Running a CEPH-Cluster from a containerized infrastructure

Use case: mySQL-database

Julius Neudecker
Bachelor of Science
julius.neudecker@haw-hamburg.de

January 2020

Contents

Setting up and operating a storage cluster with high availability is a complex task. By using containerization, it is possible to abstract away and encapsulate some repetitive tasks. Therefore this study aims to analyse the possibility, implications and findings of a multi host CEPH-Cluster, where the individual daemons are entirely running within a containerized environment. To put the findings into a frame of reference, this study utilizes a mySQL-database which has special requirements on data storage. The key points in terms of advantages and disadvantages, data integrity, performance and administion are scrutinized. Major findings are that ... [insert findings here later] ... . Therefore running a distributed CEPH-storage in a containerization environment is ... [draw conclusion] ... .

1 Introduction

In times where information is a valuable asset, it is of paramount importance to have a scalable and reliable way of storing data and information. Considerations about data throughput and IOPS[1] are also a major design parameter on modern storage solutions. These different storage solutions provide different approaches on these considerations. For any given use case, there exist several options to adress these. Depending on the architecture and scope of the problem some are better suited than others. A few major considerations are apart from scalability, reliability and speed also cost effectiveness, vendor lock-in, complexity and granular customizability.

In general hardware based solutions have advantages in terms of raw performance but they often have significant disadvantages when it comes to vendor lock-in, easy scalability or restoration of corrupted disks. Software and network based solutions are *in principle* less performant. However this can be mitigated for the most part by scaling up.

Apart from propretary cloud storage providers i.e. AWS, Azure or Google, the **free market** [**correct term?**] is heavily dominated by CEPH by more than twice as much as the next competitor:

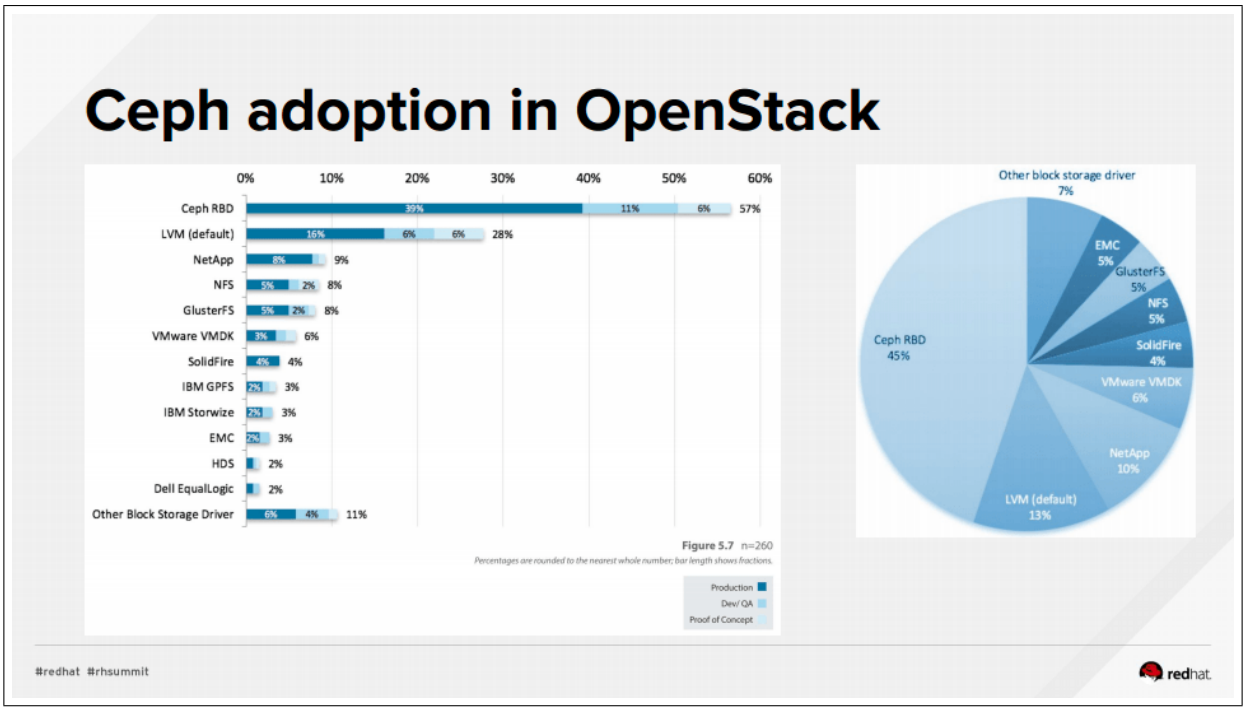

Figure 1: Adoption of CEPH in OpenStack in 2016, [**?**]

Being conceived by Sage Weil for his doctoral thesis [**?**], CEPH became part of the Linux Kernel in 2010 and was acquired by RedHat in 2014, CEPH is gaining popularity steadily since its introduction. **Source?**

Another modern important concept which is increasingly shaping modern technology stacks is *OS Level Virtualization* or *containerization* as its colloquially called. This way of deploying applications decreased the complexity, which is inherent to deploying several different applications to one single host machine. [**Since when?**] One integral property of these *containers* is that they are stateless and ephemeral. This means that they can *by principle* be created and deleted according to momentary requirements.

In this paper both technologies will be utilized to overcome some properties of these two technologies by merging them and thus using synergies [**really?**]

[1] Input/Output Operations per Second

5 Conclusion

5.1 Advantages

5.2 Disadvantages

5.3 Performance

5.4 In Summary