Software Defined Storage is the „New Normal“

Robert Grosschopff
Systems Engineer
robert.grosschopff@suse.com

Agenda

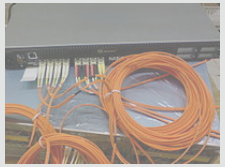
- Trends im Storage Markt
- SUSE Enterprise Storage
- Features
- Anwendungsfälle
- Konfiguration und Design
- Zusammenfassung

Trends im Storage Markt

Die Storage Evolution…

SCSI

1979



Fibre Channel

1994

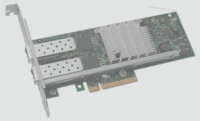
NFS
NETWORK FILE SYSTEM

NFS V3

1995 1996

CIFS

2002



iSCSI

…

ceph

CEPH (stable)

2012

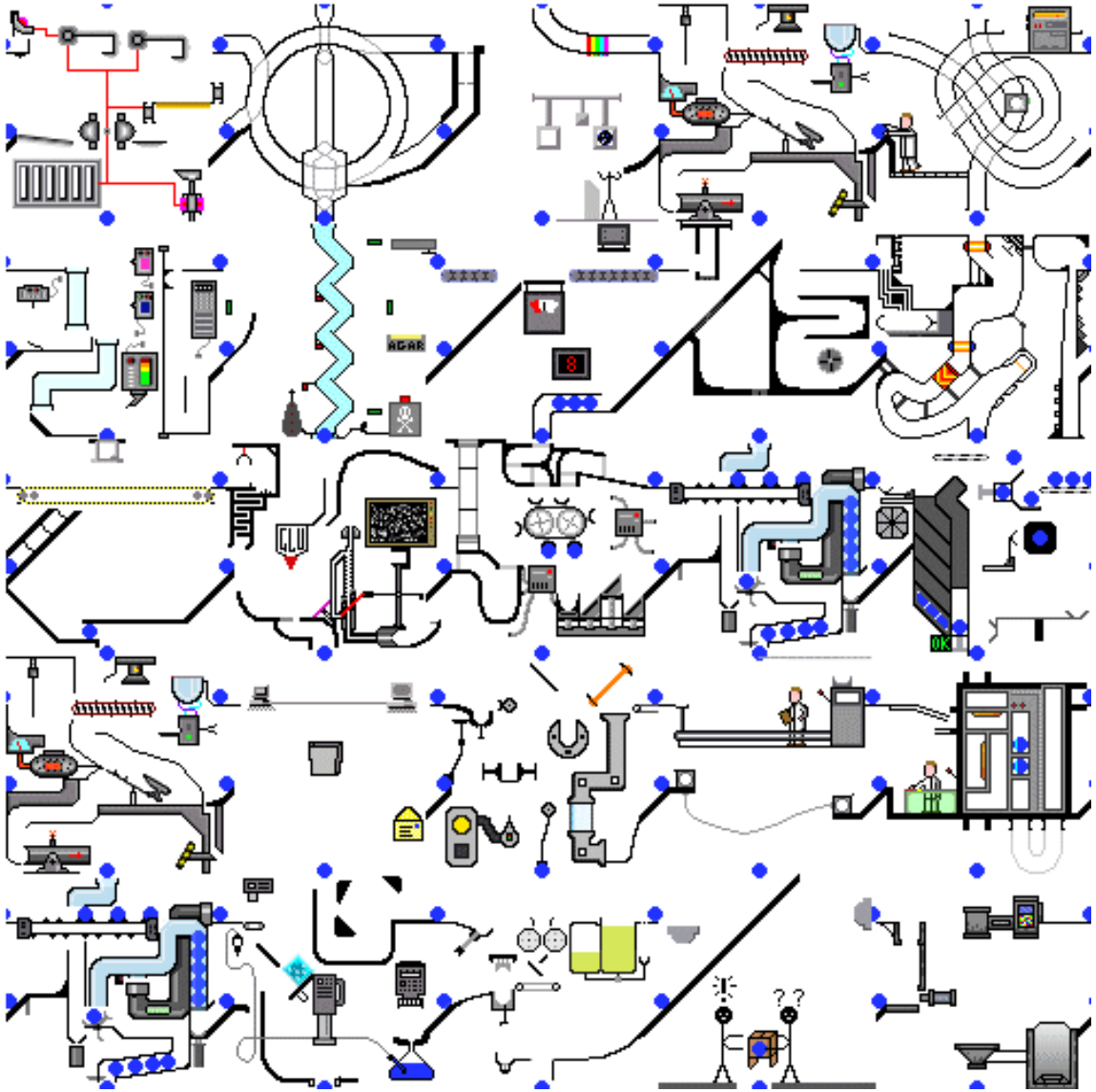
Die Storage Evolution…

Dienste im Datacenter früher…

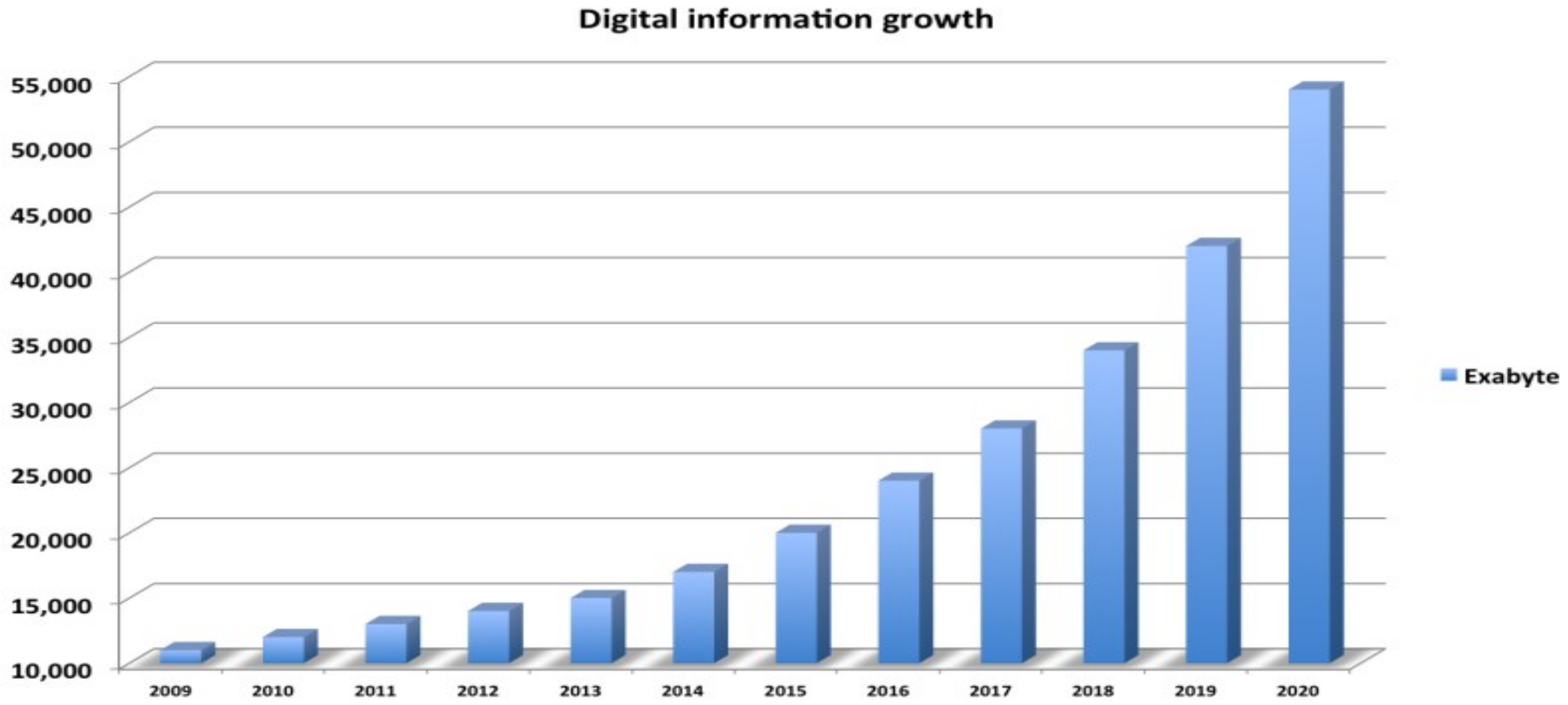
Die Storage Evolution…

Dienste im Datacenter heute…

Seafile
ShareFile
ownCloud
Nextcloud
syncplicity
Microfocus Filr
hadoop
MAPR
splunk>enterprise
d.velop
EASY SOFTWARE
WIR DENKEN SIE VORAUS
OPENTEXT
COMMVAULT
SAP
openstack
CLOUD SOFTWARE
docker
INTERNET OF THINGS
Industrie 4.0
Dev
Ops
MariaDB
MySQL
ORACLE



Und dazu das Datenwachstum…

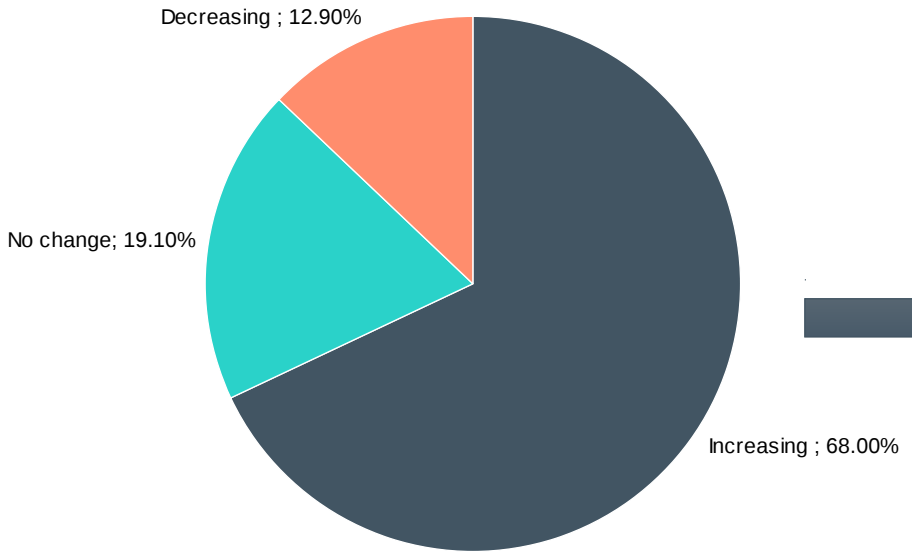


Quelle: IDC / http://www.tecchannel.de/a/software-defined-storage-wird-zum-muss,2067830

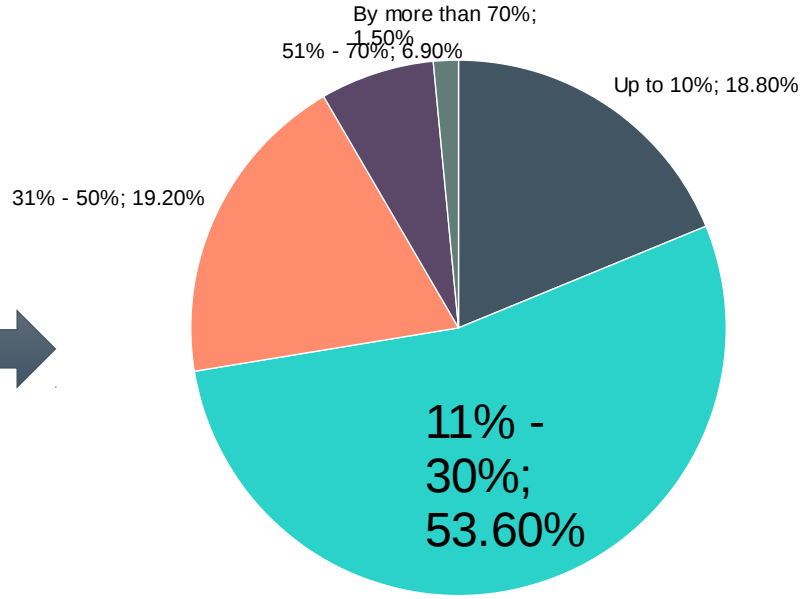
IDC
Analyze the Future

SUSE Storage Umfrage 2016 - Datenwachstum

Datenwachstum - Ja/Nein



Wachstum in %



Das durchschnittliche Datenwachstum in DACH beträgt **27%** in 2017

*1202 senior IT decision makers across 11 countries completed an online survey in July / August 2016

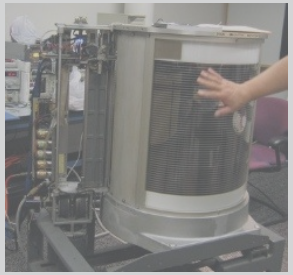
Können traditionelle Systeme die Antwort sein?

Keine nahtlose Skalierung

-

daher nicht zukunftssicher

Zu teuer



Nicht Cloud fähig

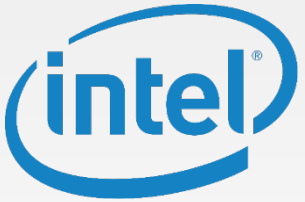
SUSE Enterprise Storage

SUSE Enterprise Storage

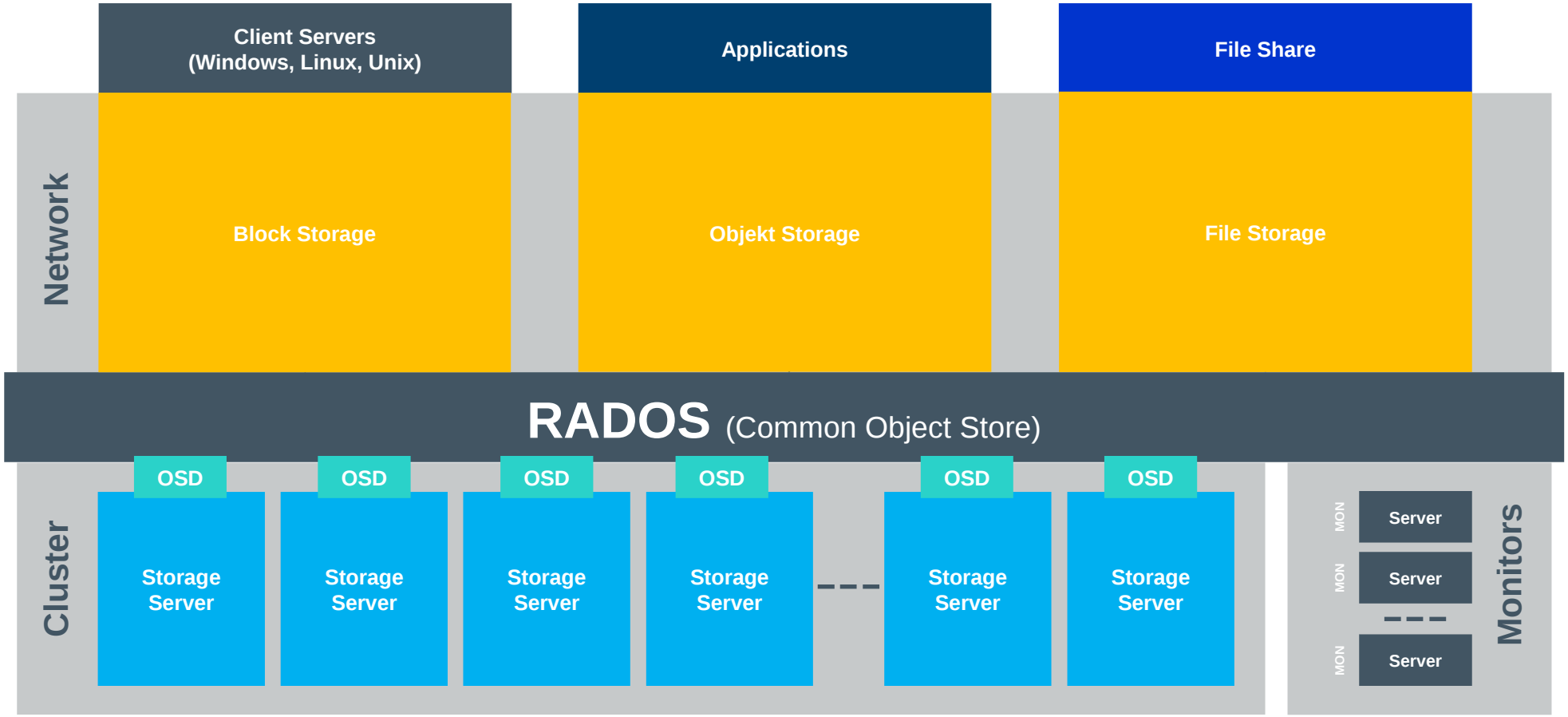
Eine **hochskalierbare, softwarebasierende** Storagelösung, die Unternehmen den Aufbau einer **kosteneffektiven** Speicherplattform, basierend auf **Standard Serverhardware** ermöglicht und zugleich **alle Enterprise Funktionen** unterstützt, die Kunden von einer derartigen Lösung erwarten.

Advisory Board

ceph



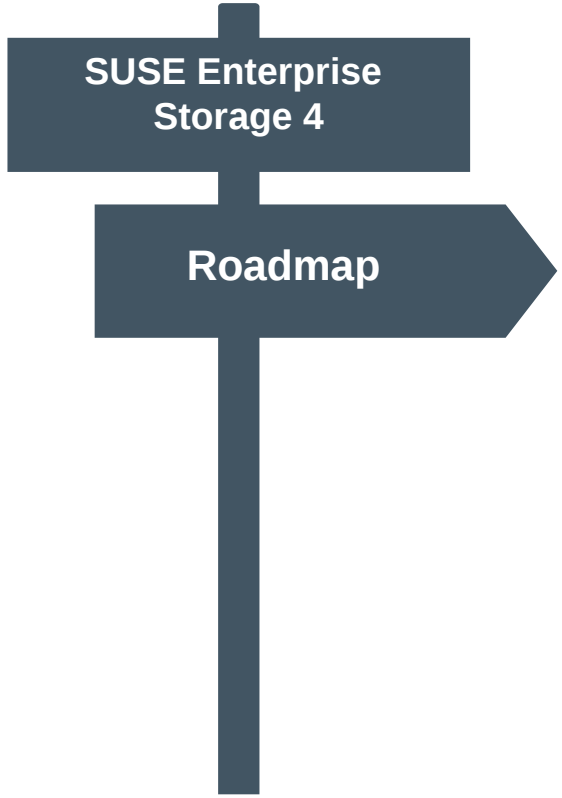
SUSE Enterprise Storage - Architektur



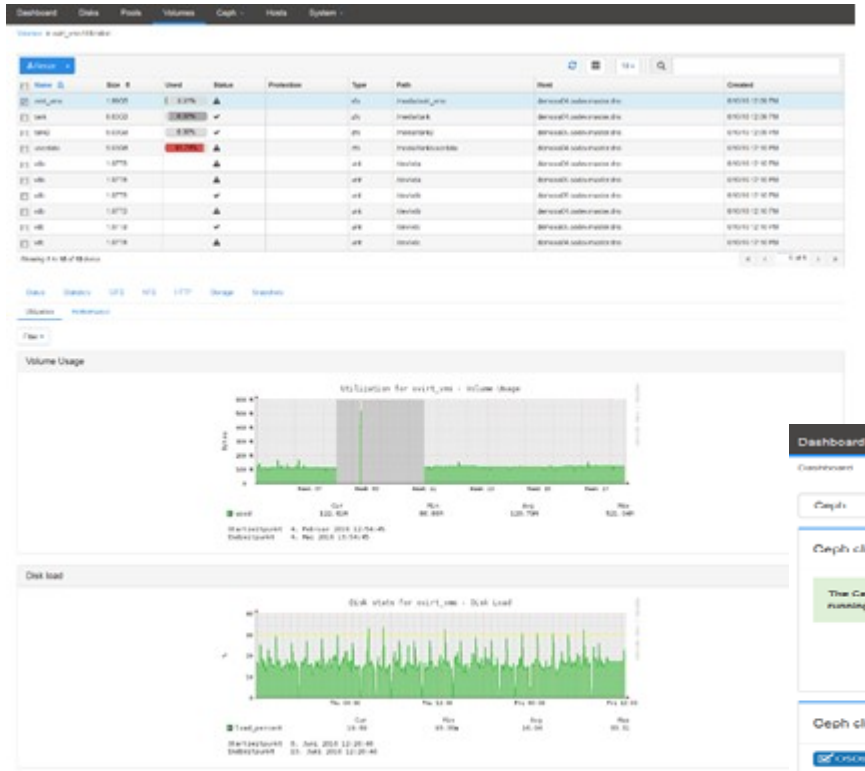
Features

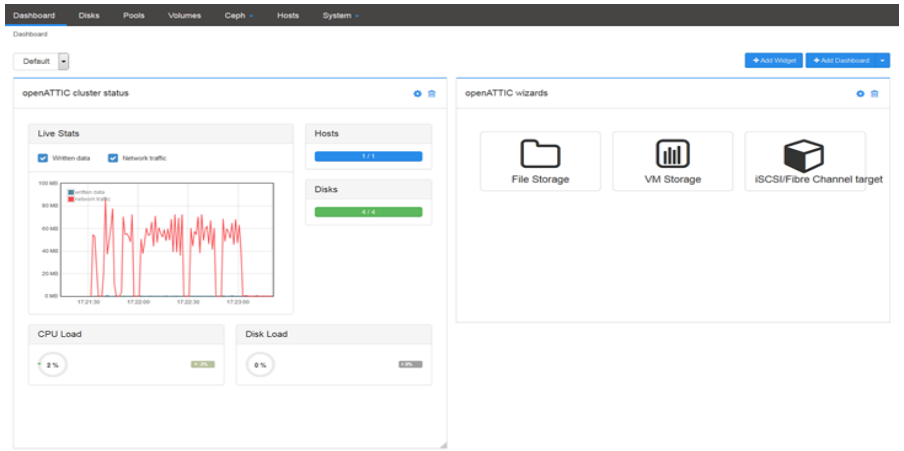
Überblick Technik

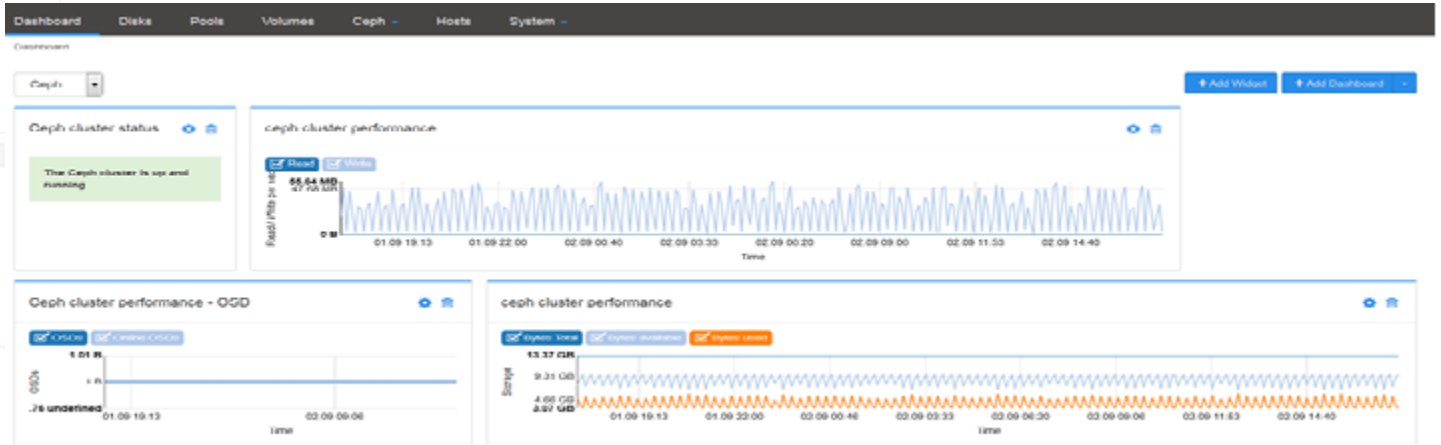
- Populärste OpenStack Distributed Storage Lösung
- Hochskalierbar von TB bis Brontobyte
- Enterprise Storage Funktionalitäten
  - No Single Point of Failure
  - Synchroner und Asynchroner Spiegel
  - “Erasure Coding” (effizientes RAID)
  - Cache Tiering
  - Unified Block, Object Interface, Filesystem
  - Thin Provisioning
- Aufgebaut auf Cluster-Servern
  - Self Healing Technology
  - Self Managing Technology

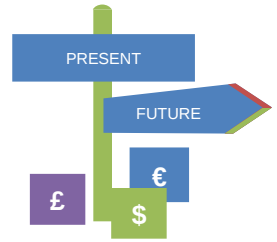


SUSE Enterprise Storage – Neues Management

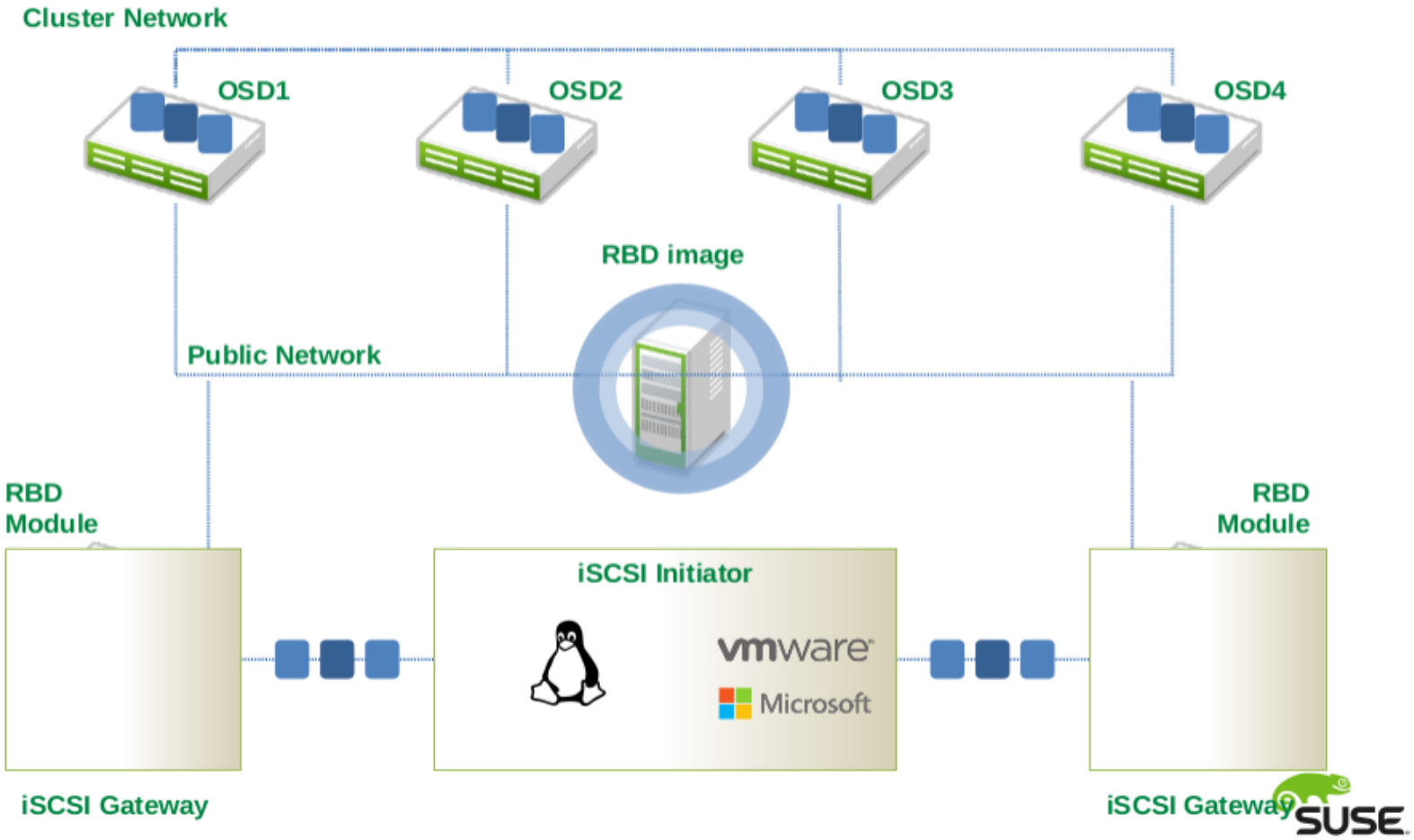








Heterogener OS Support per iSCSI



Cluster Lifecycle Management: DeepSea

- Basiert auf Salt
- Open Source
- Aufsetzen / Modifizieren / Upgraden des Clusters

Cluster Lifecycle Management: DeepSea

- Stage 0: Preparation
  - Kernel Updates, Einbinden von Repositories, …

- Stage 1: Discovery
  - Hardware Discovery
  - Erzeugen von Konfigurationsdateien für Salt
    - Cluster, Rollen, ...

Cluster Lifecycle Management: DeepSea

- Aufsetzen der policy.cfg
  - Spiegelt die Topologie des Clusters wieder
  - Zentrale Konfigurationsdatei

- Stage 2: Configuration
  - Ausrollen der Konfiguration auf die Minions

- Stage 3: Deployment
  - Installation der Ceph-Pakete, Starten von Ceph
  - Aufsetzen der MONs und OSDs
  - Erzeugt Default Pools

Cluster Lifecycle Management: DeepSea

- Stage 3: Deployment
  - Installation der Ceph-Pakete, Starten von Ceph

- Stage 4: Services
  - Starten zusätzlicher Services z.B. MDS, RGW, iSCSI Gateway, NFS Ganesha

- Stage 5: Removal
  - Entfernen auskonfigurierter Komponenten

Heterogene Zugriffe

- Technical Preview in SES4
  - NFS-Zugriff
    - S3 Buckets
    - CephFS
  - Production Ready
    - CephFS
    - S3 Buckets
    - Block Devices
    - iSCSI

Anwendungsfälle

Anwendungsfälle

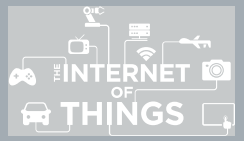
Filesync & Share

Archivierung

Videoüberwachung

Dev Ops

Backup2Disk



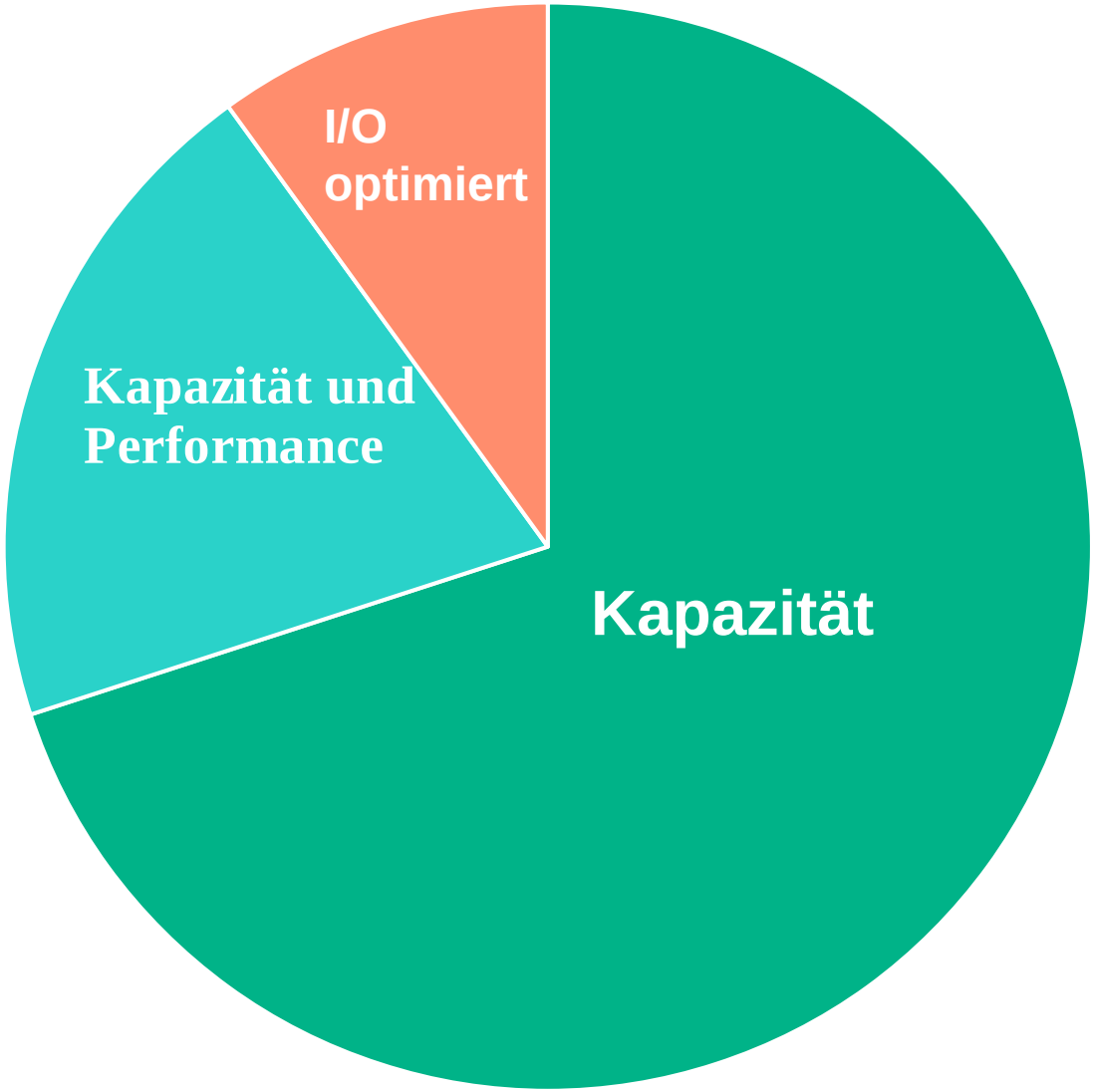
Industrie 4.0

Big Data

Cloud Storage

Ihre Anwendung?

IDC - Daten nach Storageklassen



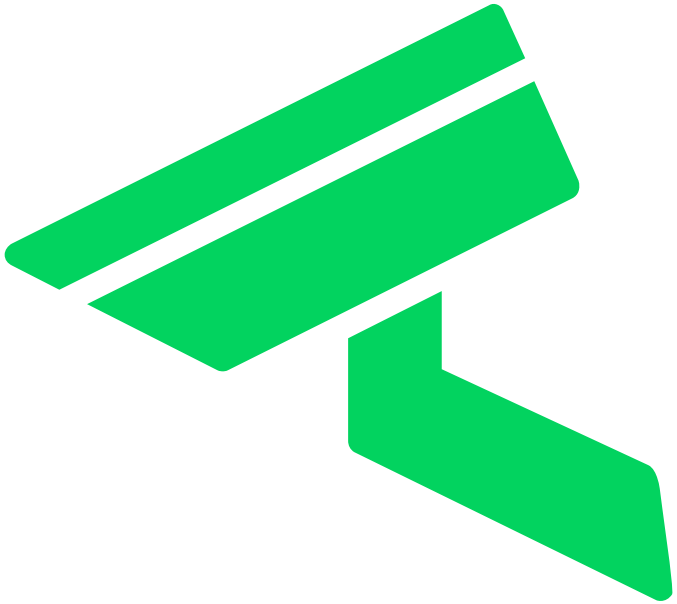
Source: IDC, 2013

Anwendungsbeispiel - Content Store

- Forschungsabteilungen
  - Meteorologische Daten, Weltraumüberwachung, Satellitendaten
  - IoT Sensordaten, Autonomes Fahren
  - HPC Daten
- Medienindustrie
  - TV, Radio
  - Webdienste (YouTube etc.)
  - Streaming Industrie (Netflix, Sky etc.)

Anwendungsbeispiel - Videoüberwachung

- Industriegebäudeüberwachung
- Verkehrsüberwachung
- Bodycams der Polizei
- Öffentliche Bereiche (Flughafen, Bahnhof)

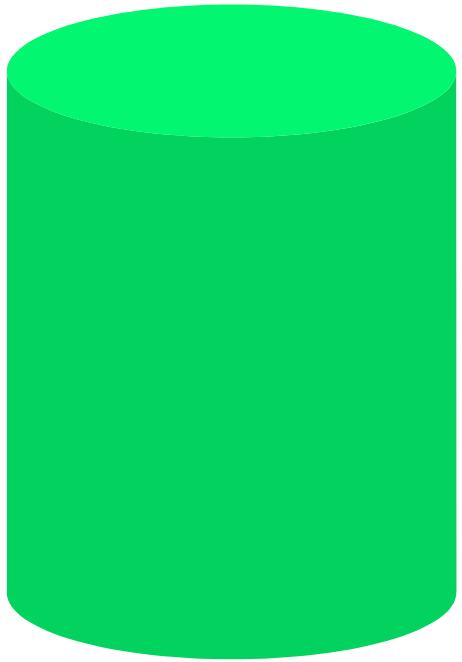


Anwendungsbeispiel – Objekt / Massenspeicher

- Daten, die konstant zunehmen
- Archivierung
- Finanzdaten, Dokumente,
- Medizinische Daten
- Big Data

Anwendungsbeispiel - Virtual Machine Speicher

- OpenStack (80% Marktanteil)
- Test-Dev VMs, “normale VMs”
  - Support via:
    - kvm – natives Blockdevice
    - Hyper-V – iSCSI
    - VMware - iSCSI



Radio Telescopes

Australian Square Kilometre Array Pathfinder

- 36 12m Parabolantennen, 188 Empfänger / 36 Beams pro Antenne
- Datenvolumen: 2 Tbps pro Antenne, 72 Tbps in Summe

280 EB pro Jahr (ca. 2.5x globaler Internetverkehr)

Daten werden deshalb vor dem Abspeichern verarbeitet

- 345 PB pro Jahr

Pictures

Massives Datenvolumen

Bilder werden bei unterschiedlichen Wellenlängen aufgenommen

- Sichtbares Licht
- Infrarot
- Radioemissionen

Daten von unterschiedlichen Quellen werden miteinander verknüpft

"hundreds of thousands of image files"

Image Manipulation

Image manipulation (FITS Files)

- 36 Mpixel Kamera: FITS file size ~436 MB (color frame)

Ein einzelnes Bild reicht nicht (Light Frames) !

- Viele Bilder müssen “gestacked” werden um das Rauschen zu vermindern und Details herauszuarbeiten
- Alle Bilder müssen registriert werden (Sterne bewegen sich ohne Nachführung scheinbar)

Alle Bilder müssen kalibriert werden

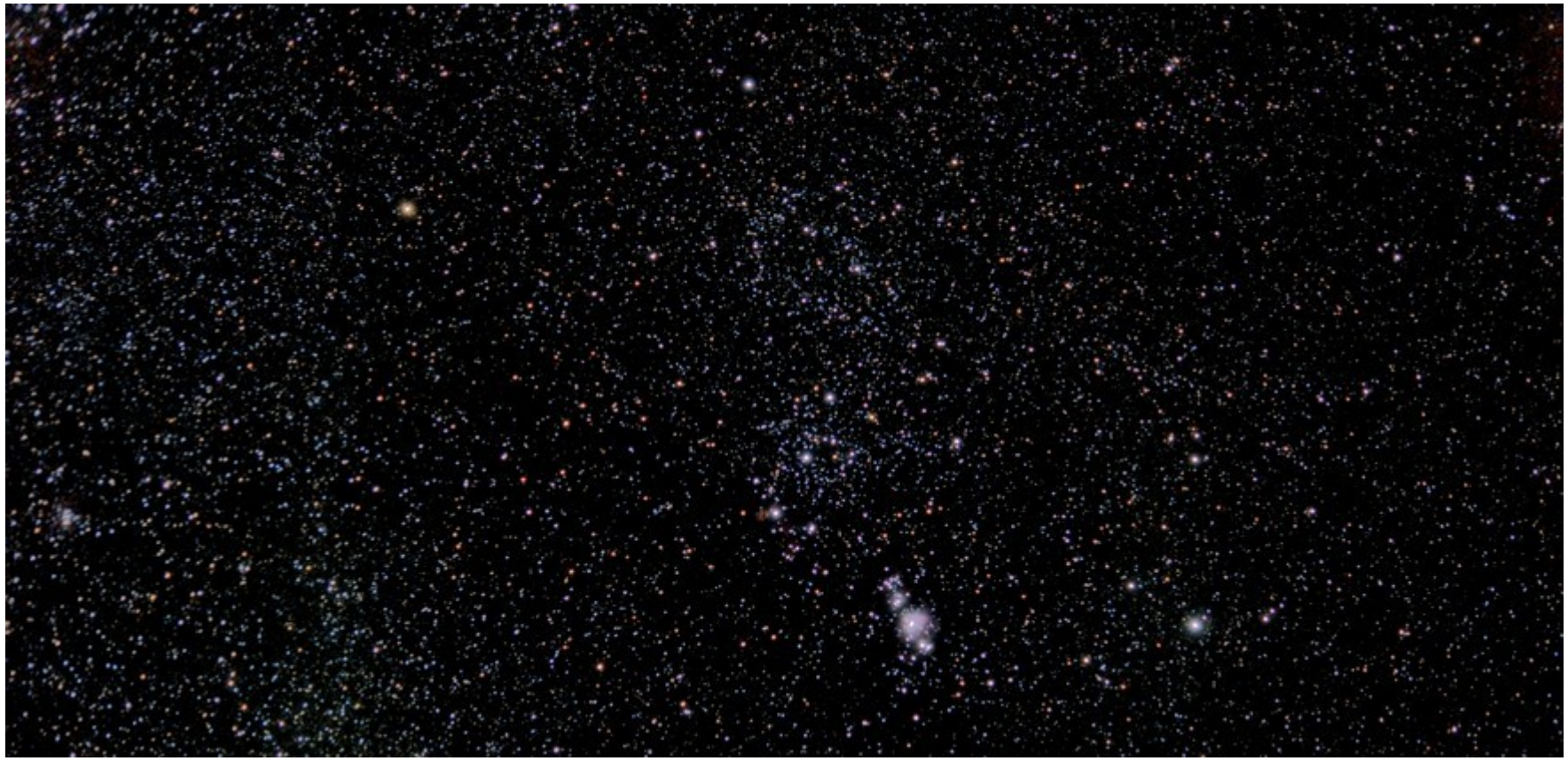
- Flat Frames (entfernt Vignettierung, Staubpartikel, etc.)
- Bias Frames (Ausleserauschen der Kamera)
- Dark Frames (Sensorrauschen)

What you got

Sehr viele davon ...

What you want

Quick’n dirty processing ...

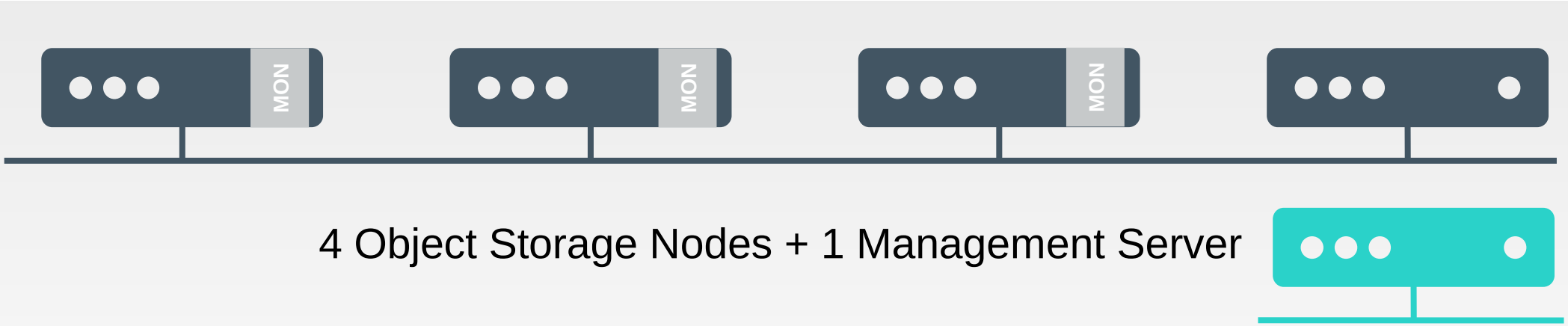


Konfiguration und Design

Konfigurationsmöglichkeiten

- Anpassbar an unterschiedliche Szenarien:
  - Performanz
  - Preis
  - Recovery
- Konfigurierbare Redundanzebenen
  - Disk
  - Knoten
  - Rack
  - Serverraum
  - …
  - Data Center

SUSE Enterprise Storage Minimalkonfiguration (1)



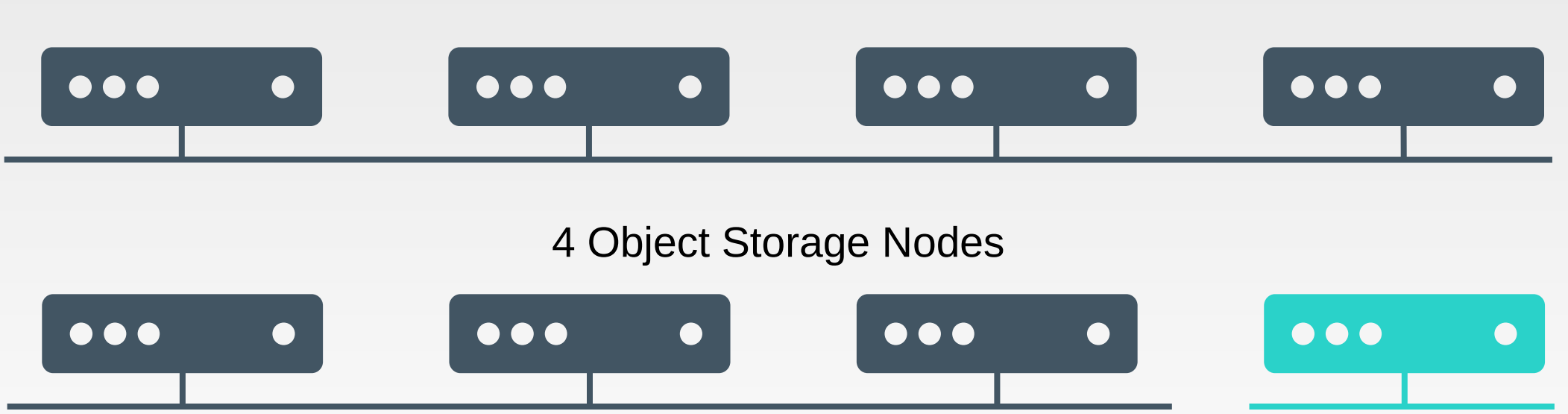
– 4 x SUSE Enterprise Storage Object Storage Nodes mit:

- Minimum 2 x 10 Gb Ethernet (1 x Storage Backbone, 1 x Client Network)
- Minimum 32 OSD HDDs im Cluster
- Dedizierte HDD für Betriebssystem
- Minimum 1 GB RAM pro TB Bruttokapazität pro Storage Node
- Minimum 1.5 GHz pro OSD pro Storage Node

– Separater Management Server

- Minimum 32 GB RAM, Minimum 4 Cores, 2 HDDs (SSD) für Betriebssystem

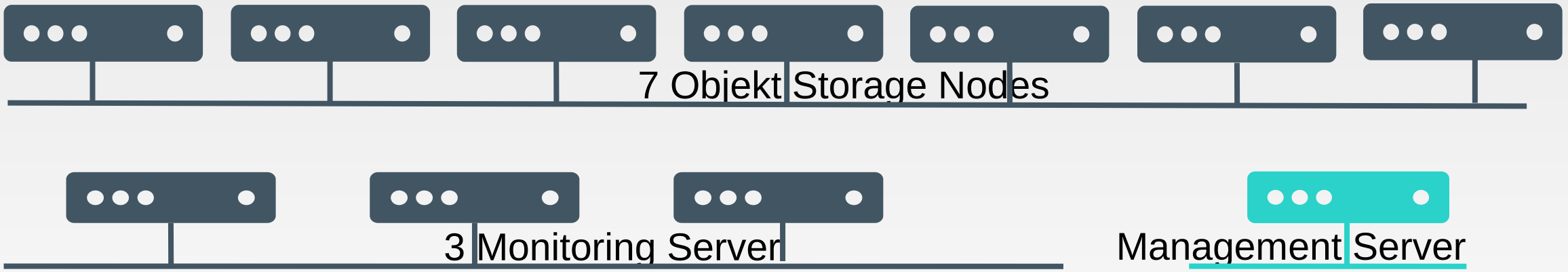
SUSE Enterprise Storage Minimalkonfiguration (2)



– Wie Konfiguration 1 zzgl. dedizierte Monitoring Server mit:

- Minimum 2 x 10 Gb Ethernet (1 x Storage Backbone, 1 x Client Network)
- Minimum 32 GB RAM
- 2 CPUs mit minimum 4 Cores
- Dedizierte HDD für Betriebssystem

SUSE Enterprise Storage Produktivumgebung



7 SES Storage Nodes

- 4 x 10 Gb Ethernet
- 56+ OSDs im Storage Cluster
- 2 SSD im RAID 1 für Betriebssystem
- SSDs für Journal (6:1 Ratio SSD Journal zu SATAs pro Node)
- 1.5 GB RAM pro TB Rohkapazität pro Storage Node
- 2 GHz pro OSD pro Storage Node

Infrastruktur Nodes:

- **Monitoring Nodes:**
  - Minimum 2 x 10 Gb Ethernet (1 x Storage Backbone, 1 x Client Network)
  - Minimum 32 GB RAM
  - 2 CPUs mit minimum 4 Cores
  - Dedizierte HDD für Betriebssystem
- **Management Node:**
  - Minimum 32 GB RAM, Minimum 4 Cores, 2 HDDs (SSD) für Betriebssystem
- **Gateway oder Metadaten Node:**
  - SES Object Gateway Nodes; 32 GB RAM, 8 core processor, RAID 1 SSDs for disk
  - SES iSCSI gateway nodes 16 GB RAM, 4 core processor, RAID 1 SSDs for disk
  - SES metadata server nodes (one active/one hot standby); 32 GB RAM, 8 core processor, RAID 1 SSDs for disk

https://www.suse.com/documentation/ses3/book_storage_admin/data/cha_ceph_sysreq.html

Zusammenfassung

Change Your Mindset

- Be prepared to lose control of knowing where exactly specific data/files are stored and distributed → Ceph takes care about this
- Build trust in Ceph to manage data distribution correctly
- Ceph uses several **new** ways to access your data
- Rethink your storage usage and carefully create Ceph Pools, Placement Groups, CRUSH Map, etc.

We adapt. You succeed.