

Q1 2026 Quarterly Breach Report

Identity Risk in the First Quarter



Table of Contents

- 3. Foreword**
- 4. About This Report**
- 5. Executive Summary**
- 6. Key Data Insights**
- 7. General Statistics**
- 10. Data Composition Analysis**
- 13. Breach Metadata Information**
- 14. Top Breaches & Geographic Impact**
- 16. Industry Sector Analysis**
- 17. Infostealer Intelligence**
- 19. Key Recommendations**
- 21. Conclusion**



IDENTITY HAS BECOME THE FRONTLINE:

Why Organizations Must Pivot Their Defense Strategy in 2026

Q1 2026 establishes our quarterly reporting baseline. The first quarter confirms that the industrialization of identity risk is now the operating standard for threat actors globally. In the first three months of 2026, Constella's threat-hunting agents identified over 229,000 breaches across the surface, deep, and dark web, yielding 9.73 billion curated records and 56.5 billion attributes now indexed in our Data Lake.

Hacked-Leaked (HL) breaches account for 72% of verified Q1 incidents, the largest category in our curated dataset. This reflects how the data is validated, not raw breach volume: Combo and URL Log sources carry high rates of recycled credentials and shrink substantially during deduplication, leaving them underrepresented in the final count. We surface HL breaches here for their organizational attribution and richer identity data, not as evidence that attackers have moved away from other sources. Threat actors target primary sources such as government registries, telecom databases, and financial systems to harvest fresh, high-value personally identifiable information (PII) at scale. Our analysis shows 95.59% of HL breaches contained data well beyond email-password pairs, including phone numbers, addresses, national IDs, and financial identifiers, complete profiles that enable account takeovers, social engineering, and targeted fraud.



Equally striking is how often credentials appear in plaintext: 42.29% of curated records contained clear-text passwords, immediately usable and requiring no effort to crack. This figure is driven largely by infostealer URL Log data captured from victims' browsers, so it measures endpoint exposure more than enterprise storage. Combined with weak hashing algorithms like MD5 (24.63% of detected hashes) and a surge in infostealer packages (31.6 million processed in Q1 from 2.77 million infected devices), the attack surface remains at unprecedented scale. These infostealers harvest session tokens, authentication cookies, and cached credentials from endpoints, bypassing password security entirely.

For years, the identity conversation treated credentials and multi-factor authentication as components of a broader strategy. That must change. Identity is no longer a supporting element of enterprise security; it is the primary attack surface, where adversaries operate most efficiently and machine-scale threats demand machine-scale response.

Identity risk is no longer peripheral. It is where defenders meet industrialized adversaries, and it demands boardroom-level commitment. In 2026, it is not only a cybersecurity issue; it is a business continuity, brand protection, and fiduciary issue.

This report provides the intelligence to understand the challenge; our recommendations offer the roadmap for response. The question is not whether your organization will face identity-based threats, but whether it will have the visibility, intelligence, and response capability to defend at scale.

Andres Andreu
CEO, Constella



About Constella's Identity Breach Report



The **Q1 2026 Quarterly Breach Report** represents the first in our 2026 quarterly series, establishing the baseline metrics for ongoing quarterly measurements. This report covers threat intelligence data collected from January through March 2026 across the surface, deep, and dark web.

This report provides a snapshot of the global identity-based risk landscape as monitored by Constella during Q1 2026. Unlike the comprehensive annual report, this quarterly update focuses on:

- Verified, deduplicated breach data ingested during Q1 2026
- High-level geographic and sectoral trends
- Infostealer activity and infection landscapes
- PII exposure patterns and password storage methodologies

Starting with the Q2 2026 report, we will introduce quarter-over-quarter and year-over-year comparative analysis.

Constella continuously monitors the Tactics, Techniques, and Procedures (TTPs) of global threat actors to identify breaches occurring across Q1 2026.

Our verification process includes:

- 1. Collection & Deduplication:** Raw data from infostealer logs, underground forums, dark web repositories, and Agentic AI-driven discovery is processed through machine learning algorithms to remove duplicate records and "zombie" data.
- 2. AI-Enhanced Validation:** Agentic AI automation performs preliminary triage, flagging sensitive information and validating domain authenticity at scale.
- 3. Expert Attribution:** Subject matter experts verify domain authenticity and breach details, attributing each incident to a specific source and sector.
- 4. Risk Scoring:** Verified breaches are scored based on password storage methodology, PII richness, and attribute density.

By applying this rigorous methodology, Constella converts raw underground data into actionable identity risk intelligence.





EXECUTIVE SUMMARY

The Q1 2026 Baseline: Scale and Composition

Q1 2026 establishes the quantitative baseline for 2026's quarterly reporting. The quarter's headline metrics:

229,472 breaches hunted across surface, deep, and dark web sources

3,685 verified breaches ingested after deduplication and quality validation

9.73 billion curated records inserted into the Data Lake

56.5 billion attributes indexed and deduplicated

31.6 million infostealer packages processed from

2.77 million unique infected devices

Data Lake Snapshot

1.41 million total breaches (cumulative, all-time)

63.95 billion records (cumulative, all-time)

486.5 billion curated attributes (cumulative, all-time)

These figures establish the foundation against which Q2 2026 and subsequent quarters will be measured.



Key Data Insights



1 | THE HACKED-LEAKED SURGE: PRIMARY DATA COMPROMISE

HL breaches make up 72.08% of verified Q1 2026 incidents, the largest share of our curated dataset. This reflects how the data is validated and deduplicated: Combo and URL Log sources are reduced heavily during processing because they carry recycled and previously observed records, so HL incidents are proportionally larger in the verified set. HL breaches are significant because they come from direct compromises of primary data sources such as government registries, telecom databases, and financial systems, and are typically tied to identifiable organizations. The SudamericaData incident (167.8M records) exemplifies this: a data broker's entire holdings were exfiltrated, exposing not just credentials but phone numbers, addresses, national IDs, and financial identifiers across Latin America.



2 | PII DENSITY OVER VOLUME

Q1 2026 saw 3,685 verified breach incidents in a single quarter, measured against 8,460 for all of FY2025. This is a single quarter compared with a full fiscal year: if the current processing pace holds across all four quarters, 2026 is on track to exceed 14,800 verified inserted breaches, so 3,685 in one quarter is a strong baseline figure. Alongside the rising count, attribute density per incident has increased significantly. Q1's average breach contains Email (39.88%), Phone (45.79%), and Address (25.35%), reflecting adversarial targeting of primary data sources such as telecom registries, DMV databases, and financial institutions rather than recycled aggregates.



3 | INFOTEALER DOMINANCE: THE NEW NORMAL

Infostealer malware (31.6M packages, 2.77M infected devices) now represents the largest vector for PII extraction. These tools harvest session cookies, cached credentials, and browser autofill data, enabling MFA bypass and account takeover without credential knowledge. With 99.61% of infostealer records containing URLs and 97.62% containing passwords, infostealers are purpose-built for automated account takeover campaigns.



4 | GEOGRAPHIC RISK CONCENTRATION

Infostealer activity is heavily concentrated in Brazil (44.52%), MENA (Egypt 24.32%), and Southeast Asia (Vietnam 24.26%, Indonesia 19.38%). Organizations with users, employees, or customers in these geographies face elevated risk.



5 | THE CRYPTOGRAPHY WEAK LINKS

Of detected hashing algorithms in Q1 breaches, only 2.42% employed robust standards (Bcrypt + SHA-256). MD5 (24.63%) and other weak algorithms dominate, enabling rapid password cracking and account takeover.



2026 Q1 General Statistics

Most exposed attributes, sectors, geographies, and key metadata



Q1 2026 Ingestion Volume and Composition

Cumulative Data Lake Volume at Q1 2026 Close.

The figures below are cumulative totals across all breach data in Constella's Data Lake as of the close of Q1 2026, not new ingestions recorded during the quarter alone. The 1,411,487 breaches, 63.95 billion records, and 486.5 billion attributes represent the full accumulated dataset, and they establish the reference baseline for measuring subsequent quarterly growth.

Ingested Breaches

1,411,487

Ingested Records

63,951,164,039

Attributes Inserted

486,517,506,806

Hunted vs. Ingested: Q1 2026



Hunted breaches represent raw threat intelligence identified through automated scanning, dark web monitoring, and intelligence partnerships. Ingested breaches are the verified, deduplicated subset that remains after curation, sanitization, validation, and quality filtering, processes that remove duplicated, recycled, and previously observed records, which are especially common in Combo and URL Log sources. The ingested dataset is what makes the data actionable: validated, deduplicated identity risk intelligence rather than raw volume.

CLASSIFICATION	BREACHES	RECORDS
Hunted (Raw)	229,472	113,145,723,596
Ingested (Verified)	3,685	9,732,490,264



Breach Type Distribution: Q1 2026

Hacked-Leaked (HL) breaches dominate at 72.08%, and 2,539 of those (95.59%) contain PII beyond credentials, indicating high-value identity intelligence. Combo breaches aggregate pre-existing credential pairs. URL Logs harvest credentials from infostealer-infected endpoints. HL breaches dominate (72.08%), and 95.59% of HL breaches contain PII beyond just credentials, indicating high-value identity intelligence.

Hacked-Leaked (HL)

2,656

+72.08%



HL with PII

2,539

95.59%



Combo

293

7.95%



URL Logs

736

19.97%

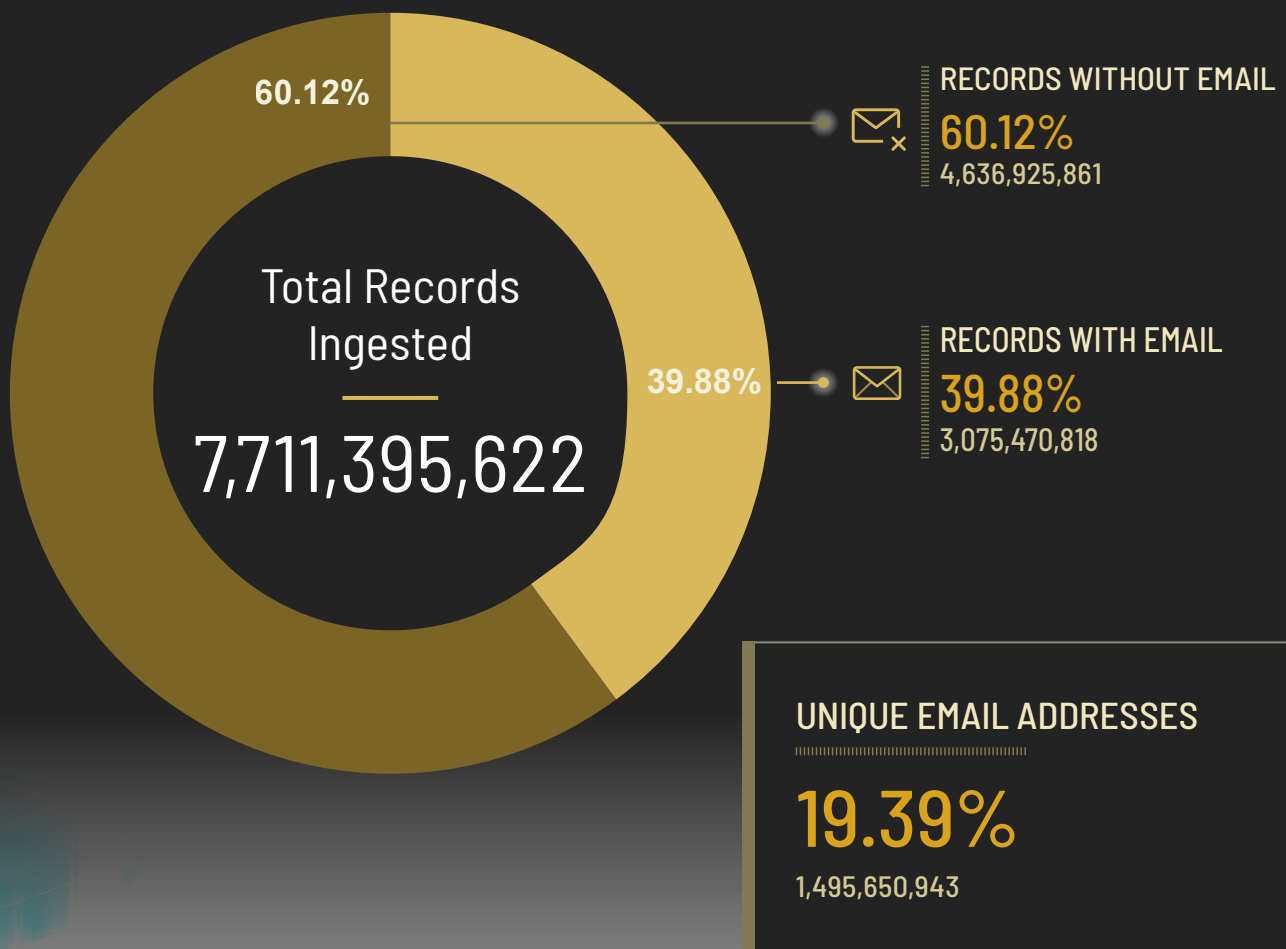




Data Composition Analysis

Records by Email Presence: Q1 2026

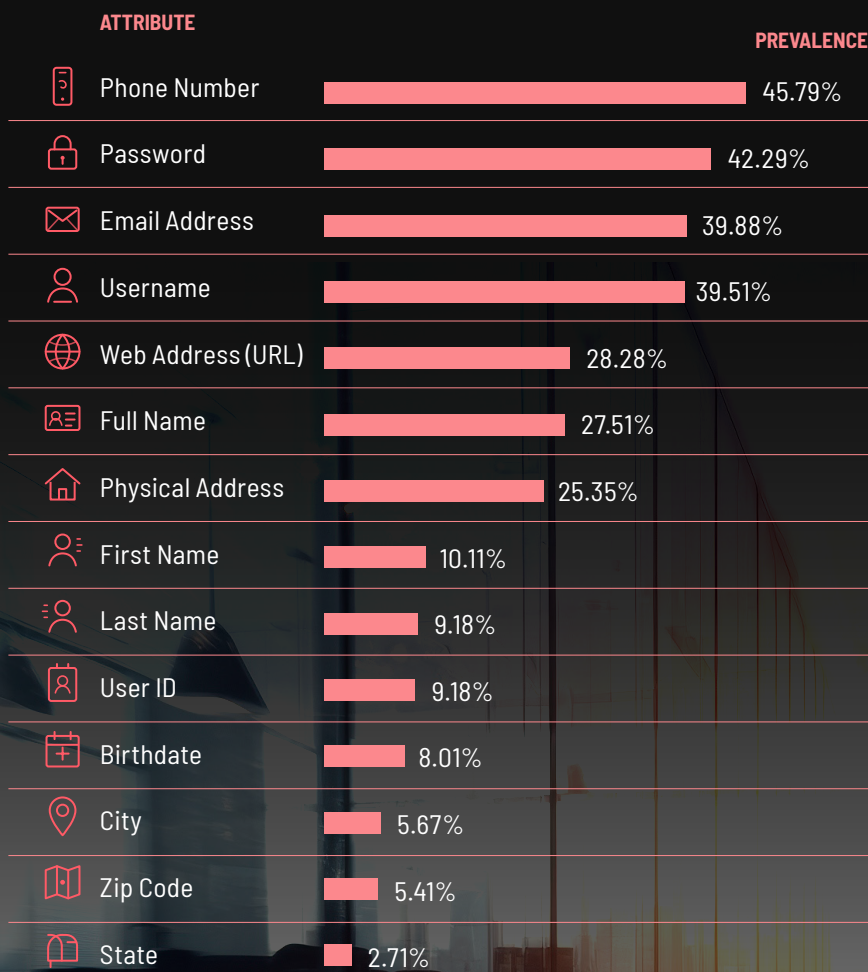
Email addresses serve as the primary pivot point for direct account takeover campaigns. The 60.12% without an email address represent supplementary PII such as phone numbers, physical addresses, identity documents, and financial information, which enable secondary targeting or enrichment of already-compromised accounts. The 19.39% unique email ratio indicates significant record-per-identity multiplicity across breaches.





Top PII Attributes by Prevalence: Q1 2026

PII prevalence across all curated HL and Combo breaches reflects adversarial targeting priorities. Phone numbers (45.79%) and passwords (42.29%) dominate, enabling SIM swap attacks and phone-based account recovery hijacking. The lower email density versus historical data reflects Q1's breach mix being dominated by large-scale country databases (India, China) that expose phones, addresses, and national IDs rather than email-centric credentials.





Social Media Exposure: Q1 2026

Hacked-Leaked breaches frequently expose social media credentials, enabling impersonation and credential-based account takeover. Telegram (59.83%) and Facebook (33.28%) dominate, reflecting widespread use in emerging markets. LinkedIn (4.03%) exposure is notable for targeting enterprise and executive account compromise. Percentages are computed as a share of all social media exposures.



PLATFORM EXPOSURE DISTRIBUTION

Telegram Profile	63,441,372	59.83%
Facebook Profile	35,285,032	33.28%
LinkedIn Profile	4,274,828	4.03%
Twitter Profile	2,716,389	2.56%
Instagram Profile	264,967	0.25%
Skype Profile	46,113	0.04%



BREACH METADATA & CRYPTOGRAPHY: Password Storage Methods: Q1 2026

The most alarming finding: 42.29% of breached passwords are stored in plaintext, requiring zero computational effort to exploit. A further 56.84% of records contained no password at all, typically non-credential PII such as phone numbers, addresses, and identity attributes. Only 0.87% employ hashing—and most of that hash uses deprecated algorithms (see next table). This plaintext prevalence creates immediate account takeover risk.

STORAGE METHOD	RECORDS	% OF TOTAL
Plaintext	3,261,234,565	42.29%
Hashed (Any Algorithm)	67,319,103	0.87%
No Password Present	4,383,843,011	56.84%

Hashing Algorithm Analysis: Q1 2026

Of 67M+ password hashes detected, only 2.42% (Bcrypt + SHA-256 combined) employ cryptographically robust algorithms. MD5 (24.63% of all hashes) is cryptographically weak and can be cracked in milliseconds. The 'Others' category (55.26%) covers application-specific implementations (Joomla, phpBB, PrestaShop), legacy algorithms (HAVAL, RIPEMD, SNEFRU, NTLM, LM), and unidentified hashes. This creates an immediate threat landscape.

ALGORITHM	DETECTIONS	% OF TOTAL
Others/Unknown	47,537,819	55.26%
MD5 (Weak)	21,184,182	24.63%
Skein	5,817,922	6.76%
Tiger	3,877,788	4.51%
Oracle	2,496,748	2.90%
Bcrypt (Robust)	1,864,894	2.17%
MySQL	1,440,223	1.67%
SHA-1	970,716	1.13%
WordPress	620,841	0.72%
SHA-256 (Robust)	213,717	0.25%



Top Breaches & Geographic Risk

Top 5 Breaches by Record Count: q1 2026

The following five breaches account for 270.4 million records—approximately 2.78% of all Q1 records. Generic country-level dumps and credential compilations are excluded to focus on company-attributable incidents. Organizations should cross-reference employee and customer populations against these entities for exposure assessment.

1

SudamericaData

Data Broker

167,879,788

Records

2

Sound Cloud

Streaming

29,892,439
Records

3

CarGurus

Automotive

25,294,065
Records

4

PhoneHouse.es

Retail

24,800,581
Records

5

Colisprive.fr

Logistics

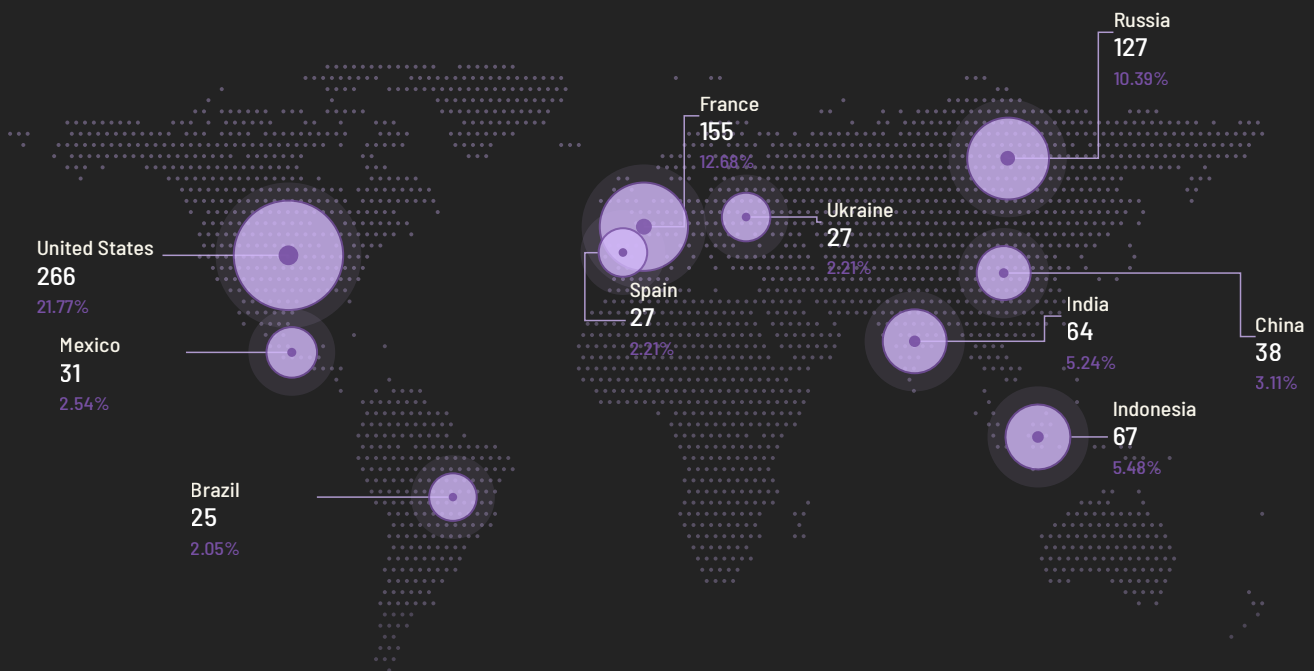
22,526,525
Records

Geographic Distribution of Breaches

Q1 2026

The top 10 countries are shown by number of breaches affecting organizations or users in that region (geographic data extracted from breached company HQ metadata). The United States (21.77%) and France (12.68%) lead incident counts, reflecting both threat actor targeting of Western infrastructure and the prevalence of internet-exposed services in these countries. The top 5 countries account for approximately 55.56% of all geographically-attributed breaches.

Geographic Distribution of Breaches



* Percentages may not total 100% because only the top 10 countries are shown.

COUNTRY	BREACHES	% OF TOTAL
United States	266	21.77%
France	155	12.68%
Russia	127	10.39%
Indonesia	67	5.48%
India	64	5.24%

COUNTRY	BREACHES	% OF TOTAL
China	38	3.11%
Mexico	31	2.54%
Ukraine	27	2.21%
Spain	27	2.21%
Brazil	25	2.05%

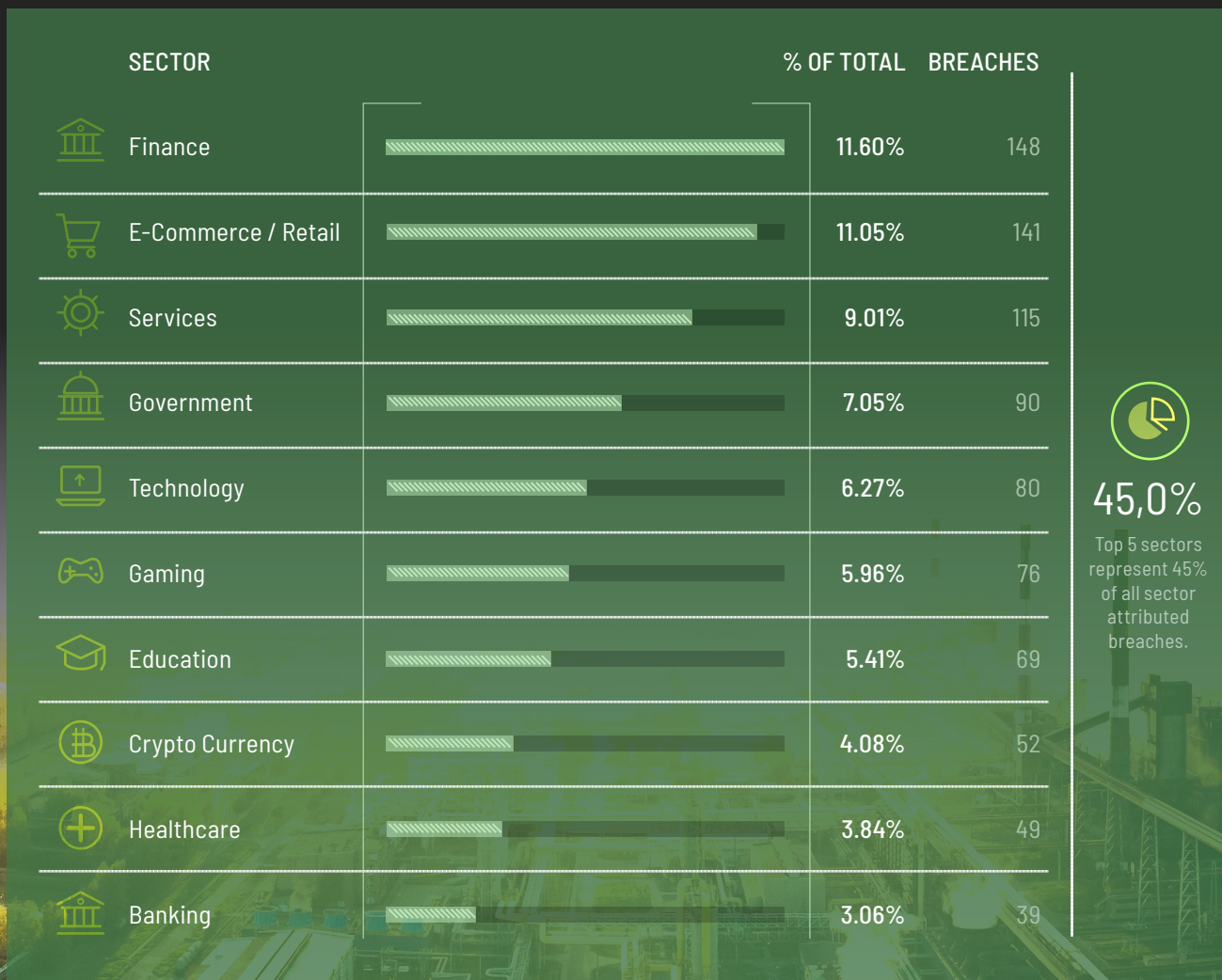


Industry Sector & Infostealer Analysis

Top 10 Industry Sectors by Breach Count

Q1 2026

Finance (11.60%) and E-Commerce/Retail (11.05%) lead breach activity, reflecting the high-value targets these sectors represent. Government (7.05%) exposure is particularly concerning due to citizen data (national IDs, SSNs, addresses) compromise. Technology sector (6.27%) breaches pose supply chain risk. The top 5 sectors (Finance, E-Commerce/Retail, Services, Government, Technology) represent 45.0% of all sector-attributed breaches.





Infostealer Intelligence

Infostealer Activity Metrics

Q1 2026

Infostealer malware is now the primary vector for industrial-scale identity theft. In Q1 2026, Constella processed 31.6 million packages from 2.77 million unique infected devices. The 99.61% prevalence of Web Addresses and 97.62% prevalence of passwords confirm that infostealers are purpose-built for automated account takeover operations. The extraction of Hardware IDs (30.31%) enables persistent device-level targeting.



31,6M

Packages Inserted,
containing infostealer data



2,77M

Unique Infected Devices,
affected by infostealer malware



1,54B

Records Inserted, extracted
from infected devices

Data Extracted



Web Addresses (URLs) Extracted



99.61%

1,538,246,450



Passwords Extracted

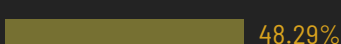


97.62%

1,507,513,356



Email Addresses Extracted

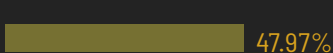


48.29%

745,758,968



Username Extracted



47.97%

740,816,897



Hardware IDs (HWID) Extracted



30.31%

468,135,280



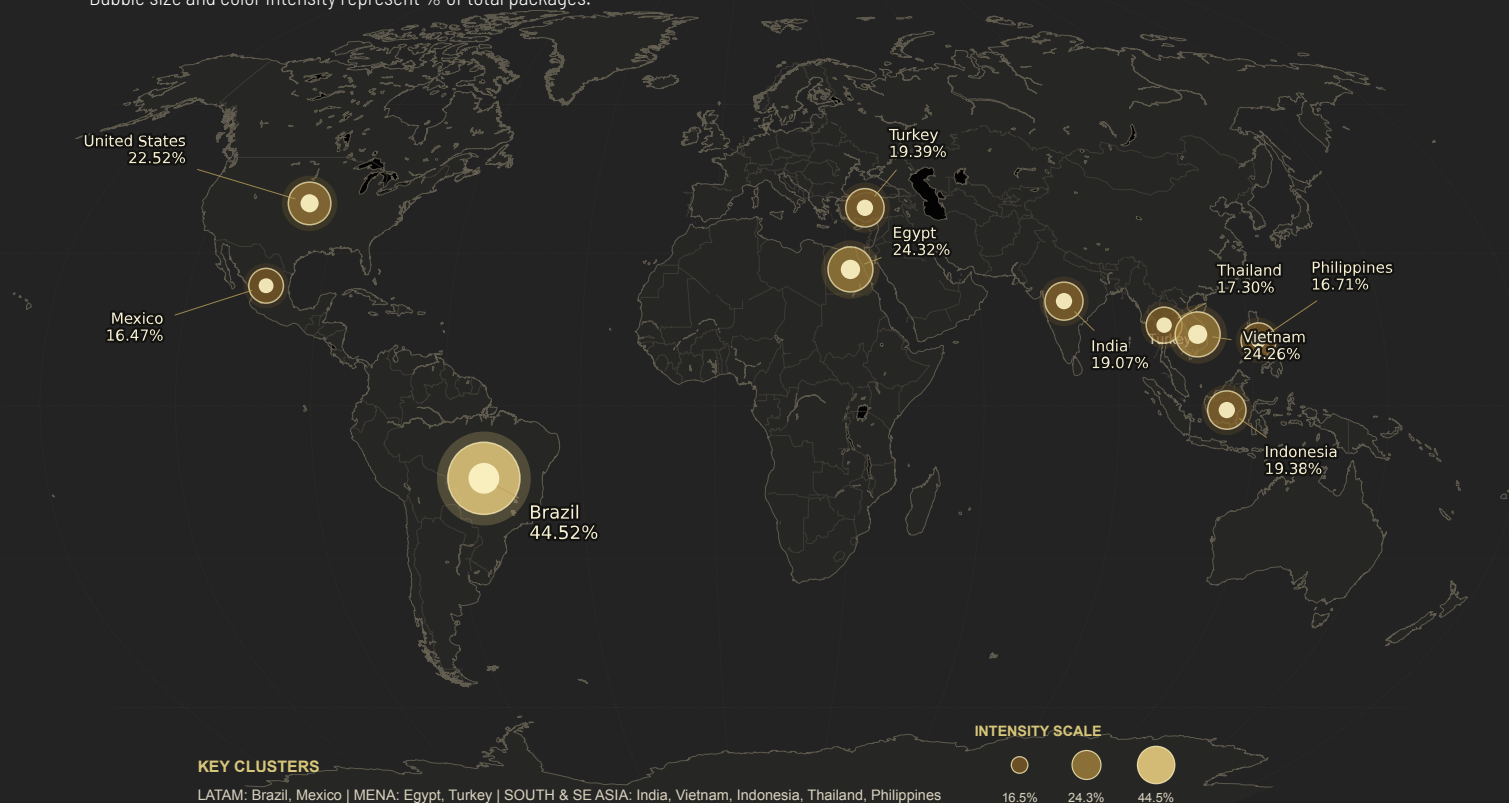
Infostealer Geographic Distribution

Q1 2026

Infostealer activity is heavily concentrated in Brazil (44.52%), MENA (Egypt 24.32%), and Southeast Asia (Vietnam 24.26%, Indonesia 19.38%), reflecting both threat actor infrastructure concentration and emerging market vulnerability. Organizations with users, employees, or customers in these geographies face elevated and asymmetric risk of endpoint compromise.

INFOSTEALER GEOGRAPHIC DISTRIBUTION

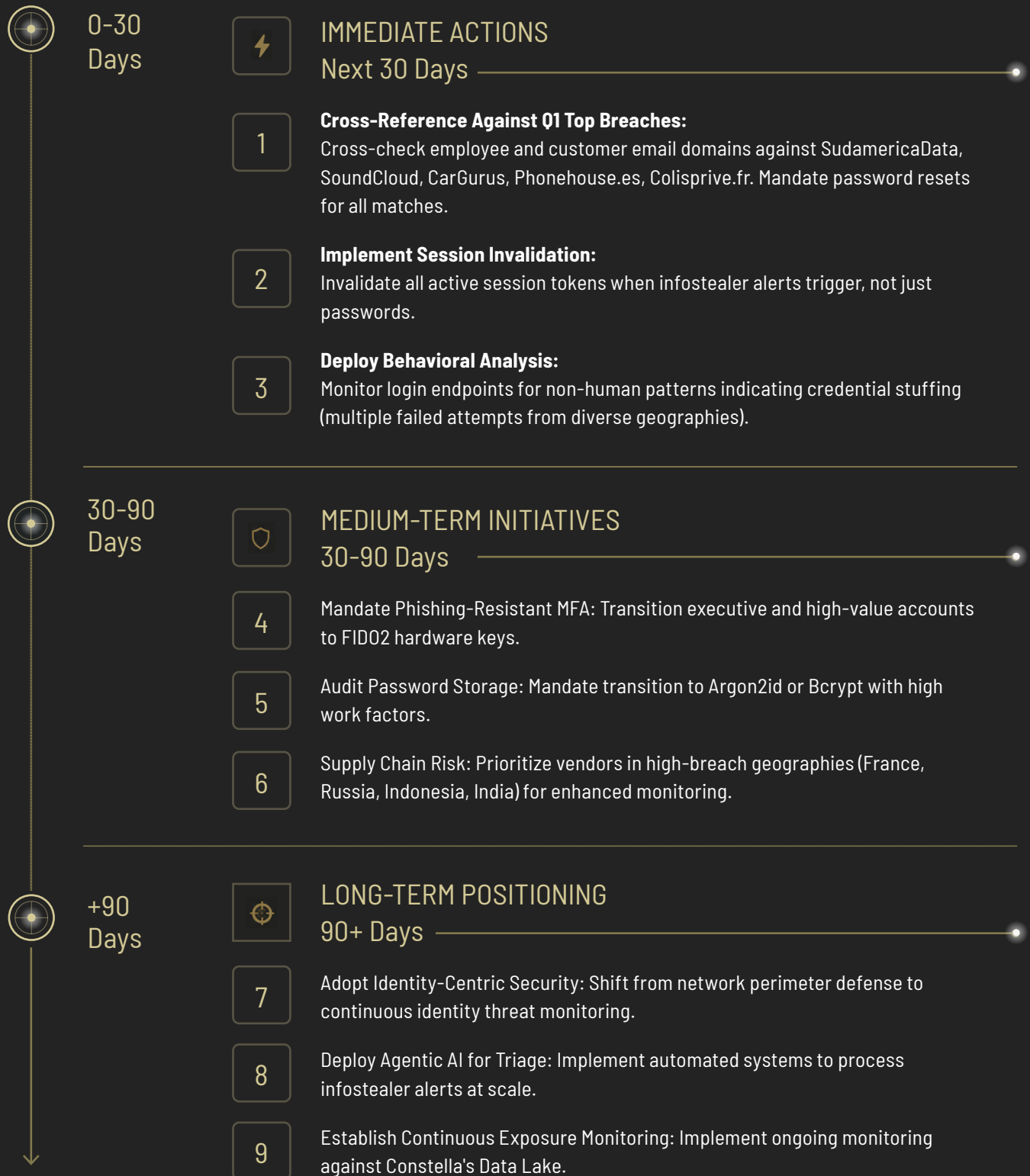
Bubble size and color intensity represent % of total packages.





Key Recommendations for Q2 2026

Strategic actions prioritized by impact and time frame





Appendix

Data Sources and Methodology

Data Collection

All data in this report is derived from Constella's Q1 2026 threat intelligence operations. Data sources include industrialized infostealer logs (31.6M packages from infected devices), underground forums and communities, dark web repositories, and agentic AI-driven discovery.

Citation Method

Every table and statistic in the report is derived from Constella's verified Q1 2026 threat intelligence dataset, processed through the collection, validation, and risk-scoring methodology described above.

Data Verification Process

Collection & Deduplication: Machine learning-based deduplication eliminates duplicate records and "zombie" data. **AI-Enhanced Validation:** Agentic AI performs preliminary triage. **Expert Attribution:** Subject matter experts verify breach authenticity and assign classifications. **Risk Scoring:** Breaches scored based on password storage methodology and PII richness.

Year-over-Year Comparison Notes

Q1 2026 establishes the baseline for quarterly reporting. Year-over-year and quarter-over-quarter comparisons will be introduced starting with the Q2 2026 report. The structure, terminology, and methodologies used here align with the Annual Breach Report 2026 (FY2025 data) for cross-reference consistency.

Limitations & Disclaimers

This report represents threat intelligence collected during Q1 2026. Geographic classifications are based on breached entity headquarters or data provenance. Infostealer data derived from IP geolocation and system locale metadata. Numbers rounded to nearest integer.



Conclusion

Q1 2026 establishes the baseline for 2026's quarterly identity risk landscape. The data confirms that the industrialization of identity theft has become the operational standard for global cybercriminals.

The composition of verified Q1 data, with Hacked-Leaked breaches at 72.08% of incidents, plaintext credentials in 42.29% of records, and 31.6M infostealer packages in circulation, points to a single conclusion: organizations face an industrialized, machine-scale identity threat. Read together, these figures describe the shape and scale of the verified dataset rather than a shift away from any one breach source, and they confirm that identity exposure now demands continuous, verified monitoring.

Organizations that meet this challenge will be those that:

01



Treat identity risk as a boardroom-level, systemic concern.

02



Implement continuous exposure monitoring against verified breach data

03



Adopt zero-trust principles for credential validation

04



Assume all credentials in circulation are compromised and act accordingly

The Q2 2026 Quarterly Breach Report will provide comparative metrics and trend analysis. Until then, the Q1 baseline serves as the foundation for measuring quarterly progress and identity risk evolution throughout 2026.



About Constella Intelligence

Constella is a global leader in Identity Risk Intelligence. Powered by the world's largest breach and infostealer data lake, spanning over one trillion attributes across 125+ countries and 50+ languages, Constella empowers organizations to detect, investigate, and respond to threats linked to exposed personal data. Enterprises, managed service providers, and law enforcement agencies worldwide rely on Constella to strengthen identity posture, fuel threat intelligence, and defeat digital risk.



Where Data & Identity Intersect

Powered by

✓ **THE WORLD'S LARGEST VERIFIED
IDENTITY DATA LAKE**