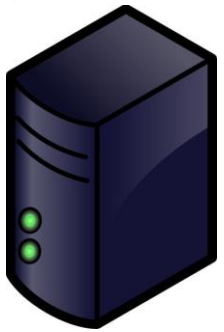
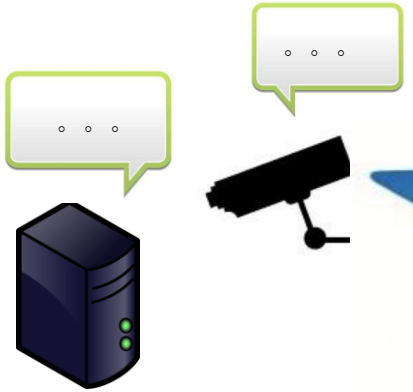
CRH 大数据监控 & 存储解决方案

郭威

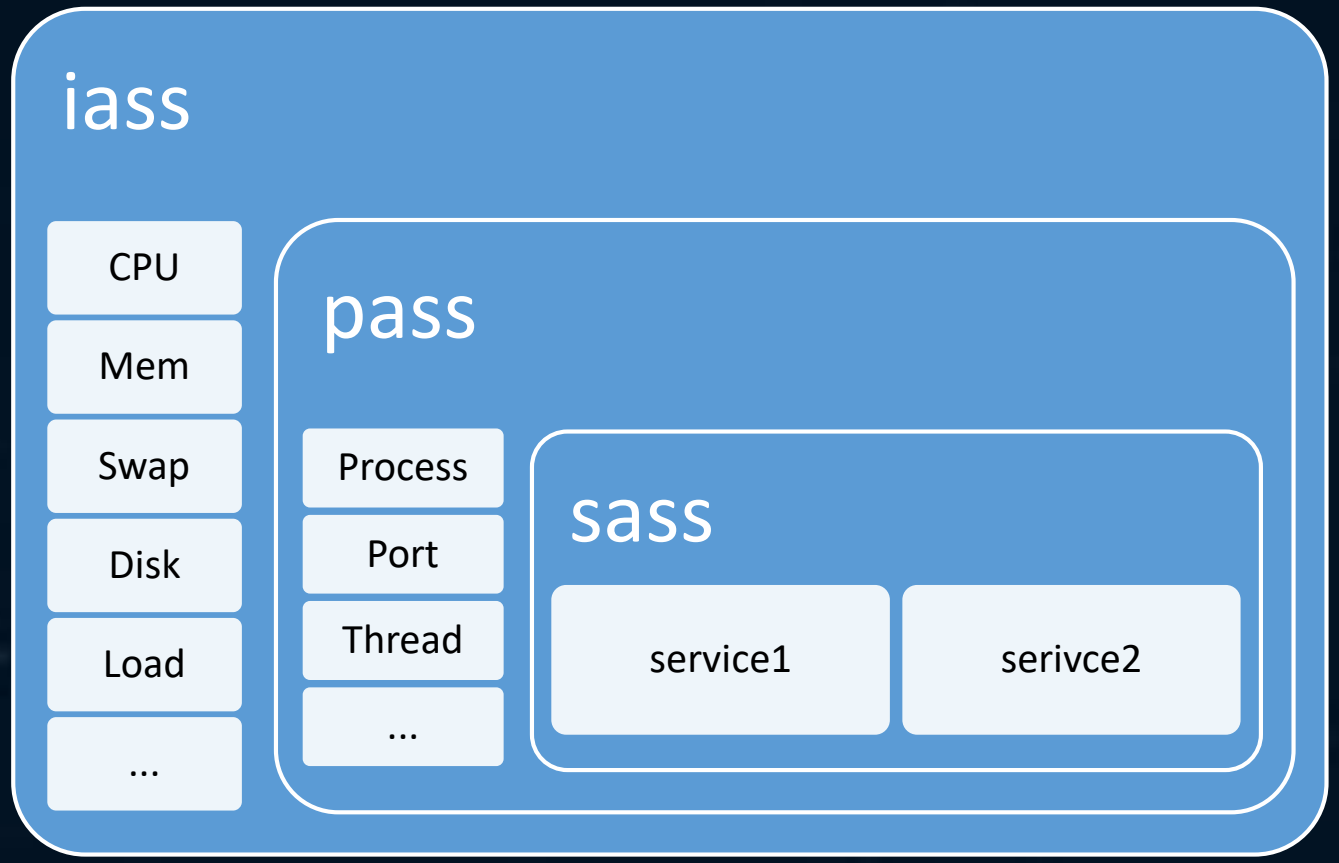
监控


加载失败
Error...


。。。
。。。


未恢复的报警
[P0 #2/2] localhost-2/load.1min
all(#3) 0<4 负载低于4 3 minutes ago [config | solved]
[P0 #2/2] localhost/load.1min
all(#3) 0.01<4 负载低于4 3 minutes ago [config | solved]
全选/反选
标记选中部分为已解决

如果有一套完善的监控解决方案来简化运维成本呢？

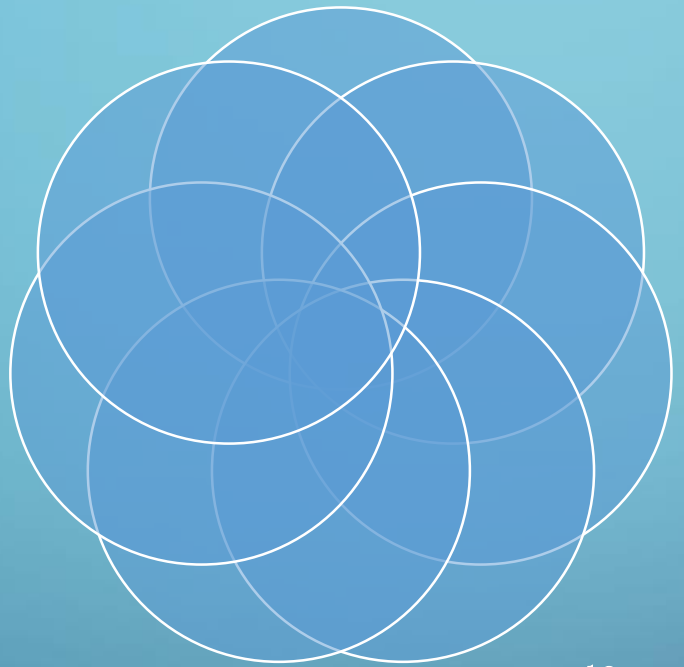


HADOOP 指标

基础的 HDFS/YARN
指标项分类如下：

MetricsSystem
#指标系统

JvmMetrics
#Hadoop 集群的JVM相关信息



UgiMetrics
#用户组指标信息

RPC
#Hadoop RPC 通信相关指标

yarn
context(ClusterMetrics,
QueueMetrics,
NodeManagerMetrics)
#yarn 资源调度相关信息

dfs context(namenode,
FSNamesystem,
JournalNode, datanode)
#dfs存储相关信息

HADOOP全方位监控

DFS 存储相关指标

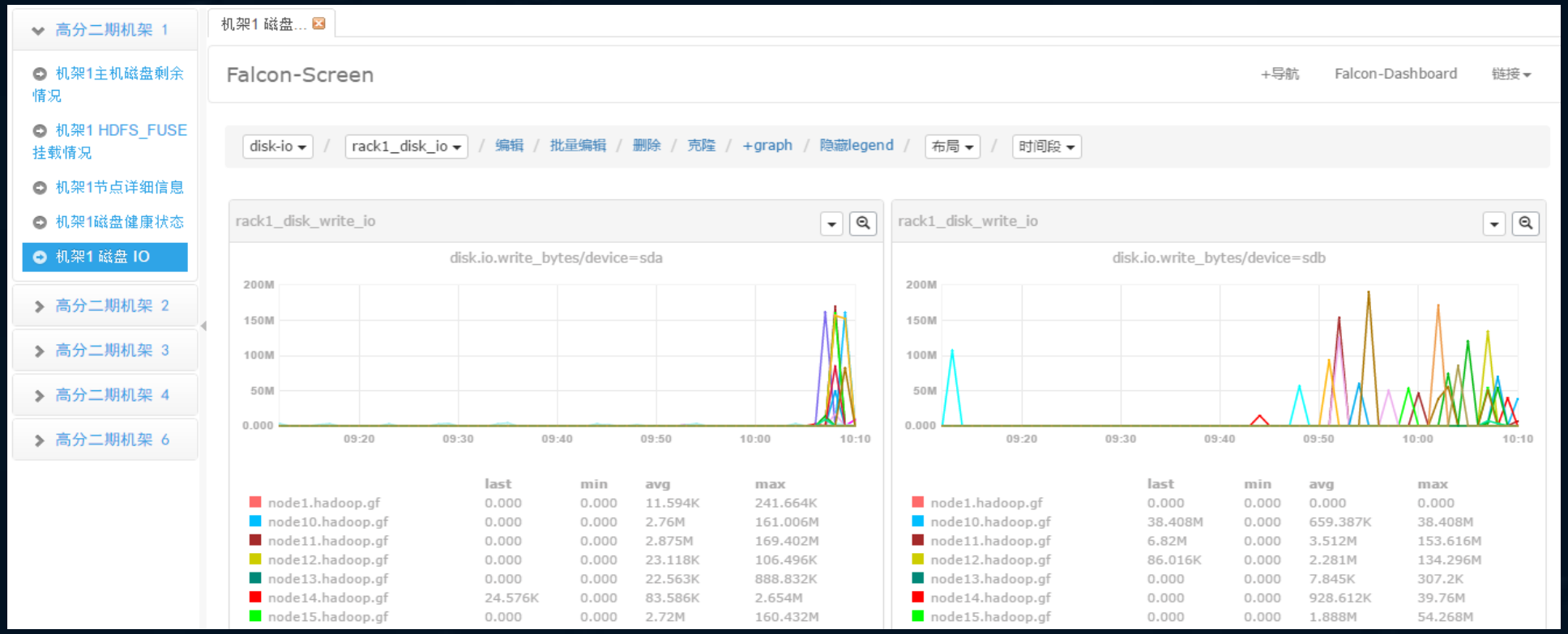
dfs context

namenode

Each metrics record contains tags such as ProcessName, SessionId, and Hostname as additional information along with metrics.

Name	Description
CreateFileOps	Total number of files created
FilesCreated	Total number of files and directories created by create or mkdir operations
FilesAppended	Total number of files appended
GetBlockLocations	Total number of getBlockLocations operations
FilesRenamed	Total number of rename **operations** (NOT number of files/dirs renamed)
GetListingOps	Total number of directory listing operations
DeleteFileOps	Total number of delete operations
FilesDeleted	Total number of files and directories deleted by delete or rename operations
FileInfoOps	Total number of getFileInfo and getLinkFileInfo operations
AddBlockOps	Total number of addBlock operations succeeded
GetAdditionalDatanodeOps	Total number of getAdditionalDatanode operations
CreateSymlinkOps	Total number of createSymlink operations
GetLinkTargetOps	Total number of getLinkTarget operations
FilesInGetListingOps	Total number of files and directories listed by directory listing operations
AllowSnapshotOps	Total number of allowSnapshot operations
DisallowSnapshotOps	Total number of disallowSnapshot operations
CreateSnapshotOps	Total number of createSnapshot operations
DeleteSnapshotOps	Total number of deleteSnapshot operations
RenameSnapshotOps	Total number of renameSnapshot operations
ListSnapshottableDirOps	Total number of snapshottableDirectoryStatus operations
SnapshotDiffReportOps	Total number of getSnapshotDiffReport operations
TransactionsNumOps	Total number of Journal transactions
TransactionsAvgTime	Average time of Journal transactions in milliseconds
SyncsNumOps	Total number of Journal syncs
SyncsAvgTime	Average time of Journal syncs in milliseconds
TransactionsBatchedInSync	Total number of Journal transactions batched in sync
BlockReportNumOps	Total number of processing block reports from DataNode
BlockReportAvgTime	Average time of processing block reports in milliseconds
CacheReportNumOps	Total number of processing cache reports from DataNode
CacheReportAvgTime	Average time of processing cache reports in milliseconds
SafeModeTime	The interval between FSNameSystem starts and the last time safemode leaves in milliseconds. (sometimes not equal
FsImageLoadTime	Time loading FS Image at startup in milliseconds
FsImageLoadTime	Time loading FS Image at startup in milliseconds
GetEditNumOps	Total number of edits downloads from SecondaryNameNode
GetEditAvgTime	Average edits download time in milliseconds
GetImageNumOps	Total number of fsimage downloads from SecondaryNameNode
GetImageAvgTime	Average fsimage download time in milliseconds
PutImageNumOps	Total number of fsimage uploads to SecondaryNameNode
PutImageAvgTime	Average fsimage upload time in milliseconds

某航天客户CRH集群监控方案



CRH Cluster

falcon

Falcon-Plus

- 指标监控

Eagle

Eagle

- 审计监控

LogSearch

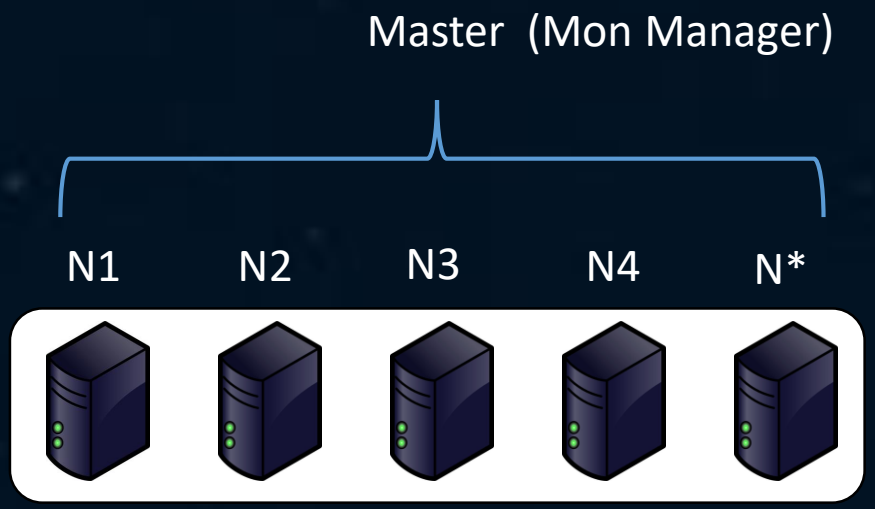
LogSearch

- 日志监控

CRH 未来版本规划

节点交付即监控

CRH Cluster



- 高耦合集成 openfalcon,eagle, logsearch
- dashboard 监控报表视图展现
- 自定义大数据集群报警策略
- 智能化定位/解决报警

存储

大数据时代，应用数量以百万计算，数据终端则以千万或亿计算。对各种大数据的近实时和实时分析，逐步成为IT创造价值的主要运营模式。

那么，对于从技术层面来说，存储，该如何去做呢？

数据？？

非结构化数据？ 半结构化数据？ 结构化数据？

CRH Storage

- 数据清洗
- 数据分析
- 数据挖掘
- 数据展示

HDFS 存储

优点	处理大文件
	流式访问数据
	硬件成本低

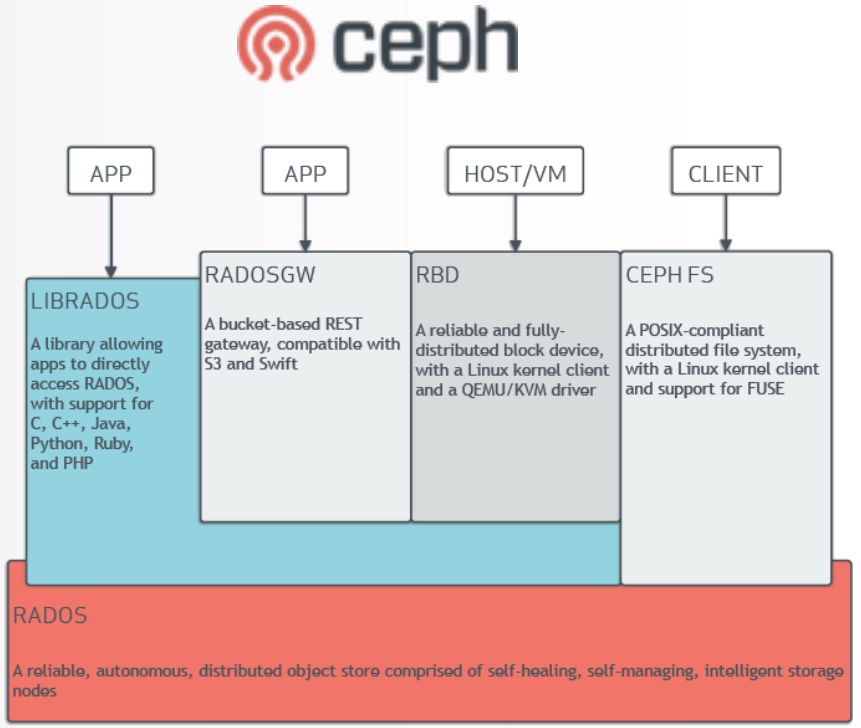
缺点	无法高效存储大量的小文件
	文件系统无法多用户自定义权限控制
	不适合低延迟数据访问

对象存储

内存加速

Ceph是一个开源的PB级文件系统
同时兼容CephFS(文件系统)，RADOS(对象存储)，RBD(块存储)

其中RADOS是一个支持海量存储对象的分布式对象存储系统。RADOS最大的特色是使用CRUSH算法维护存储对象与存储服务器的对应关系，并采用了无Master设计， 对象复制，集群扩容，数据迁移，故障检测和处理能复杂功能由OSD提供，既避免了单点失败，又提升了集群扩展性。

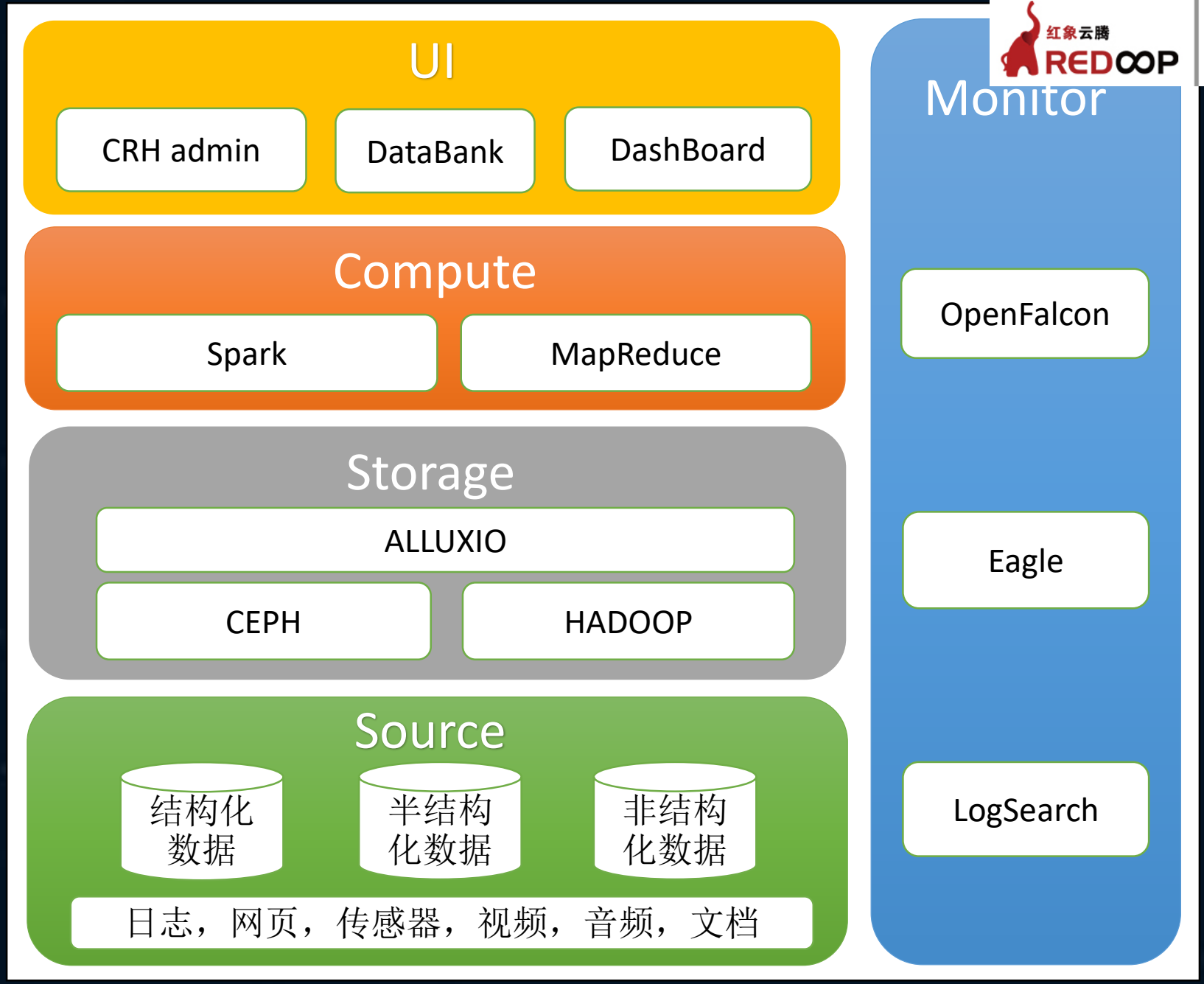


它是基于PG和CRUSH的存储对象定位方式是RADOS比较重要的一个创新点

Alluxio（之前名为Tachyon）是世界上第一个以内存为中心的虚拟的分布式存储系统。它可以统一数据访问的方式，连接上层计算框架和底层存储系统。应用只需要连接Alluxio即可访问存储在底层任意存储系统中的数据。Alluxio的以内存为中心的架构使得数据的访问速度能比现有常规方案快几个数量级。

REDOOP

CRH



案例：某航天机构 存储规模

10PB

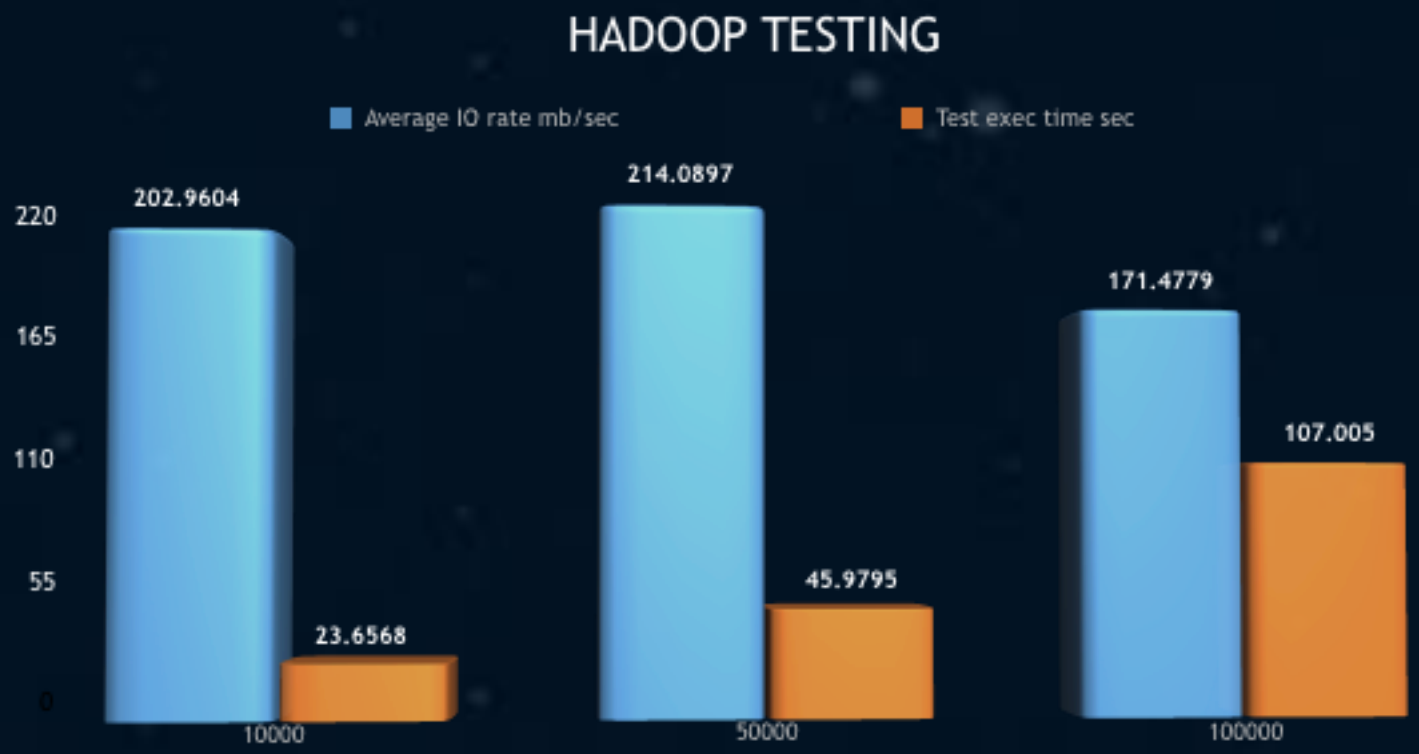
效率6倍

成本1/4

200台

5+卫星数据

稳定运行1年



----- TestDFSIO ----- : write

Total MBytes processed :
10000 、 50000 、 100000

Number of files: 10

FPGA & Hadoop 硬件加速

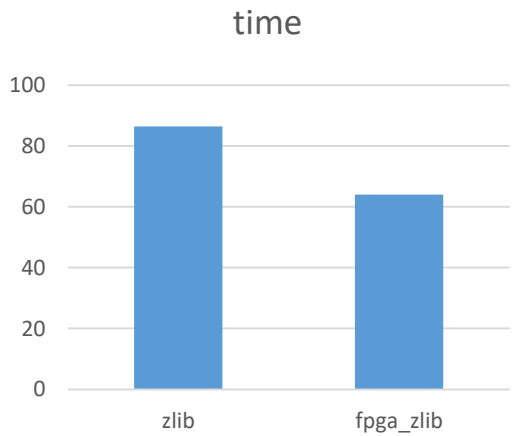
REDOOP Hadoop + 加速云 SPEED CLOUDS FPGA

红象云腾 联合 加速云

针对Hadoop二次开发。用 FPGA 处理压缩/解压缩

对 Hadoop 进行加速

100W 数据量测试结果



真正的数据高铁

REDOOP

专注产品　全面合作　开放自由