第一章 大数据背景

概要

当今世界数据量的产生已经巨大，这些数据中包含了人类社会各种各样的信息，这就是所谓的大数据。考虑到数据的生成量，实时处理成为了许多机构要面对的挑战。现在许多公司用来进行大数据分析的工具是hadoop，而hadoop通常用作静态数据的大批量处理。另一个非常有效的开源实时计算工具就是storm。Storm由twitter公司开发，现在已经被许多机构采用。Storm通常被比作“实时的hadoop”，然而storm远比hadoop来的简单，因为用它来处理大数据不会带来新老技术的交替。

当前互连网飞速发展，许多公司的日常运营会产生TB级的数据，数据的增长速度也越来发挥快。据了解，（素材）。数据的来源包括了互连网装置可以捕获的所有类型数据，网站、社交媒体、交易型商业数据以及其他商业环境中创建的数据。

Storm的出现背景

过去10年中，伴随互连网的高速发展，企业产生和积累的数据越来越大，越来越多。随着Hadoop、Google MapReduce等相关技术的出现，对大数据的处理变得容易起来。但是这些系统是针对大批量数据的处理，并不是实时的。而实时的计算系统和批处理系统的模型也有着本质区别。随着企业业务的快速增长，针对大规模数据的实时（real-time）处理变得越来越有必要。缺少“实时的大数据计算系统”已经成为整个大数据生态系统中的一个巨大缺失。Storm，正是在这样的背景下出现的，storm也能很好应对这一需求。以往对于实时计算的大数据任务，开发者需要手动维护一个消息队列和消息处理者组成的实时网络。消息处理者从消息队列中取出消息进行处理，然后更新数据库，再发送消息给其他队列。所有这些操作都需要开发者自己实现，这对开发者而言是巨大的负担。开发者需要花费大量的精力去维护消息如何发送，消息发送到哪里，如何部署消息的处理者，如何部署消息处理的中间节点等。而使用storm进行开发，只需要很少的消息处理逻辑，就可快速开发一个实时计算系统。这样开发者就可以专注于业务逻辑的开发，从而提高了大数据实时系统的开发效率。

Storm是什么

全量数据处理使用最多的是hadoop或者hive。Hadoo作为一个批处理系统，以吞吐量大、自动容错等优点，在海量数据处理上得到了广泛的使用。

Hadoop框架对于数据的处理流程

Hadoop是磁盘级计算，也就是说，进行计算时，数据在磁盘上，系统进行运算需要进行磁盘io操作。

相比hadoop的批处理，storm是个实时的、分布式以及具备高容错的计算系统。同hadoop一样storm也可以处理大批量的数据，然而storm处理数据更加实时，并同样保证数据计算的高可靠性。数据的实时计算保证所有的信息都会被处理。Storm还具备容错和分布计算的特性，这让storm可以部署到不同的机器上进行大批量数据处理。

Storm与hadoop区别

Hadoop是基于磁盘进行计算，需要进行磁盘读写；storm基于内存计算，数据通过网络导入内存。Storm的网络直传、内存计算，其时延必然比hadoop通过HDFS传输低得多，当计算任务适合用流式模型时，宜选用storm的流处理。Storm的流式处理，省去了批处理的数据收集时间，同时也省去了作业调度的时延。所以从时延上来说，storm要快于hadoop。

从原理上讲，hadoop基于HDFS，需要进行切分数据、产生中间文件、排序、数据压缩、数据复制等，效率较低。Storm基于ZeroMQ这个高性能消息通讯库，不会持久化数据。

在消耗资源相同的情况下，一般来说storm的延时低于mapreduce，但是吞吐量也低于mapreduce。Storm是典型的实时流计算系统，mapreduce是典型的批处理系统。

总的来说，hadoop使用磁盘作为数据交换的媒介，storm的数据是一直在内存中流转的。两者面向的领域也不同，hadoop是面向批量处理的，吞吐量大，基于任务调度；storm是实时处理，基于数据流。以水为例，hadoop可看作纯净水，要一桶一桶地搬；而storm是水管，只要预先接好水管，用的时候打开水龙头，不就源源不断地出来了。在这个类比中，“水”就是我们需要处理的数据，“桶装纯净水”和“水管”就是数据的来源。

数据处理流程

数据处理流程大致分三个阶段：

1. 数据采集与准备。目前典型的处理策略，数据一般出自页面打点和解析db的log。流计算系统将数据采集到消息队列，批处理系统一般将数据采集进分布式文件系统（比如HDFS ），有的也使用消息队列。我们暂且将消息队列和文件系统称为预处理存储，二者在延时和吞吐上没有太大差别。接下来从这个预处理存储到数据计算阶段有很大区别，流计算实时读取消息队列，将数据放入流计算系统进行相应运算。批处理系统一般人积累一大批数据批量导入到计算系统，这里就有了延时的区别。
2. 数据计算，涉及中间计算媒介，通过中间媒介导入数据。Hadoop使用磁盘作为数据交换媒介，storm通过网络将数据直接导入内存，使用内存作为媒介。Storm里程是常驻的，有数据就可以进行实时的处理。Storm每个计算单元之间通过网络直接进行数据传输，而mapreduce进行map任务后运算的结果要写入到HDFS，相对来说多了磁盘读写，比较慢。对于复杂运算，storm的运算模型直接支持DAG（有向无环图）。
3. 数据结果展现。流计算的计算结果一般直接反馈到最终的结果集中（展示页面，数据库，搜索引擎索引），而mapreduce一般需要整次运算结束后将结果批量导入到结果集中。实际上，流计算和批处理系统并没有本质的区别，像storm的trident也有批量概念，而mapreduce可以将每次运算的数据集缩小（比如每隔几分钟启动一次），facebook的puma就是基于hadoop做的流计算系统。

下面对storm的特性作具体分析

易于扩展：一般来说，商业机构希望系统是可以扩展的，因为伴随业务发展，公司运营产生的数据量会越来越大，数据的计算量相应地也越来越大。为了处理更多的数据，需要数据处理系统可以动态扩展，方便地增加机器来应付更多的计算量。Storm使用hadoop zookeeper来进行集群协调，这样可以保证集群系统的良好运行。

实时：每条信息都进行处理。系统的设计也保证了消息能够得到快速处理，使用ØMQ作为底层消息队列。

基于内存：storm是内存级计算，数据通过网络或其他途径导入内存。计算机读写内存的速度相比磁盘快几个数量级。进行磁盘的延迟约为内存访问的75000倍，因此，storm进行实时处理速度快得多。

易于管理：storm的集群管理非常简单和方便。

简单的编程模型：类似于MapReduce降低了并行批处理的复杂性，storm降低了实时处理的复杂性。

可以使用各种编程语言：storm默认支持Clojure、Java、Ruby和python，要增加对其他语言的支持，只需实现一个简单的storm通信协议即可。

可靠的消息处理：storm保证每个消息至少得到一次完整处理，任务失败时，会从消息源重试消息。

容错机能：storm的topology一旦进行提交，storm就会一直运行直到topology废弃或关闭。当在执行时出现错误时，storm会重新分配任务，也就是说，集群中的一个节点挂了不会影响应用进程。Storm也会管理工作进程和节点的故障。

低延迟：storm是实时计算系统，相比批处理系统延迟低很多。

本地模式：storm有一个“本地模式”，可帮助用户完全模拟storm集群，这可让你进行快速开发和单元测试。

Storm的流处理

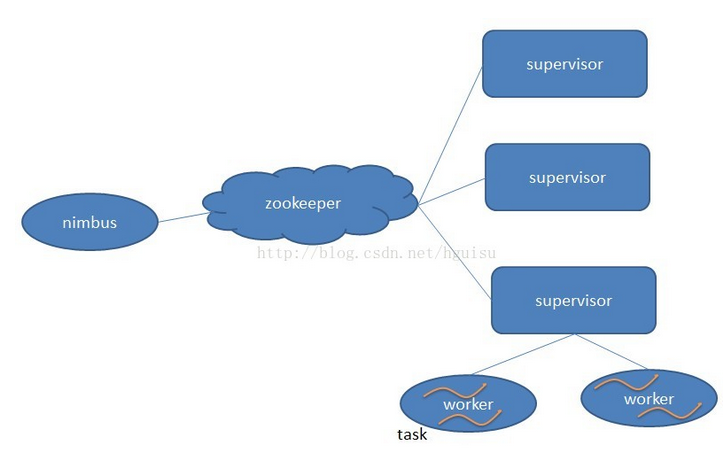
对于处理大批量数据的map/reduce程序，任务完成之后就停止了，而storm是用于实时计算的，当数据处理任务完成后，程序并不会停止，而是会等待任务，当有新的任务时，则再次执行，这个过程不断循环，直至手动终止程序。

Storm是针对在线业务的计算平台，如统计用户的交易量、生成某个用户的推荐列表等实时性较高的需求。因此，storm是一个“流处理”框架。什么是流处理？就是说，storm将数据以stream流的方式，按照topology顺序，依次处理并最终生成计算结果。

Storm的组件

Storm集群表面上看与hadoop集群很类似，在hadoop上运行的是mapreduce任务，而storm上运行的是拓扑（topology）。这两者是很不一样的，其中一个关键区别是：mapreduce任务最终会结束，而topology会一直执行下去（除非手动停掉）。

Storm集群由一个主节点和一群工作节点组成，主节点通常为nimbus后台程序，工作节点称为supervisor，它们通过zookeeper进行集群协调。



主节点（master node）

主节点运行一个守护进程（daemon），Nimbus，它负责在集群中分发代码，分配工作任务并监测失败情况。Nimbus类似hadoop里的JobTracker。

工作节点（worker node）

工作节点同样运行一个守护进程，它运行的守护进程为Supervisor。Supervisor监听工作任务，根据要求运行工作进程。每个工作节点执行一个拓扑（topology）的子集。

Zookeeper

Zookeeper负责保持主节点和工作节点的协调工作。实时应用程序的逻辑则被封装进storm的拓扑（topology）。Storm中的拓扑是一组由spouts（数据源）和bolts（数据操作）通过stream groupings进行连接的图。下面对这些概念进行更加深入的分析。

Spout

简单来说，spout从数据源读取数据供storm拓扑使用，一个spout可以是可靠的，也可以是不可靠的。可靠的spout保证在数据处理失败时将元组（tuple）进行数据重发。不可靠的spout在数据发送后并不追踪元组数据。在Spout中最主要的方法为nextTuple()，这个方法向拓扑发送一个新的元组数据或者在没有数据时直接返回。

Bolt

Bolt负责storm拓扑中所有的处理过程。Bolt可以做任何事情，比如：连接过滤，数据聚合，与文件、数据库通讯，等等。Bolt从spout接收数据并处理，也可能在遇到复杂流处理时进一步将元组数据发送给另外一个Bolt。Bolt中最主要的方法是execute()，它接收一个元组（tuple）作为输入。不管是在spout还是在bolt中，将一个元组发送给多个流，都可以在declareStream()中指定这些流。

Stream grouping

流分组（stream grouping）定义了流在bolt任务中是如何切分的。Storm提供了内置的流分组方法，包括：随机分组（shuffle grouping）、字段分组（fields grouping）、全部分组（all grouping）、全局分组（global grouping）、无分组（none grouping）、直接分组（direct grouping）。还可以实现CustomStreamGrouping接口来定制自己需要的分组。

实现

在本项目中，设计了一个由spout和bolt组成的拓扑，可以处理海量数据（日志文件），日志文件中包含分布式服务器集群所有进程和服务的运行状况（stat），根据这个stat日志信息，实时读取并交由storm拓扑进行分析处理，最终得到有价值的分析结果。