



Insider Threats packing their bags with your data

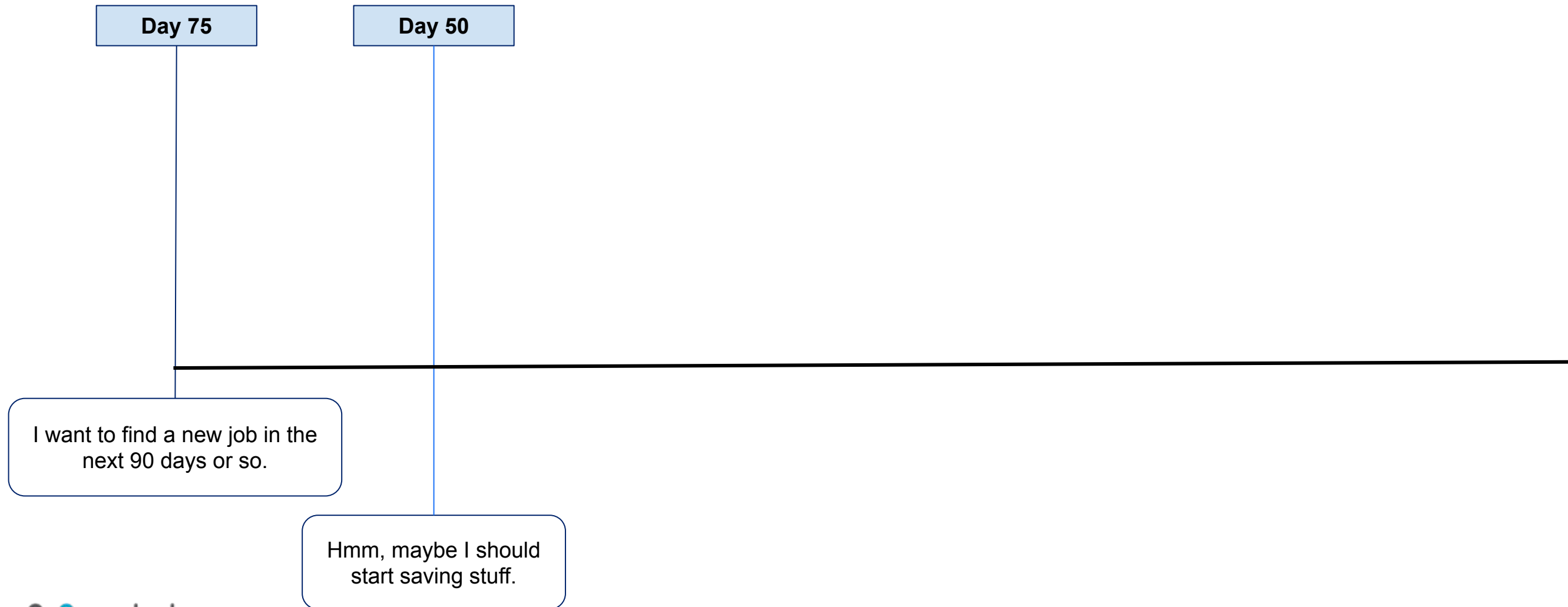
Colin Estep
Netskope Threat Labs

Insider Story

Day 75

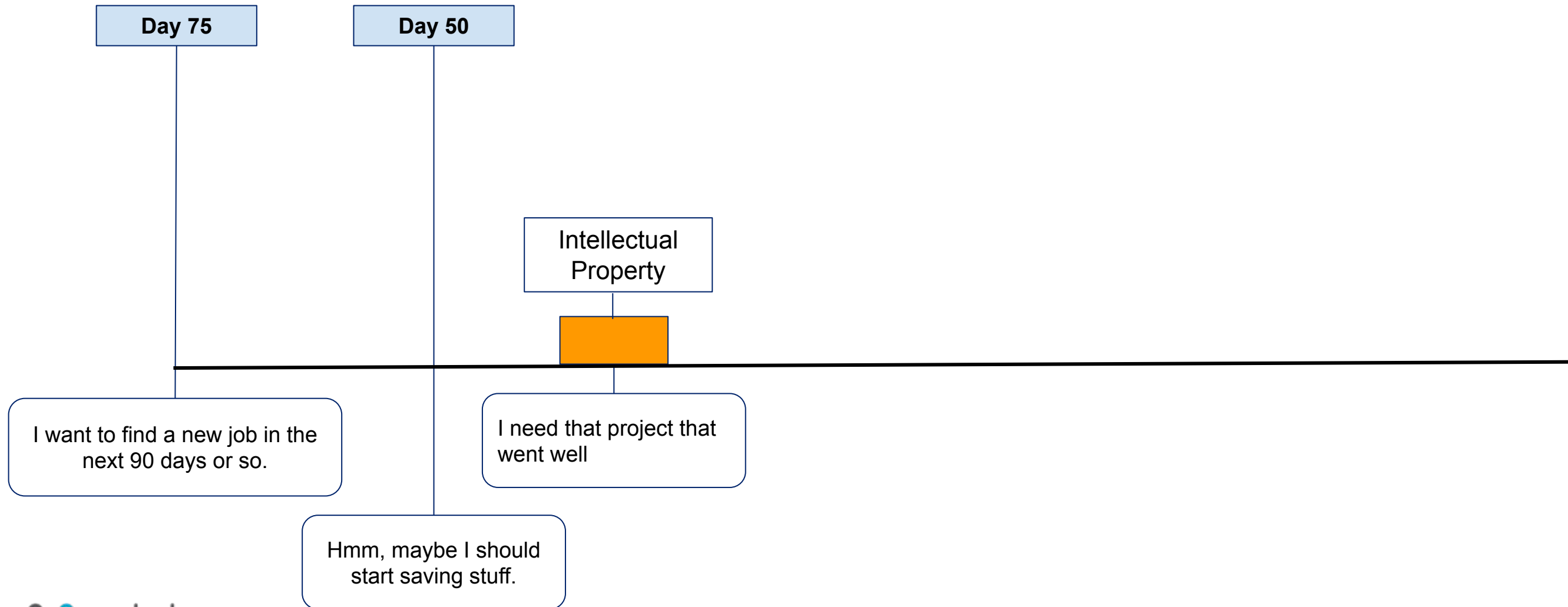
I want to find a new job in the next 90 days or so.

Insider Story



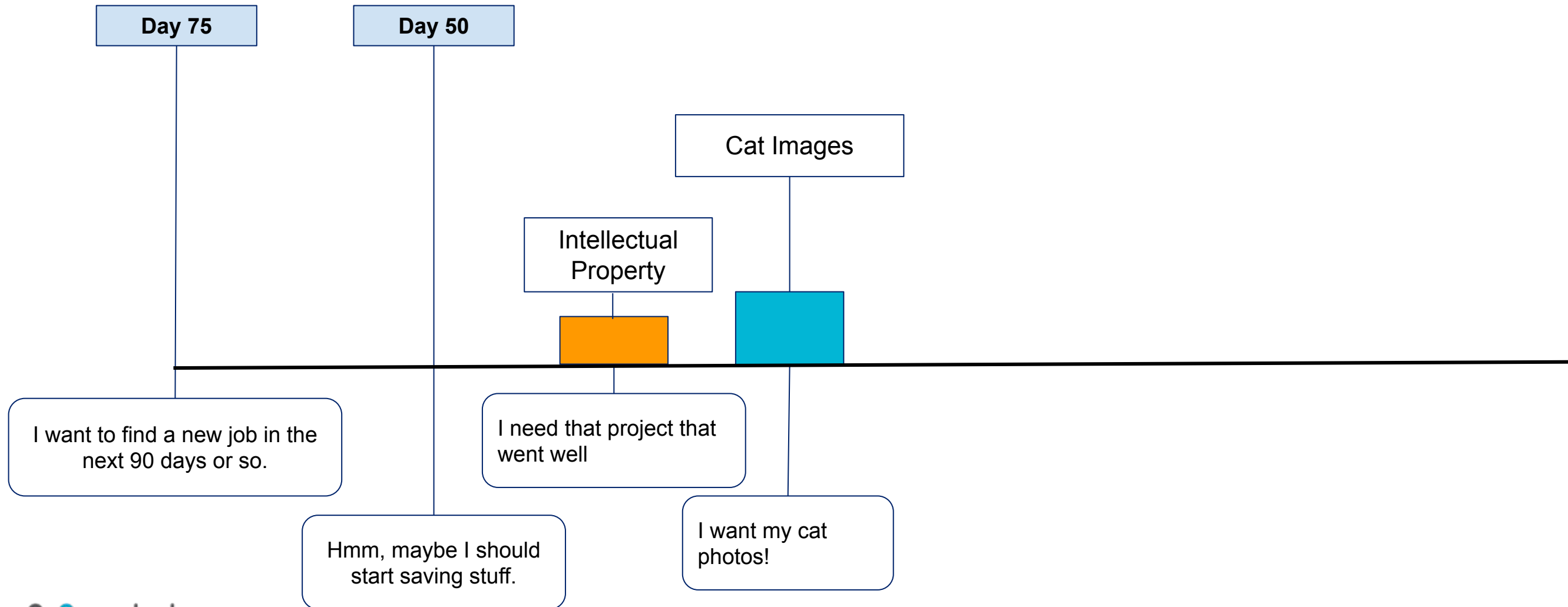
Insider Story

 = Uploads to personal Google Drive

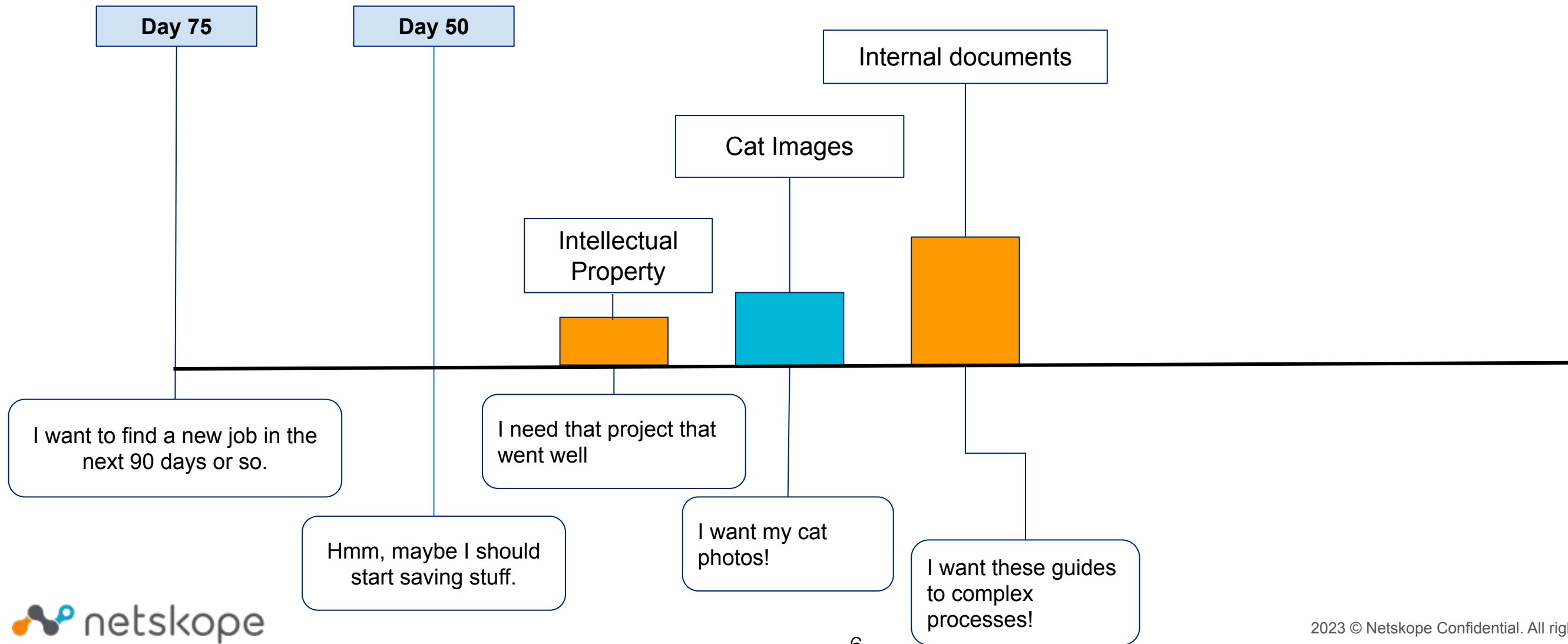
Insider Story

 = Uploads to personal Google Drive

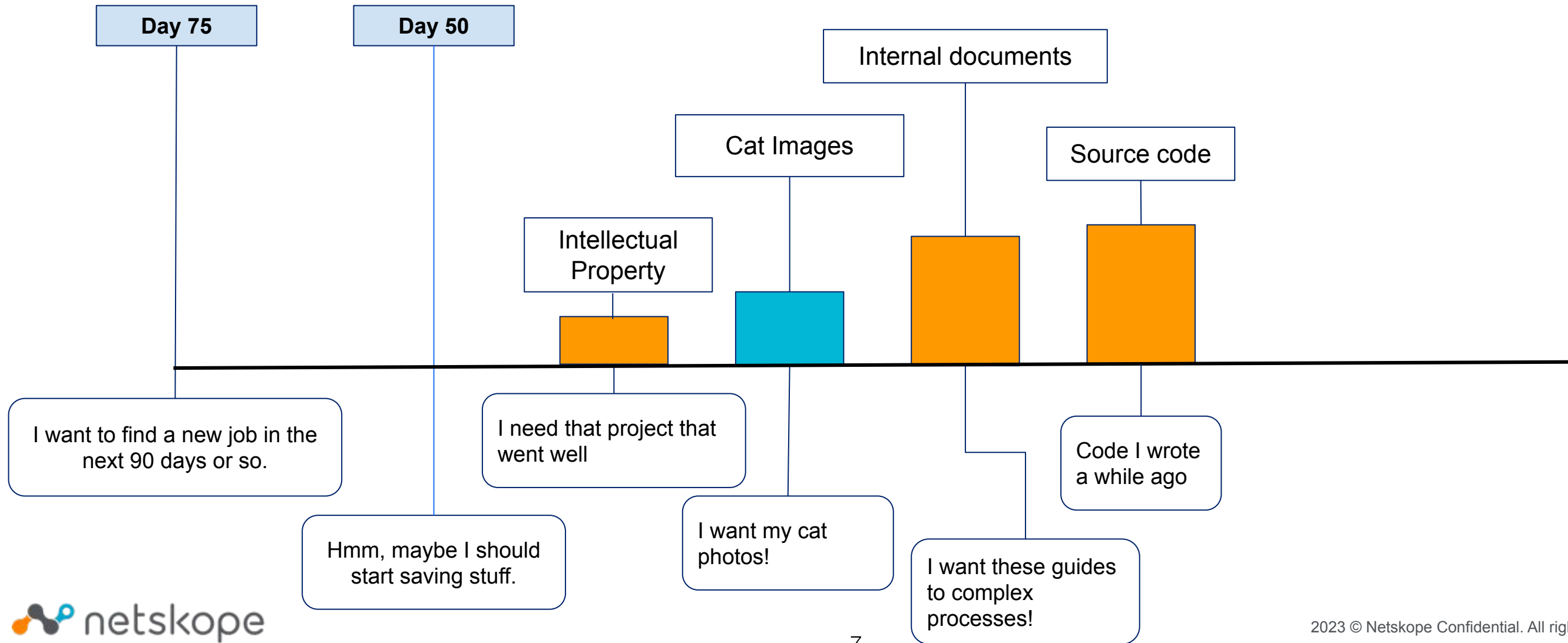
Insider Story

 = Uploads to personal Google Drive

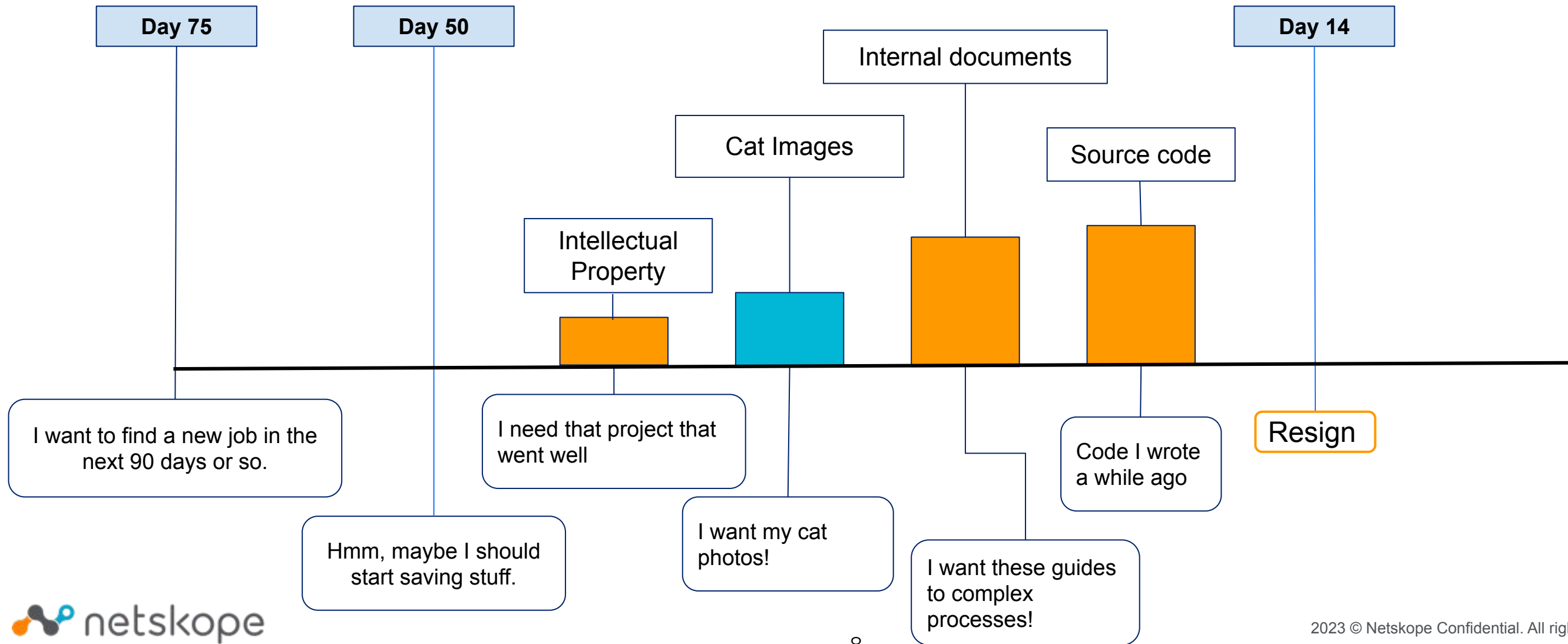
Insider Story

 = Uploads to personal Google Drive





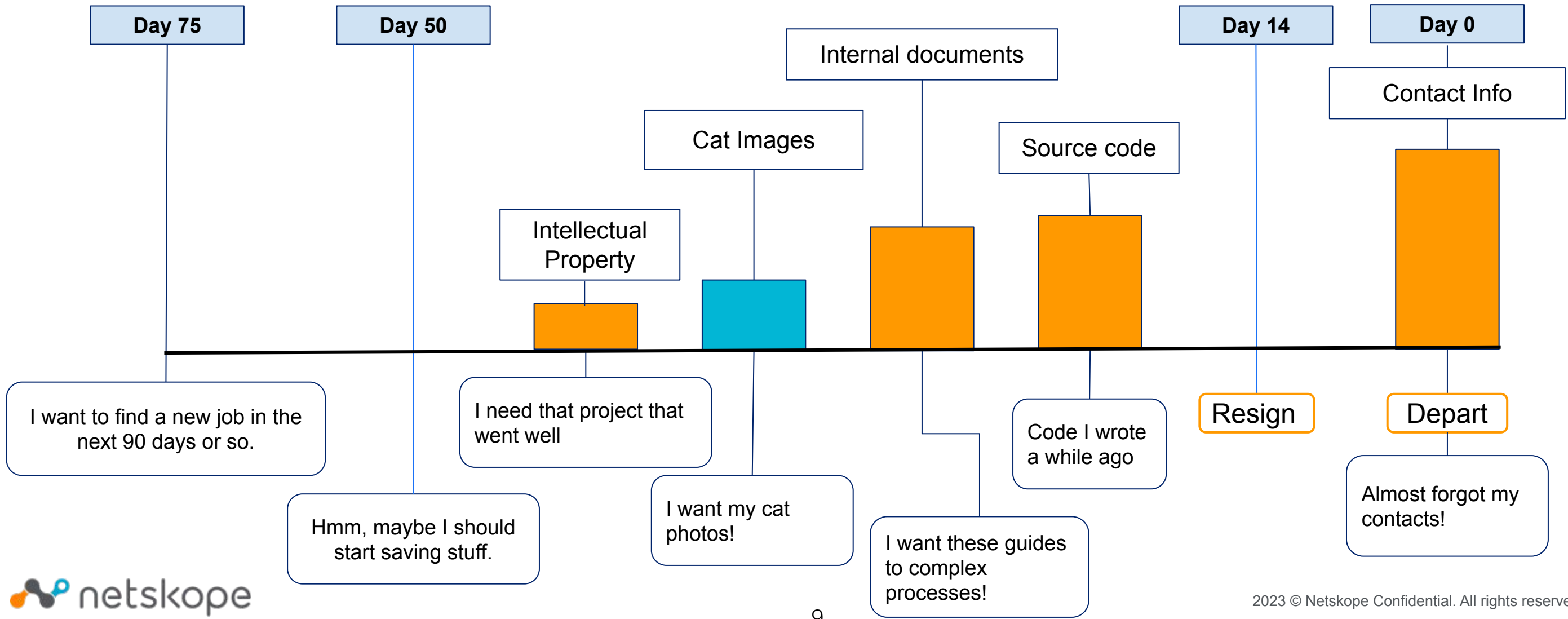
Insider Story

 = Uploads to personal Google Drive

Insider Story

 = Uploads to personal Google Drive

Why listen to me?

Our Data

Timeline:
July 2022 to April 2023



207 organisations



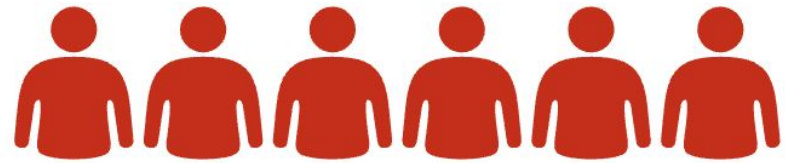
4.7M active users

58,314 individuals left their employment

Information presented in this talk is based on anonymized usage data collected by the Netskope Security Cloud platform relating to a subset of Netskope customers with prior authorization

Our Findings

100%



58,314 users that left

15%



Moved data to personal apps

2%



Mishandled corporate data

Intellectual Property
and PII accounted for
70% of the data taken

Agenda

- The problem
- Overview of our approach
- Results of our study
- Finding exfiltration
- What's next?
- Takeaways



The problem

The problem

An insider who has exfiltrated sensitive corporate data using cloud apps.

Sensitive Data refers to data that could hurt the organization if it is exposed externally

The scope of an insider for this presentation is:

- Not using a USB drive
- Not printing out documents and walking out of the building with them
- Not taking pictures of a monitor with their phones

Why is this important?

Insiders

- A 2020 Securonix Insider Threat Report found that 60% of Insider Threats involve "Flight Risk" employees
- Every organization has "flight risk" employees

Data Exfiltration

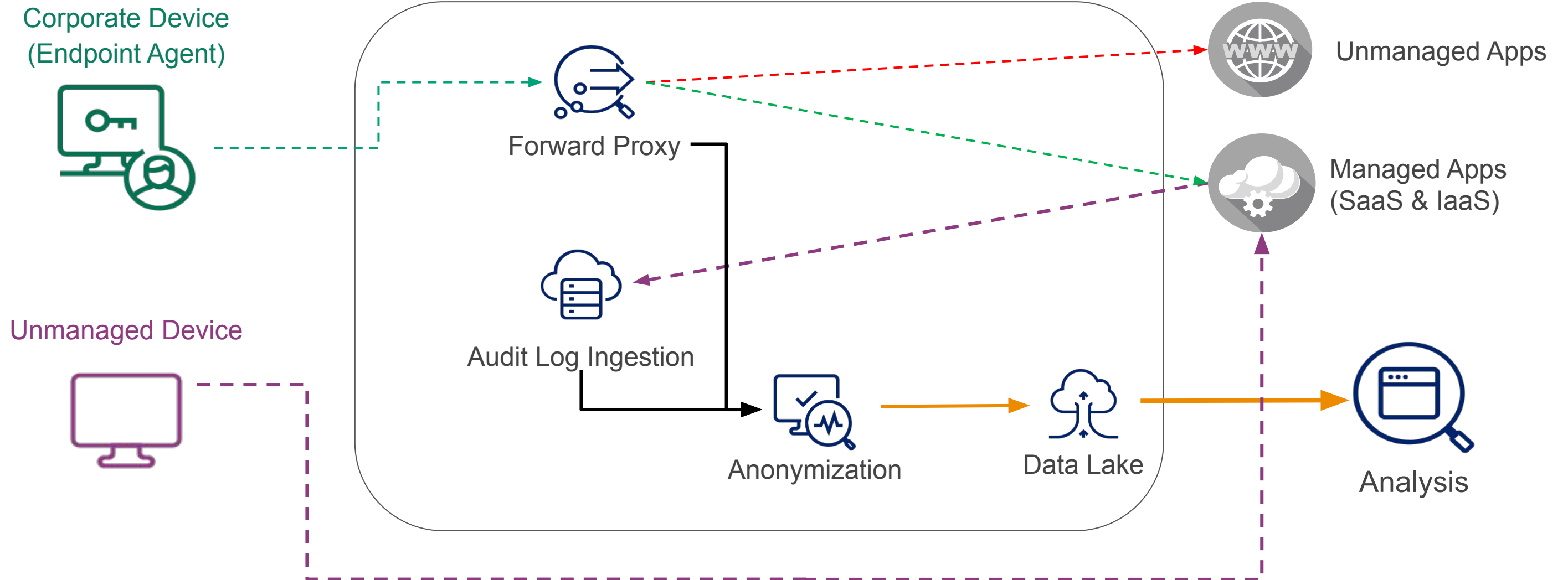
- More organizations than ever have Personally Identifiable Information (PII) and other sensitive data
- Liability around data breaches are typically on the organization itself

Every organization should have a strategy to address this threat

Overview of the approach

Architecture

Monitoring Systems



3 Signals from the Data



Direction: Are users are saving data to their own personal cloud storage?



Nature: What files contain sensitive corporate information?



Volume: Which users are downloading or uploading more than usual?

Direction Signal: Labeling Applications

The domain associated with a cloud application, which indicates who controls that particular application, is an instance. We use some heuristics to label the instances as data comes in for analysis.

Application	Domain	Label	Percentage of Traffic
Google Drive	netskope.com	Business	50%
Google Drive	gmail.com	Personal	15%
Google Drive	foobar.com	Unknown	35%

Nature Signal: Labeling Data

We need a way to label the files that contain an organization's sensitive information.

DLP policies can alert us when something contains the following:

- Intellectual Property
- Data in scope for compliance (PCI-DSS, GDPR, etc.)
- Secrets

The DLP violations provide us a nice signal about the nature of the data involved.

What events look like

User	App	App Instance label	Activity	File Name	DLP Violation
user@gmail.com	Google Drive	personal	upload	black_project.docx	Secret project code names

Results of our study

Results: Departures

Our Data

Timeline:
July 2022 to April 2023



207 organisations

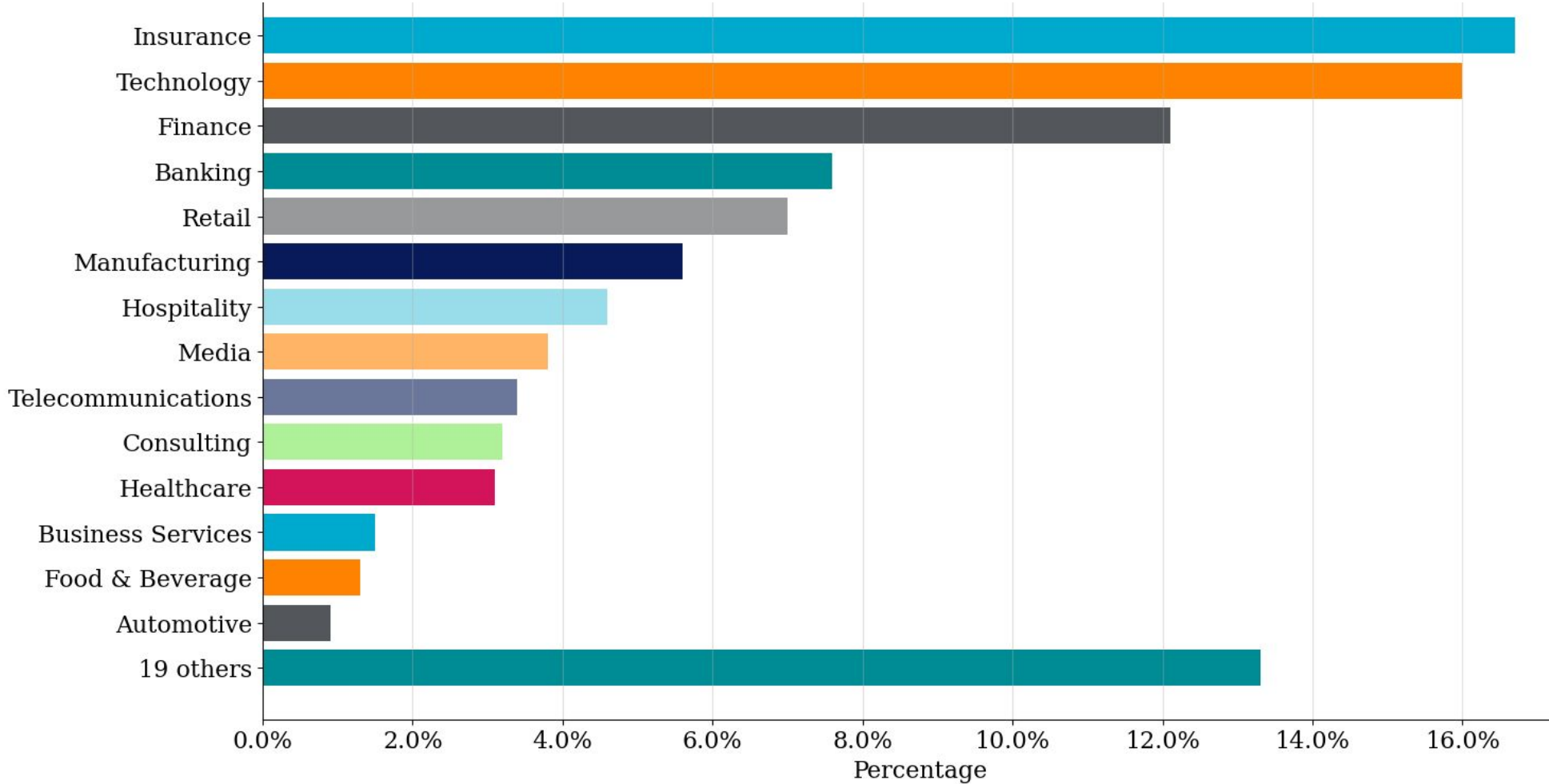


4.7M active users

58,314 individuals left their employment

Information presented in this talk is based on anonymized usage data collected by the Netskope Security Cloud platform relating to a subset of Netskope customers with prior authorization

Industry breakdown for departures



How many people move data to personal apps?

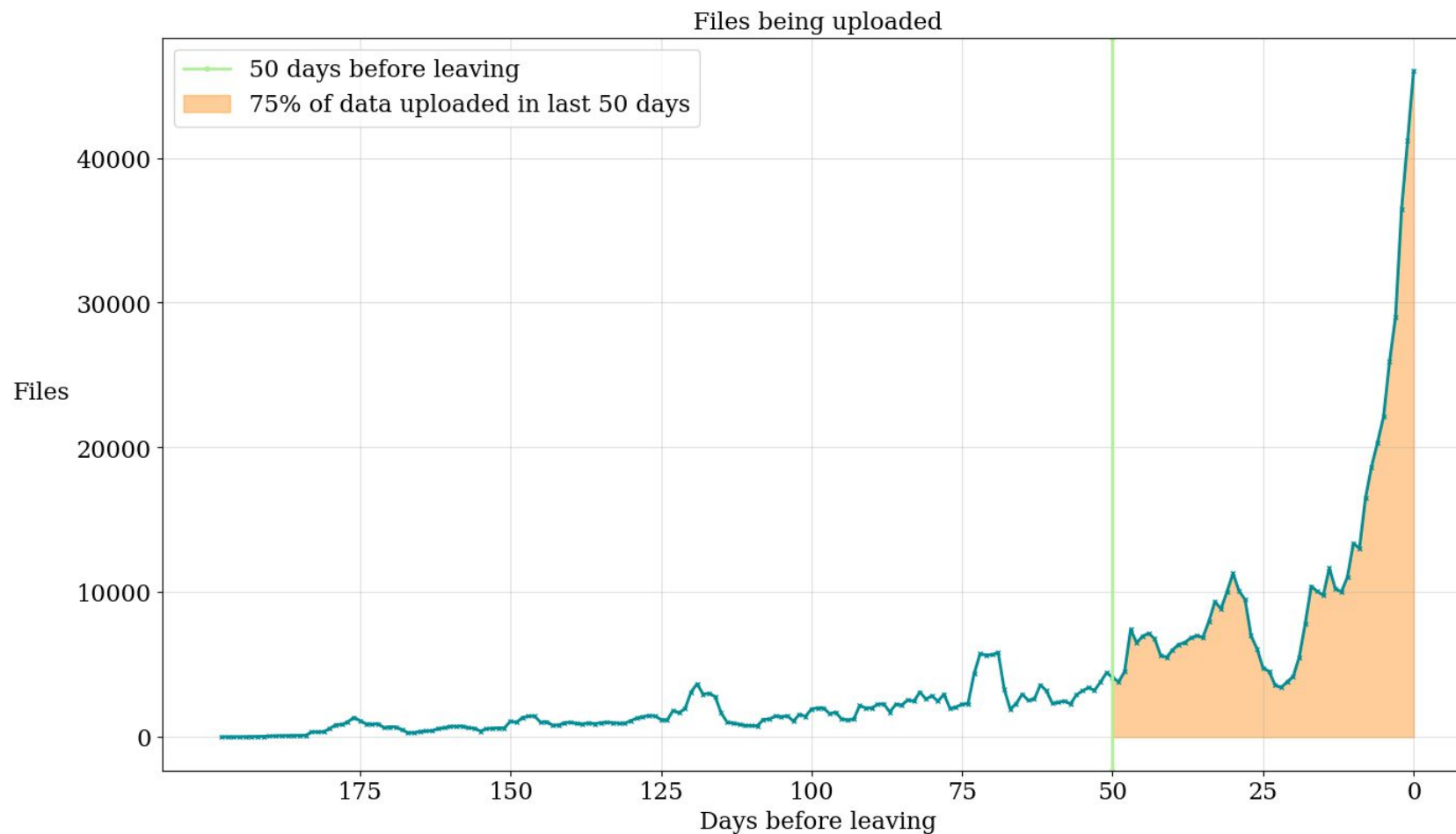
85% of flight risks did not move data

15% of flight risks moved some kind of data

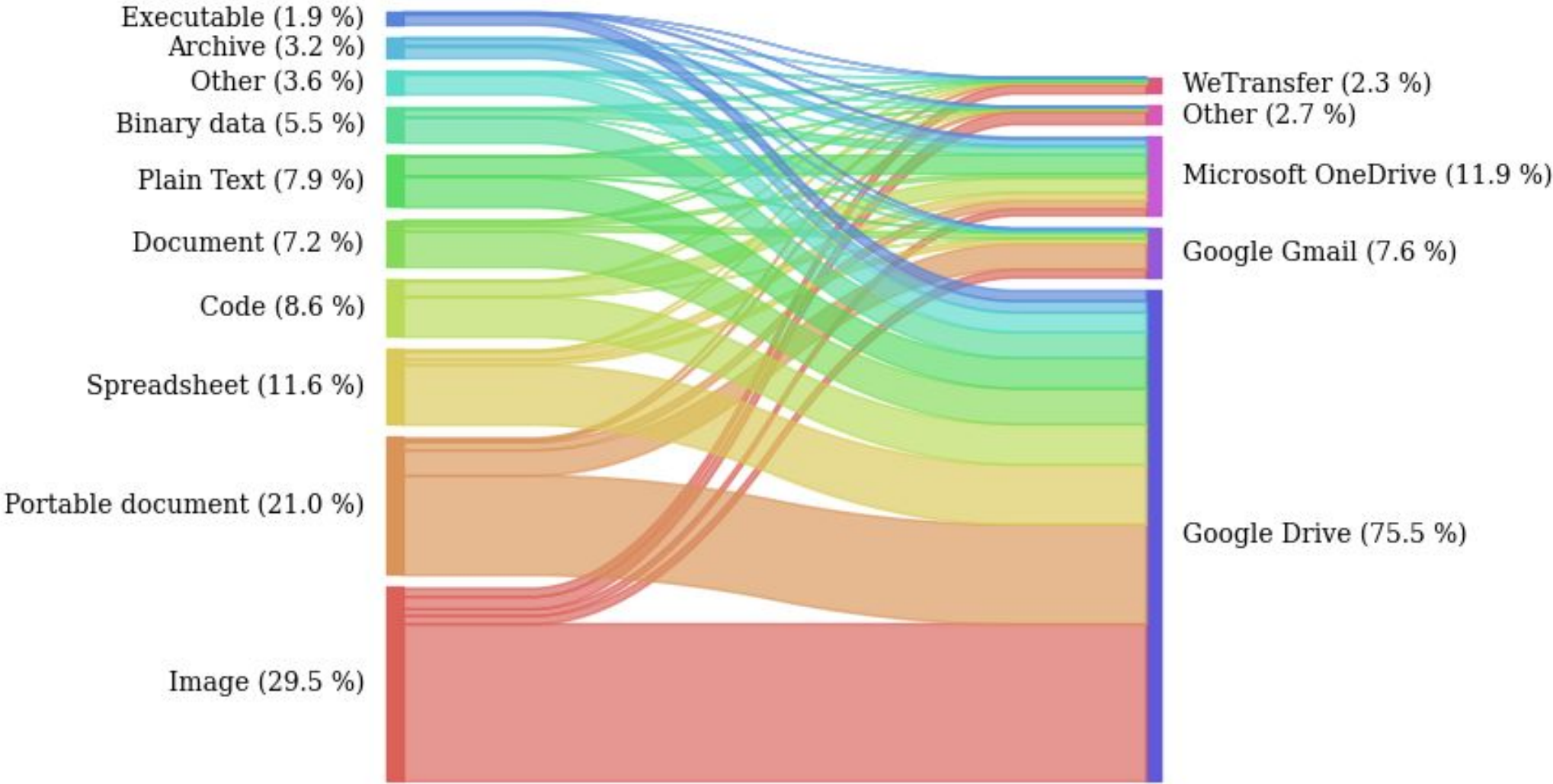
(this includes benign cases)

When is the data moved to personal apps?

75% of all files uploaded to personal apps were uploaded in the last 50 days



What sort of data gets moved?



Files moved in the last 50 days

Results: Data Exfiltration

What kind of data exfiltration?

An insider who has exfiltrated sensitive corporate data using cloud apps.

Sensitive Data refers to data that could hurt the organization if it is exposed externally

Exfiltration by departing employees

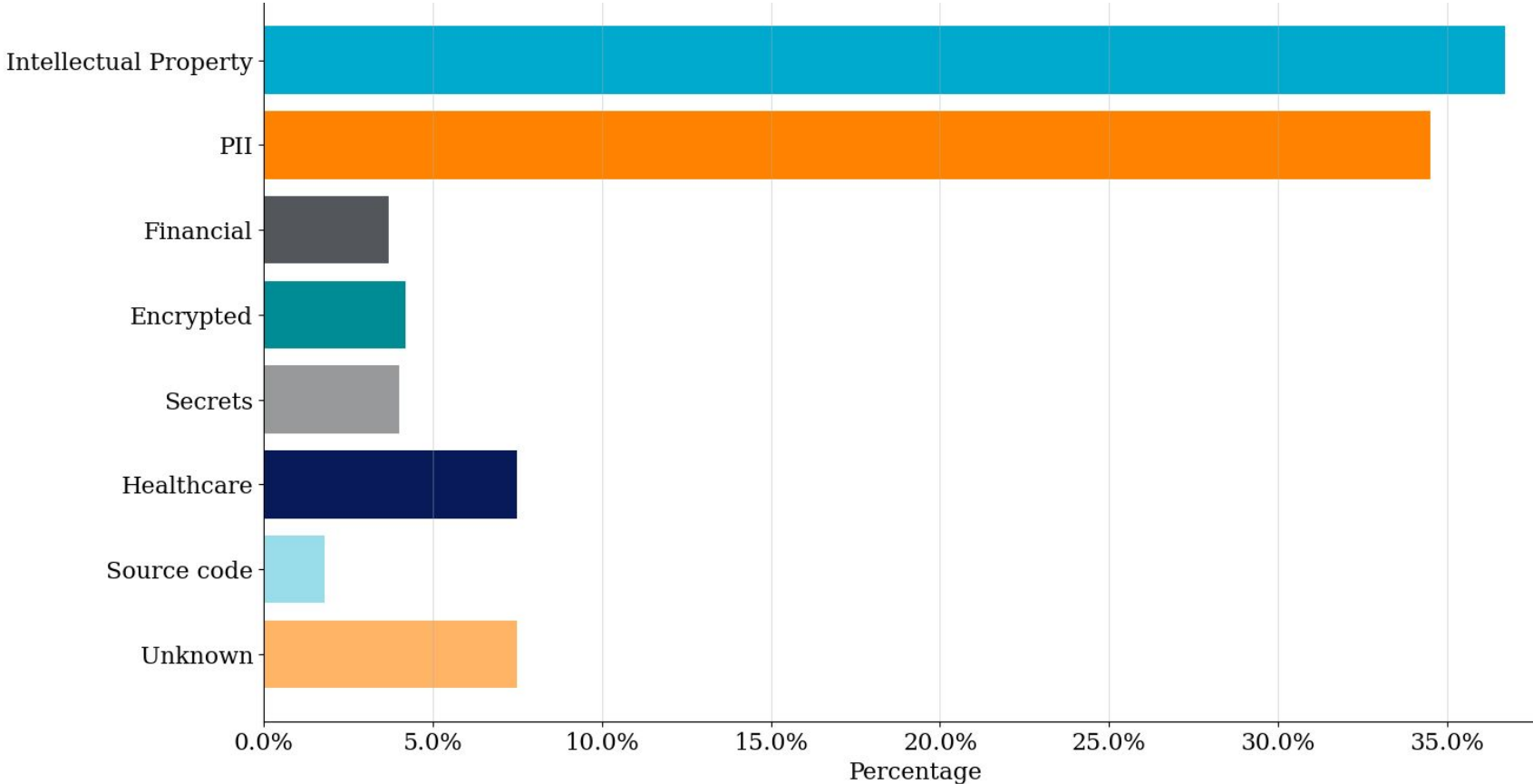
2% exfiltrated corporate data via cloud apps

Percentage of **sensitive files** uploaded:

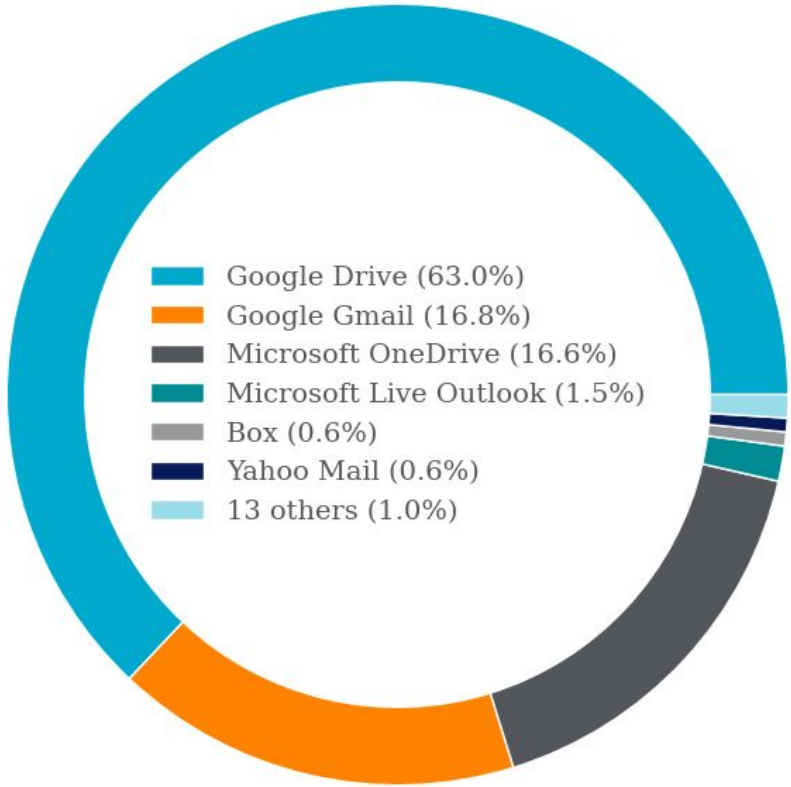
- 94% in the last 91 days
- 84% in the last 49 days
- 74% in the last 28 days

If you monitor the last 30 days of employment, you may get around 75% of the files being mishandled before someone leaves.

Data Targeted



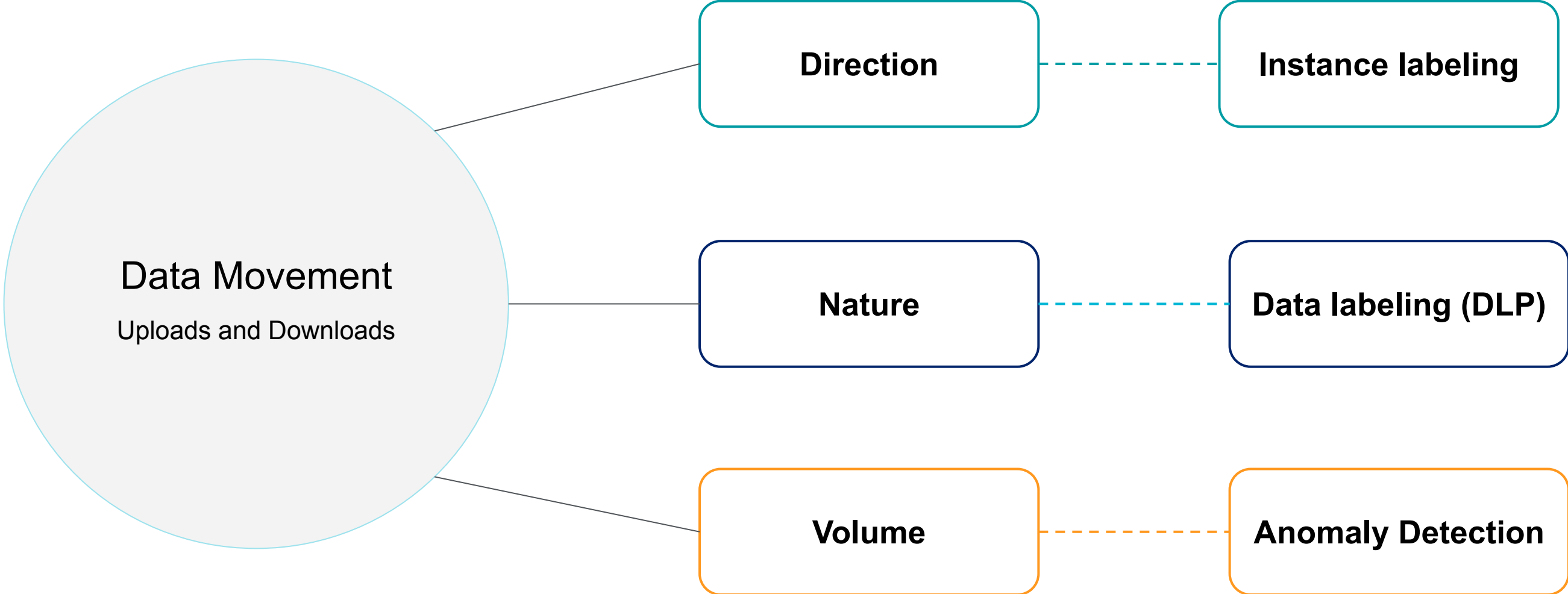
Policies violated



Apps used

Finding Exfiltration

The signals and their components



Detection Categories

	Heuristic	Anomaly Detection	Anomaly Detection + Data Labeling
Volume: Behavior Baseline	X	✓	✓
Direction: Application Labels	X	✓	✓
Nature: Data Labels	X	X	✓
Example	More than 100 files uploaded	More uploads than usual to personal app	A lot of corporate secrets uploaded to personal app

Detection efficacy

What is the relative signal strength of each type of detection to find someone who is going to leave?

Data Movement Detection	Improvement	Example # of alerts
Heuristic	Baseline	215
Anomaly Detection	15.6 x	14
Anomaly Detection + Data Labeling	43.0 x	5

Derived from organizations with 3,000+ daily active users

Finding Exfiltration: Anomaly Detection

Volume Signal: Anomaly Detection



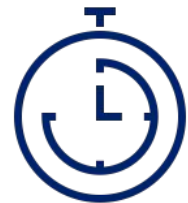
Spikes

User downloaded more files than normal



First Occurrence

Nobody has ever uploaded to this S3 bucket



Rare Occurrence

User logged in to Slack for the first time in 90 days

Model Levels

User Models



Baseline for this **user only**

Lower Severity

Peer Group Models



Baseline for a **group**

Moderate Severity

Organization Models



Baseline for the **organization**

Higher Severity

Building Models

Build Models

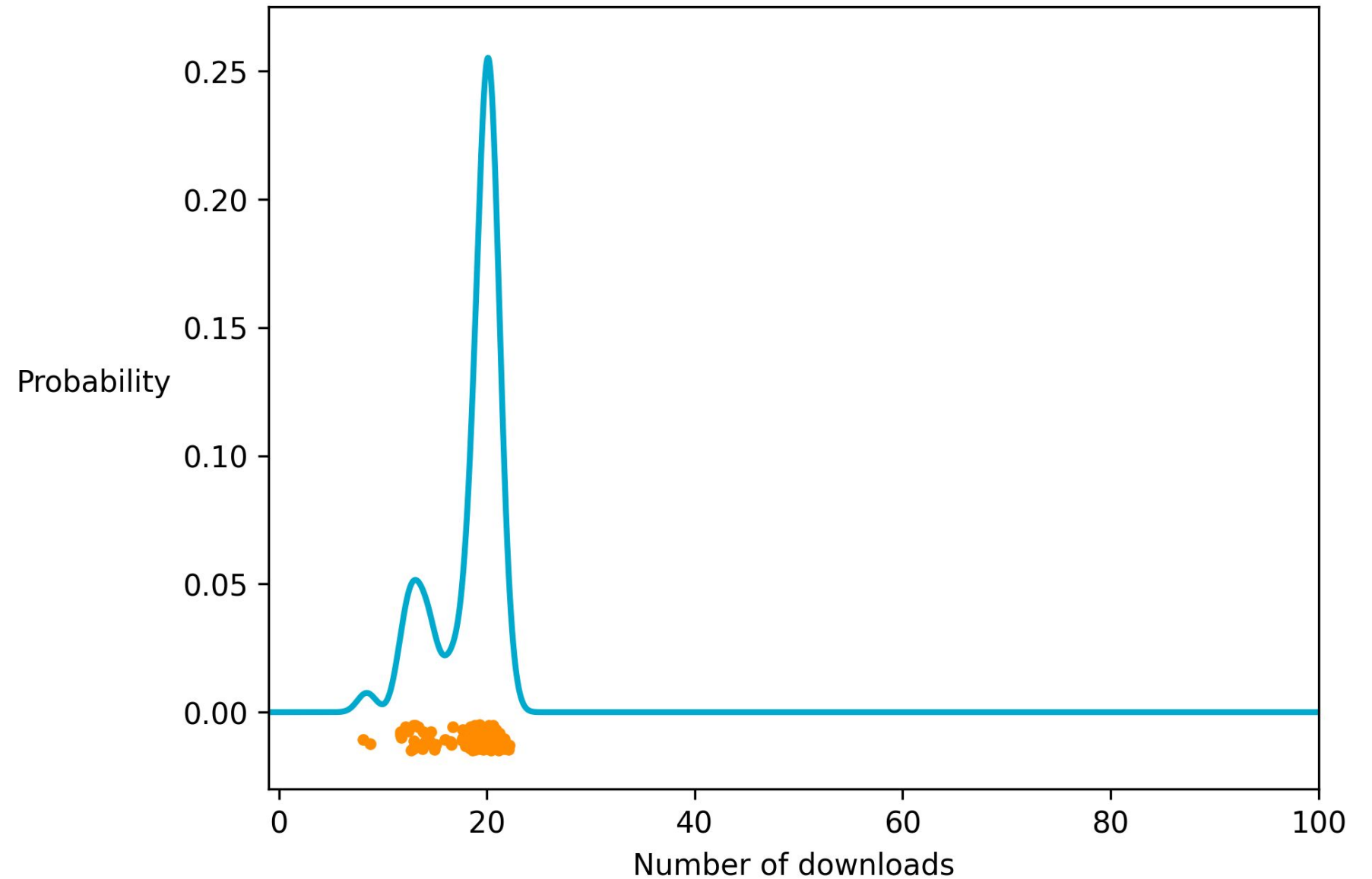
Select features

(files downloaded per day)

Fit data into a distribution

Track over time

(6 months)



Building Models

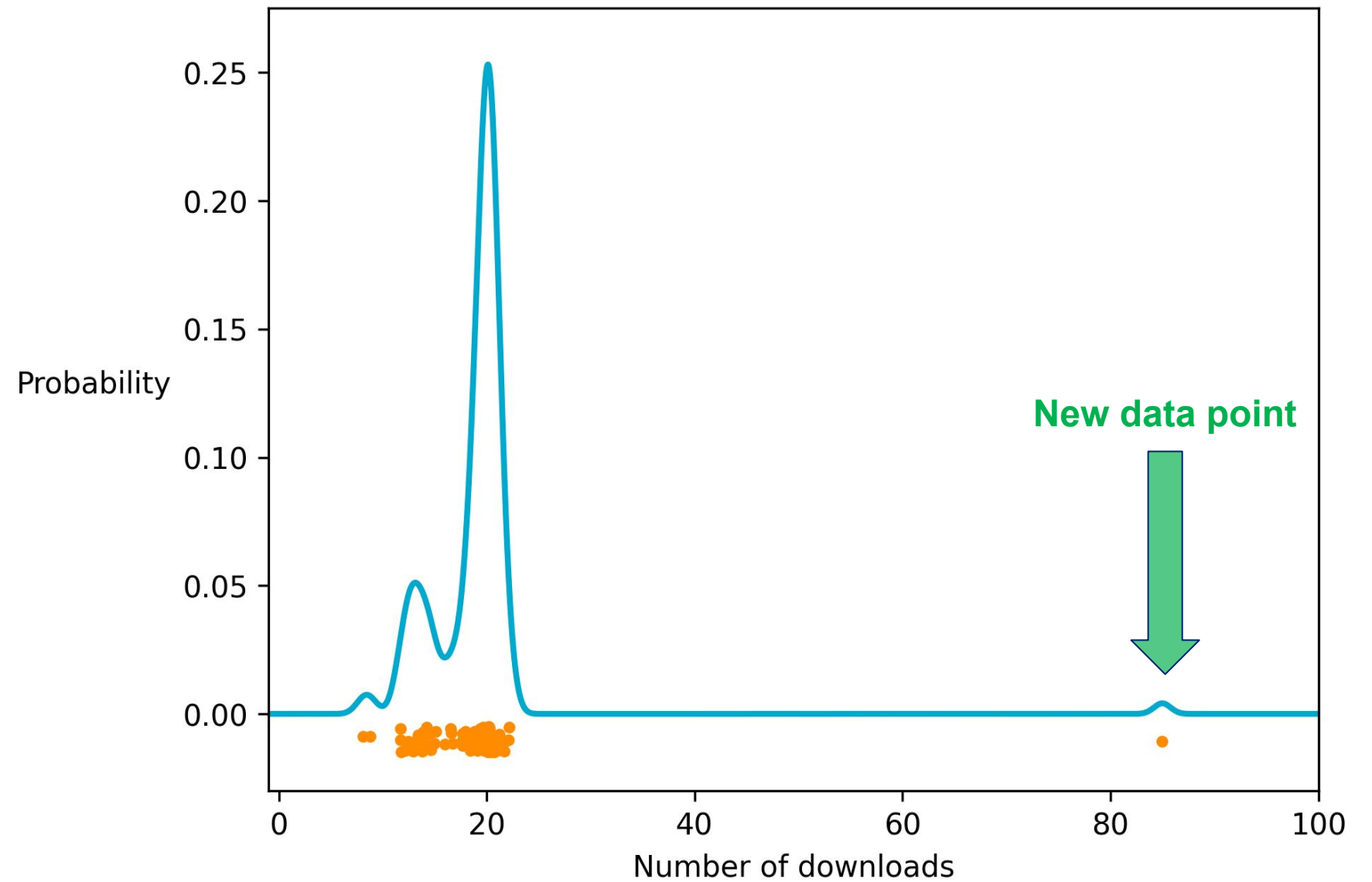
Analyze Relevant Events

Filter events

(feature and entity)

Save it to the model

Find the probability



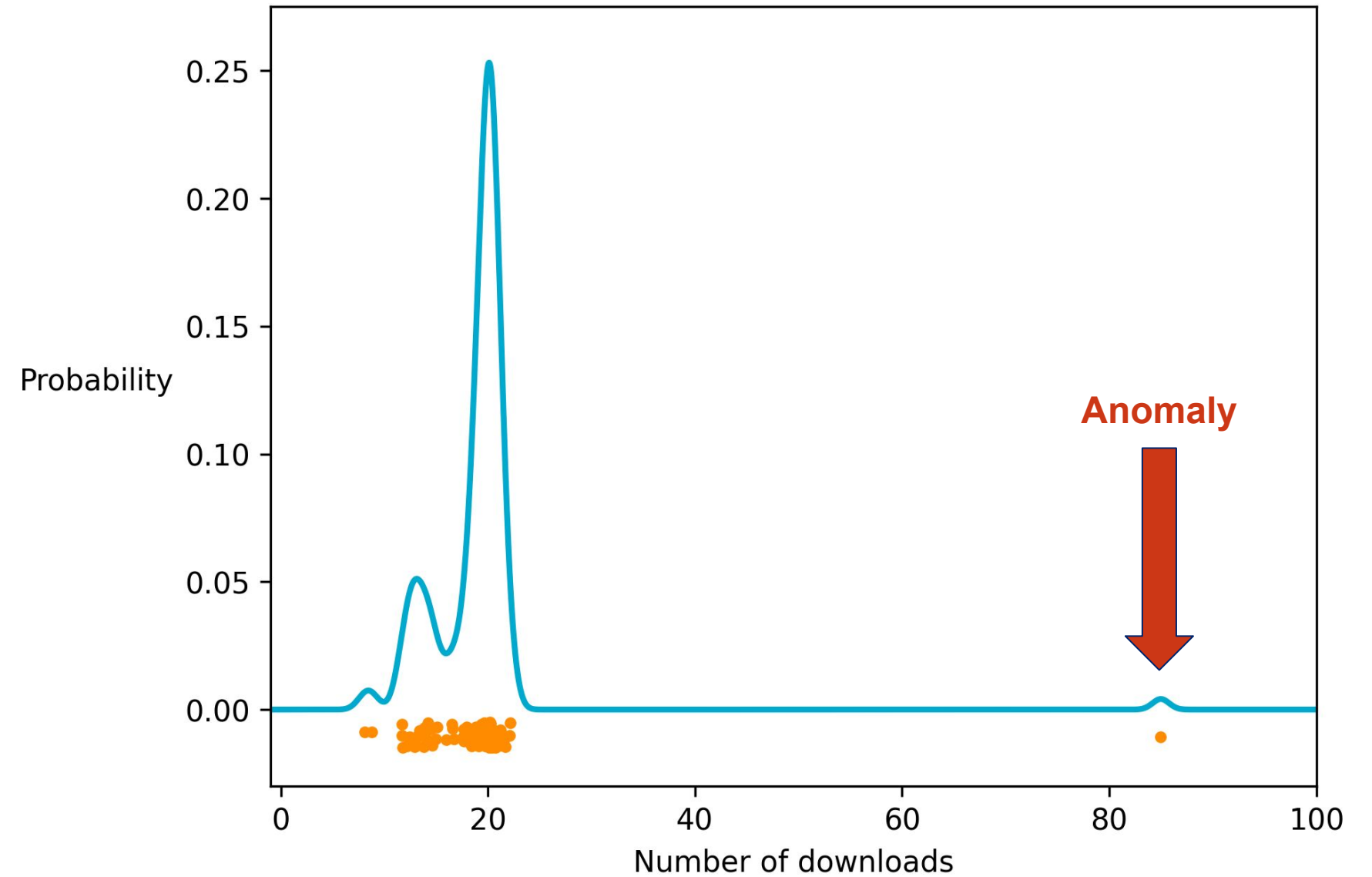
Triggering Anomalies

Trigger Anomalies

Mature models only
(minimum data requirement)

All conditions are met

Spikes / First / Rare



Anomaly Examples

Behavior	Signals	Anomaly
Download Spike	Volume + Direction	A user-based spike in data/files downloaded
Upload Spike	Volume + Direction	A user-based spike in data/files uploaded to personal apps
DLP + Download Spike	Volume + Direction + Nature	A user-based spike in sensitive data/files downloaded
DLP + Upload Spike	Volume + Direction + Nature	A user-based spike in sensitive data/files uploaded to personal apps
DLP + Download Spike + Upload Spike	Correlated Volume + Direction + Nature	Potential sensitive corporate data movement

Deployment

- We've deployed **70 different models** to our production environment
- The models have learned from hundreds of organizations
- Some models have been running for 2 years

Investigation Steps

1. Triage correlated data movement anomalies to find the user
(single digit volumes)
2. Examine the user's DLP violations
3. Examine the files being moved more closely

This is a very manageable process

Finding Exfiltration: Case Studies

Case Study #1: Employee Departure

Confirmed Insider

Confirmation	Behavior	Signals
✓	Spike of 2,700 files uploaded to personal Google Drive	Volume + Direction
✓	Spike of 1,500 DLP violations	Nature
✓	First authentication to personal Google Drive	First Occurrence

12 days prior to departure

Case Study #2: No Departure

Confirmed Insider

Confirmation	Behavior	Signals
✓	Spike of 1,900 files uploaded to personal Google Drive	Volume + Direction
✓	Spike of 100+ DLP violations for Patents, PII, and more	Nature
✓	Files contained legal and financial information	Nature

Case Study #3: Benign Activity

Benign

Confirmation	Behavior	Signals
✓	Spike of 500 files uploaded to personal Google Drive	Volume + Direction
✓	Spike of 300+ DLP violations for PII	Nature
✓	User was uploading their own tax records, bank statements, and images	Nature



What's next?

Current Limitations

- Only analyzed data movement via cloud applications
- Scope was insiders that end up leaving the organization, but there are ones that do not
- Unknown traffic (neither personal or business) was primarily excluded from our analysis

Future Development

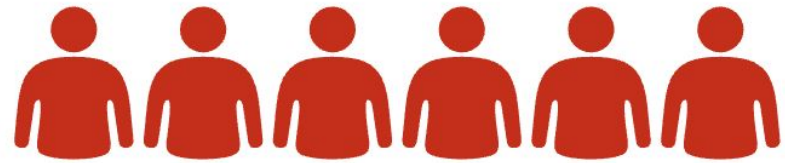
- Develop “Flight Risk” signals without data exfiltration
 - Job hunting activities?
 - Reduction in business related activity?
- Incorporate additional metadata about the files that are downloaded
 - Owner
 - Location (folder structure, shared drives, etc.)



Takeaways

Our Findings

100%



58,314 users that left

15%



Moved data to personal apps

2%



Mishandled corporate data

Intellectual Property
and PII accounted for
70% of the data taken

Takeaways

- Monitor 50 days of activity if you can
- 3 signals are critical for data movement:
 - Direction
 - Nature
 - Volume
- Investigating alerts that combine the 3 signals is very manageable



Check out our [blog](#):

