# Kafka

## 一.简介

### 1.kafka的定义

**kafka是一种高吞吐量的分布式发布订阅消息系统**

kafka是一个分布式消息系统，由LinkedIn使用Scala编写，用作LinkedIn的活动流（Activity Stream）和运营数据处理管道（Pipeline）的基础，具有高水平扩展和吞吐量。

### 2.目前使用的公司：

阿里，百度，Twitter，亚马逊

### 3.优势

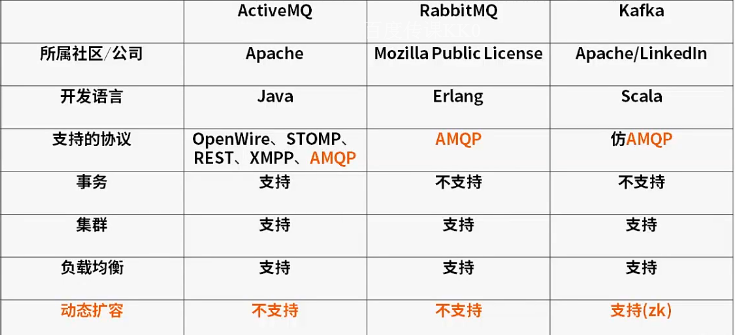
因为kafka具有**高水平扩展和高吞吐量**的能力，支持Apache Flume（日志收集），Apache Storm（实时数据处理），Spark（内存数据处理），elasticsearch（全文检索）等的集成。

### 4.分布式消息系统对比

Scala和java是运行在jvm上的两种语言

Erlang是一种通用的面向[并发](https://baike.baidu.com/item/%E5%B9%B6%E5%8F%91)的编程语言，它由[瑞典](https://baike.baidu.com/item/%E7%91%9E%E5%85%B8)电信设备制造商爱立信所辖的CS-Lab开发，目的是创造一种可以应对大规模并发活动的编程语言和运行环境

**分布式消息系统中的事务：**对应到消息队列中，多个消息一起发送，要么全部发送成功，要么全部失败。



## 二.遵循协议

### 1.AMQP定义

AMQP，即Advanced Message Queuing Protocol,一个提供统一**消息服务**的**应用层**标准高级[**消息**](https://baike.baidu.com/item/%E6%B6%88%E6%81%AF/1619218)**队列协议**,是[应用层](https://baike.baidu.com/item/%E5%BA%94%E7%94%A8%E5%B1%82/4329788)协议的一个开放标准,为面向消息的中间件设计。基于此协议的客户端与消息中间件可传递消息，并不受客户端/[中间件](https://baike.baidu.com/item/%E4%B8%AD%E9%97%B4%E4%BB%B6)不同产品，不同的开发语言等条件的限制。

### 2.AMQP的特点

多信道、协商式、异步、安全、跨平台、中立、高效。

分布式消息系统一些基本概念

## 三.kafka支持的语言

C、C++、Erlang、java、.net、perl、PHP、Python、Ruby、Go、javascript可以使用以上任何一种语言与kafka服务器进行通信（编写自己的consumer和producer程序）。

## 四.kafka中的术语

**消费者（Consumer）**:从消息对客中请求消息的客户端应用程序；

**生产者（Producer）**：向Broker发布消息的客户端应用程序；

**AMQP服务器（Broker）**：用来接收生产者发生的消息并将这些消息路由给服务器中的队列；

kafka中的消息是按照**主题（topic）**来组织的。

**主题（Topic）**：一个主题类似新闻中的体育、娱乐、教育等分类感念，在实际工程中通常一个业务一个主题。

**分区（Partition）**：一个topic中的消息数据按照多个分区组织，分区是kafka消息队列的最小单位，一个分区（Partition）可以看做先进先出的数据缓存器（FIFO）队列。

主题和分区的关系：一个主题（Topic）中的数据是按照分区（Partition）进行组织

zookeeper：一个提供分布式的状态管理，分布式配置管理，分布式锁服务管理的集群。

**@author** CGB

@createtime 2018年2月27日22:00:44

@Public email cgb2499404424@163.com

@Location Beijing China

@github https://github.com/chenguangbo