How Software Defined Storage Can Help To Solve Retail Industry Challenges

Luis Rico - Senior Storage Specialist Solution Architect, EMEA
Jorge Tudela - Senior Cloud Consultant, Iberia
José Ángel de Bustos - Senior Cloud Consultant, Iberia

07/May/2019

AGENDA

AGENDA

- Retail Industry Challenges
- Red Hat Ceph Storage Product Overview
- Object Storage Use Case for E-Commerce Platform
- Fastest Red Hat Ceph Object Storage
- Conclusions
- Q&A

RETAIL INDUSTRY CHALLENGES

Retail Industry Challenges

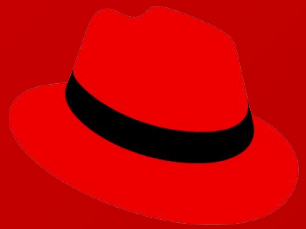
- Availability
  - Global Geo-Availability 24x7
- Scalability
  - Aggressive growth year by year
  - ~2x growth or more for many companies
- Performance
  - Support high traffic peaks during specific dates
    - Christmas sales, Black Friday, etc
  - E-commerce store sensitive to web experience (lag, delays, etc)
    - Requires fastest R/W operations
- Data location is important
  - To comply with National legislations:
    - In some countries, bills have to be stored physically inside the country

Retail Industry Challenges

- Data retention
  - Retail companies must comply with local National legislations
  - Data retention policies are dictated by National legislations
  - For example, European GDPR Article 5:
    - "Personal data shall be kept in a form which permits identification of data subjects for **no longer than is necessary**"

- High SLAs, specially e-commerce platform
  - Small service outages are worth $$. E-commerce cost of downtime:
    - amazon.com revenue loss per minute **$220,318.80**
    - walmart.com revenue loss per minute **$40,771.20**
    - nike.com revenue loss per minute **$5,685.60**

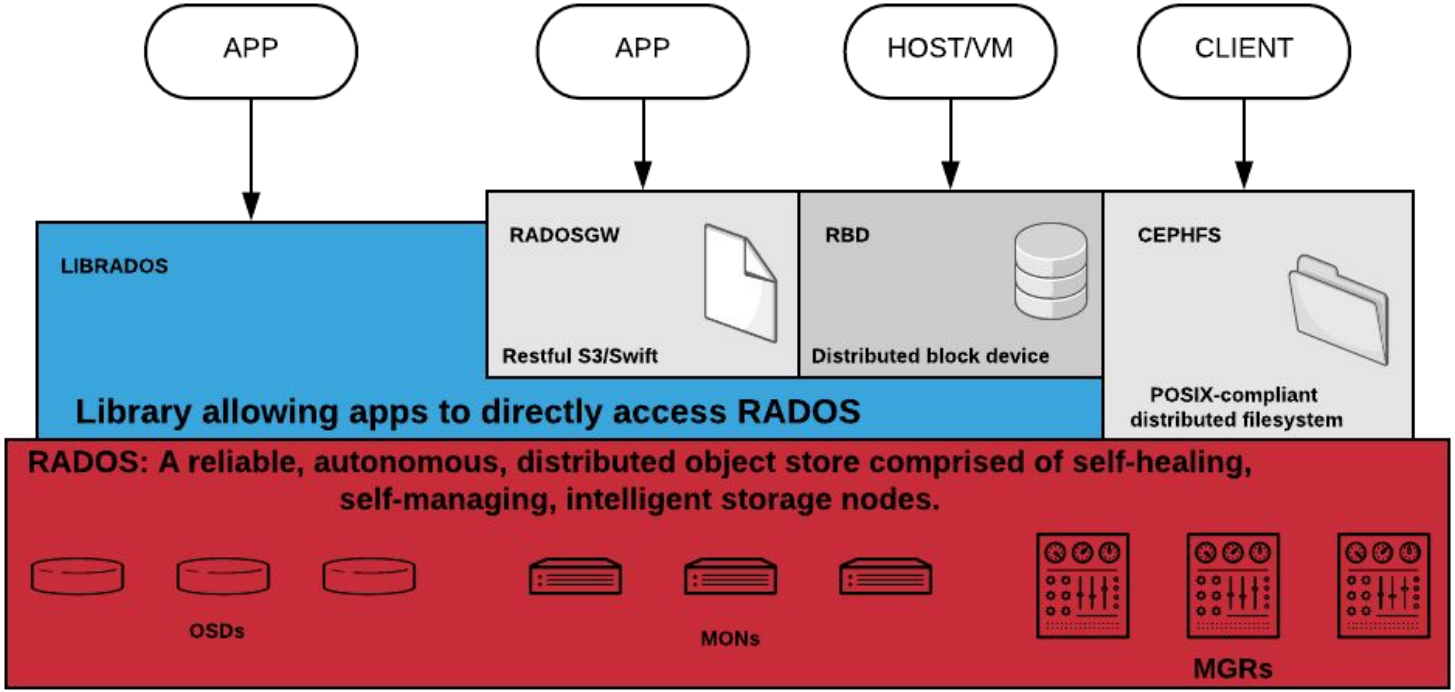
* Source: https://www.gremlin.com/ecommerce-cost-of-downtime/



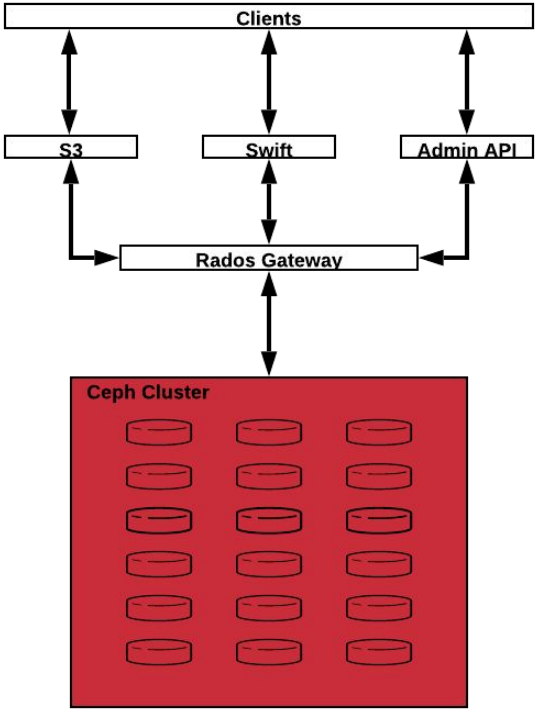
Red Hat Ceph Storage

PRODUCT OVERVIEW

Red Hat Ceph Architecture

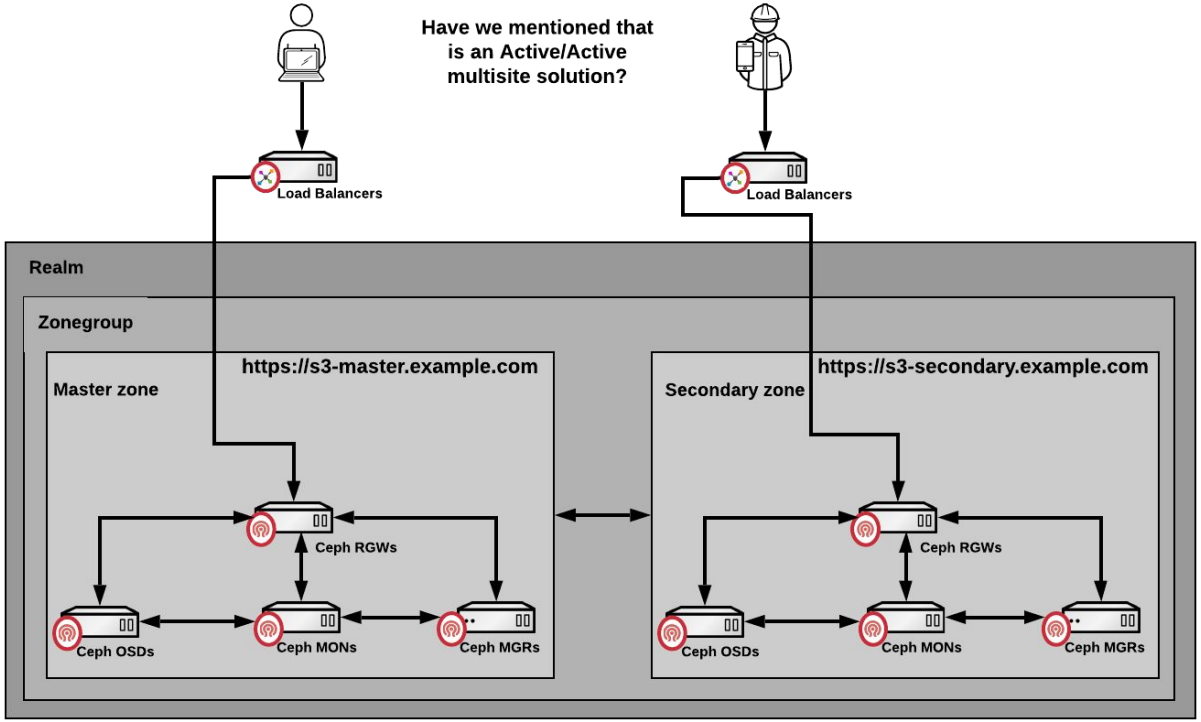


Red Hat Ceph Object Architecture



- S3 like API & Swift API
- Objects are stored in buckets
- Bucket index can be sharded into multiple parts for better performance

Red Hat Ceph Active/Active Multi-site Architecture



OBJECT STORAGE USE CASE FOR E-COMMERCE PLATFORMS

CUSTOMER REQUIREMENTS

Customer's requirements

- **Store bills in PDF issued by e-commerce store worldwide**
- Some numbers:
  - ~80,000,000 bills per year
  - ~64Kb PDF size
  - ~200K request during first sales hour
    - peaks of ~6000 purchases per minute
  - ~15.000.000 bills during Black Friday
- ~2x growth year by year!!!
- In some countries, e-bills have to be stored there
- High Availability and Disaster Recovery

- Currently stored in traditional NAS not able to geo-scale

WHY CEPH FOR THE RETAIL INDUSTRY?

Why Red Hat Ceph Storage was chosen? (I)

- Bills are stored in unique PDFs
  - PDF is an object --> Object Storage
- Ceph can scale to many millions of objects
- Easily and massively scalable:
  - Scale out process is simple
  - From one disk or one server with disks
- Flexibility and freedom to customize commodity HW
  - Freedom to choose any x86 hardware vendor
  - Disk technology to satisfy performance
- Open Source vs Proprietary

Why Red Hat Ceph Storage was chosen? (II)

- Highly Available
  - Distributed architecture
  - No SPoF
- Easy maintenance
  - No outages when upgrading & operating
- Data durability via erasure coding or replication
- Able to meet performance requirements
  - Scale out
  - Customized architecture: CPU, RAM, disks, networking

Why Red Hat Ceph Storage was chosen? (III)

- Object Storage Rest API compatible with Amazon S3 API
  - Based on the de-facto industry standard-proprietary API (S3)
  - Commonly used with any object storage: No vendor lock-in
- Ceph Multi-site architecture
  - Complies with Geo-distribution of bills
  - Business continuity + Disaster Recovery
- Successful PoC that demonstrates the features!!

- Competition:
  - EMC Elastic Cloud Storage (ECS)

Red Hat
Consulting

FASTEST CEPH OBJECT STORAGE

ARCHITECTURE

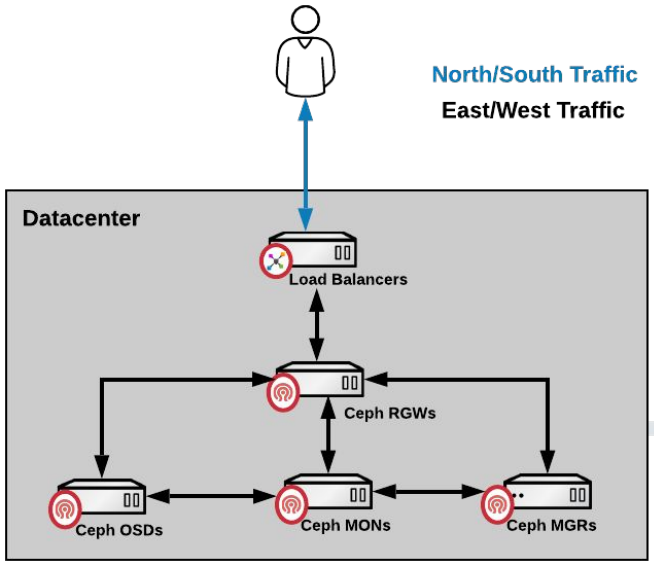
Customer Architecture

- Why is this solution unique?
  - Red Hat Ceph Storage 3.0
  - Full flash NVMe disks
  - No SPoF
  - Active/Active Multi-site replication between 2 DCs
  - Collocated & Containerized Ceph daemons MONs, OSDs & RGWs
  - RGWs perform both tasks, attend **customer requests** and **data replication**.

- Two Ceph production clusters, each cluster:
  - 4 servers for storage. 10 NVMe per server. 40 NVMe disks per cluster.
  - 3 servers for MONs/RGWs.

Customer Architecture

- App traffic (North/South)
  - F5 LBs layer to load balance RGWs
  - Expose RGWs APIs (S3) to the Apps
- Ceph cluster replication traffic (East/West)
  - RGWs inter DC sync is point to point, no LB involved
  - RGWs communicate to each other across DCs



IMPLEMENTATION DETAILS

Implementation Details

- First worldwide deployment of its kind:
  - Full flash NVMe
  - Object Storage Multisite Active-Active Architecture
  - Containerized Ceph Services
  - Red Hat Ceph Storage 3.0!!
    - Was release 3 months ago ;)

- Strong collaboration
  - Customer
  - Red Hat Ceph Engineering
  - Red Hat Ceph Support
  - Red Hat Storage Business Unit
  - Red Hat Consulting

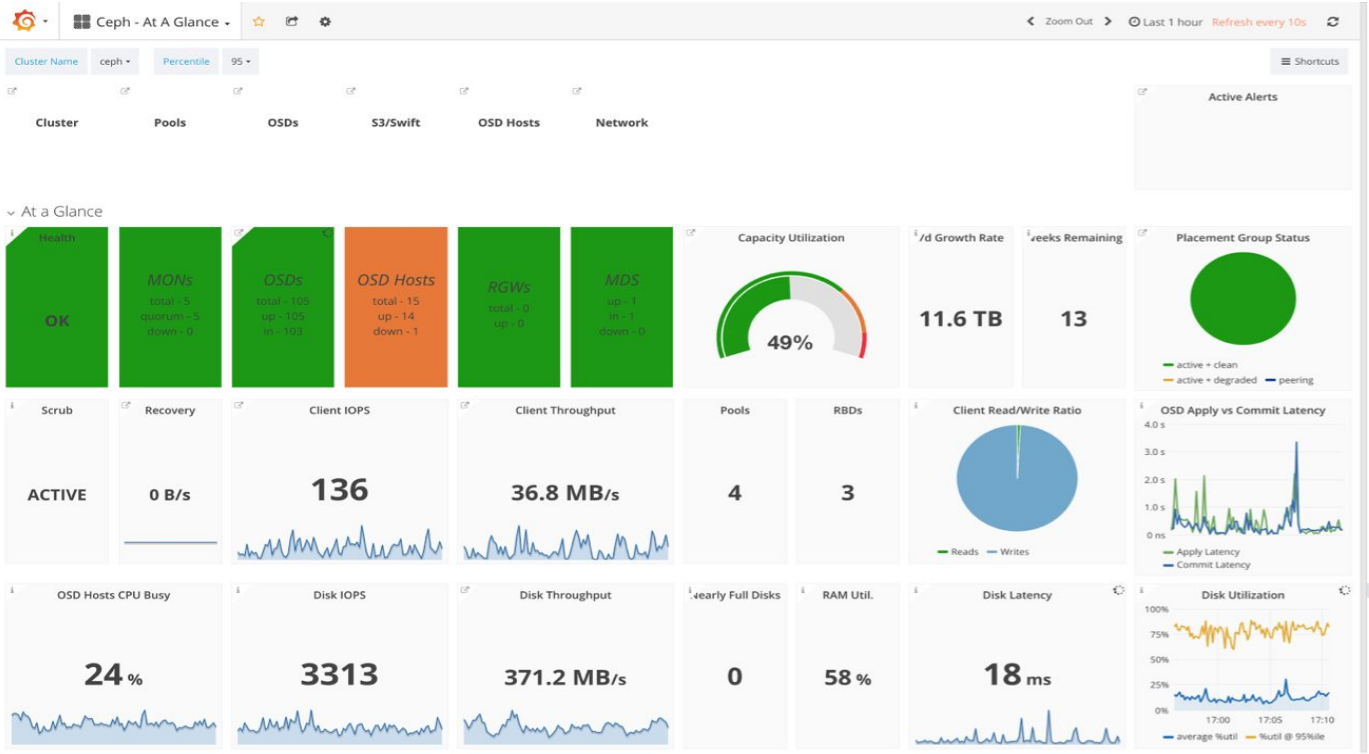
Implementation Details

- Containerized installation using ceph-ansible tool
  - Supported, easy and fast
- 2nd day Operations performed with ceph-ansible
  - Upgrades, add & remove disks, etc
- Ceph daemons running in containers
  - Installed just a few packages
  - New version of Ceph -> New container image
  - Ceph operations have to be done inside the container!

Implementation Details

- Ceph metrics, visually monitors various metrics in a Ceph cluster
  - Comes with Ceph Ansible installer
  - Real time monitoring tool!!!
  - Very easy to install
- Key to visualize and analyze benchmark results
  - Gathers many key metrics: I/O, Network, latency, etc.
- Before Ceph metrics, monitoring a Ceph cluster was a DIY effort.

Implementation Details



BENCHMARKING

Customer's requirements

- **Store bills in PDF issued by e-commerce store worldwide**
- Some numbers:
  - ~80,000,000 bills/objects per year
  - ~64Kb Object size
  - ~200K request during first sales hour
    - peaks of ~6000 purchases per minute
  - ~15.000.000 bills during Black Friday
- ~2x growth year by year!!!

Benchmarking - Single Cluster

- CosBench tool to "try" to stress the cluster
  - A benchmark tool for cloud object storage service
- We really could NOT stress the disks/RGWs nor disks ;)
  - We did many tests
  - We saturated the network
  - We saturated the CosBench nodes
- 88.000.000 objects(64k) digested in the cluster, in 11 hours!!!!
  - Only to one cluster, no replication active yet.
  - Customer requirements exceeded in the first test, with no tuning!!
- Cluster filled with 240.000.000 objects (64K)
  - Close to cluster full capacity
  - Not performance degradation!

Benchmarking - Multisite Replication

- Simulated customers needs for Multisite replication benchmark
  - Filling the cluster with objects, with no cleanup
  - Last benchmarks performed with cluster full of objects
  - No performance degradation!!

- Performance test is executed from 4 external CosBench nodes, directly to the LBs.
  - All RGWs nodes as LBs backends.
  - RGWs perform both tasks, attend **customer requests** and **data replication**.

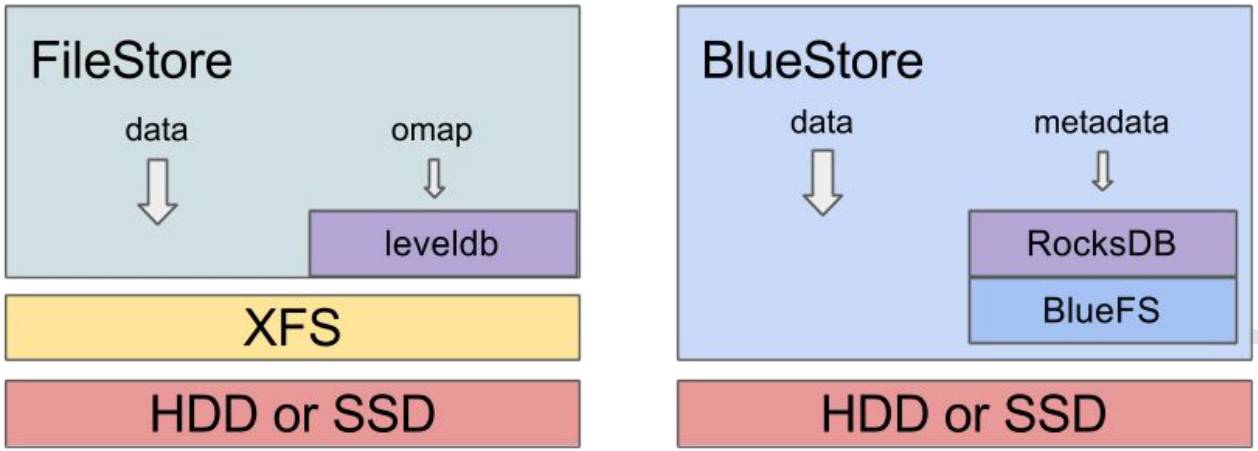
- CosBench execution time for 10M objects (64k):
  - 2 hours and 35 minutes (9300 seconds)

- Performance:
  - 10M requests / 9300 = **1075 requests/second!!!**

FUTURE IMPROVEMENTS

Migrate to BlueStore Backend

- Red Hat Ceph Storage 3.2 supports BlueStore
- BlueStore is a new Ceph Backend
  - Replaces current backend: Filestore



Migrate to BlueStore Backend

- Significant performance improvements for Block and Object.
- Already public benchmarks.
- 4M Objects - 100% writes
  - 88% increase in throughput
  - 47% decrease in average latency
- 4M Objects - 70% read / 30% write
  - 64% increase in throughput
  - 40% decrease in average latency

Source: https://ceph.com/planet/bluestore-unleashed/

Cold Backup Cluster

- Avoid malicious or accidental buckets/objects deletion.
  - Data is critical!!
  - Requirement to keep all objects (including history) in a separate area.
- Storing every object in a full flash NVMe cluster is expensive ;)
- So syncing objects to a cold backup cluster is the solution chosen.
- New archive zone feature coming in Nautilus!!!
  - **Archive zone federation enables full preservation of all objects (including history) in a separate zone (cluster).**
- Separate tool to restore objects from the cold backup cluster.

Conclusions

Conclusions

- Object storage is able to satisfy requirements traditional NAS storage is not capable to accomplish in the retail industry
- Red Hat Ceph Storage is an open, flexible and scalable object storage solution
- Hardware to run Ceph can be customized and adapted to fulfill any performance requirements
- Ceph multi-site architecture provides geographical async replication between clusters in active-active mode

Conclusions

- Ceph is flexible enough to accommodate other use cases in the future for this customer:
  - Store web images for online stores
  - Store millions of WhatsApp attachments for customer supporting returns and refunds

- New use cases for Ceph as storage solution beyond providing storage to OpenStack:
  - Persistent storage for OpenShift and Kubernetes with Rook.io
  - Data analytics and Shared Data Lake for Big Data through S3A
  - Massively scalable Object storage for IoT, Machine Learning and AI

Team members

- Sales team:
  - Mar Santos, Key Account Manager
  - Ramón Gordillo, Solution Architect
  - Luis Rico, Storage Specialist Solution Architect EMEA

- Red Hat Consulting team:
  - Mariola Ramos, Technical Project Manager
  - Daniel Domínguez, Cloud&Storage Architect
  - Jorge Tudela, Cloud&Storage Consultant
  - Maurizio Garcia, Cloud&Storage Consultant
  - José Ángel de Bustos, Cloud&Storage Consultant
  - Eric Goirand, Storage Architect EMEA

Q&A

RED HAT SUMMIT

THANK YOU


plus.google.com/+RedHat


facebook.com/redhatinc


linkedin.com/company/red-hat


twitter.com/redhat


youtube.com/user/RedHatVideos

Sources

- https://ceph.com/community/new-luminous-scalability/
- https://www.gremlin.com/ecommerce-cost-of-downtime/
- http://www.privacy-regulation.eu/en/article-5-principles-relating-to-processing-of-personal-data-GDPR.htm
- https://www.redhat.com/en/success-stories/?f%5B0%5D=taxonomy_product%3ARed+Hat+Ceph+Storage