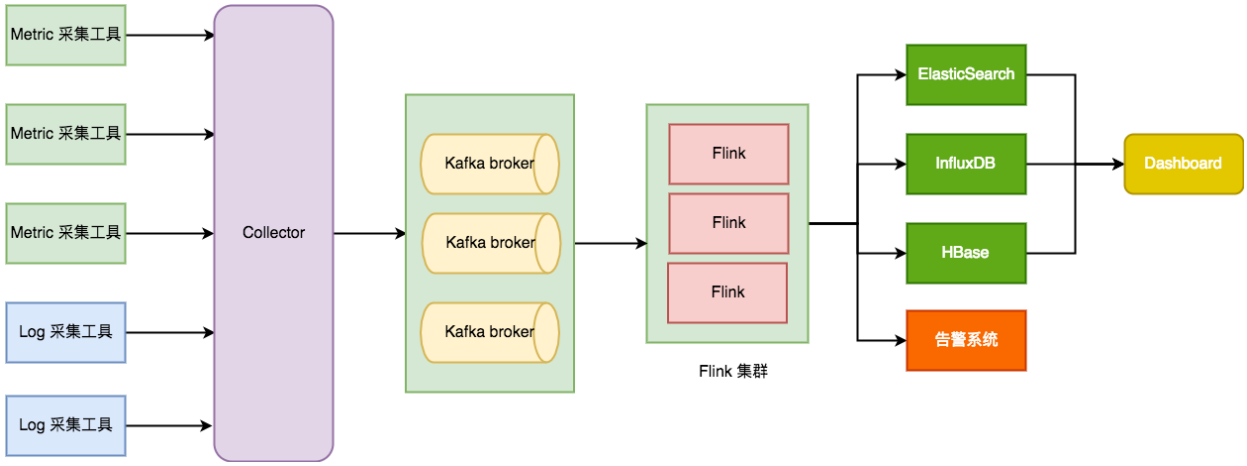
1、监控平台是什么架构 ， 多大的数据量 监控需要用到flink?



并不是说数据量要达到多少量级就需要用 Flink，而是看是否有实时场景的需求，比如上面的告警就需要的

采集数据的工具可以采用开源，可以自己写，看公司的人力资源够不？如果不够可以在开源的基础上修改维护

采集metric的话 比如 telegraf，日志的话logstash，filebeat等

2、有没有一些成熟运用flink的开源项目可以供大家学习参考?

1、https://github.com/zhisheng17/flink-learning 我自己的 Flink 学习记录目录，应该是全网第一篇这么全的，后面会把 cep、sql、window、state、connectors 等学习案例放在这里，最后也会把对 Flink 的监控数据采集、存储、告警放在这里

2、https://github.com/DTStack/flinkStreamSQL 基于开源的flink，对其实时sql进行扩展；主要实现了流与维表的join，支持原生flink SQL所有的语法

3、https://github.com/DTStack/flinkx 基于flink的分布式数据同步工具

第一个是自己的项目，第二、三个是袋鼠云在github开源的

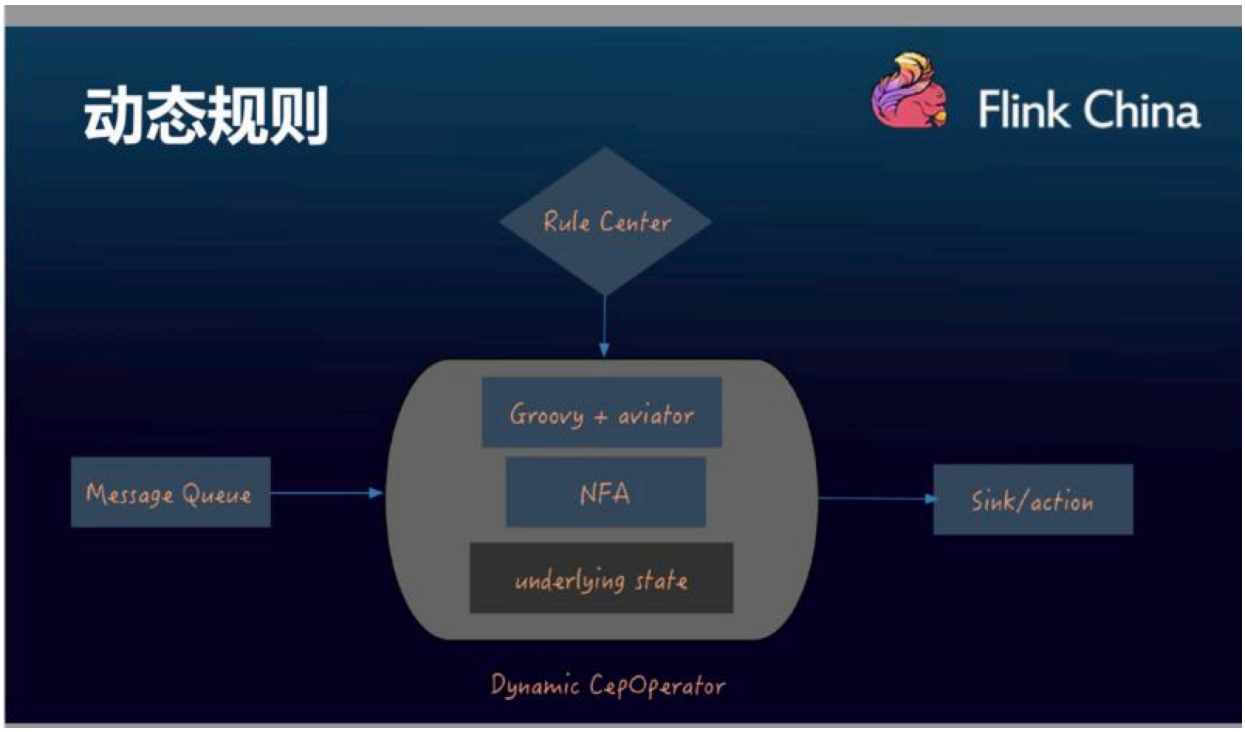
其他的就没发现运用比较成熟的开源项目了，另外就是建议大家有时间可以稍微阅读一下源码

3、flink 动态cep怎么做？有没有相关的解决方案及案例

关于 Flink CEP 这块，处理复杂事件确实方便，但是它的弊端就是不能动态更新，就比如阈值，不能够动态的调整，所以如果是能够将这个固定的地方或者整段 cep pattern 代码放在外部，然后代码里面能够做到自己去加载然后解析，这样是否就能够做到动态解析呢?

之前有调研过这块，发现滴滴里面好像有改造过这块，它的实现如图一，在这个 PPT 里露出来了实现的 cep json ， 可以看下 PPT，我上传到百度云了。链接:https://pan.baidu.com/s/1t2mopgkWF2uAK1MGTkV7KQ 密码:9ix5

这个动态加载规则昨天我和 Blink 团队三个大佬交流的时候也谈到过这块，cep 动态规则这块确实比较难弄，如果做到 job 不重启就能够动态的加载规则。



4、Flink相对spark有哪些优势?

这个我一般也不讲他们谁好谁坏，只能说各有千秋，Flink 更适合处理流数据、Spark 出现的比较早、生态可能比较火，包括scala都估计是他带火的，更适合 离线计算，当然现在spark streaming 也说是支持流流计算，但是本质其实还不是的，还是以微批计算的，未来我觉得他们会越来越像的

5、用户会实时产生一些数据，我希望在数据产生后，发送到kafka，然后flink消费kafka，实时计算得出多个维度的统计数据，flink支持这样的场景吗?

这个统计方式是支持的

6、flink应用部署生产的方案和注意事项，如何根据数据量等估算资源，怎么支持多租户

因为flink有很多种部署方式，standalone、on yarn 、on mesos、on k8s，不同的方式可能部署会有点不同，确实得注意一下，官网的资料也有讲一些部署的情况

另外的话如何分配资源了，这个应该要根据你的数据量进行压测后才可以估量要划分多少资源

https://ci.apache.org/projects/flink/flink-docs-release-1.8/ops/deployment/cluster_setup.html

官网这个地方有部署在不同的上面的讲解的

因为不同的公司底层调度系统可能不一样，有的用yarn，有的用k8s

Flink has a mor
monitoring API

The monitoring

Overvi

The monitoring
be configured i
currently the sa

In the case of r
information abo

Develo

The REST API
`org.apache.f`

We use *Netty* a
combination ha

To add new rec

- add a ne
- add a ne
- add the

A good exampl

7、flink和 blink有什么区别

Blink 这个就是 阿里从Flink 基础上改的，然后开源出来了，后面会把代码合到Flink master分支上去的

8、 能否介绍下智能推荐或相关推荐是如何实现的?

无

9、实时计算的逻辑由flink来完成，那这段逻辑是用flink来完成还是使用其他的consumer应用来完成，有什么区别？是效率吗？为什么?

Flink 来完成的，注意你的 Flink job 其实就是一个消费者，你说的这是同一个东西，没什么区别，所以也就没对比性了

10、Flink 相对 Spark 有哪些优势?

这个我一般也不讲他们谁好谁坏，只能说各有千秋，Flink 更适合处理流数据、Spark 出现的比较早、生态可能比较火，包括scala都估计是他带火的，更适合 离线计算，当然现在spark streaming 也说是支持流流计算，但是本质其实还不是的，还是以微批计算的，未来我觉得他们会越来越像的

11、我看到 Github-Flink 的源码大多数是 Java 编写的，是不是用 Java 开发 Flink 更好？阿里内部用 Scala 多吗?

同样，这个问题我昨天和Blink 团队的大佬们交流过这个问题，目前Flink 主干分支的代码确实有些是Scala写的，但是绝大多少是Java，因为不同的人可能喜好不一样，但是阿里那边他们几乎是用Java，开发 Flink Job 的话应该是 Java 和 Scala 都可以的，他们底层都是交给 JVM 来管理的，估计差别不大，一般看团队的成员语言是否共通，最好团队里面是统一的

12、我们现在消费kafka数据进行过滤计算，都是通过java项目进行的，这样如果数据量大的话是不是必须切换为大数据工具，比如flink

不是必须，现在有压力了吗？没有压力的话可以先不用，后面等有压力可以上Flink

13、大数据是不是都是用java语言开发的，能用golang python吗

这个大数据目前scala偏多吧，为什么scala偏多呢？可能就是因为spark用的多了，哈哈哈，而且scala用的爽，但是现在java8也支持stream和lambda

14、dashboard 我可以用grafana吗

可以、一般grafana用的比较多

15、flink和blink具体区别，blink的主要变化，基于数据标准化映射及关系归并的实时流处理平台建议用blink还是开源的flink

如果你们等得急的话建议等blink合到主干分支去、然后用新版本的flink就行、不然到时候还得折腾迁移带来的麻烦

16、比如每日10亿的交易数据，建议用多大的服务器配置， cpu，内存，存储

一般做jobmanager HA 的话，起码就需要两个 jobmanager，再加上 你的 TaskManager ，最关键的是怕你会将一些 state 数据存储在内存中，那样就会比较耗内存，因为如果多个job去消费同一份数据的话，多个job也是很占用资源多，所以你最好把前提条件都说一下吧，具体的我估计还是得你们实地压测才能估量出来