What is CEPH?

CEPH is an open source distributed storage system, designed to have no single point of failure, it can scale to an infinite number of nodes, and nodes are not coupled with each other, while traditional storage systems have some components shared between controllers (cache, disks)

CEPH is built to organize data automatically using an algorithm called CRUSH (Controlled Replication under Scalable Hashing) to ensure that data is evenly distributed across the cluster and that all cluster nodes are able to retrieve data quickly without any centralized bottlenecks and then uses the nodes of the cluster as the managers of those data. There is no shared component between servers, even if some roles like Monitors are created only on some servers, and accessed by all the nodes.

CEPH does not use technologies like RAID or Parity, redundancy is guaranteed using replication of the objects, any object in the cluster is replicated at least twice in two different places of the cluster. If a node fails, the cluster identifies the blocks that are left with only one copy, and creates a second copy somewhere else in the cluster. CEPH can be dynamically expanded or shrunk, by adding or removing nodes to the cluster, and letting the Crush algorithm rebalance objects.
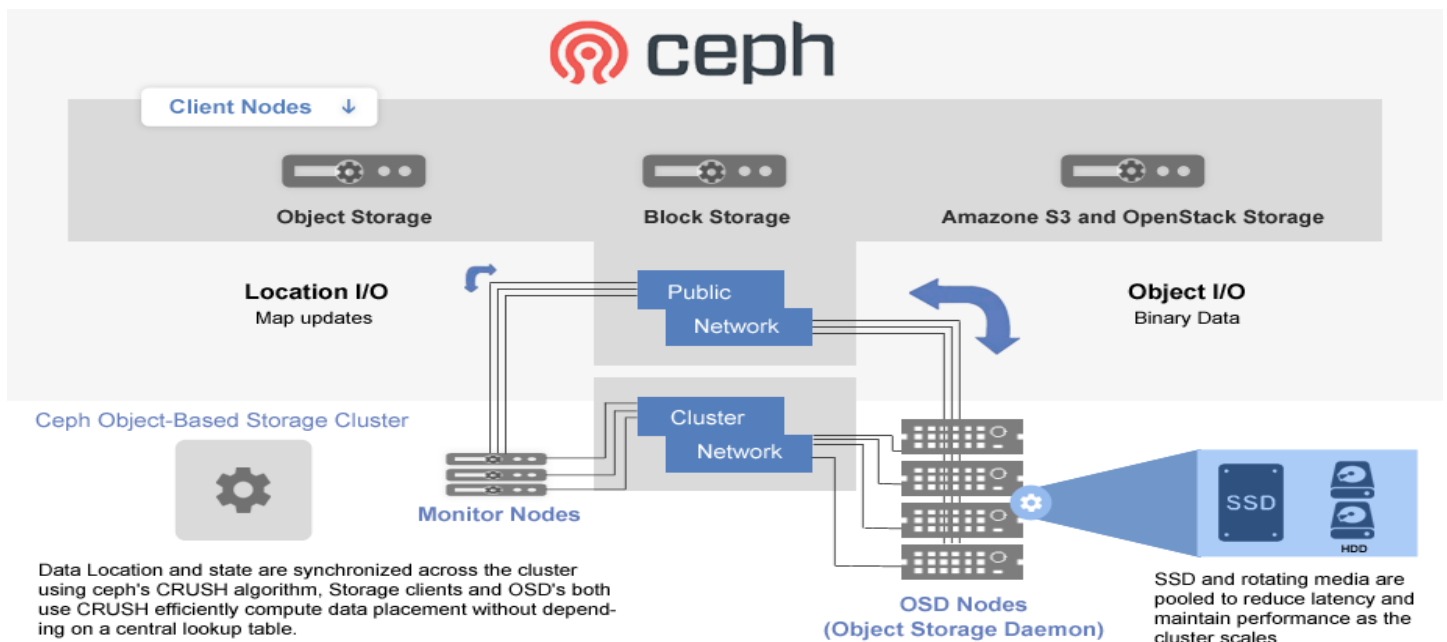
Each object typically includes the data itself, a variable amount of metadata, and a globally unique identifier. Each file entering the cluster is saved in one or more objects and the object is saved multiple times in the cluster. The process is reversed when data needs to be accessed. Object, block and file storage delivers high performance, high scalability and high availability.

CEPH object storage is accessible through Amazon Simple Storage Service (S3) and OpenStack Swift REST APIs and a native API for integration with software applications.

CEPH Block-storage makes use of a CEPH Block Device, which is a virtual disk that can be attached to bare-metal Linux-based servers or to virtual machines. The CEPH Reliable Autonomic Distributed Object Store (RADOS) provides block storage capabilities such as replication of data. The CEPH RADOS Block Device (RBD) is integrated to work as a back end with OpenStack Block Storage.