OneAPM 的 Druid 应用实践

刘麒赟
OneAPM 大数据架构师
13810556729
liuqiyun@oneapm.com

声明

- 本文的一些图片、文字、数据来自 Imply.io 公司、Druid.io 的 Druid 介绍文档，以及互联网的一些公开资料，在此对相关作者表示感谢！
- 本文中出现的一些相关于 OneAPM 公司的数据与技术介绍仅出于演讲者的个人理解，不代表 OneAPM 公司的官方态度
- 本次演讲的所有言论仅代表演讲者的个人言论

概 览

- 选择 Druid 的背景
- 应用实践

选择 Druid 的背景

我们的数据特点

- 时间序列数据：按一定时间频率收集数据
- 性能指标数据：多 Dimensions ，多 Metrics
- 数据增长快

我们对数据引擎的要求

- 大数据量实时处理与历史存储，并且提供同时快速查询历史数据与实时数据的能力
- 能对数据进行丰富的 OLAP 分析：多维度聚合、过滤、 groupBy 等
- 良好的扩展性与高可用性

应用实践

OneAPM 利用 Druid 存储查询 SaaS 数据

- 目前 Druid 服务于 3 个产品的 SaaS 版本：AI(Application Insight),BI(Browser Insight),MI(Mobile Insight)

OneAPM Druid 集群概况
（by 2016 Feb.19）

- 集群分类：生产集群、测试与监控集群
- 集群物理位置：阿里云
- 总的机器数量：> 50 台
- 总的 DataSource 数量：> 10 个
- 每天实时处理 event 数：> 7 billion
- 每天实时处理数据量：> 3 TB
- 在线 Segment 的总数据量：> 700 GB
- 查询数量：峰值 > 8000 QPH
- 查询平均响应时间：< 150 ms
- 查询响应时间分布：
    * 0 - 100 ms : 70.63%
    * 100 — 200 ms : 16.87%
    * 200 — 300 ms : 5.43%
    * 300 — 800 ms : 6.77%
    * > 800 ms : 0.3 %

OneAPM 金字塔结构的 Druid DataSources

- 相关重要原理：
  * 查询粒度（queryGranularity）：对 Druid Segment 的最低查询粒度，指导 Druid 对数据的聚合以及 Segment 的生成
  * Druid 能够提供在低查询粒度 DataSource 的数据基础上做高粒度聚合查询的功能
  * 从存储成本上看：聚合粒度越低时，同样单位时间所需的存储空间越高
  * 影响 Query 效率的主要相关因素有：
    - 返回的数据量：数据量越大，速度可能越慢
    - 数据源本身的聚合粒度：本身的聚合粒度越接近聚合查询指定的粒度，查询的速度会更高
    - 查询在内存中命中率：若命中率低，则会花大量时间将 Segment 从磁盘加载到内存

OneAPM 金字塔结构的 Druid DataSources

- 相关需求：
  对于某个 Kafka Topic 数据源，我们希望对用户提供不同的查询粒度（比如，1 分钟、十分钟、1 小时），并且为不同的查询粒度提供不同的查询跨度（比如，6 小时、十分钟、1 小时）。

- Straightforward 的方案：
  总共创建一个 DataSource，按最低查询粒度保存数据

- 我们选择的方案：OneAPM Druid 金字塔结构的 DataSources—— 创建 3 个不同的 DataSources，它们都消费同一个 kafka topic，但同时有几个主要不同之处：
  * 消费同一个 topic 的不同的 group
  * queryGranularity 分别为：1 min, 10 min, 1 hour
  * Rules 设置的数据存留时间不同：6 hour, 2 day, 1 month
  * 使用不同的 Realtime Nodes & History Nodes

OneAPM 金字塔结构的 Druid DataSources

	1 DataSource 方案	3 DataSource 金字塔方案
Historical Node 上的数据冗余	无	在一些重合的时间段有数据冗余
Realtime 的数量	少，1 套 Realtime Nodes 即可	多，3 套 Realtime Nodes
1 分钟粒度查询支持时间跨度	长，1 个月	短，6 小时
总的数据量大小	大	小。可以达到 1/6 左右空间
查询速度（同样集群规模）	一般（所需内存更多，命中率较低）	较快（所需内存更少，命中率较高）

OneAPM Druid 集群监控方案

- Druid Metrics ：
  Druid 通过 metrics 记录集群运行时产生的各个主要指标。
  Druid 可以发送 metrics 到 log ，或者通过 http 往外发送

- OneAPM Druid 集群监控方案一：
  通过 http 将 broker 等 service 产生的 metrics 发送到一个 http proxy(druid-metrics-to-kafka) ，然后该 proxy 再将 metrics push 到对应的 Kafka topic 。最后，在用作监控的 Druid 集群里创建一个独立的 DataSource ，让其消费该 Kafka topic 上的数据，并保存到 Druid 集群中。此时，便可以通过查询该 DataSource 的内容得到具体的 metrics 信息——查每一条详细信息、聚合结果、平均值等

- OneAPM Druid 集群监控方案二：
  从 Kafka topic 上将 Metrics 数据拉到 OneAPM 的数据管理平台 Cloud Insight(CI) 上。CI 兼顾 IT 基础设施和平台服务监控，能够很好地对 Druid Metrics 进行分析与展示，也可以进一步接上报警系统实现报警功能。

Thanks