

Kafka基础和應用

- 什么是kafka
- Kafka特性
- Kafka解决了什么问题
- Kafka基本概念和名词
- Kafka基本操作
- Kafka支持的语言
- Kafka -python

什么是kafka?

Kafka是最初由Linkedin公司开发，是一个分布式、支持分区的（partition）、多副本的（replica），基于zookeeper协调的分布式消息系统，它的最大的特性就是可以实时的处理大量数据以满足各种需求场景：比如基于hadoop的批处理系统、低延迟的实时系统、storm/Spark流式处理引擎，web/nginx日志、访问日志，消息服务等，用scala语言编写，Linkedin于2010年贡献给了Apache基金会并成为顶级开源项目。

- 高吞吐量、低延迟：kafka每秒可以处理几十万条消息，它的延迟最低只有几毫秒，每个topic可以分多个partition，consumer group 对partition进行consume操作。
- 可扩展性：kafka集群支持热扩展
- 持久性、可靠性：消息被持久化到本地磁盘，并且支持数据备份防止数据丢失
- 容错性：允许集群中节点失败（若副本数量为n,则允许n-1个节点失败）
- 高并发：支持数千个客户端同时读写

www.mindg.cn

kafka解决了什么问题?

知乎

首页

发现

话题

搜索你感兴趣的内容...



Kafka

kafka解决了什么问题?

请高手指明一下kafka解决了什么问题，什么场景下使用？消息订阅和发布吗，好像redis也支持，功能是否有重叠？

关注问题

写回答

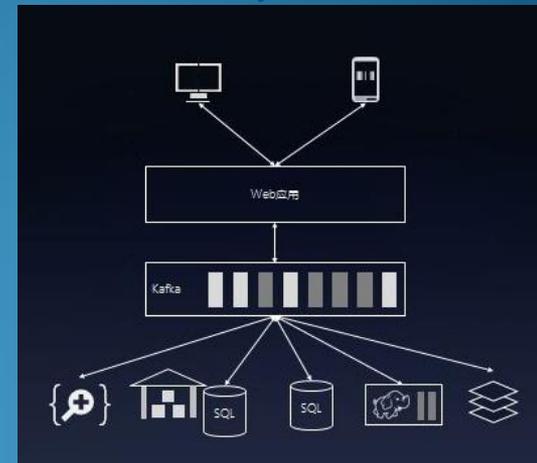
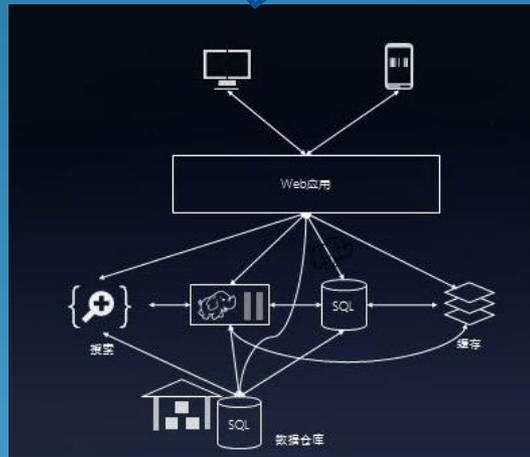
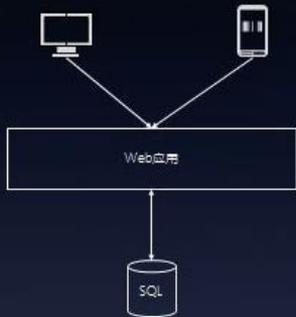
添加评论

分享

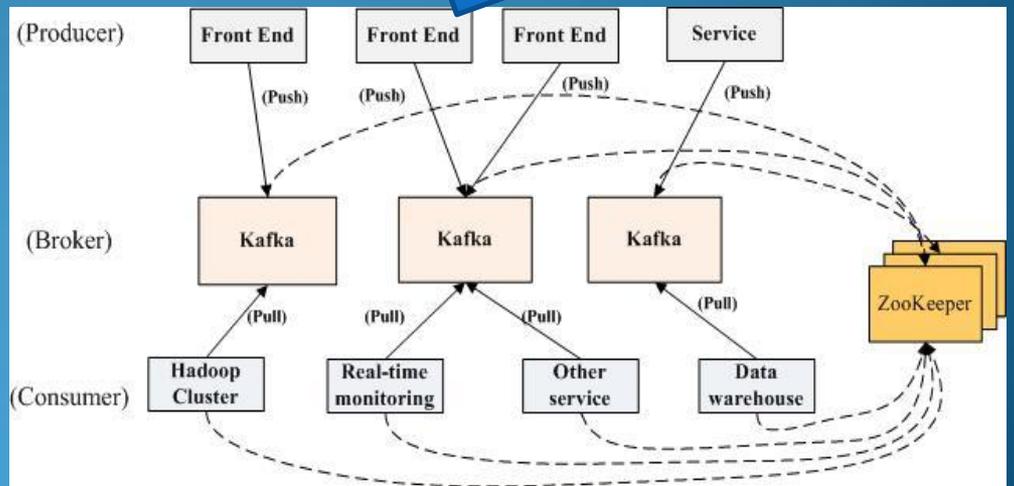
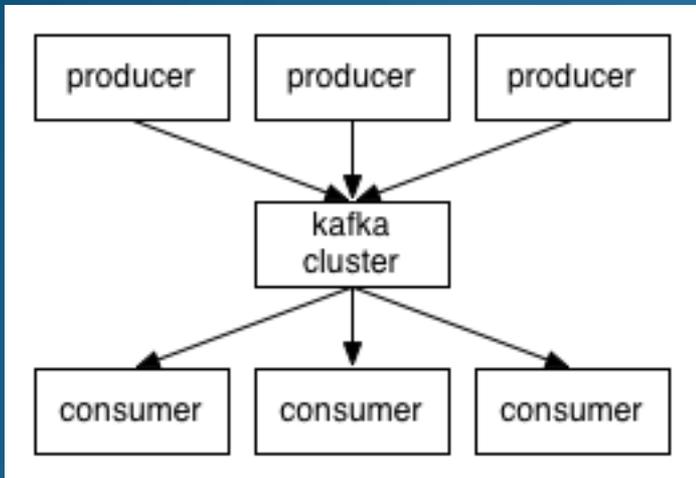
邀请回答



kafka解决了什么问题?

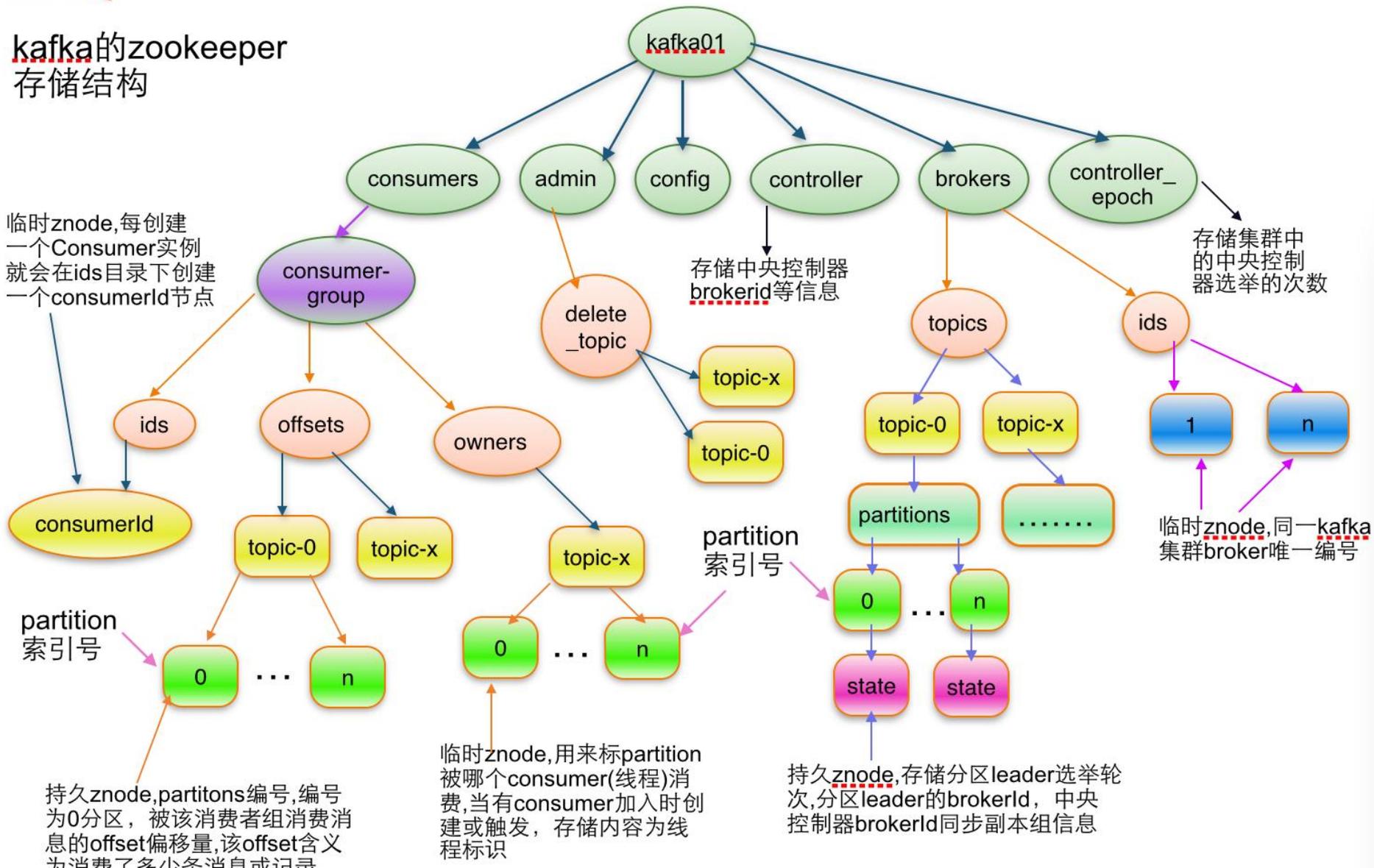


Kafka数据处理流程图



www.mindg.cn Kafka和zookeeper关系

kafka的zookeeper存储结构



Kafka基本概念和名词

1) Broker

Kafka集群包含一个或多个服务器，这种服务器被称为broker。broker端不维护数据的消费状态，提升了性能。直接使用磁盘进行存储，线性读写，速度快；避免了数据在JVM内存和系统内存之间的复制，减少耗性能的创建对象和垃圾回收。

2) Producer

负责发布消息到Kafka broker

3) Consumer

消息消费者，向Kafka broker读取消息的客户端，consumer从broker拉取(pull)数据并进行处理。

4) Topic

每条发布到Kafka集群的消息都有一个类别，这个类别被称为Topic。（物理上不同Topic的消息分开存储，逻辑上一个Topic的消息虽然保存于一个或多个broker上但用户只需指定消息的Topic即可生产或消费数据而不必关心数据存于何处）

5) Partition

Partition是物理上的概念，每个Topic包含一个或多个Partition.

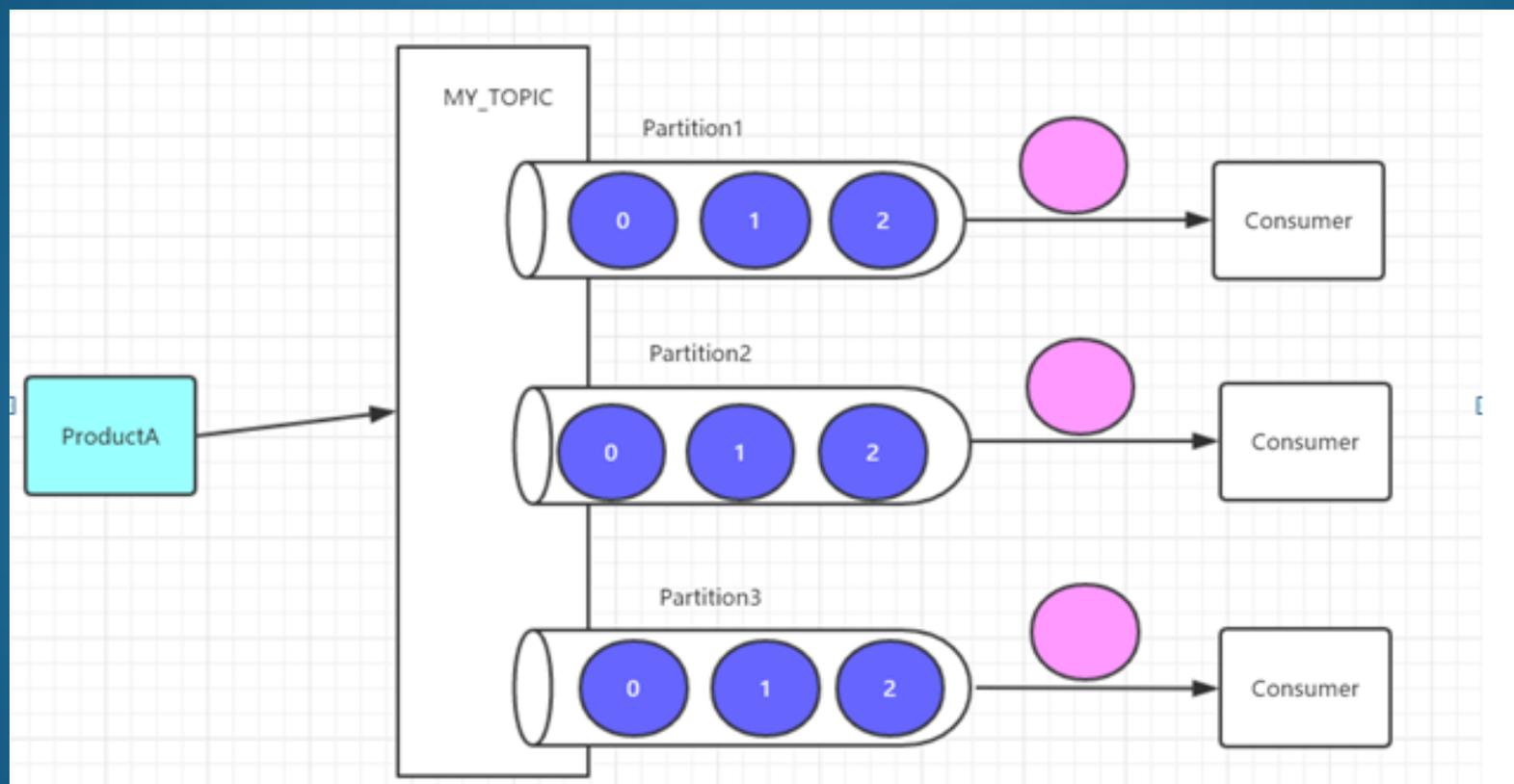
6) Consumer Group

每个Consumer属于一个特定的Consumer Group（可为每个Consumer指定group name，若不指定group name则属于默认的group）

7) Topic & Partition 关系

Topic在逻辑上可以被认为是一个queue，每条消费都必须指定它的Topic，可以简单理解为必须指明把这条消息放进哪个queue里。为了使Kafka的吞吐率可以线性提高，物理上把Topic分成一个或多个Partition，每个Partition在物理上对应一个文件夹，该文件夹下存储这个Partition的所有消息和索引文件。

Kafka概念之间关系



Kafka基本操作

查看topic列表:

```
1.kafka-topics.sh --list --zookeeper 10.1.14.39:2181,10.1.14.40:2181,10.1.14.41:2181
```

创建topic:

```
2.kafka-topics.sh --create --zookeeper 10.1.14.39:2181,10.1.14.40:2181,10.1.14.41:2181 --replication-factor 3  
--partitions 1 --topic my-replicated-to
```

Pic

查看topic详细信息

```
3.kafka-topics.sh --describe --zookeeper 10.1.14.39:2181,10.1.14.40:2181,10.1.14.41:2181 --topic
```

生产消息:

```
./kafka-console-producer.sh --broker-list 10.1.14.39:9092,10.1.14.40:9092,10.1.14.41:9092 --topic nglogs
```

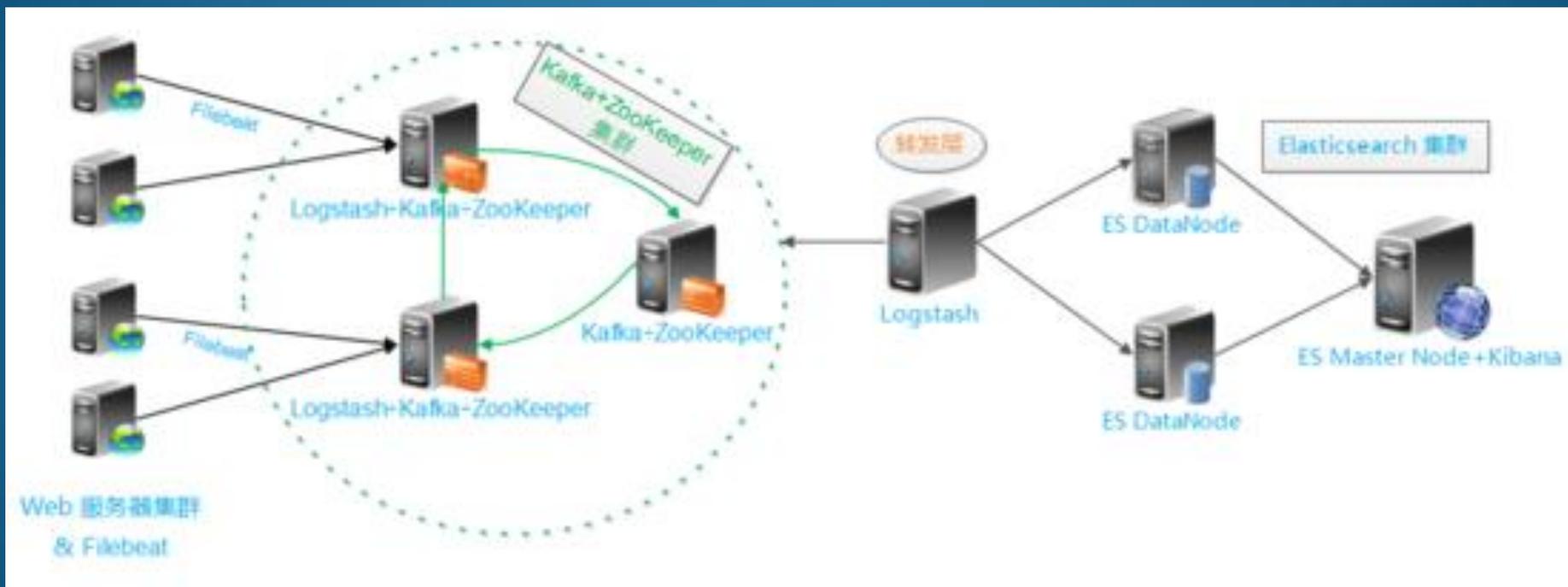
消费消息

```
./kafka-console-consumer.sh --bootstrap-server 10.1.14.39:9092,10.1.14.40:9092,10.1.14.41:9092 --  
from-beginning --topic nglogs
```

Kafka 客户端支持的语言

kafka支持的客户端语言：Kafka客户端支持当前大部分主流语言，包括：C、C++、Erlang、Java、.net、perl、PHP、Python、Ruby、Go、Javascript可以使用以上任何一种语言和kafka服务器进行通信（即辨析自己的consumer从kafka集群订阅消息也可以自己写producer程序）

Kafka应用



www.mindg.cn

完