Infrastructure and Scaling Arkime

slido

Arkime Experience

ⓘ Start presenting to display the poll results on this slide.

Arkime Components

Capture - monitors network traffic, creates PCAP on disk or S3, generates meta data saved in OpenSearch/Elasticsearch

Viewer - node.js application that serves the UI and provides an API

Cont3xt - contextual intelligence gathering tool for support of technical investigations

Parliament - Tool for management multiple Arkime clusters

WISE - Intelligence feeds aggregator and enrichment interface for capture

OpenSearch/Elasticsearch - Database & Magic

slido

Which extra pieces are you using?

ⓘ Start presenting to display the poll results on this slide.

Architecture

arkime.com/architecture

On prem or cloud?

Size?

Multiple clusters or one large cluster?

Network Packet Broker

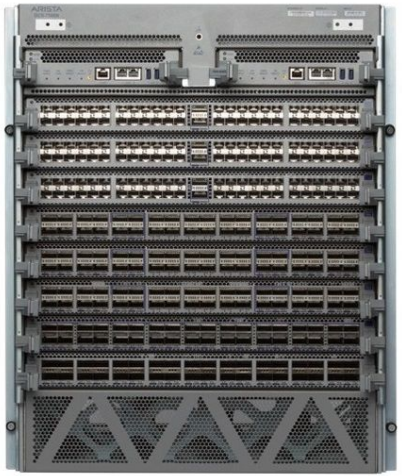
Load Balancing - Distribute the flows across security tools hosts evenly

Scaling - Network and security tools can scale differently

Aggregation - Security tools like getting the whole flow

Seperation of Duty - Network owns inputs, Security owns output

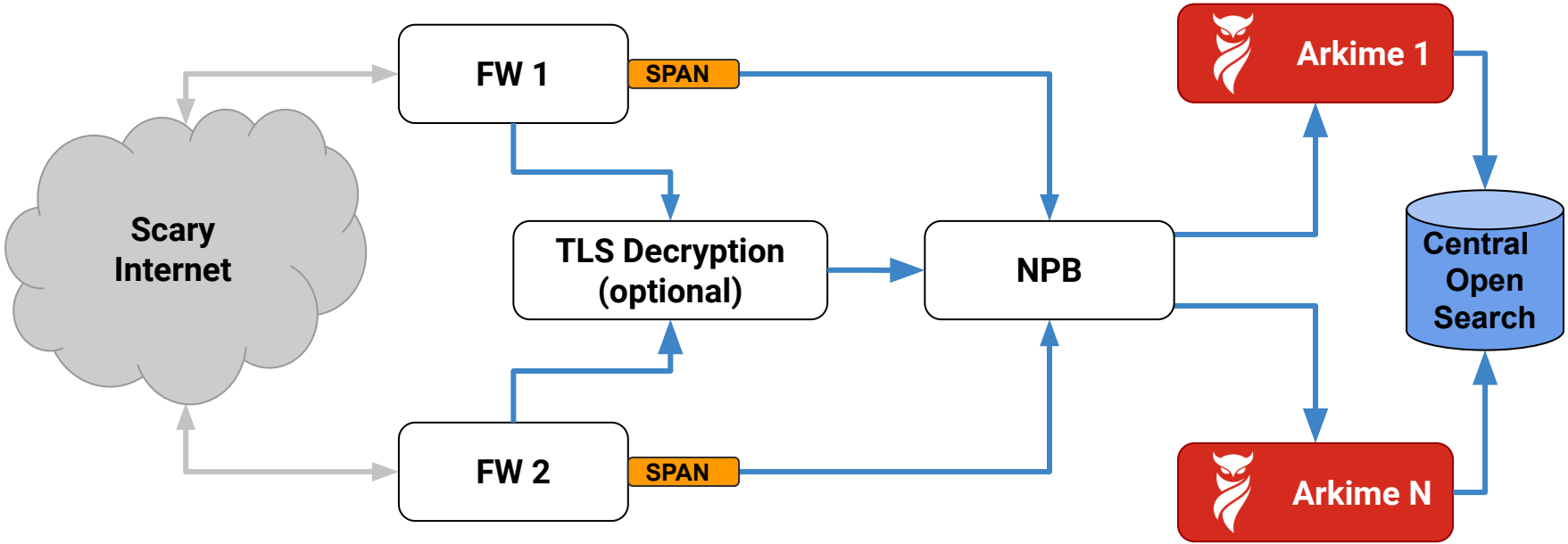
Filtering - Can reduce traffic Arkime needs to look at



Which tap to use

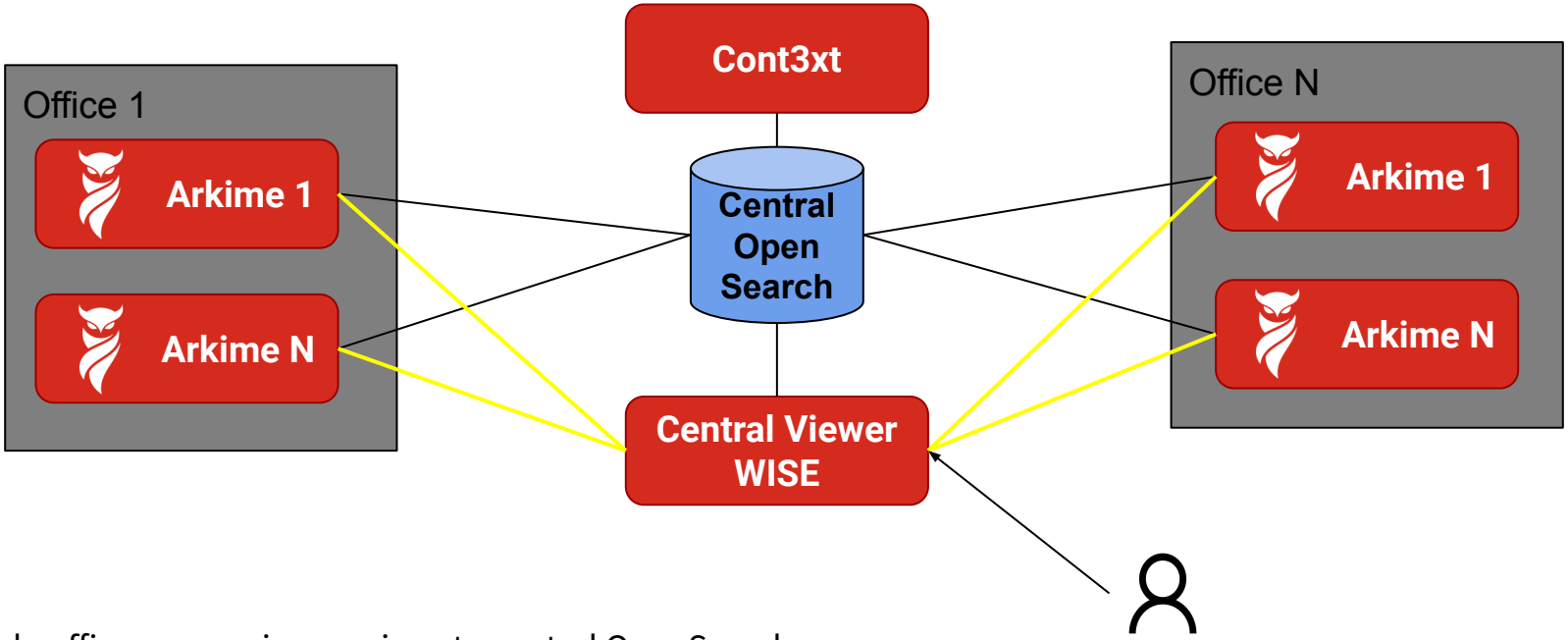
SPAN Tap	Optical Tap
Active - Based on switch config the switch makes a copy of packets to send out special SPAN port to NPB	**Passive** - A prism “steals” some of the light of each fiber link to send to the NPB
+ Only need to tap each switch	- Every link needs to be tapped x2
- Busy switch may drop packets or overload NPB connection	+ Capture everything, can’t overload NPB connection
- Depends on config/humans	+ It is either there or not
+ Cheaper	- More expensive
+ Link is full light power	- Link only gets partial light (20%)

Centralized OS Deployment (1)



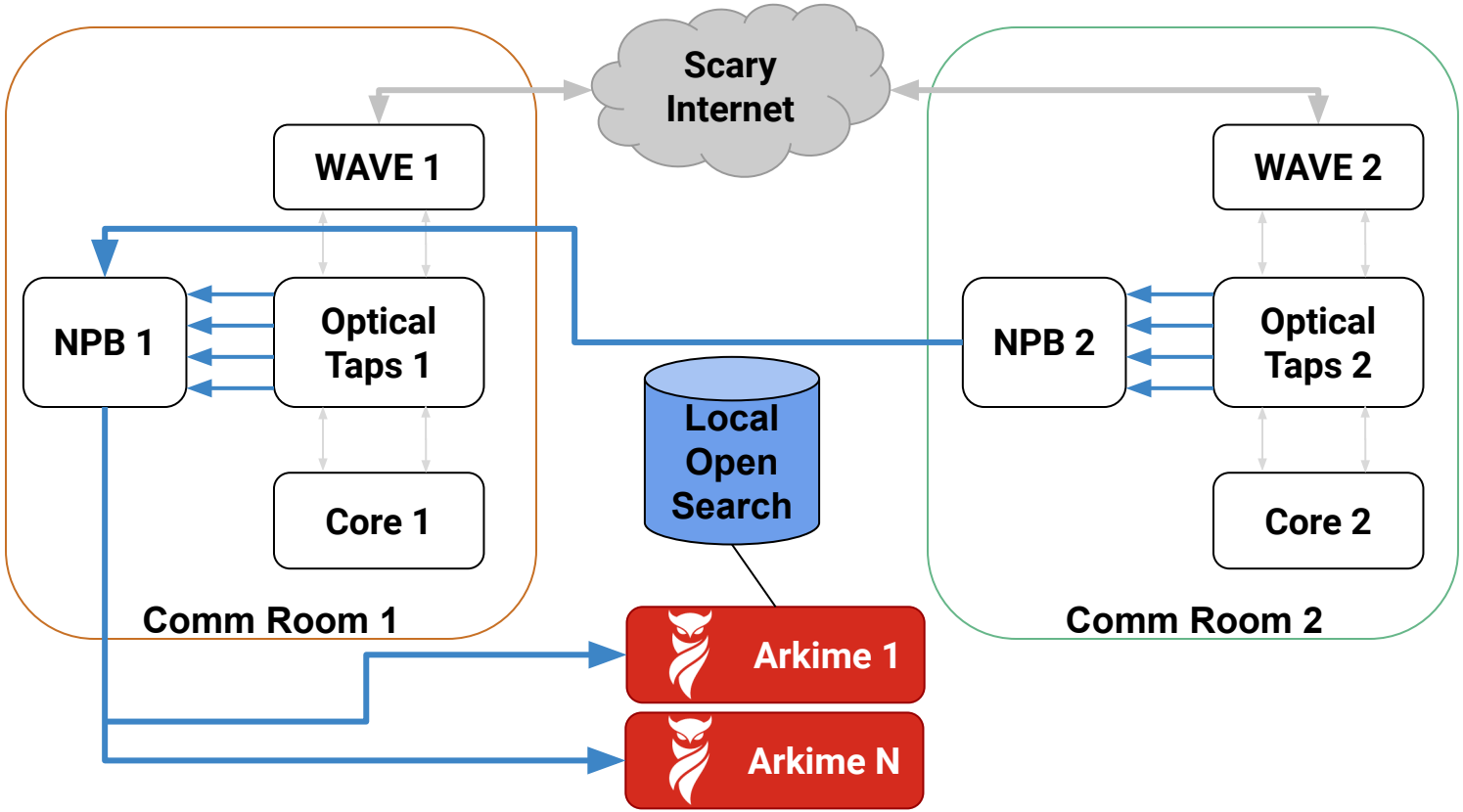
1) Packets flow eventually to a main FW from users' devices
2) SPAN port duplicates packets to NPB
3) NPB load balances packets to Arkime hosts
4) Metadata sent to centralize OpenSearch/Elasticsearch cluster

Centralized OS Deployment (2)

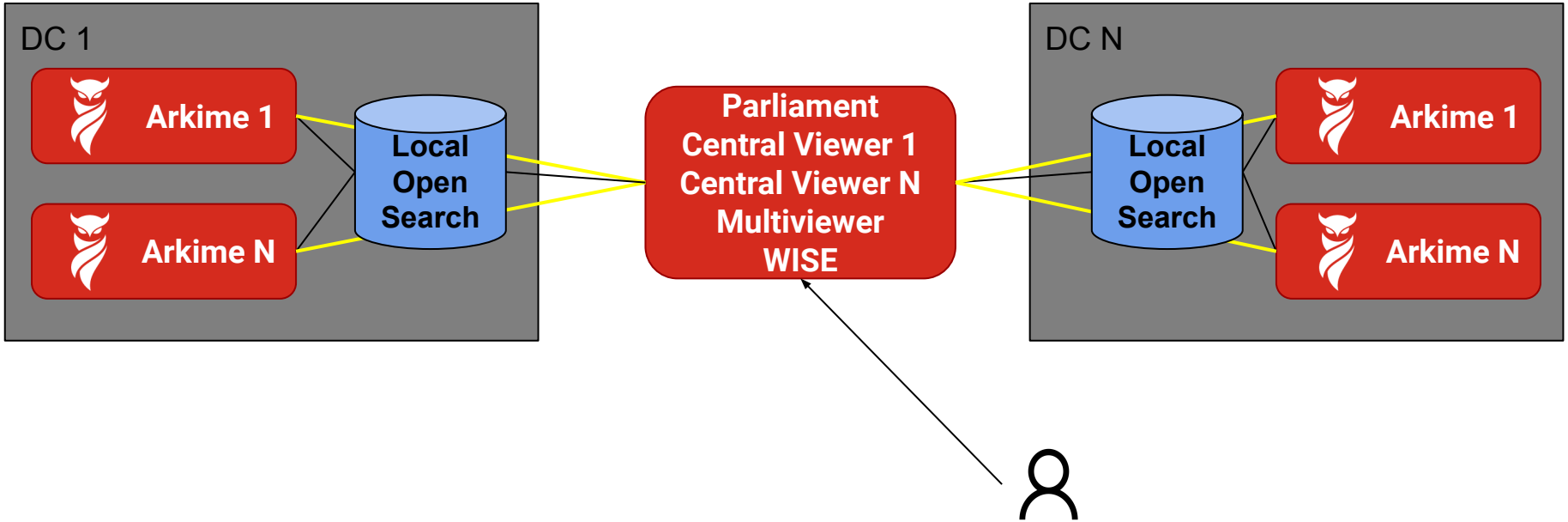


Multiple offices are saving sessions to central OpenSearch
User interacts with a central viewer

Distributed OS Deployment (1)



Distributed OS Deployment (2)



OpenSearch localized with Arkime
Central Viewer & Multiviewer for users

Estimators

arkime.com/estimators

Provides a starting point for sizing

Capture

Average gigabits per second	9

PCAP Days	14	Disk Size	12 TB	Disks per machine	20	TLS %	25%	Compression %	20%	Max per machine	4 Gbps

Space Required	All disks for data RAID 0	One disk extra RAID 5	Two disks extra RAID 6 or RAID 5 + Hot Spare
817 TB	4 hosts / 864 TB	4 hosts / 821 TB	5 hosts / 972 TB

OpenSearch

Average gigabits per second	9

Retention Days	30	Disk Size	12 TB	Disks per machine	4	Nodes per machine	1	Replication	0 Replicas

	Total Space Required	All disks for data RAID 0	One disk extra RAID 5	Two disks extra RAID 6 or RAID 5 + Hot Spare
Average traffic mix	73 TB	2 hosts	3 hosts	4 hosts
High DNS/HTTP traffic	103 TB	3 hosts	4 hosts	5 hosts
Pathological traffic mix	190 TB	5 hosts	6 hosts	9 hosts

Authentication & Authorization

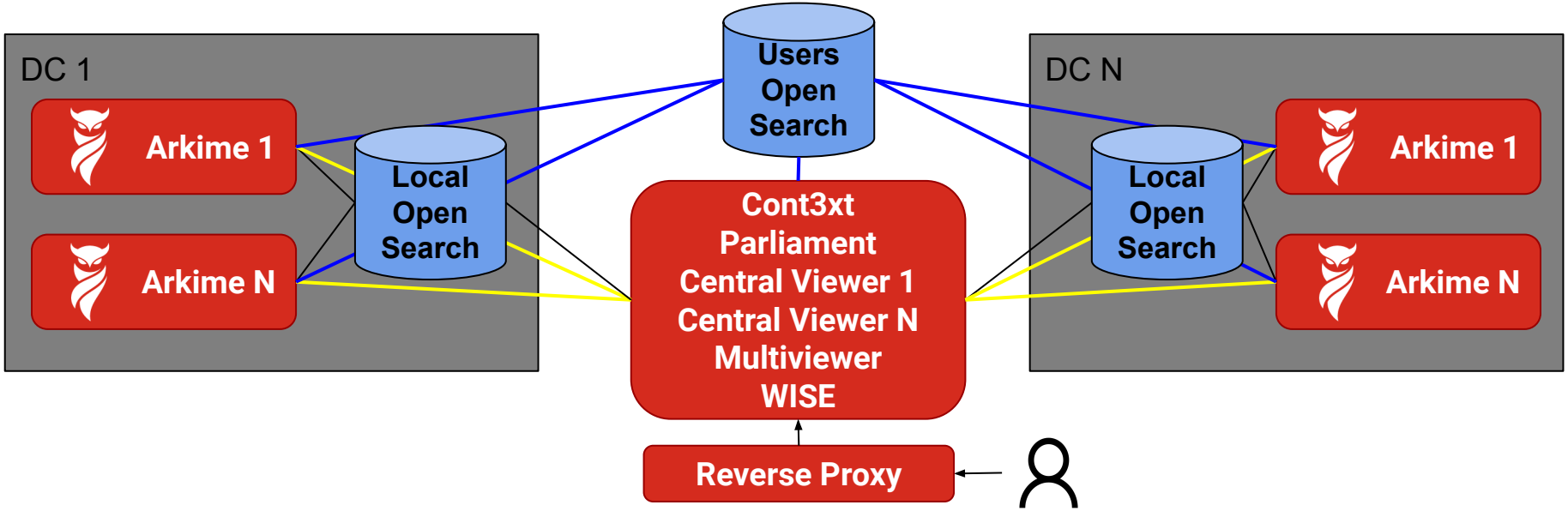
arkime.com/roles
arkime.com/settings#security

Authentication - authn - none, digest, header, oidc

Authorization - authz - builtin only, but can initialize from outside data

Users OpenSearch - Store user data central location

Distributed Deployment (3)



ALL viewers will connect to same users OpenSearch
usersElasticsearch=https://users-opensearch.example.com:9200
Supports syncing of settings/views/shortcuts across clusters

Important Auth Settings (1)

passwordSecret - The shared key used to encrypt the **md5 hashed** password and **cont3xt settings** before storing in OpenSearch.

httpRealm - The auth realm used for digest and **md5 hashed** password

```
iv = randomBytes(16)
ha1 = md5(`${userId}:${httpRealm}:${userPassword}`)
store = "$iv." + aes256(passwordSecret, ha1, iv)
```

Important Auth Settings (2)

serverSecret - The shared key used to encrypt data sent between viewers

userAuthIps - A comma separated list of CIDRs users are allowed to authenticate from. In header mode defaults to localhost, since a header is spoofable, other modes wide open

userNameHeader - Specifies both the authentication mode and what header to use :(

- digest, oidc, anonymous, s2s are accepted modes
- everything else is the header for reverse proxy

userAutoCreateTmpl

Danger!!! Use to auto create users, has access to http headers.

```
userAutoCreateTmpl={
  "userId": "\${this['x-forwarded-email']}",
  "userName": "\${this['x-forwarded-name']}",
  "enabled": true, "webEnabled": true,
  "headerAuthEnabled": true, "emailSearch": true,
  "createEnabled": false, "removeEnabled": false,
  "packetSearch": true,
  "roles": ["cont3xtUser", "arkimeUser"]
}
```

Capture Tuning

arkime.com/settings#capture
arkime.com/settings#reader-afpacket

Doing more with less

Performance Settings

magicMode=basic - libmagic is slow, use the smaller built in one

pcapReadMethod=tpacketv3 - AFPacket recommended for packet acquisition

tpacketv3BlockSize - Buffer size to acquire packets in

tpacketv3NumThreads - Threads to use to acquire packets

packetThreads - Threads that process the packets after acquisition

Space Saving Settings

rulesFiles - Rules that can be used to reduce traffic

Gallery at arkime.com/rules

Please contribute

enablePacketDedup=true - Drop duplicate packets before processing/saving

simpleCompression=zstd - Compress pcap when writing to disk

Rules

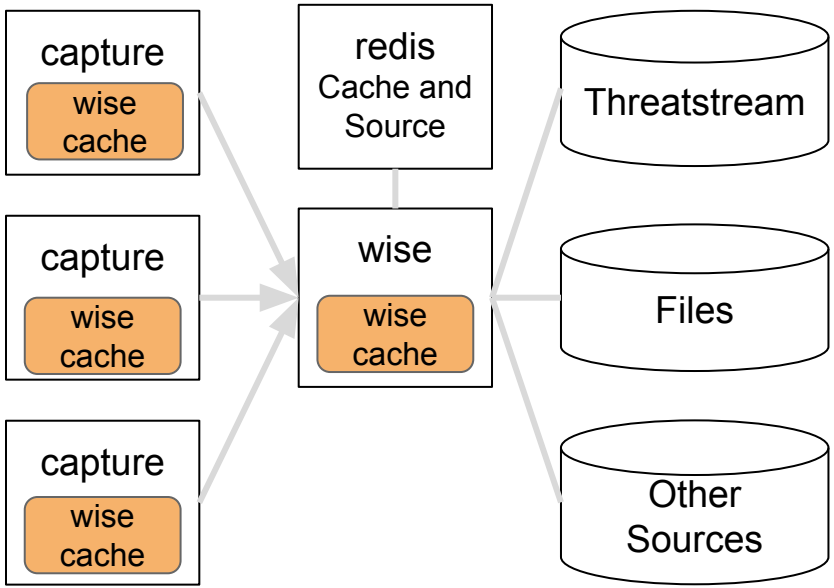
```
name: "Truncate Encrypted PCAP"
  when: "fieldSet"
  fields:
    protocols:
      - tls
      - ssh
      - quic
  ops:
    _maxPacketsToSave: 20
```

```
name: "Drop syn scan"
when: "beforeFinalSave"
fields:
  packets.src: 1
  packets.dst: 0
  tcpflags.syn: 1
ops:
  _dontSaveSPI: 1
```

PCAP Encryption

Arkime Encryption	**Disk Encryption**
Can’t use tools on files directly	Can use packet tools on file
File access isn’t enough to copy data	File access is enough to copy data
Password in config file and DEK in ES	Password at boot or TPM
Requires OpenSearch	Self contained
Potentially Less Secure Encryption	Potentially More Secure Encryption
Just a config change	More complex to setup

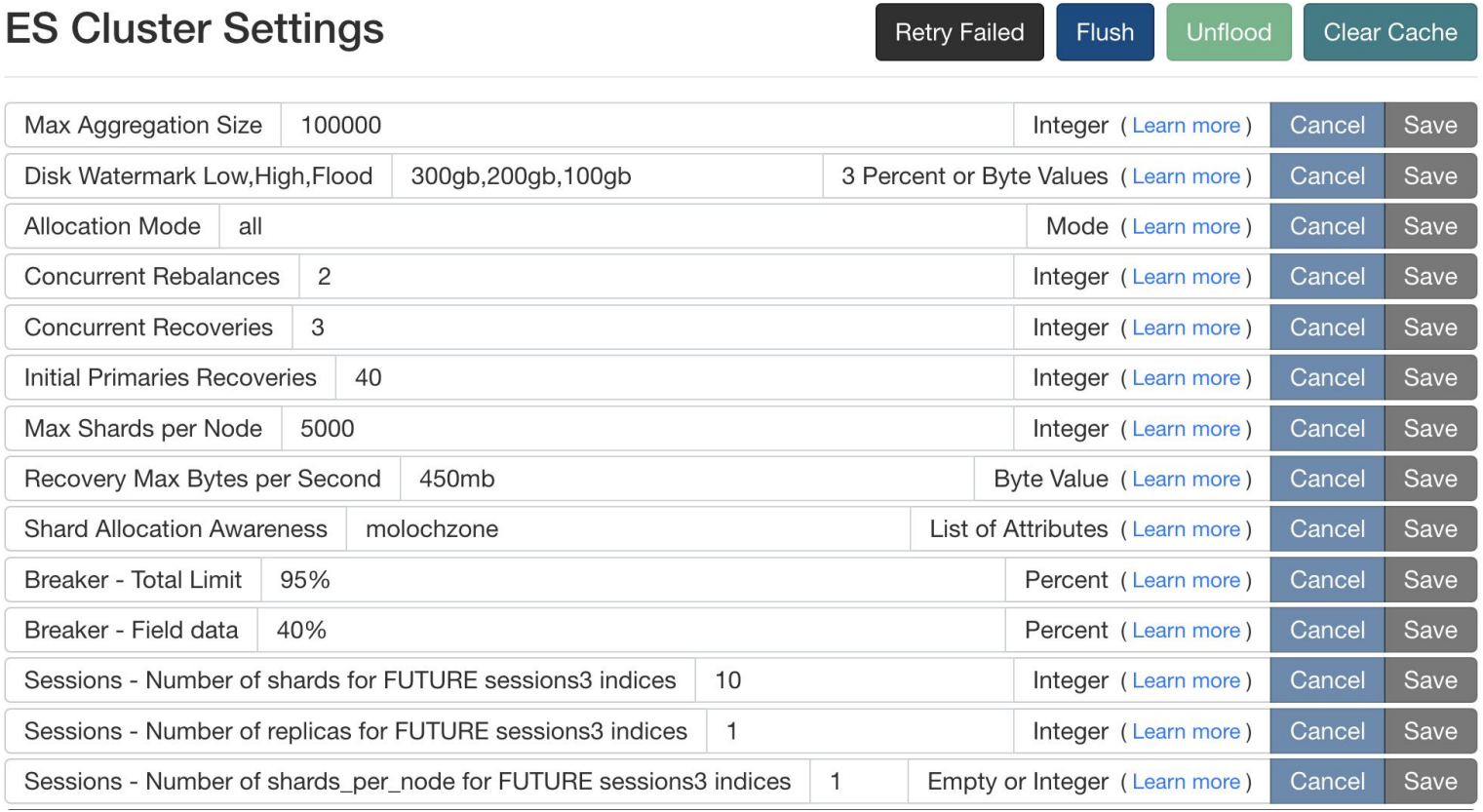
WISE Architecture



For performance reasons lookups are cached at multiple layers.

1) Check wise cache in capture (ALWAYS)
2) Check wiseService cache (for some sources)
3) Check redis cache (if configured)
4) Query the data source for information

Stats -> ES Admin Tab



slido

Audience Q&A Session

ⓘ Start presenting to display the audience questions on this slide.