Kubernetes集群资源配置数据的备份、恢复和自动化

1、概述

1.1 Kubernetes是什么？

“Kubernetes（常简称为K8s）是用于自动部署、扩展和管理容器化（containerized）应用程序的开源系统。该系统由Google设计并捐赠给Cloud Native Computing Foundation（今属Linux基金会）来使用。”这是维基百科上对Kubernetes的介绍。

“它将组成应用程序的容器组合成逻辑单元，以便于管理和服务发现，Kubernetes 构建在 Google 15 年生产环境经验基础之上,并结合来自社区的最佳创意和实践。”这是Kubernetes官网中文版写在扉页上的寄语。

简单来说，Kubernetes能帮你把应用封装在容器里，想让它运行多少个副本就运行多少个。你的应用也许服务于1万个客户，部署10个应用容器，就能满足需求。也许你的应用服务于1000万客户，可能部署1000个应用容器就能满足需求。Kubernetes能帮你实现业务规模的自动水平伸缩。当然，Kubernetes能做的远不止这些。

Kubernetes对云计算的重要性，不亚于二十年前Java语言随第一波互联网浪潮掀起的革命性风暴。Java语言能让程序员的源代码“编写一次，运行在任何地方。”而Kubernetes引领的云计算革命，能让应用程序经过编排、调度组合，组成收放自如的复杂计算机应用系统。全自动绣花机把一根根线头、一个个色块、一个个印象元素，编织出绚丽的彩色图案。

某年双十一前，某宝的高管豪言，系统已经就绪，只需一杯清茶，再大的访问流量（N亿级用户），运维团队也能应对。难道他们没有用到可伸缩的容器编排？

1.2 为什么备份？

备份的目的是防止数据永久丢失，快速恢复系统可用。数据为什么会丢失呢？用户手潮误操作、不完美的版本升级、单点失效的硬件故障，地震、台风、洪水、海啸等自然灾害，都可能影响系统正常运行，造成数据部分丢失甚至全部灭失。

Kubernetes也是软件，软件运行的中间状态和运行结果也是以数据保存的。Kubernetes的运行数据存储在etcd，etcd是以键值对Key/Value格式存储，并提供读和写服务。

Etcd一般以多点集群形式出现，发生单点失效的可能性很小，但是用户误操作、升级失败的可能性还是存在的。

1.3 备份方法介绍

数据备份大致可以分为两类，物理备份和逻辑级备份。

物理备份不区分数据的内在逻辑关系，把数据存储作为一个整体来备份，恢复时也是作为整体恢复，不能只恢复一部分数据。逻辑备份按照数据的内在逻辑关系，选择性提取部分数据或全部数据，恢复时可以选择恢复一部分数据。

数据的物理备份，一般来说对人类是不可读的，只有原软件系统才能读取识别。Oracle、MySQL的Dump文件、重做日志和数据快照，etcd、Redis数据文件的直接拷贝或者数据快照，都可以看做是物理备份。

从物理备份恢复数据的优点是明显的，要恢复完整数据很快。缺点也很明显，可能会丢失上次备份以来的数据更新，但是可以通过全量备份配合增量备份来弥补，缩小数据丢失的范围。物理备份也不能做部分恢复，那怕只是一小点数据更新。

从逻辑备份恢复数据的优点也是明显的，恢复部分数据很快，想要恢复哪部分就恢复哪部分。缺点是，逻辑备份的速度慢一些，遇到数据量大时耗时相对较长。

物理备份与逻辑备份的优缺点正好相反，在实践中不妨结合起来使用，扬长补短，发挥各自的优势。物理备份适合系统崩溃后的快速恢复重建，而逻辑备份适合更加精细化的局部修补。

本小文要介绍的Kubernetes数据备份是一种逻辑备份方法，需要深入到Kubernetes内部的逻辑结构，探寻奥秘，条分缕析，庖丁解牛，找到一种逻辑备份新途径。

2、模型

2.1 Kubernetes集群

2.2 Kubernetes层级模型

Kubernetes

命名空间：

几个系统命名空间

应用命名空间

资源从属于命名空间

资源类型：

服务 service

部署 deploy

配置 configmap

加密配置 secret

任务 job

定时任务 cronjob

副本集 replicaset

驻留任务集 daemonset

有状态集 statefulset

service deploy configmap secret job cronjob replicaset daemonset statefulset

2.3 备份/恢复模型

时间标尺快照

备份模型

恢复模型

3、备份

3.1 全量备份

3.2 优先备份

3.3 备份流程

3.4 备份代码详解

3.5 备份的备份

4、恢复

4.1 准备数据

4.2 执行恢复

4.3 恢复检测

4.4 恢复代码详解

5、应用

5.1 故障恢复

5.2 历史跟踪

持续跟踪资源配置历史

5.3 时点比对

时点比对配置变化

5、自动化

5.1 备份自动化

数据备份的自动化

5.2 监测自动化

重点配置项监测的自动化。

6、多集群

6.1 多集群模型

6.2 多集群备份

6.3 多集群恢复

7、小结

7.1 工作小结

7.2 未来展望