DATA IS POTENTIAL

Building the Ultimate Object Store for 175 ZBs of 2025, one step at a time

Gregory Touretsky, Principal Product Manager

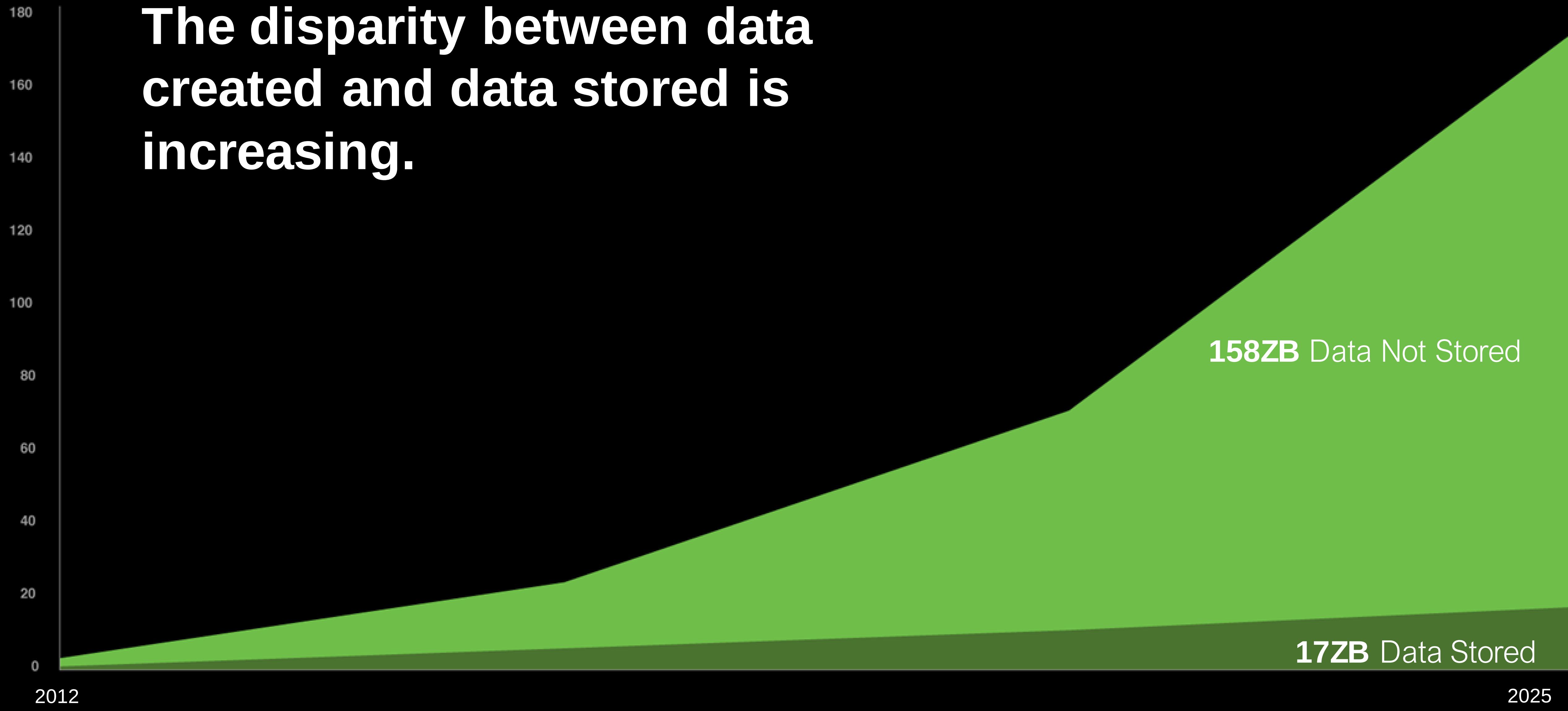
@gregnsk

SEAGATE

DATA IS IN OUR DNA

Why are we doing it?

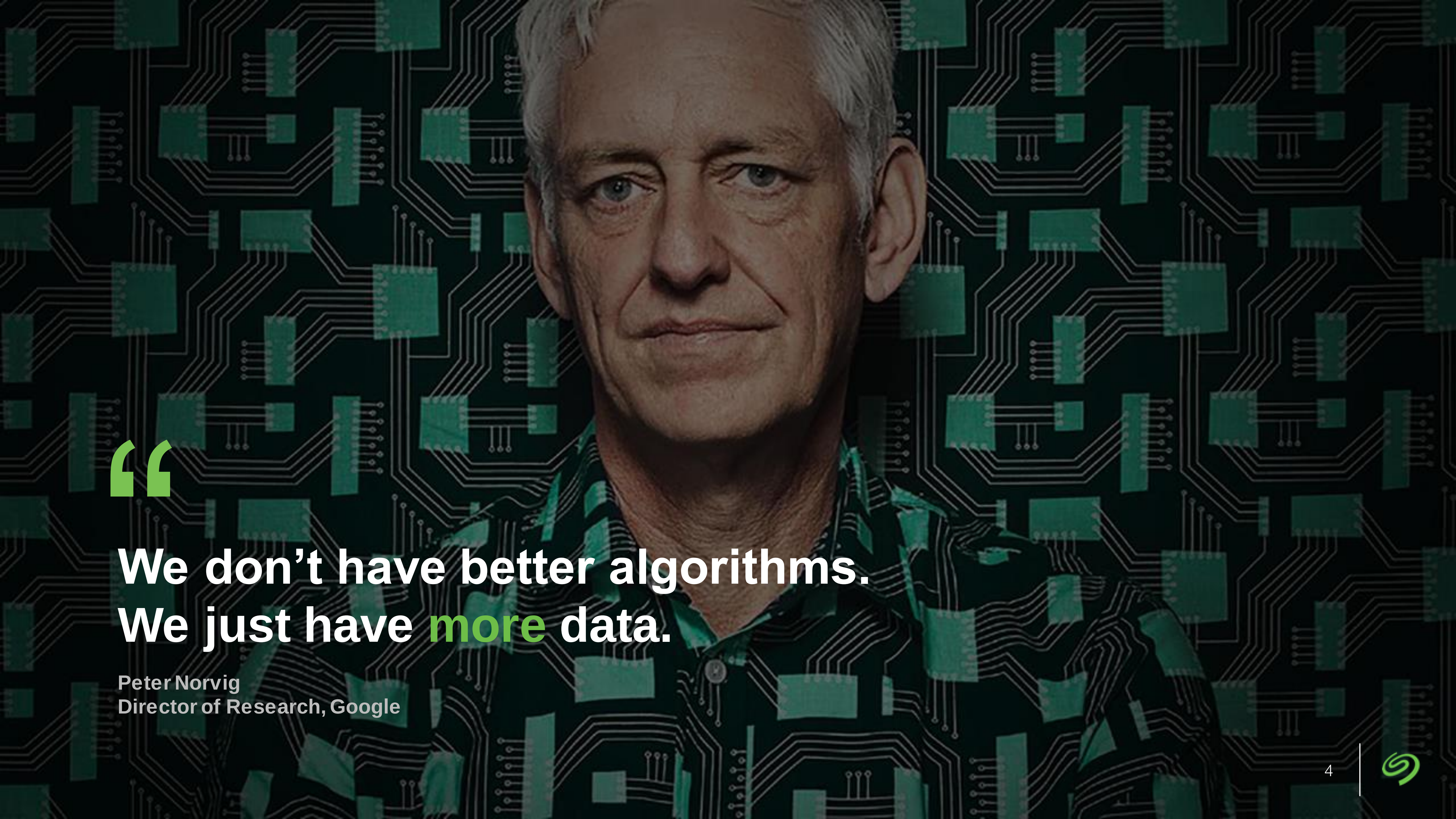
The disparity between data created and data stored is increasing.



158ZB Data Not Stored

17ZB Data Stored

Source: Data Age 2025, sponsored by Seagate with data from IDC Global Datasphere, Nov 2018. ZB = Zettabyte.



“

**We don’t have better algorithms.
We just have more data.**

Peter Norvig
Director of Research, Google

Enterprises are forced to compromise in the data economics equation.

COST OF STORING MORE DATA

Infrastructure OPEX
Infrastructure CAPEX
Human Resources

VS

VALUE OF STORING MORE DATA

Customer Insights
Operational Efficiencies
New Revenue Opportunities

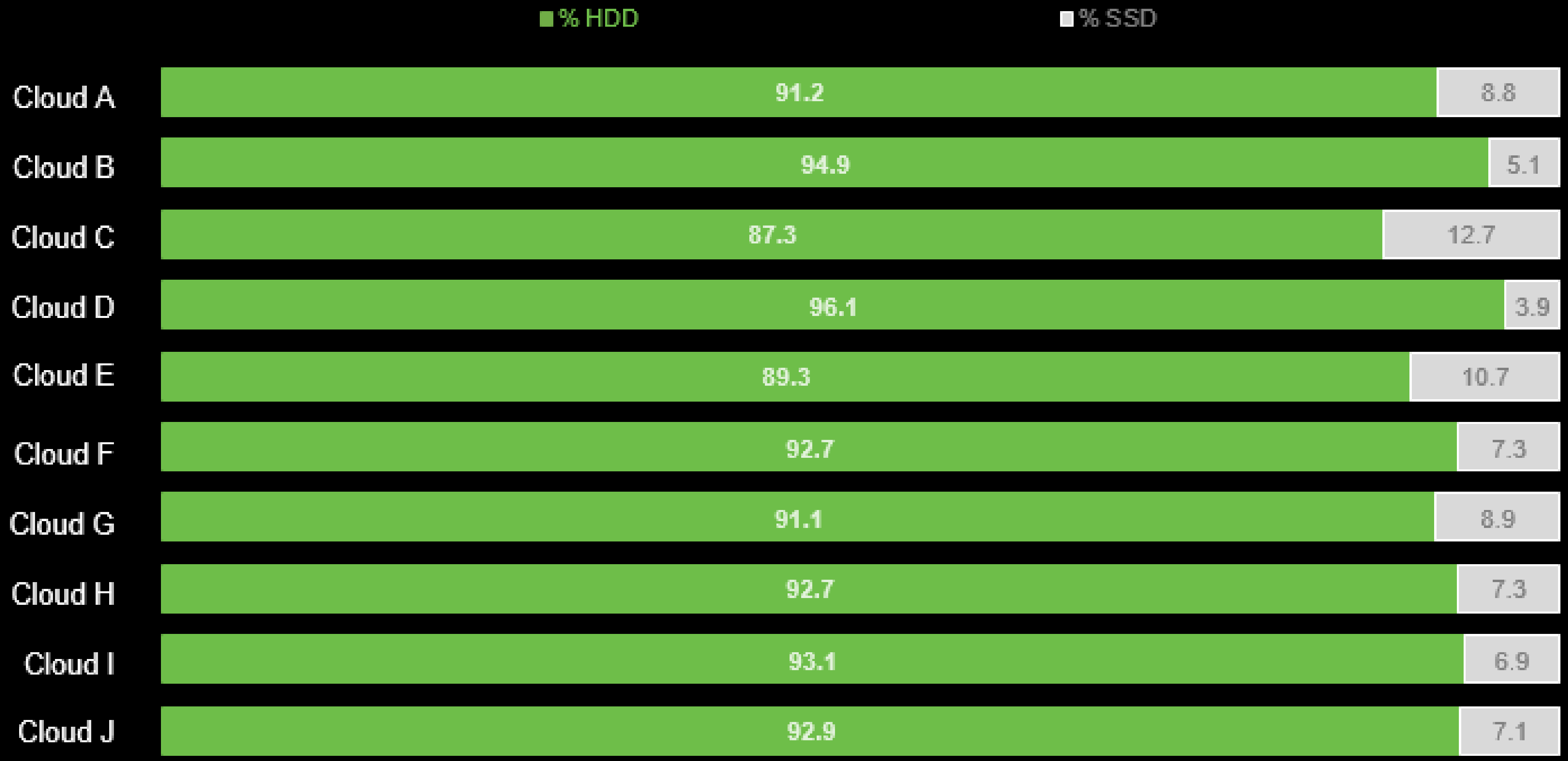
Hyperscalers have the optimal stack for mass unstructured data

- **“Software-defined everything”** with proprietary **object storage software** and high-leverage of open source software

- **Rapid adoption** of higher-capacity storage devices and advancements because of the TCO advantage

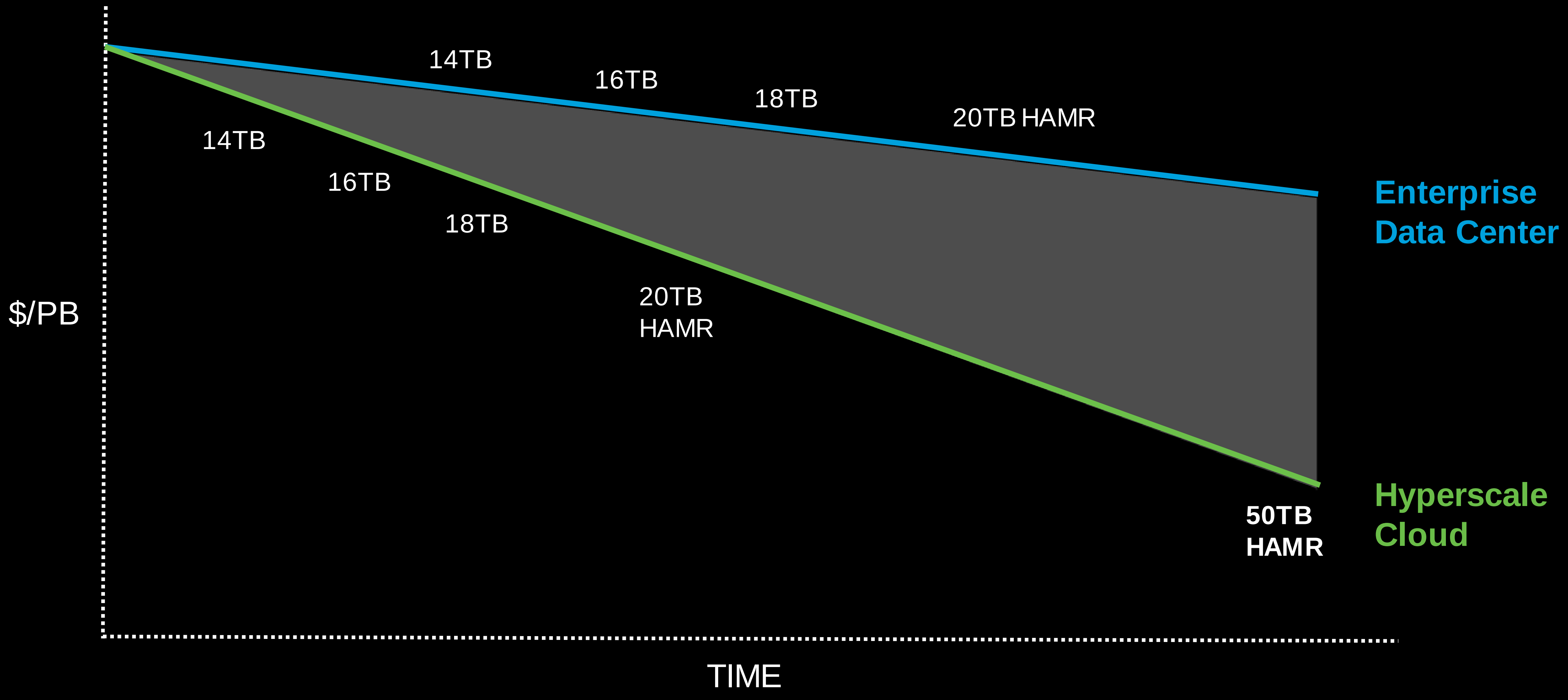
- **Industry-standard hardware** optimized for cloud-scale efficiency, scale, performance

Hyperscale is optimized for mass capacity



Source: IDC Cloud Infrastructure Index 2019

Rapid, continuous adoption of highest-capacity HDD underpins a sustained cost advantage.



Object storage: two categories

- **Mass Capacity**
- ~90% of on-prem object storage capacity
  - New workloads depend on Mass Capacity
  - Training Workloads for AI/ML
  - Archive, backup, etc
- Cost per GB is important
  - Density matters like cost
  - Segment also known as *cheap and deep* in the market
  - Wide spread of pricing in the market

High performance

- ~10% of on-prem object storage capacity
  - More common in the public cloud
- Performance (latency) is important
  - Azure Premium Blob storage: $150/TB/month

DATA IS IN OUR DNA

CORTX today

What is CORTX™

- An S3-compatible object storage platform
  - 100% open source project on GitHub

License Apache 2.0 | code quality A | codacy-analysis-cli passing

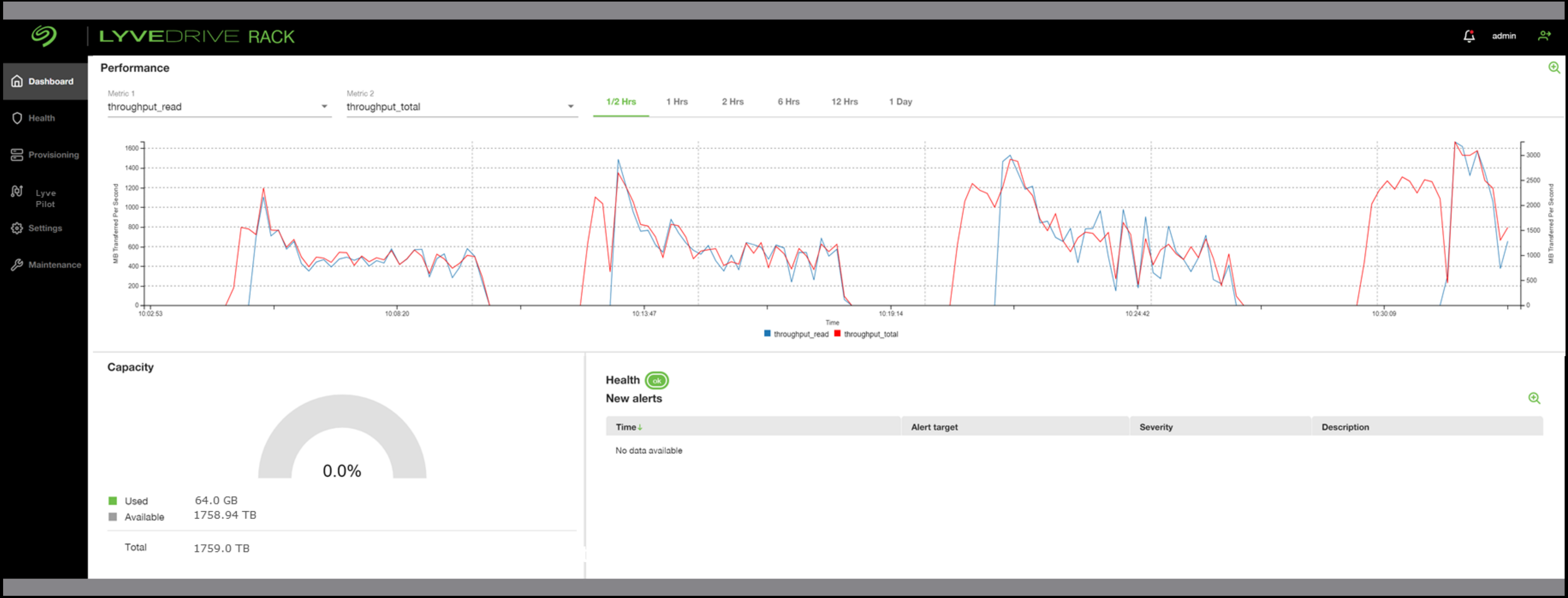
- Pre-built VM image for testing and quick start
  - 15 minutes to launch your own CORTX instance
  - Functionality preview only, not for production

- Supported by the community

Join the CORTX community

https://github.com/Seagate/CORTX

CORTX - GUI



Performance

Metric 1: throughput_read

Metric 2: throughput_total

Capacity

0.0%

🟩 Used	64.0 GB
▨ Available	1758.94 TB
Total	1759.0 TB

Health ok

New alerts

Time↓	Alert target	Severity	Description
No data available			

Supported API calls

Swagger: https://<managementIP>:28100/api-docs

S3 APIs

Account operations:

- GET Account

Bucket operations:

- DELETE Bucket
- DELETE Bucket Policy
- GET Bucket
- GET Bucket ACL
- GET Bucket Policy
- HEAD Bucket
- GET multipart uploads
- PUT Bucket
- PUT Bucket ACL
- PUT Bucket Policy
- GET Bucket Tagging

Object operations:

- DELETE Object
- DELET Object Tagging
- DELETE Multiple Objects (POST)
- GET Object
- GET Object ACL
- GET Object Tagging
- HEAD Object
- PUT Object
- PUT Object ACL
- PUT Object Tagging
- *PUT Object (Copy) - WIP*
- Initiate Multipart Upload (POST)
- Upload Part (PUT)
- Complete Multipart Upload (POST)
- Abort Multipart Upload (DELETE)
- List Parts (GET)

IAM APIs

Account operations:

- Create Account
- Delete Account
- List Accounts

User operations:

- Create User
- Update User
- Delete User
- List Users
- Change Password

Key operations:

- Create Access Key
- List Access Keys
- Delete Access Key
- Update Access Key

Misc operations:

- Get Temp Auth Credentials
- Create Account Login Profile
- Update Account Login Profile
- Get Account Login Profile
- Create User Login Profile
- Update User Login Profile
- Get User Login Profile

CSM APIs

User operations:

- GET /csm/users/{user_id}
- PATCH /csm/users/{user_id}
- POST /csm/users
- DELETE /csm/users/{user_id}
- GET /permissions

Alerts operations:

- GET /alerts/{alert_id}
- PATCH /alerts/{alert_id} (ACK)
- GET /alerts_history/{alert_id}
- GET /alerts/{alert_id}/comments
- POST /alerts/{alert_id}/comments

Health operations:

- GET /system/health/summary
- GET /system/health/node
- GET /system/health/components
- GET /system/health/resources

Misc operations:

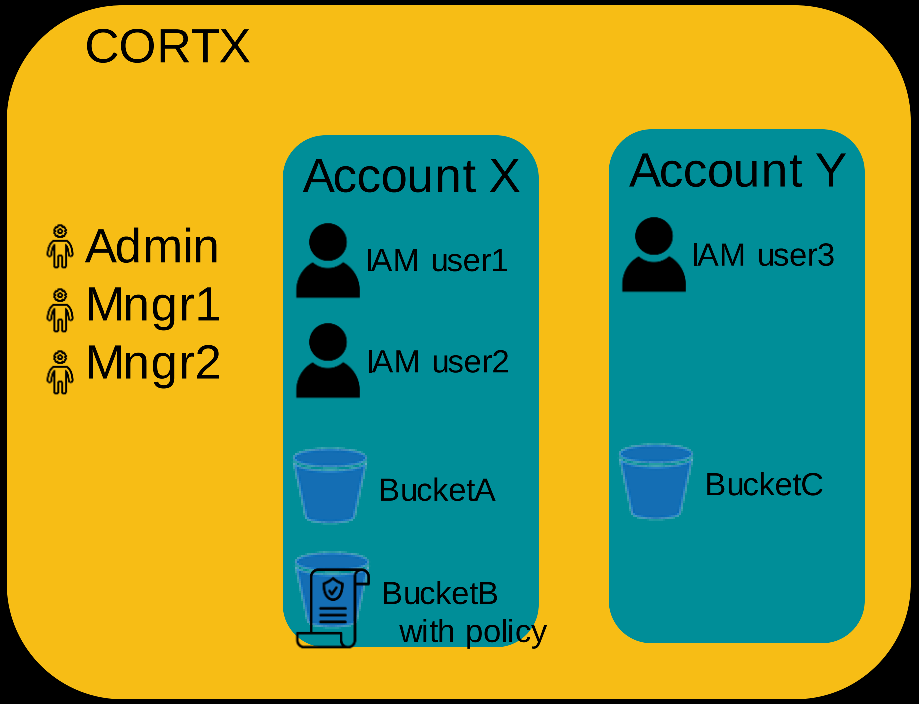
- GET /product_version
- GET /stats
- GET /capacity
- POST /login
- POST /logout
- GET /auditlogs/show/{component}
- GET /auditlogs/download/{component}

CORTX accounts

- Administrative accounts
    - Control CORTX system
    - Roles:
        - Admin (superuser)
        - Manage (modify)
        - Monitor (readonly)
    - Attributes:
        - Username, email, password, role

- S3 accounts (namespaces / tenants)
    - Attributes:
        - Account name, email, password
        - One or more access key/secret key pairs
    - At least one is required to store data
    - Each may control zero or more IAM accounts and buckets

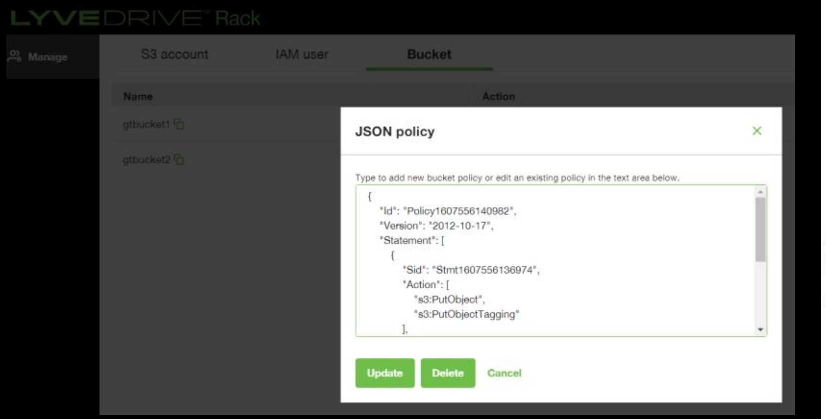
- IAM users
    - Attributes:
        - Username, password
        - User id
        - ARN
        - One or more access key/secret key pairs



Bucket Policies

- Attached to buckets
- Specify what actions are Allowed or Denied for which Principal
- Apply to all objects within the bucket
- May include conditional statement
- Written in JSON using AWS access policy language
  - Up to 20KB

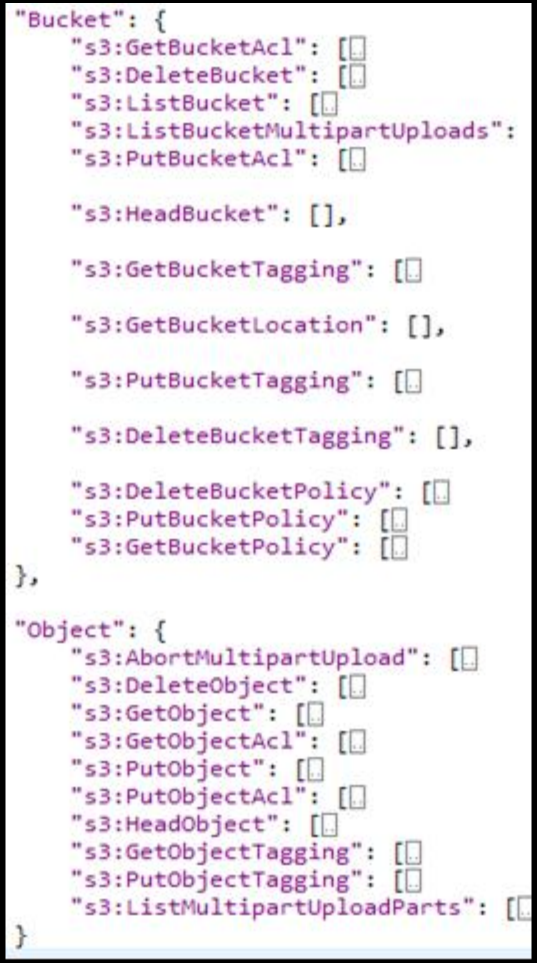
```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Stmt1607556136974",
      "Action": [
        "s3:PutObject",
        "s3:PutObjectTagging"
      ],
      "Effect": "Allow",
      "Resource": "arn:aws:s3:::gtbucket2/*",
      "Principal": {
        "AWS": [
          "arn:aws:iam::847912992506:user/gtiamuser"
        ]
      }
    }
  ]
}
```



Type to add new bucket policy or edit an existing policy in the text area below.

```
{
  "Id": "Policy1607556140982",
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Stmt1607556136974",
      "Action": [
        "s3:PutObject",
        "s3:PutObjectTagging"
      ],
```

Supported S3 Actions:



```
"Bucket": {
    "s3:GetBucketAcl": [□
    "s3:DeleteBucket": [□
    "s3:ListBucket": [□
    "s3:ListBucketMultipartUploads":
    "s3:PutBucketAcl": [□

    "s3:HeadBucket": [],

    "s3:GetBucketTagging": [□

    "s3:GetBucketLocation": [],

    "s3:PutBucketTagging": [□

    "s3:DeleteBucketTagging": [],

    "s3:DeleteBucketPolicy": [□
    "s3:PutBucketPolicy": [□
    "s3:GetBucketPolicy": [□
},

"Object": {
    "s3:AbortMultipartUpload": [□
    "s3:DeleteObject": [□
    "s3:GetObject": [□
    "s3:GetObjectAcl": [□
    "s3:PutObject": [□
    "s3:PutObjectAcl": [□
    "s3:HeadObject": [□
    "s3:GetObjectTagging": [□
    "s3:PutObjectTagging": [□
    "s3:ListMultipartUploadParts": [□
}
```

NotPrincipal, NotResource and NotAction are not supported

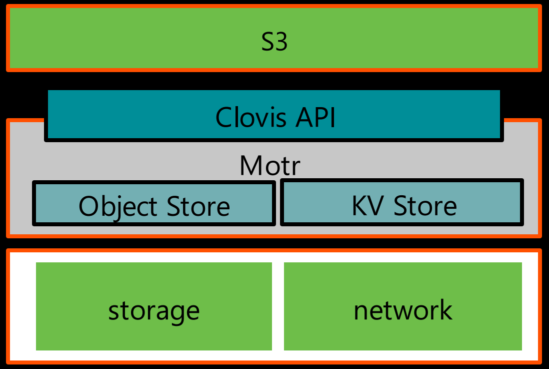
ACLs

- Legacy access control mechanism
- Attached to buckets and objects
- Evaluated if Bucket policy isn’t present on bucket or Policy is not concluding
- Default ACL: Full Control for the resource owner
- Authorization decision = union of all the S3 bucket policies and S3 ACLs that apply in accordance with the principle of least-privilege

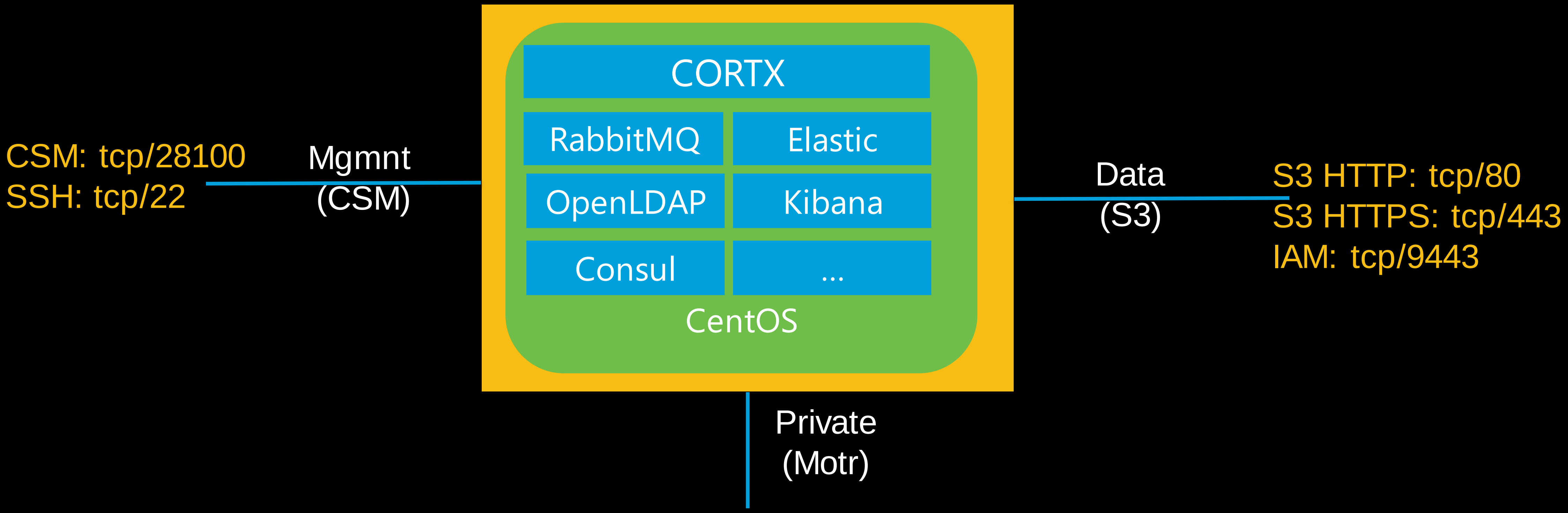
```
$ aws --endpoint-url http://172.16.8.16 s3api put-object-acl --bucket gtbucket1 \
 --key node_manifest.json --grant-full-control \
 id=a9202d6a64d94fa1ac6b6d09a902ae2b84390e4557004d19b4a471cd2525f429 \
 --grant-read emailaddress=gregory@seagate.com
```

Motr

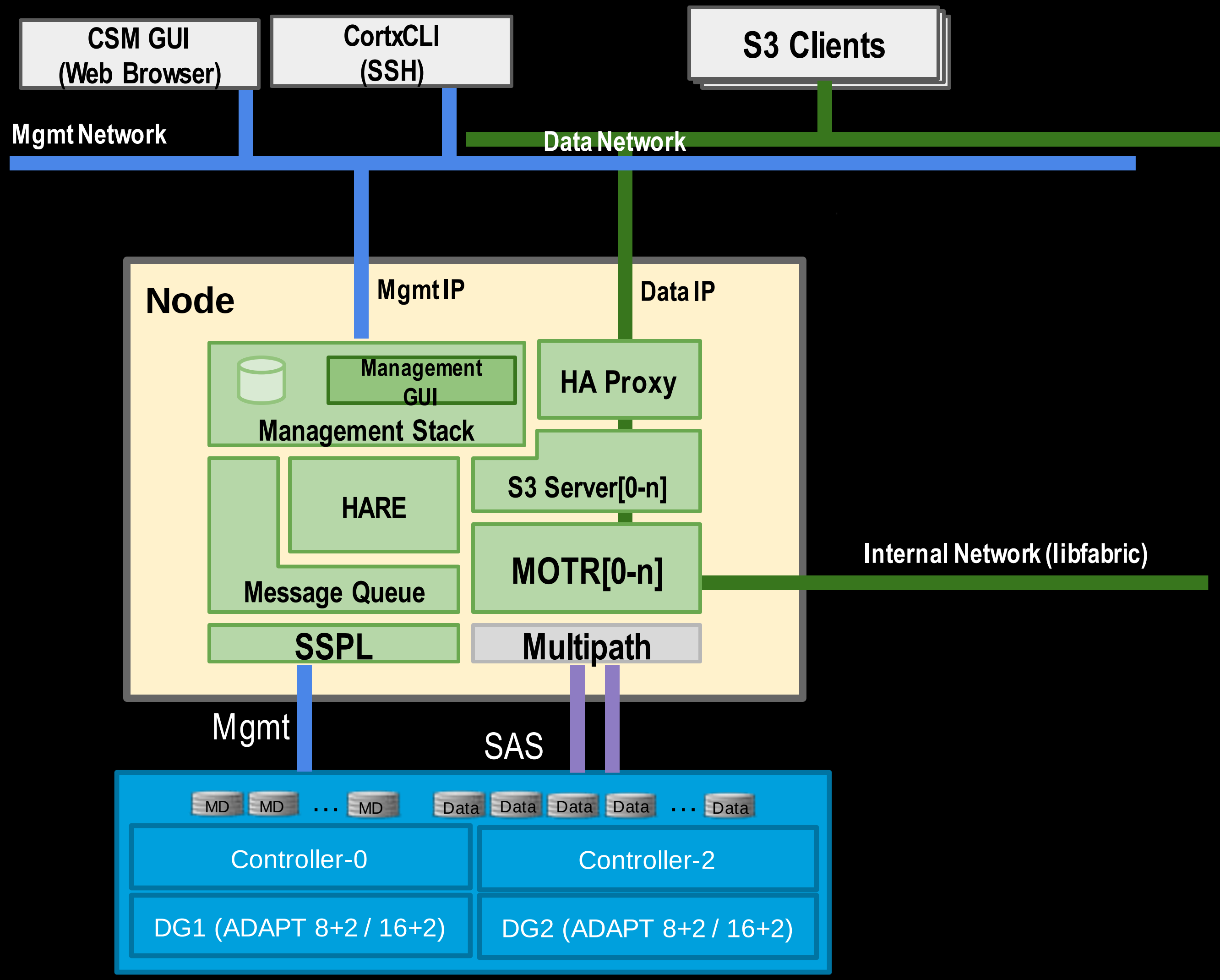
- The “Heart” of CORTX
- Scalable
  - Horizontal scalability
    - grow system by adding more nodes
    - no meta-data hotspots, shared-nothing IO path
    - Extensions running on additional nodes.
  - Vertical scalability
    - more memory and CPU on the nodes.
- Fault-tolerant:
  - flexible erasure coding taking hardware and network topology into account
  - fast network RAID repairs
- Observable: built-in monitoring collecting detailed information about system behavior
- Extensible
  - extension interface
  - flexible transactions
- Portable: runs in user space on any version of Linux



CORTX node – 30,000 ft view



CORTX/LR node – 10,000 ft view



- CSM (Component Service Management)
- SSPL (Seagate Storage Platform Library)
- HARE
  - HA for Motr and S3 server
  - hctl interface
- S3 server
- Motr
  - IO service (Object store)
    - Create/Write/Read/Unlink
  - Index service (KV store)
    - Idx Create/List/Lookup/Drop
    - KV Put/Get/Next/Delete

DATA IS IN OUR DNA

CORTX usage

SEAGATE

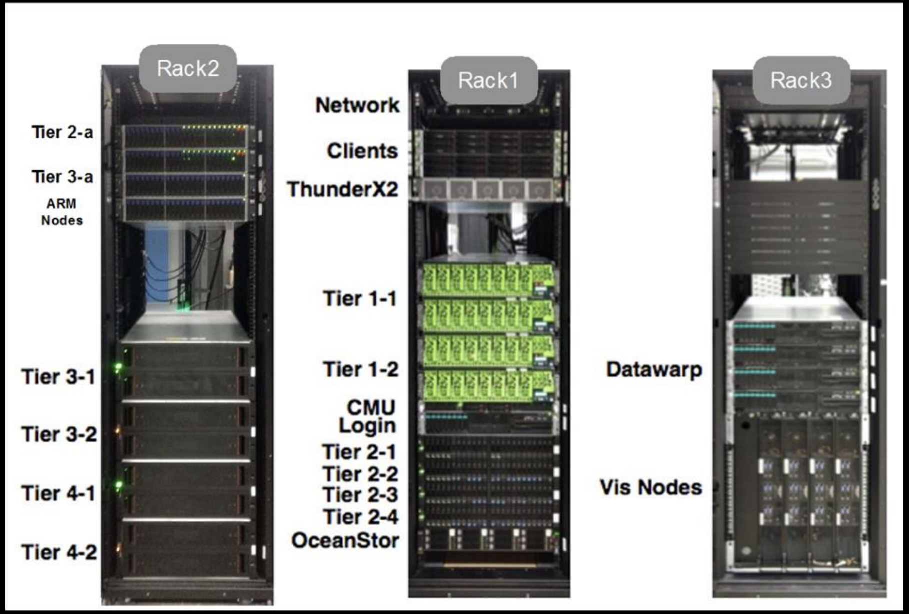
What is Sage2

- European Research project
- Percipient Storage for Exascale Data Centric Computing 2
    - Unified data storage platform for AI, Deep Learning, Big Data analysis and High-Performance Computing workloads
    - CORTX for HPC/AI - Motr & Motr API
    - Usage of multiple tiers

https://sagestorage.eu/

SAGE2 status and plans

- 22-node Motr cluster at Juelich Supercomputing Center, Germany
- Focus on Application Porting
- Completion of Prototype Implementations
- Detailed Performance analysis of CORTX on SAGE
- Multiple POCs:
    - QoS (HSM and Performance throttling)
    - Arm porting
    - dCache on Motr API
    - 3DXPoint NVDIMM interoperability
    - Slurm CORTX burst buffer plugin
    - Motr Function shipping



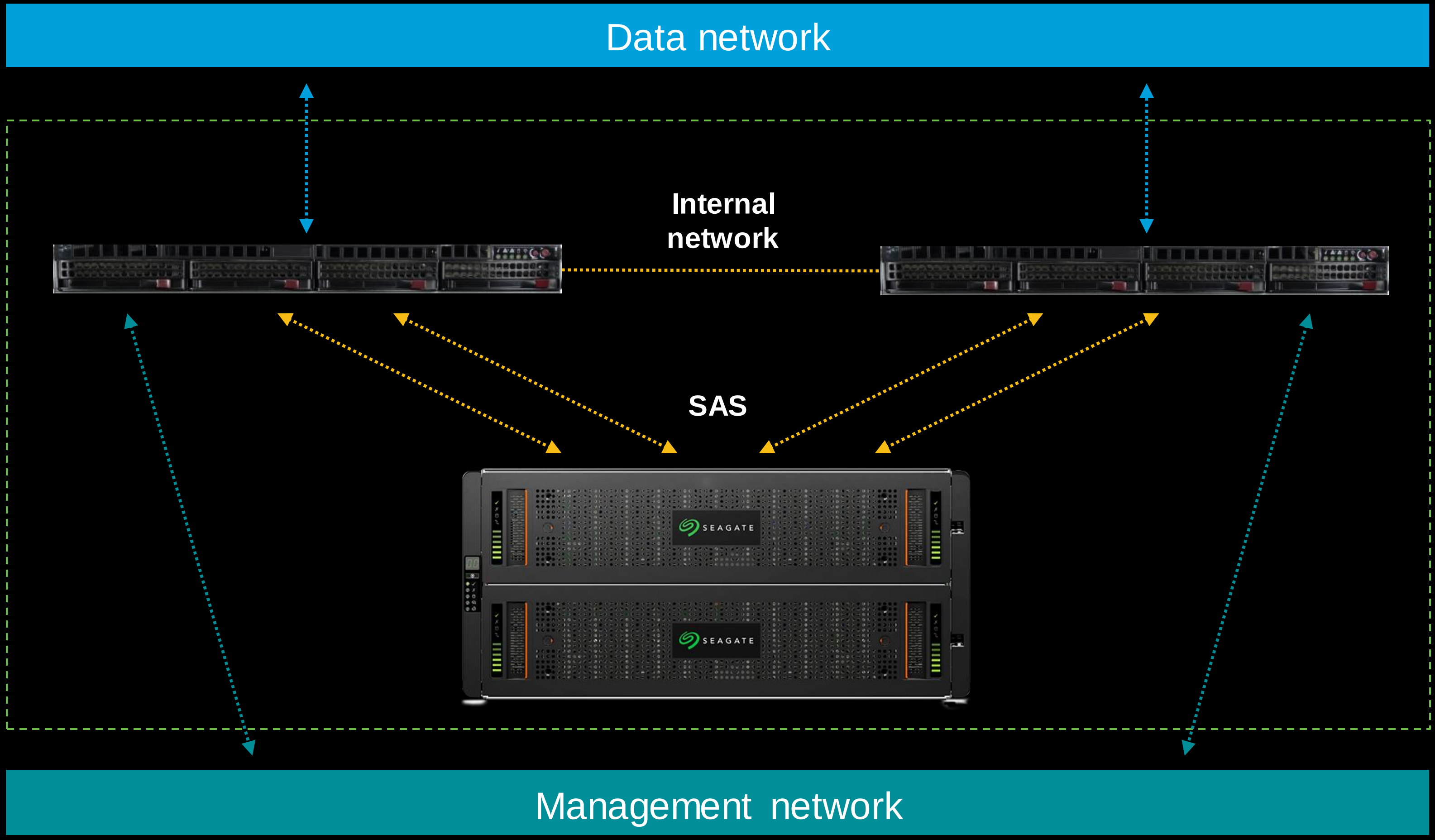
What is LYVE Rack

- HW + SW reference architecture for Enterprise customers
- Powered by 100% open source CORTX
- Tested and supported by Seagate
- Available via selected partners

Lyve Rack R1 Reference Architecture (Edge)

Supported enclosures	5U84
Controllers	Dual 1U servers
External network	2x50Gbps (data) 2x1Gbps (management)
Protocols	S3 compatible
Data protection	Seagate ADAPT

- Powered by CORTX
- Active-Active High Availability (HA)



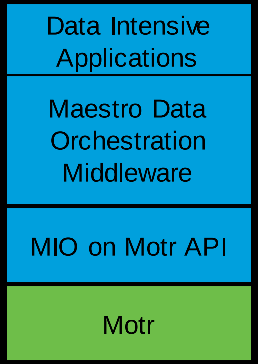
DATA IS IN OUR DNA

Where are we heading?

SEAGATE

CORTX in the Research projects

- Data-aware middleware for extreme scale applications
- https://www.maestro-data.eu



Data Intensive Applications
Maestro Data Orchestration Middleware
MIO on Motr API
Motr

- Exascale weather and climate simulations
- https://www.esiwace.eu/

- Data management platform suitable for Exascale

- Deduplication for CORTX
- https://research.zdv.uni-mainz.de/deduplication-for-cortx/

Perpetual Storage Platform (aka Lyve Rack R2)

Not a single project, it’s an evolving product

Scalable clustered S3-compatible on-premises object storage solution that delivers market-competitive capabilities at the lower price-point.

- Renewable cluster
- Easy to use
- Scalable
- Reliable
- Upgradable
- Affordable

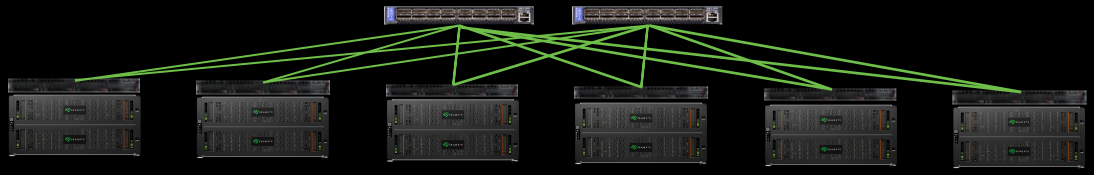
https://www.cloudoye.com/blog/cloud-storage/rising-popularity-cloud-object-storage-and-limitations-of-scale-out-storage

Live Rack R2 components

- Node = Seagate Smart Enclosure + Server + OS + CORTX

- StorageSet = smallest building block
  - X nodes
  - StorageSets may have different HW (ex: newer generation)

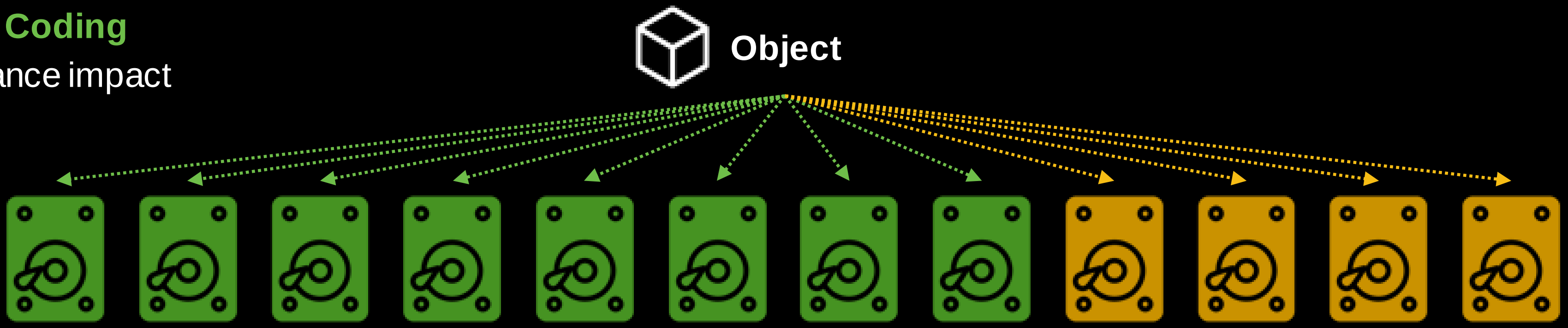
- Cluster = an instance of Lyve Rack R2
  - Consists of 1 or more StorageSets.
  - **“Perpetual storage”**
    - Same level of support for the entire cluster (NBD or 24x7)
    - Support contract follows the StorageSet
    - Graceful addition / removal of StorageSets



Hierarchical Erasure Coding

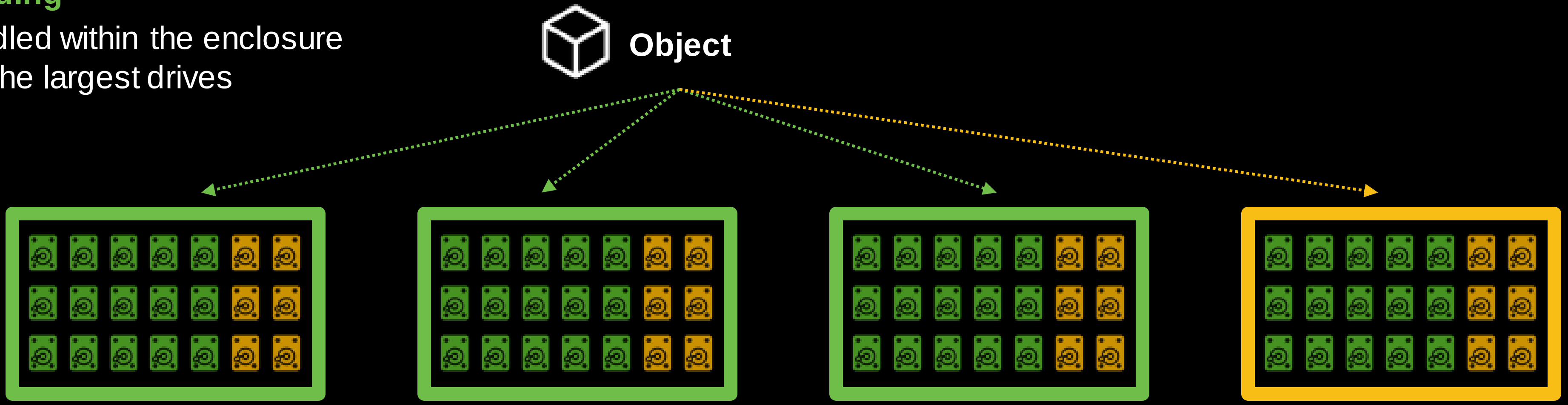
Single Level Erasure Coding

Disk failure = performance impact

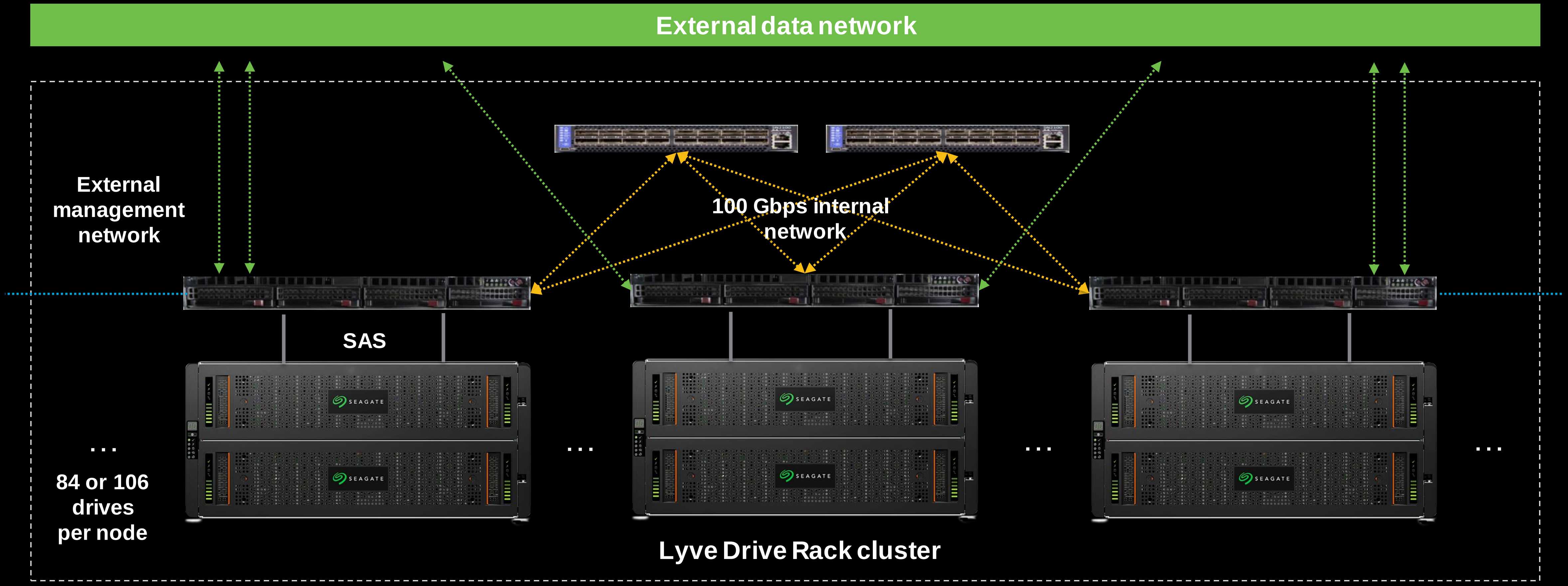


Hierarchical Erasure Coding

Most disk failures are handled within the enclosure
Fastest rebuild even with the largest drives



Lyve Rack R2 Reference Architecture


External data network
External management network
100 Gbps internal network
SAS
. . .
84 or 106 drives per node
. . .
. . .
. . .
Lyve Drive Rack cluster

DATA IS IN OUR DNA

Questions

SEAGATE