基于 Flume/Kafka/Spark 的分布式日志流处理系统的设计与实现

陈任飞，吕玉琴，侯宾
（北京邮电大学电子工程学院，北京 100876）

摘要：随着移动互联网技术的发展和普及，企业的日常生产和交易活动伴随着大量日志的产生。如何使用新型的技术处理传统技术无法处理的海量日志数据，提取相应的商业信息，已经成为目前诸多行业的企业急需解决的问题。分布式计算框架的出现为解决这个问题提供了一种思路。Flume 是一个可靠的用于日志收集、聚合和传输的分布式系统；Kafka 是一个高吞吐的分布式发布/订阅消息系统；Spark 是继 Hadoop 之后的新一代大数据分布式数据处理框架，Spark Streaming 是 Spark 专门服务于流式数据的处理框架。本文基于 Flume、Kafka 和 Spark 构建了一个分布式日志流处理系统。通过这个系统，企业可以高效、实时、可靠地获取和分析日志流数据，获得可用于辅助企业做出相关商业决策的信息，从而提高企业的服务质量和竞争力。

关键词：分布式系统；日志流；Kafka；Flume；Spark
中图分类号：TP391

Design and Implementation of Distributed Log Streams Processing System Based on Flume/Kafka/Spark

CHEN Renfei, LV Yuqin, HOU Bin
(School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876)

Abstract: With the development and popularization of mobile Internet technology, daily production and trading activities of enterprises bring about massive amounts of log data. How to deal with these logs which can't be processed by traditional log system with a new method and how to extract relevant business information has become an issue that needs to be addressed urgently in many industries. The emergence of the distributed computing framework provides a train of thought to solve this problem. Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data. Kafka is a high-throughput, distributed publish-subscribe messaging system. Spark is a new generation distributed big data computing framework after Hadoop. Spark Streaming is a framework specially for stream-oriented computation. We have designed and implemented a distributed log streams processing system based on Flume, Kafka and Spark. With this system, enterprises can fetch and analyze log streams data and obtain the decision making information for enterprises business efficiently, real-timely and reliably, thereby the service quality and competitiveness of enterprise can be improved.

Key words: distributed system; log stream; Kafka; Flume; Spark

0 引言

信息时代，数据呈现出爆炸性增长的态势。企业在生产和交易等诸多环节都会产生日志，日志数据的规模也从传统数据库时代的 GB 数量级跃升到 TB 甚至 PB 这样的数量级。面对海量的日志，传统的日志处理系统框架已经无法满足企业目前的需求。与此同时，企业业务对日志处理的实时性需求也逐渐提高。传统的流数据处理框架其吞吐量和容错性存在先天的

作者简介：陈任飞（1990-），男，硕士研究生，主要研究方向：大数据信息处理
通信联系人：吕玉琴（1944-），女，教授，主要研究方向：智能信息处理. E-mail: lvyq@263.net

缺陷，不适用于目前类似于互联网这类行业的高速扩展的业务需求。

业务需求的改变加剧了数据处理技术的革新需求。目前的日志系统缺乏分布式的设计考虑[1]，或者只是部分单一组件有分布式的设计[2]，并没有形成一套整体的分布式方案。除此之外，在日志的收集过程中数据的可靠性没有很好的保障，日志的转发吞吐性能也在高并发的海量日志情境下显得不足，对于流式日志数据的处理[3]，以及其系统的容错性也没有很好的考虑[4]。针对这一系列尚待改进的问题，本文设计并实现了一套分布式的日志流处理系统，它具有分布式、易扩展、大吞吐、高可靠的特性，能够准确并且实时地为企业根据日志内容提取目标信息，帮助企业采取相关营销措施，为用户提供及时的个性化服务，从而提高企业的竞争力。

1 架构设计与实现

该系统分为三个模块：日志收集模块、日志分发模块、日志分析模块。日志收集模块负责从不同的日志源收集日志数据，并转发给日志分发模块。当日志分发模块收到日志数据后，根据不同的主题把日志流进行分组，再把对应的日志流发送给订阅了该主题的日志分析模块。日志分析模块收到特定主题的日志数据后进行特定的数据分析。系统的三个模块之间无缝衔接，具有分布式、高扩展性、高可靠性、实时性的特点要求。

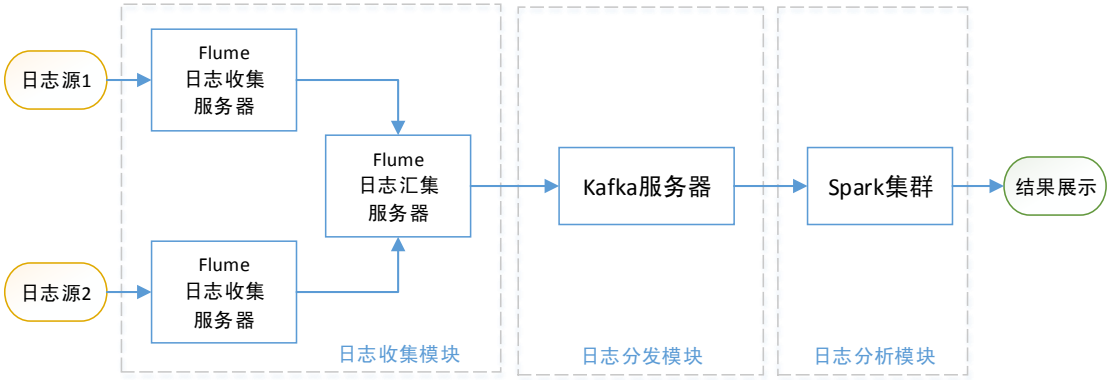


图 1 系统架构

Fig. 1 System architecture

1.1 日志收集模块

日志收集模块需要一个分布式的、可靠的、高可用的、能处理海量日志的框架，并且需要支持多源采集且集中存储。在诸多的日志收集系统中，本文选择的是 Apache Flume。从 Flume OG 进化到 Flume NG，它拥有诸多先进的特性[5]。Flume 有一个基于流媒体数据流的简单而灵活的架构。Flume 使用基于事务的数据传递方式来保证事务传递的可靠性。数据的源和宿被封装进一个事务。事务被存放在通道中直到该事务被处理，通道中的事务才会被移除，这是 Flume 提供的点到点的可靠机制。Flume 支持用户建立多级流，具有高可扩展性，支持多层次多项扩展，整个 Flume 模块的拓扑结构就是根据具体逻辑需求由一个或者多个代理组成。从多级流来看，前一个代理的宿和后一个代理的源同样由它们的事务来保障数据的可靠性。因此，在可靠性上 Flume 比其他如 Scribe 等日志收集框架要优秀得多。本文涉及的日志收集模块结构图如下：

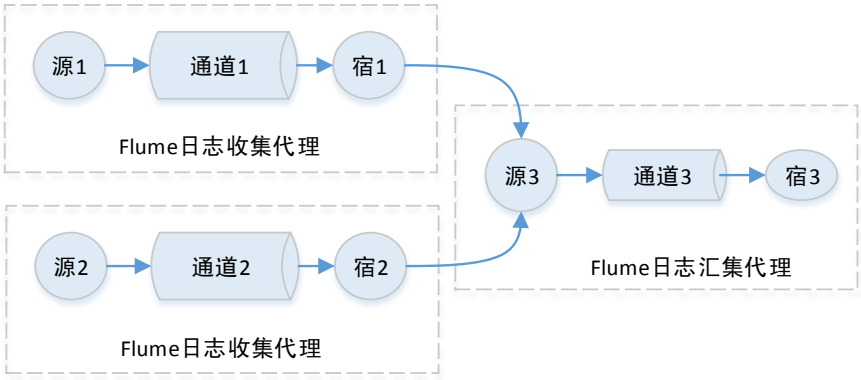


图 2　日志收集模块

Fig. 2　Log collection module

Flume 的“代理”是数据流的基本单元，由源、通道、宿三部分组成。图 2 中的源 1 和源 2 属于 Exec 类型，用于监听日志数据文件，当文件末尾有新增数据时，Flume 将实时获取到新增的部分[6]。通道实现了消息队列的功能，能够缓存到内存中，也可以实现数据的持久化。宿 1、宿 2、源 3 属于 avro 类型，它们之间的数据传输基于 avro 事件，可实现跨节点的远程传输，通过配置地址及端口实现数据的发送和接收。宿 3 是自定义类型，作为一个 Kafka 消息生产端，实现了将 Flume 收集到的日志数据发送给 Kafka 的功能。

1.2　日志分发模块

现有的消息队列系统如 ZeroMQ 等，能够很好的处理实时或者近似实时的应用，但未处理的数据通常不会写到磁盘上，这对于 Hadoop 或 Spark 之类的离线应用而言，可能存在问题。Kafka 是一种基于发布/订阅的消息系统[7]，支持 Hadoop 或 Spark 数据并行加载，对于像 Hadoop 或 Spark 的一样的日志数据和离线分析系统，但又要求实时处理的限制，这是一个可行的解决方案。Kafka 实现了以时间复杂度为 O(1)的方式提供消息持久化能力，即使对 TB 级以上数据也能保证常数时间复杂度的访问性能。Kafka 还是一个完全的分布式系统，代理服务器、消息生产端、消息消费端都原生自动支持分布式，自动实现负载均衡，三者的数量可以是一个也可以是多个。多个代理服务器、消息生产端、消息消费端可以运行在一个大的集群上，作为一个逻辑整体对外提供服务。并且通过使用 ZooKeeper（一个分布式应用管理框架）来解决分布式应用中的数据管理等问题。Kafka 基于 Java 开发，能够跟诸多其他框架友好地衔接，结构灵活，扩展性好，优于诸如使用 Erlang 编写的非常重量级的 RabbitMQ。本文中日志分发模块的架构图如下：

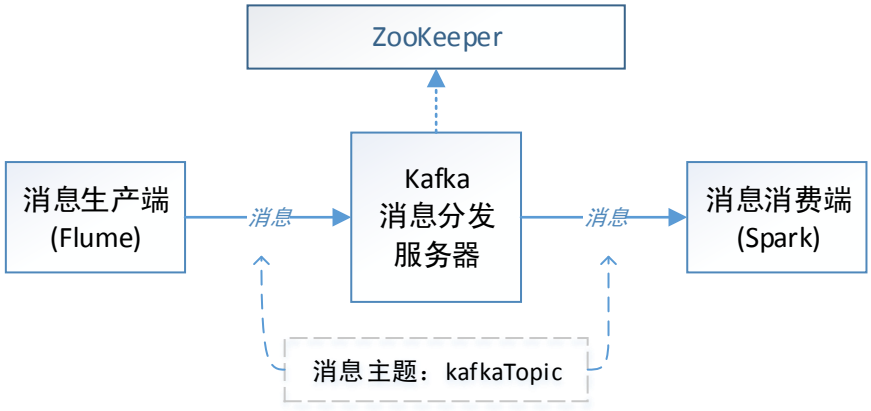


图 3　日志分发模块

Fig. 3　Log dissemination module

1.3 日志分析模块

Spark 是一个高效的分布式计算系统，利用其基于内存的特点，特别擅长迭代式和交互式数据处理[8]。Spark Streaming 是建立在 Spark 上的实时计算框架，通过 Spark 提供的丰富的 API、基于内存的高速执行引擎，可以实现可扩展的、高吞吐的、具有容错性的实时数据流处理[9]。它将输入数据按照一定的批量大小分成 DStream（离散数据流），每一段都转换成 Spark 中的 RDD（弹性分布数据集），然后将 Spark Streaming 中对 DStream 的转换操作变为针对 Spark 中对 RDD 的转换操作，将 RDD 经过操作变成中间结果保存在内存中。整个流式计算根据业务的需求可以对中间的结果进行叠加，或者存储到外部设备。对于流式计算来说，容错性至关重要。每一个 RDD 都是一个不可变的分布式可重算的数据集，其记录着确定性的操作继承关系，只要输入数据是可容错的，那么任意一个 RDD 的分区出错或不可用，都是可以利用原始输入数据通过转换操作而重新算出的。这保证了 Spark Streaming 的容错性。本文的日志分析模块数据处理流程图如下：

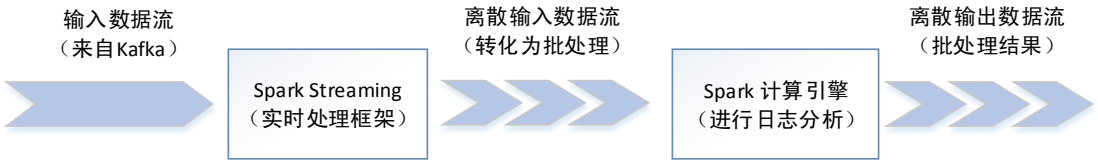


图 4 日志分析模块

Fig. 4 Log analysis module

2 系统测试

2.1 实验环境

表 1 实验硬件及软件环境

Tab. 1 The hardware and software environment of experiment

虚拟机节点	节点 1	节点 2	节点 3
硬件配置	内存：4G 硬盘：100G 处理器：双核	内存：2G 硬盘：100G 处理器：双核	内存：2G 硬盘：100G 处理器：双核
操作系统	ubuntu-12.04.5-amd64	ubuntu-12.04.5-amd64	ubuntu-12.04.5-amd64
软件版本	spark-1.3.1-bin-hadoop2.4		
	kafka_2.10-0.8.2.1		
	apache-flume-1.5.2-bin		
集群角色	主节点、计算节点	备份主节点、计算节点	计算节点

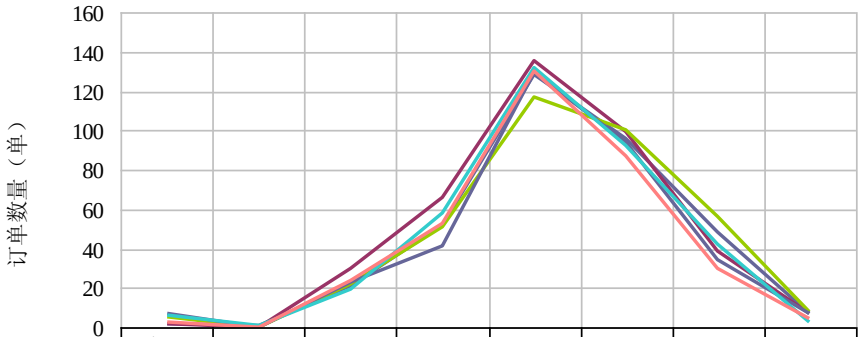
2.2 实验数据

实验数据来源于阿里巴巴天池数据实验室。该数据集包含“天猫”匿名用户在“双 11”前六个月的购物记录。本文在试验中使用该数据集模拟企业营业日志数据，利用交易日志中信息进行分析，获取订单量与用户年龄、性别的相互关系。试验中利用脚本将数据转化为高并发的数据流发送到该系统中，然后对日志流数据进行实时分析。

2.3 实验结论

实验过程中，对日志流数据进行的实时分析，获得如下实验结果。

2.3.1 订单数量与客户年龄的关系



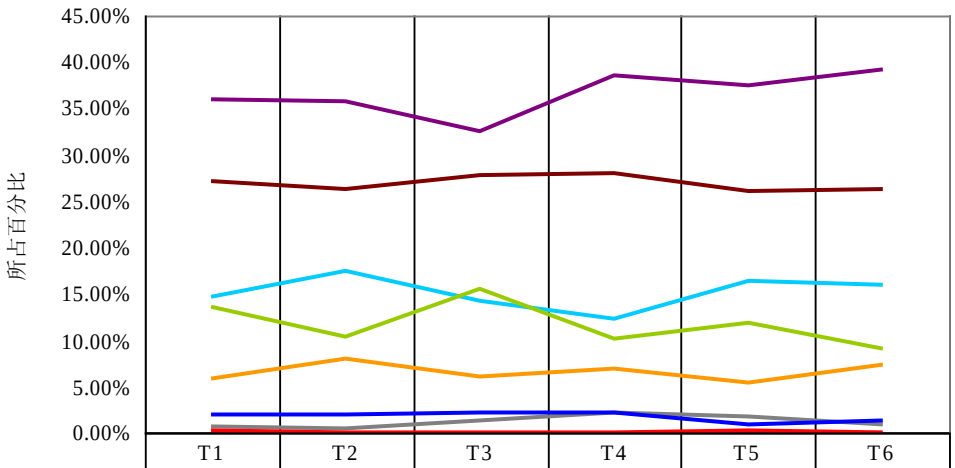
	未知	<18	[18,24]	[25,29]	[30,34]	[35,39]	[40,49]	>=50
T1	2	1	21	52	128	96	48	7
T2	2	0	30	66	135	99	39	7
T3	5	0	22	51	117	100	56	8
T4	7	0	23	41	129	94	34	7
T5	6	1	19	58	132	92	42	3
T6	3	0	24	53	130	87	30	4

年龄分布（岁）

图 5 客户订单数量关于年龄的分布

Fig. 5 The age distribution of customer order quantity

从获得的实验结果数据，我们可以清晰地看到订单数量关于年龄段的分布情况。客户主要分布在 18 岁到 49 岁这个年龄范围内，其中 30 岁到 34 这个年龄段的用户订单数最多。再根据这些实验数据我们可以很快的得出各年龄段所在比例。



	T1	T2	T3	T4	T5	T6
未知	0.56%	0.53%	1.39%	2.09%	1.70%	0.91%
<18	0.28%	0.00%	0.00%	0.00%	0.28%	0.00%
[18,24]	5.92%	7.94%	6.13%	6.87%	5.38%	7.25%
[25,29]	14.65%	17.46%	14.21%	12.24%	16.43%	16.01%
[30,34]	36.06%	35.71%	32.59%	38.51%	37.39%	39.27%
[35,39]	27.04%	26.19%	27.86%	28.06%	26.06%	26.28%
[40,49]	13.52%	10.32%	15.60%	10.15%	11.90%	9.06%
>=50	1.97%	1.85%	2.23%	2.09%	0.85%	1.21%

取样时间点

图 6　各年龄段客户订单数量所占比例

Fig. 6　The proportion of every age group

2.3.2　订单数量与客户性别的关系

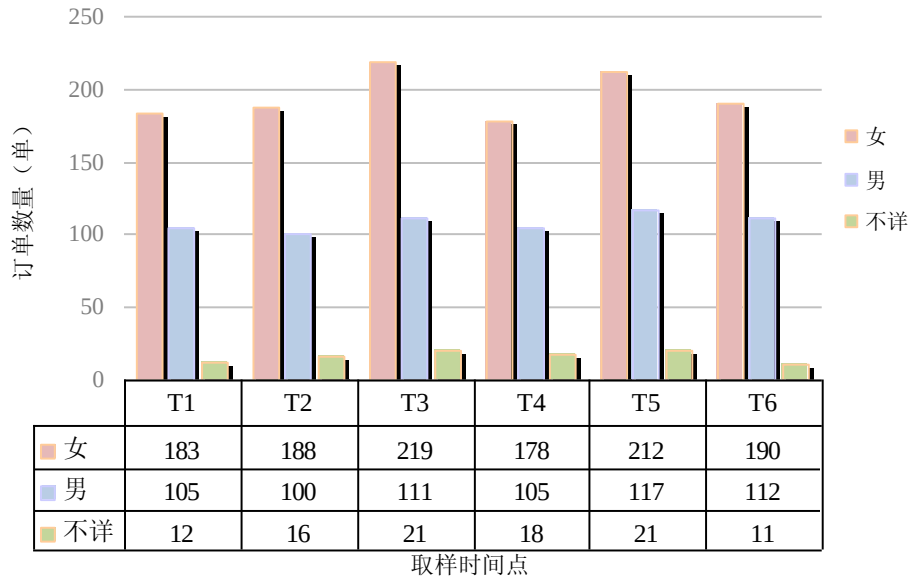


图 7　客户订单数量关于性别的分布

Fig. 7　The gender distribution of customer order quantity

从获得的实验结果数据，我们可以清晰地看到订单数量关于客户性别的分布情况。女性客户的订单数量明显多于男性客户的订单数量。同样，我们也可以根据这些实验数据得出性别数据对应的比例。

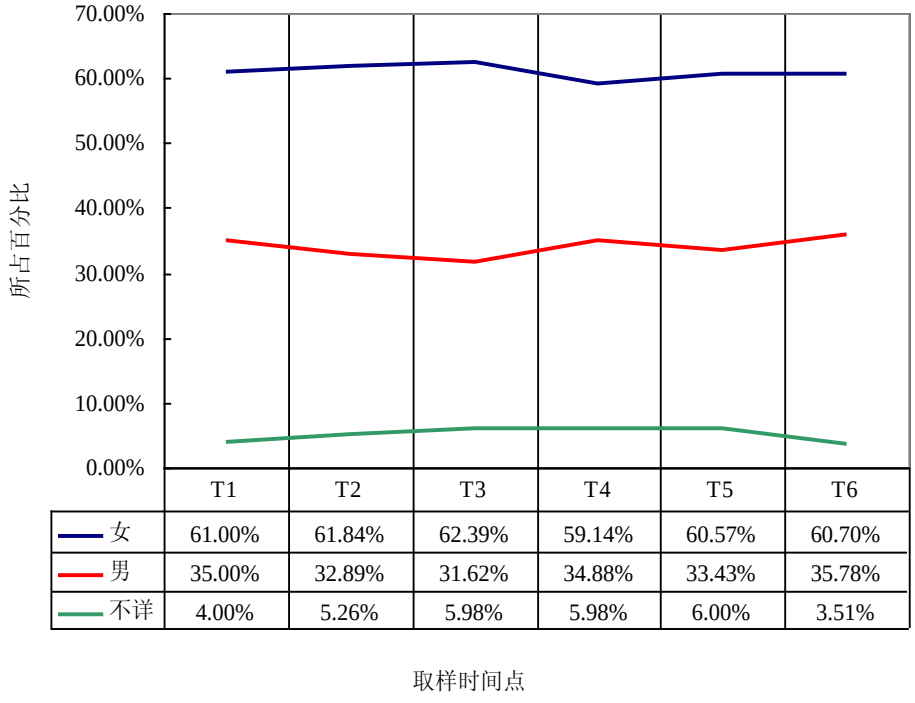


图 8　不同性别客户订单数量所占比例

Fig. 8　The proportion of every gender group

3 结论

本文给出了一种基于 Flume/Kafka/Spark 的分布式日志流处理系统。能够实时高效准确地处理大规模的日志数据。利用 Flume 的高扩展性和高可靠性，可将多个数据源的日志数据准确地收集汇总，并且易于新增扩展；Kafka 的高吞吐、可扩展、分布式的特性满足了海量日志数据的分发需求；Spark Streaming 提供了一套高效、可容错的实时大规模流式处理框架，能够迅速地对大量的日志流数据进行分析计算。这一分布式日志流处理系统架构新颖，可实现对日志流数据收集、分发、分析的整体处理需求。企业可根据该分析结果采取调整广告投放目标人群等相关措施。该系统高效准确地完成了对日志流数据的处理，通过对订单日志的分析，提取相关信息，可帮助企业调整营销策略，提高业绩。

[参考文献] (References)

[1] 崔星灿,禹晓辉,刘洋,吕朝阳. 分布式流处理技术综述[J]. 计算机研究与发展,2015,02:318-332.
[2] 卓海艺，赵文深，吕玉琴等. 基于 Hadoop/CloudBase/MySQL 的日志分析系统的设计与实现[OL]. 中国科技论文在线 http://www.paper.edu.cn//releasepaper/content/4503064
[3] Xiuqin LIN,Peng WANG,Bin WU. LOG ANALYSIS IN CLOUD COMPUTING ENVIRONMENT WITH HADOOP AND SPARK[A]. IEEE Beijing Section.Proceedings of 2013 5th IEEE International Conference on Broadband Network & Multimedia Technology[C].IEEE Beijing Section:,2013:4.
[4] 王润华. 基于 Hadoop 集群的分布式日志分析系统研究[J].科技信息,2007,15:60-109.
[5] 陈芳芳. Flume NG：Flume 发展史上的第一次革命[OL]. [2014-04-08]. http://www.ibm.com/developerworks/cn/data/library/bd-1404Flumerevolution/index.html
[6] Apache Flume. Flume Homepage[OL]. http://flume.apache.org/.
[7] Apache Kafka. Kafka Homepage[OL]. http://kafka.apache.org/.
[8] Apache Spark. Spark Homepage[OL]. http://spark.apache.org/.
[9] 王家林. 大数据 Spark 企业级实战[M]. 北京：电子工业出版社，2015.