知识点阐述：

1. ceph架构

ceph是c/s架构的分布式存储系统。提供rbd、rgw、cephfs存储服务。对象寻址通过客户端分片生成oid，然后对oid做hash（+pool id）生成pg id，得到pg id后做crush计算得到一组osd集合，找到这个对象存放的位置。

Crush是一种伪随机hash算法。实现pg到osd的映射。Crush包含crush map和crush rule。Crush map是对存储集群的抽象，通过bucket代表设备，将集群设备组成一个查找树。最顶层是root作为入口，下层可以是room，rack、host等，也可以是自定义设备对象，最终叶子节点是osd，osd具有自己的权重。Crush rule定义了在crush map中查找的规则。

简单来说，crush描述根据生产环境设备树形结构查找osd过程。Crush采用straw抽签算法，算出一个随机数，然后再乘以osd的权重，得到结果。这样的话尽量选择权重大的设备。上层父节点的权重大小是叶子osd节点的权重和，所以当需要选出多个结果，则是根据choose选择的类型和个数。

Take root

Choose[3,host] 挑选3个节点

Choose[1,osd] 每个节点挑选一个osd

emit

1. bluestore理解