XRootD at Qualys

Artem Harutyunyan,
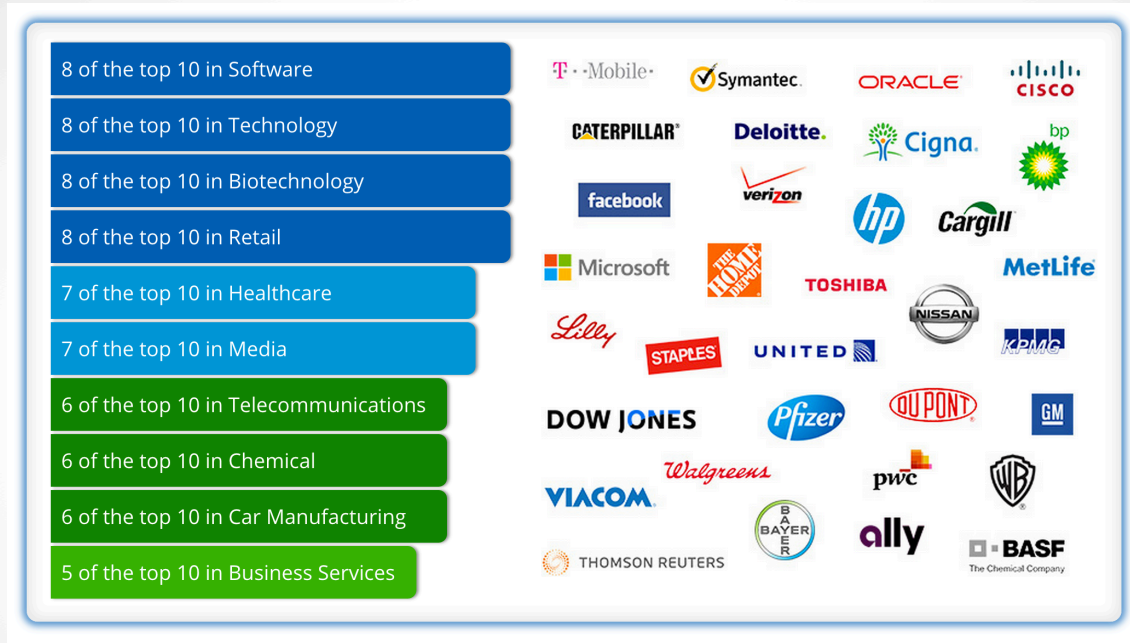Engineering Director

Who am I?

- ALICE Offline (2004 – 2010)

  - AliEn (porting to Windows, authentication, scheduling, etc.)

- Fellow at PH/SFT (2010 – 2012)

  - CernVM Co-Pilot – distributed computing framework powering LHC@home 2.0

  - CernVM development and infrastructure

  - CernVM File System infrastructure

- Architect/Engineering Director at Qualys (2012 – present)

Qualys®
CONTINUOUS SECURITY

About Qualys

- Provider of cloud security, compliance and related services for large corporations as well as small and medium-sized enterprises

- 6700 customers in more than 100 countries

- Engineering offices across US (Silicon Valley, Denver, Wisconsin, Seattle), in France (Paris), Switzerland (Geneva), India (Pune)
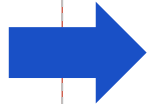
QUALYS®
CONTINUOUS SECURITY

About Qualys Customers

8 of the top 10 in Software	
8 of the top 10 in Technology	
8 of the top 10 in Biotechnology	
8 of the top 10 in Retail	
7 of the top 10 in Healthcare	
7 of the top 10 in Media	
6 of the top 10 in Telecommunications	
6 of the top 10 in Chemical	
6 of the top 10 in Car Manufacturing	
5 of the top 10 in Business Services	

- Some of the challenges
  - Scanning an entire class A network in under 4 hours
  - Manage O(1,000,000) assets per company
  - Manage O(10,000) web applications per company

How it works?

Network scanning
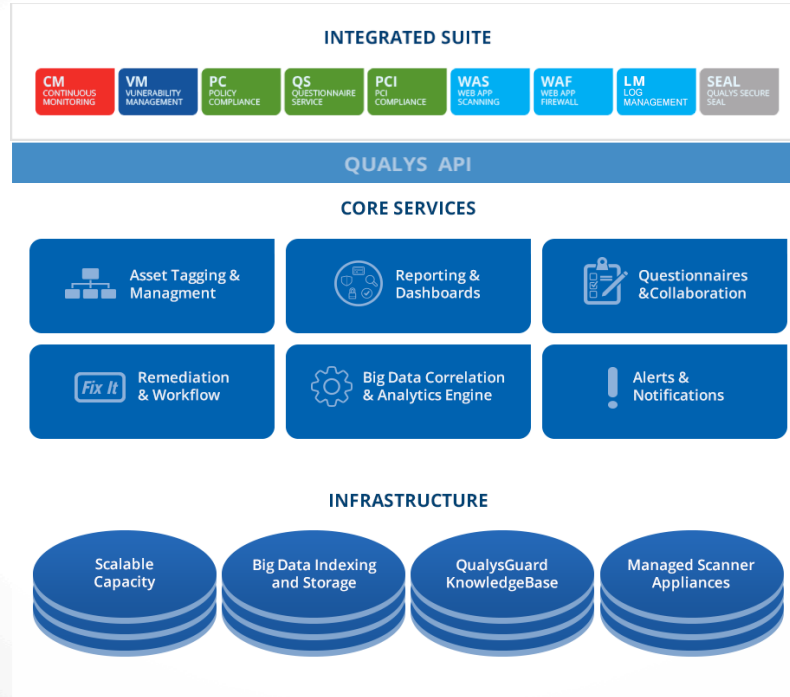and data collection

Storage and
Processing

Analysis and
Correlation

Customer's network

Qualys data center

QUALYS®
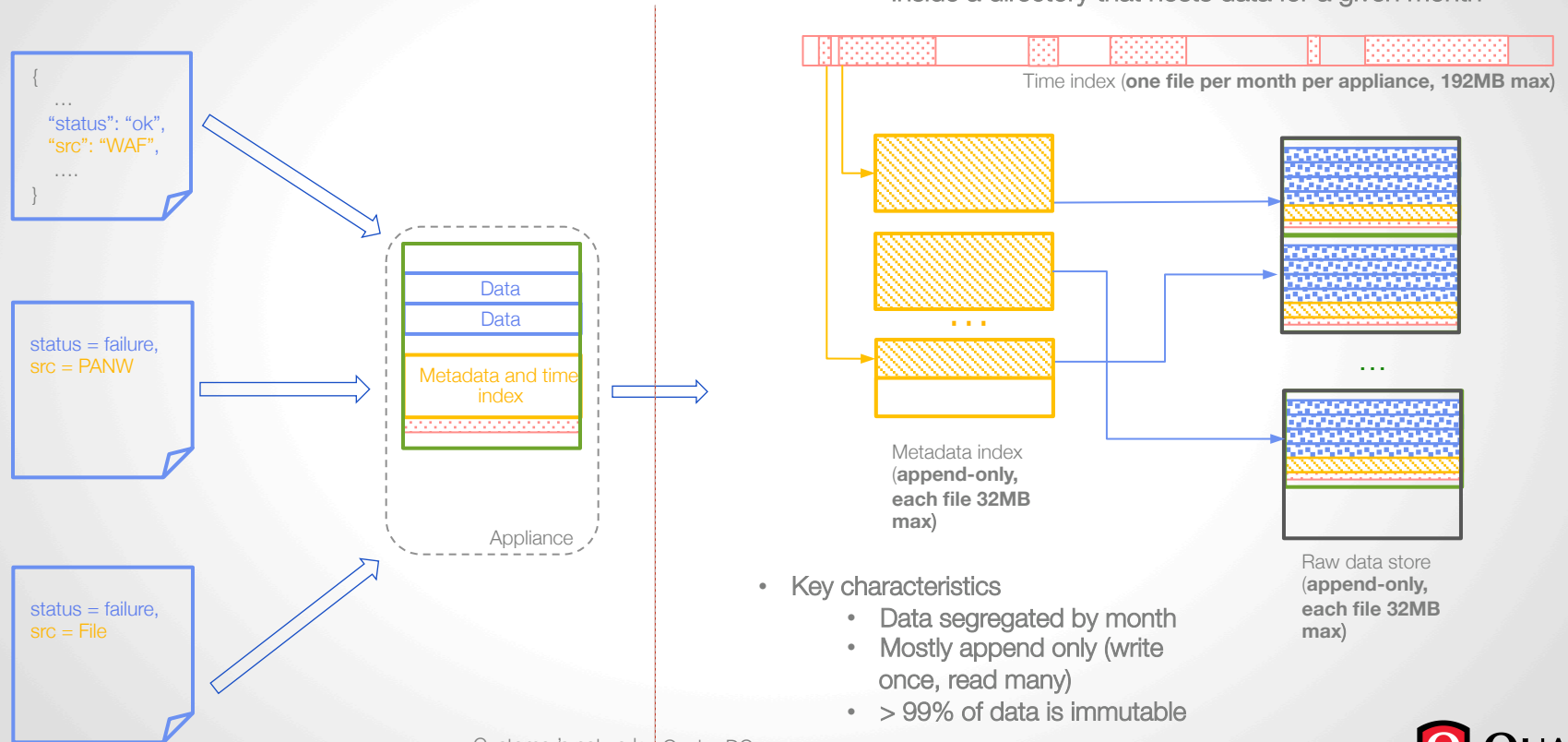CONTINUOUS SECURITY

What does it do?

How do we use XRootD?

- XRootD is used to store

  - Compressed raw data

  - Time and metadata indices

- Data is accessed in parallel by a set of distributed processing nodes

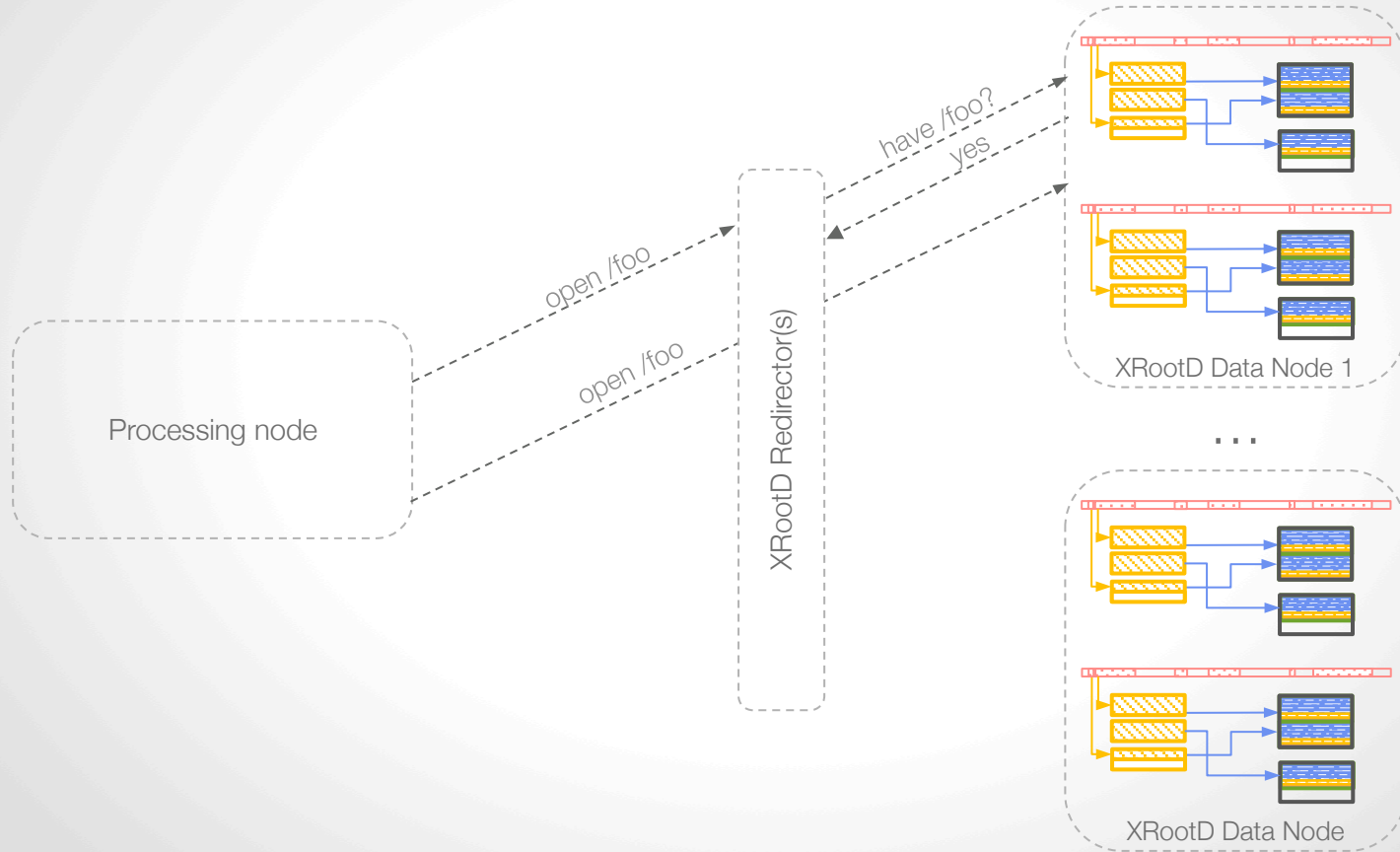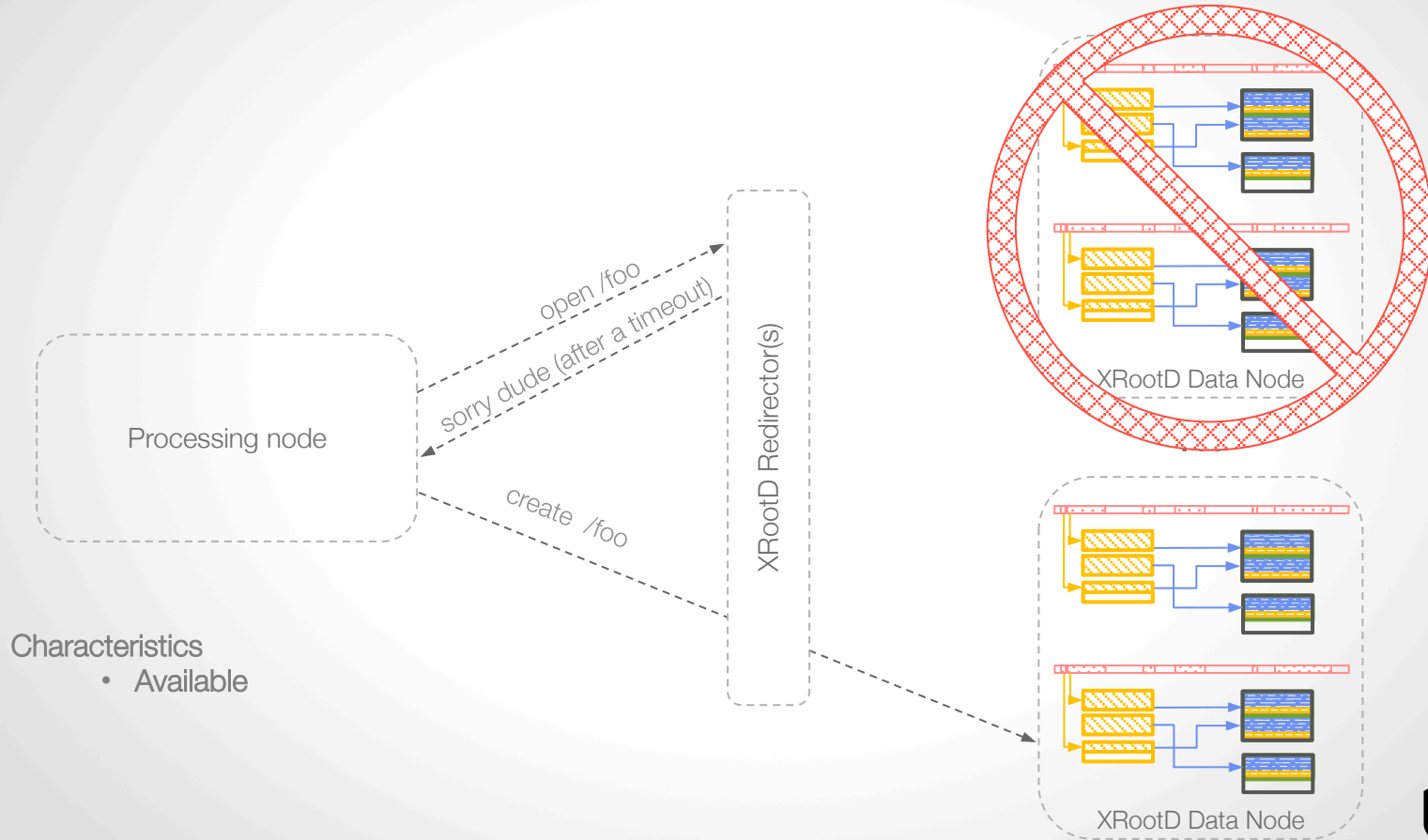- All the data resides within the same cluster/datacenter

QUALYS®
CONTINUOUS SECURITY

Data collection and storage

```
{
   …
   "status": "ok",
   "src": "WAF",
   ….
}
```

status = failure,
src = PANW

status = failure,
src = File

Appliance

Data
Data
Metadata and time index

Inside a directory that hosts data for a given month

Time index (**one file per month per appliance, 192MB max**)

. . .

Metadata index (**append-only, each file 32MB max**)

. . .

Raw data store (**append-only, each file 32MB max**)

- Key characteristics
  - Data segregated by month
  - Mostly append only (write once, read many)
  - > 99% of data is immutable

Customer's network | Qualys DC
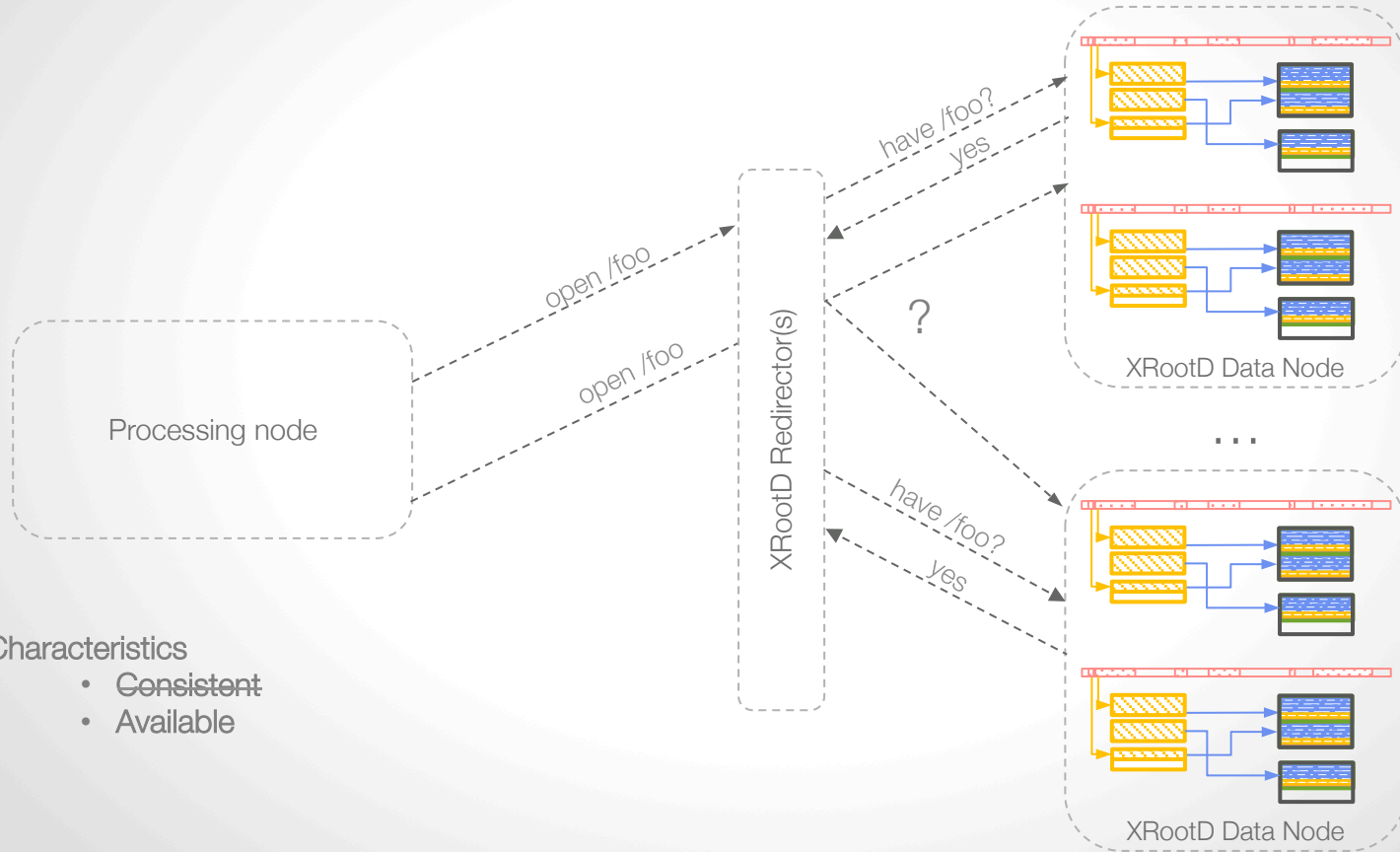
8

QUALYS®
CONTINUOUS SECURITY

Data access with XRootD

Data access with XRootD

Processing node

open /foo

sorry dude (after a timeout)

create /foo

XRootD Redirector(s)

XRootD Data Node

XRootD Data Node

Characteristics
- Available

10

Data access with XRootD

Processing node

open /foo

open /foo

XRootD Redirector(s)

have /foo?

yes

?

XRootD Data Node

. . .

have /foo?

yes

XRootD Data Node

Characteristics
- ~~Consistent~~
- Available

11

Data access with XRootD: 'Fixed' FS plugin

open /foo

goto MD5(/foo) % no. of nodes

Processing node

open /foo

XRootD Redirector(s) **with Fixed FS Plugin**

XRootD Data Node

. . .

XRootD Data Node

Characteristics
- Consistent (single copy)

Data access with XRootD: 'Fixed' FS plugin

Processing node

open /foo

goto MD5(/foo) % no. of nodes

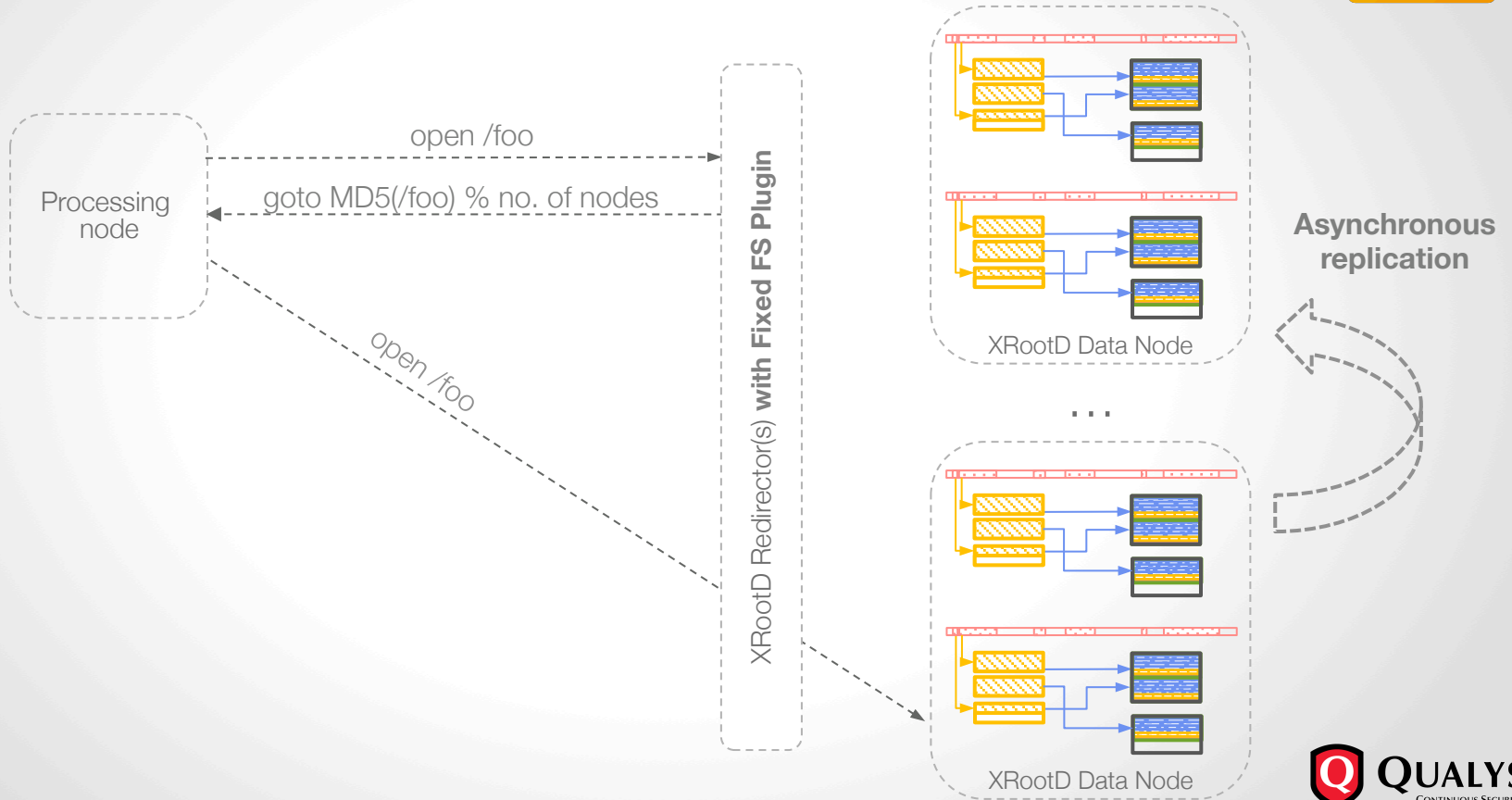open /foo fails

XRootD Redirector(s) **with Fixed FS Plugin**

XRootD Data Node
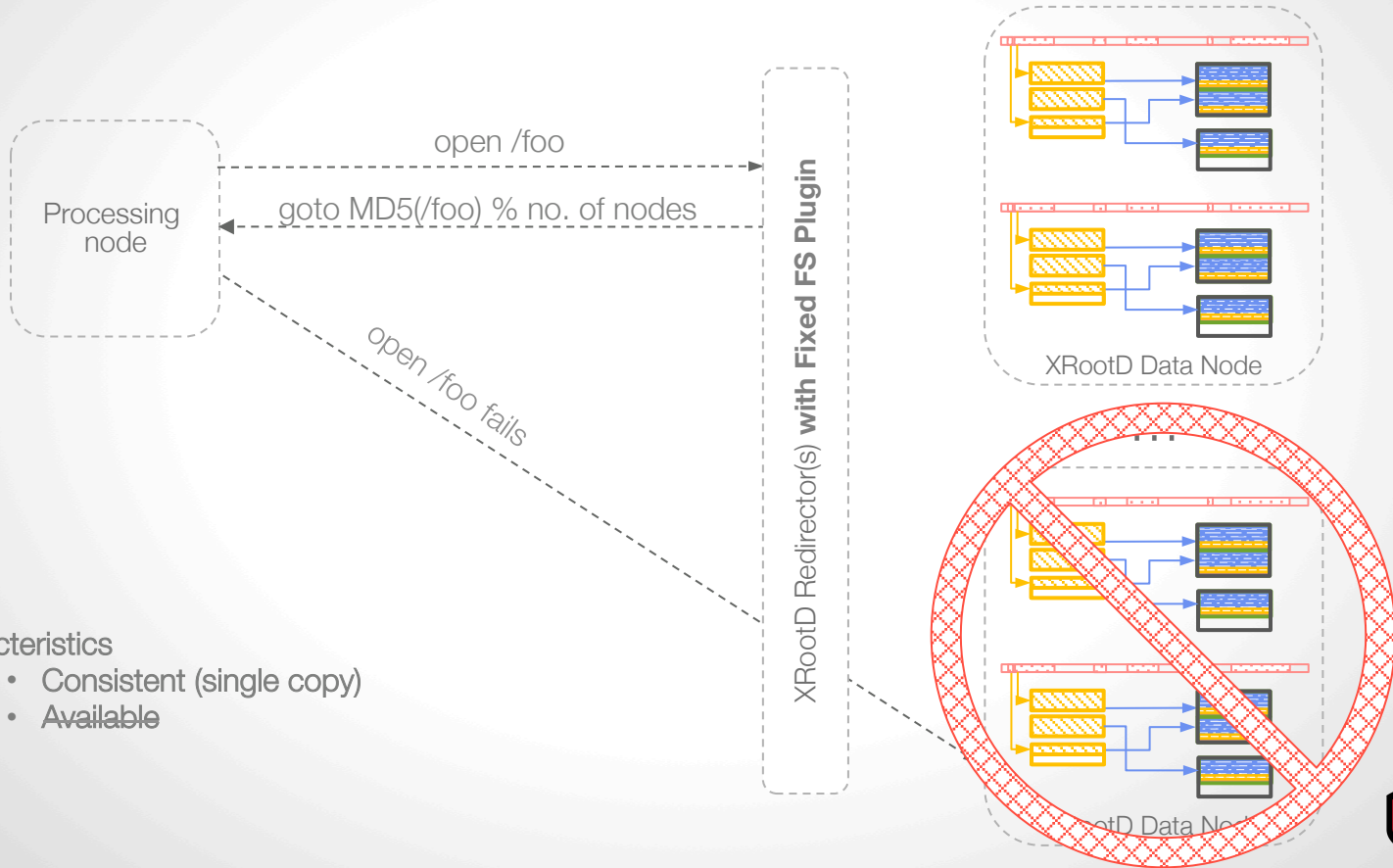
XRootD Data Node

Characteristics
- Consistent (single copy)
- ~~Available~~

Data access with XRootD: 'Fixed' FS plugin + replication

open /foo

goto MD5(/foo) % no. of nodes

open /foo

Processing node

XRootD Redirector(s) **with Fixed FS Plugin**

XRootD Data Node

. . .

XRootD Data Node

Asynchronous replication

Data access with XRootD: 'Fixed' FS plugin + replication

Processing node

open /foo

goto MD5(/foo) % no. of nodes

read /foo works

write /foo fails

XRootD Redirector(s) **with Fixed FS Plugin**

XRootD Data Node

XRootD Data Node

Asynchronous replication

Characteristics
- Somewhat consistent
- Somewhat available
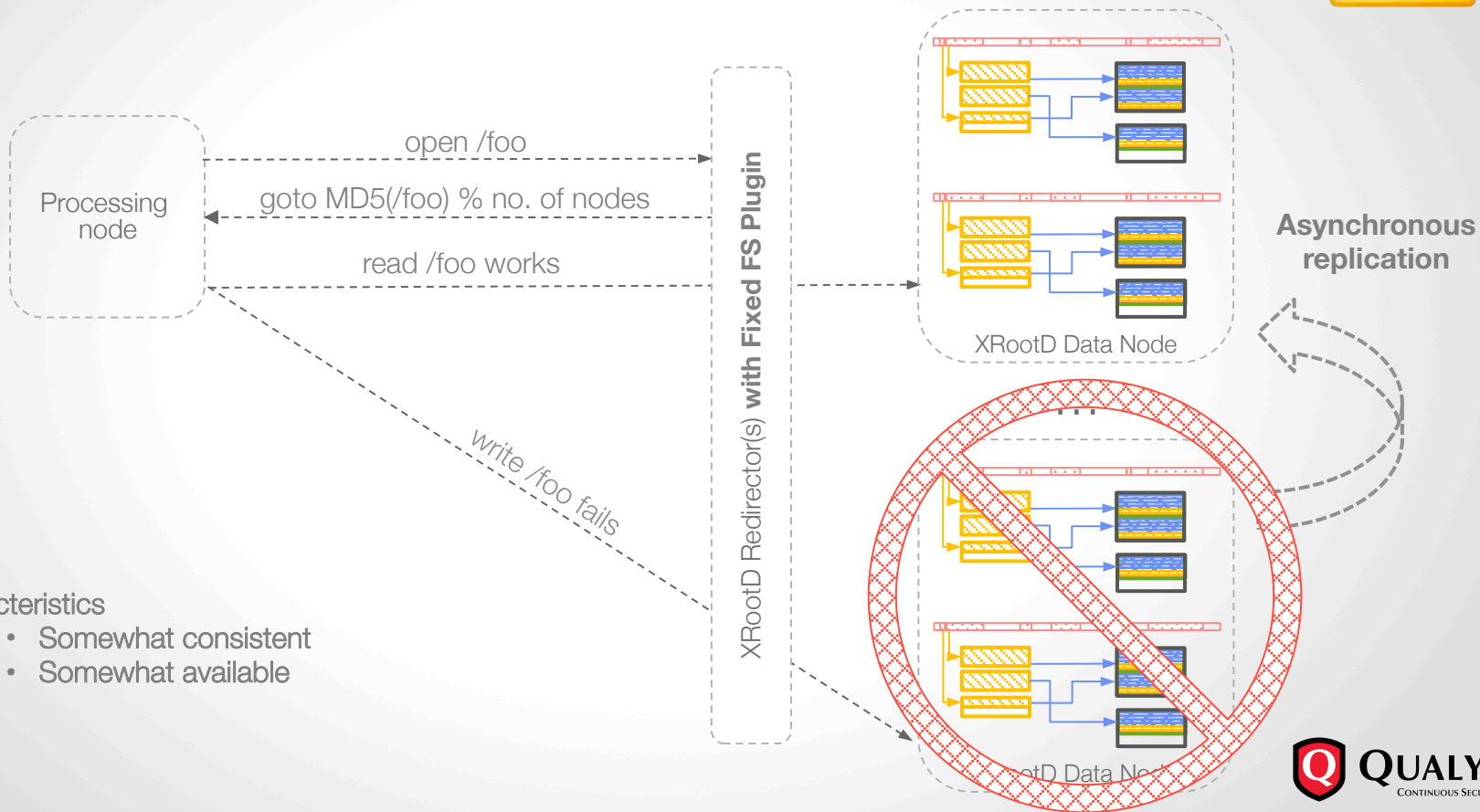
What about scaling?

- Changing the number of data nodes will require relocation of the (entire) dataset. Can be done, but better be avoided

- When the current cluster gets near it's storage capacity, the new one will be created and Fixed FS redirector plugin will redirect requests to the right cluster (based on the month/year which is always present in the path)

QUALYS®
CONTINUOUS SECURITY

Questions?

Thank You

hartem@qualys.com
twitter.com/hartem
linkedin.com/in/artemharutyunyan