves security posture
- Enables proactive defense

Negative Impact:

- Attackers map attack surface
- Information leakage risks
- Targeted attacks become easier

Recommendations & Mitigation Techniques

Security Measures:

- Hide server banners
- Implement Web Application Firewall (WAF)
- Disable unnecessary services
- Use DNS security (DNSSEC)
- Regular vulnerability scanning

Best Practices:

- Monitor reconnaissance attempts
- Rate-limit scanning requests
- Apply least-privilege principle
- Conduct regular penetration testing

Conclusion

- The project demonstrated the importance of reconnaissance as the first phase of cybersecurity assessment.
- Ethical passive reconnaissance tools were used to gather publicly available information about Flipkart.com.
- The analysis showed that Flipkart exposes minimal technical details, indicating a strong security posture.
- Proper use of security mechanisms such as HTTPS, security headers, and CDN reduces information leakage.
- The project concludes that responsible reconnaissance helps organizations identify exposed information and improve security defenses.

References

- OWASP Foundation – Web Security Testing Guide (Information Gathering) - https://owasp.org
- Nmap Project – Network Discovery and Security Auditing Tool - https://nmap.org
- Wappalyzer – Web Technology Detection Tool - https://www.wappalyzer.com
- DNSDumpster – DNS Reconnaissance & Research Tool - https://dnsdumpster.co
- GeeksforGeeks – Cybersecurity and Reconnaissance Concepts - https://www.geeksforgeeks.org
- DomainTools – WHOIS Lookup Service - https://whois.domaintools.com
- HackerOne – Ethical Hacking & Bug Bounty Platform - https://www.hackerone.com
- Google Developers – Robots.txt Specification - https://developers.google.com/search/docs/crawling-indexing/robots/robots_txt
- Flipkart Official Website - https://www.flipkart.com