|  |  |  |
| --- | --- | --- |
| 技术交底书  案号：  日期： | 申请人： |  |

|  |  |
| --- | --- |
| 交底书名称 | **一种面向大数据集群的日志和监控存储机制** |
| 发明人 |  |
| 交底书撰写人 | 孙杰 |
| 撰写人办公电话 | 15165285929 |
| 撰写人手机号码  （急案必须填写） | 15165285929 |
| 撰写人E-mail | fantom1001@163.com |

|  |  |
| --- | --- |
| 撰写指导意见  (IRP填写) |  |

缩略语和关键术语定义

**数据中心**：指用于安置计算机系统及相关部件的设施，例如电信和储存系统。一般它包含冗余和备用电源，冗余数据通信连接，环境控制（例如空调、灭火器）和各种安全设备。

**服务器集群**：服务器集群就是指将很多服务器集中起来一起进行同一类服务。集群可以利用多个计算机进行并行计算从而获得很高的计算速度，也可以用多个计算机做备份，从而使得任何一个机器坏了整个系统还是能正常运行。

**任务调度节点（主节点）**：一个服务器集群的中央节点，负责资源分配、任务调度和状态监控。

**工作服务器**：也称工作节点，负责执行具体的任务。

1. **相关技术背景**

**1.1背景技术：**

近年来，物联网、云计算和移动设备迅速普及，各行各业对于大数据汇聚和处理的需求逐渐显现。为更快获取处理结果，人们通常部署和维护用于处理大数据任务的服务器集群（简称大数据集群），并通过集群管理软件分发任务、监控资源使用情况，通过集群服务器的协同工作，提高效率。在大数据集群的管理和使用过程中，如何获取和维护整个集群的海量监控数据是一个不可避免的难题。同时，同一服务器上存在监控数据重复采集的问题，这从长远来看，造成了巨大的资源浪费。

大数据集群的监控数据主要记录了集群中所有的服务器运行状态，具体包括**服务器工作日志**、**服务器资源使用量监控数据**、**任务完成情况监控数据**等几个部分。这些监控数据有不同的特征，根据其特征被分布在工作服务器或集中式任务调度节点上。

一个大数据集群通常由一个或几个主节点，以及大量的工作节点构成。运行状态监控数据记录了集群中所有服务器的实时运行情况，多由主节点采集和存储。服务器工作日志记录了服务器执行任务的详细信息，包含了程序级的调试数据、警告数据、错误数据等输出，多为文本形式且数据量较大，因此通常存放在服务器本地磁盘；在需要做日志分析的时候，日志分析工作通常也在服务器本地完成。服务器资源使用量监控数据记录了服务器的CPU、内存、磁盘、网络等主要资源的实时用量，为任务资源申请和调度提供依据。任务完成情况监控数据记录了某项大数据处理任务在工作节点的工作状态。综上所述，大数据集群的监控数据表现多元化，其来源、数据量、数据语义各不相同，某些数据反映集群的长期工作状态，对其存储和处理具有重要的意义；而其他一些数据的价值较低，长期存放和处理的性价比不高。因此，将不同类型的监控数据区分对待，根据监控数据的特性分别对其进行存储和处理，可降低大数据集群的维护开销，合理发挥监控数据的总体价值。

1. **本发明技术方案的详细阐述（发明内容）**

本发明公开一种面向大数据集群的监控数据采集和存储方法。其中，该方法涉及：根据其来源或用途对监控数据进行了分类；根据不同类型监控数据的价值确定的数据采集方式；根据不同类型监控数据的规模确定的数据存储方式和策略；一种大数据集群监控模块的设计方法等。

**2.1、工作节点监控数据**

一个大数据集群包括了分布式的大数据服务器，采用主-从结构搭建，由主节点提交任务，由工作节点执行任务，任务执行过程中可通过主节点查询任务状态，任务完成后通过主节点汇集和展示结果。目前大数据集群的监控数据来源多样化，监控数据采集的发起者较多，造成了数据的重复采集。例如，重要监控数据在服务器操作系统日志中进行了记录，但是某些系统程序仍然会记录自己产生的日志。同时，虽然操作系统日志粒度较细，但是由于数据量较大，不利于快速分析。因此，对关键监控数据进行个性化的记录，以便进行快速的监控数据分析，具有重要的意义。随着持续的运行，大数据集群中每一台服务器上产生的日志是不断积累的，过期的日志会持续存储，造成了存储空间资源的严重浪费。



针对这些问题，本发明将工作节点本地的监控数据分为系统日志、任务日志和资源使用量日志三个部分：

（1）系统日志。系统日志存放在固定的目录，粒度较细，记录服务器的运行状态、资源使用量现状、工作状态及所有应用程序的运行状态。系统日志是系统管理员进行系统运维的重要依据。特别的，本发明按照小时为单位组织系统日志文件，即每小时生成一个新的系统日志文件，用于存储此后一小时的所有系统日志；同一服务器上的所有系统日志存放在同一个目录下。

（2）任务日志。大数据集群含有大量的工作节点服务器，工作节点需为其执行的大数据处理任务维护日志，即为任务日志。特别的，本发明按照任务为单位组织任务日志，即为每一个任务生成一个任务日志文件。同一服务器上的所有任务日志存放在同一个目录下。

（3）资源使用量日志。大数据集群中工作节点的资源使用量实时状态是大数据集群进行任务分配和资源调度的依据。虽然系统日志中可能记录资源使用量的现状，但是由于此项信息使用频繁，因此本方法将资源使用量冗余记录，单独存放冗余的资源使用量日志可支持高效的日志处理。

|  |  |
| --- | --- |
| 日志类型 | 日志特征 |
| 系统日志 | 实时更新，按小时生成新的日志文件 |
| 任务日志 | 实时更新，按任务生成新的日志文件 |
| 资源使用量日志 | 实时更新 |

工作节点中上述三种不同类型的日志特征明显，根据这些特征，设计了高效的服务器日志管理机制，提出了本地日志的存储和处理系统。包括以下几个模块：

（1）本地日志预处理模块

本地日志预处理模块的任务是从本地日志中提取关键数据并存储。通过下面（3）所述的数据查询接口，对主节点提供监控数据。

上述关键数据包括：

①服务器运行状态及资源使用情况。记录服务器运行时长、服务器CPU使用情况、服务器内存使用情况、服务器网络使用情况、服务器磁盘使用情况、服务器缓存占用情况及缓存命中情况。

②服务器当前任务执行情况。记录服务器当前的任务状态（包括正在运行的任务数量及任务的运行进度）。

③服务器历史任务执行情况。记录服务器此前执行的任务的总数、成功完成的任务的数量、失败的任务的数量等情况。

④本地日志中，级别为“ERROR”的报警信息文本（细粒度）或个数（粗粒度）。

（2）本地日志压缩模块

随着大数据集群的运行，本地日志逐渐累积，占用大量存储空间。本发明根据本地日志的使用规律，设计了本地日志压缩策略，通过本地日志压缩模块删除不常用日志，节省存储空间。对于已成功运行完的任务，与其相关的详细日志信息被再次使用的频率极低，而对于正在运行的任务，其相关的进度信息和运行输出可能被频繁利用，用以估计任务执行进度和任务执行状态。因此，本地日志压缩模块应当利用日志使用的规律，删除一定时间之前的历史日志，回收日志存储空间。

需要指出的是，对于本地日志，根据集群的配置，每隔48小时进行一次日志压缩任务。压缩模块直接将48小时之前的日志文件删除。**而对于本地日志预处理模块产生的关键数据**，压缩模块将每隔48小时定期将其打包、使用压缩算法压缩后，将其拷贝到“本地历史日志目录”。

（3）数据查询接口

工作节点接受来自主节点的数据查询请求，将本地存储的重要监控信息通过数据查询的返回值的方式，反馈到工作节点。数据查询接口采用远程过程调用（RPC）的方式或将其设置为符合RESTful设计规范的API的方式，由集群主节点暴露接口，提供给集群管理员和集群用户直接调用。

**2.2、主节点监控数据**

主节点存储了整个集群的服务器状态及任务状态等监控数据。与工作节点的本地数据相比，主节点存储的数据粒度较粗，但是这些数据将被用来直接呈现给用户的数据，因而具有很高的价值。由于大数据集群中的工作节点数量一般很多，因此监控数据量较大。对主节点的监控数据存储和处理方法进行优化设计的主要目标是提高监控数据检索效率。针对此，本发明提出以下几个方面的设计方案：

（1）工作节点监控数据获取机制

主节点的后台进程根据预定的数据获取时间间隔，使用工作节点提供的数据查询接口，从分布式工作节点拉取监控数据。默认的监控数据拉取频率为1分钟/次。主节点需对所有有效的工作节点拉取数据，而不同工作节点的数据拉取互不干涉，因此这个过程是可并发执行的。

（2）多粒度监控数据存储方法

为减少需要存储的数据量，设计了多粒度监控数据存储方式。监控数据按照其粒度分别存储在不同的数据表中。例如，若采用默认的监控数据拉取频率，即每分钟拉取一次监控数据，则：分钟级监控数据存放在分钟表中，采用循环存储的方式进行存储——即按照某条监控数据属于第几分钟，将其存储到对应序号的监控记录里；同时满足每台工作节点对应的监控数据在分钟表中的最大记录条数为60条（0号-59号记录即对应到小时中的第0分钟-第59分钟），下一小时某指定分钟的监控记录会覆盖上一小时相同分钟的监控数据，例如2:39分的监控数据会覆盖1:39分的监控数据。此外，还有小时表，日表，月表等几个不同粒度的监控数据存储表，其数据存储原理相同。主节点从分布式工作节点拉取到监控数据后，首先存放到分钟表中。预设的监控数据拉取频率（默认1分钟）即为最细的监控数据存储粒度。然后，在整点时刻，即当前时刻表盘的分针为0时，（例如1:00,2:00）计算分钟表中当前监控指标的各项平均值，将其填入对应的小时表记录中；在整天时刻，计算小时表中当前监控指标的各项平均值填入对应的日表记录中；在整月时刻，计算日表中当前监控指标的各项平均值填入对应的月表记录中。月表为最粗粒度的数据。按照这种组织方式，将主节点上的监控数据的总体数量大幅压缩。

此外，为展示当前集群总体资源利用率，主节点会单独计算当前集群中所有工作节点资源利用率的平均值，作为当前集群总体资源利用率数据，并将其单独存储。此外，由于总体资源利用率通常不需要细粒度的监控数据，因此仅设计小时级、日级和月级三种粒度的总体资源利用率数据。

（3）多粒度监控数据展示方法

监控数据的展示方式分为两种：第一种为单独展示指定的某一工作节点资源监控数据，第二种为展示当前集群总体资源监控数据。

监控数据的展示的粒度与监控数据存储的粒度一一对应，即工作节点资源利用率的监控数据展示粒度分为分钟级、小时级、日级、月级等四个等级，集群总体监控数据展示粒度分为小时级、日级、月级三个等级。

**3、本发明所要解决的技术问题**

第一点：缓解大数据集群工作节点日志的存储开销

大数据集群工作节点的日志日积月累、不断产生，这包括系统日志、任务日志等多个部分，其中由于配置不合理，某些日志可能还被重复存储，在多个日志文件中冗余出现，这产生了巨大的存储开销，若不加处理，可能造成大数据集群工作节点的存储空间耗尽。

第二点：提供基于日志的监控数据提取方法

大数据集群的工作节点的重要监控数据混合分布在大量的日志记录中，针对这种监控数据的有效提取和组织方式是提高监控数据获取效率的重要问题。此外，本发明还指定了监控数据在集群中的传输机制，即通过主节点定期查询的方式，从工作节点传输到主节点之间的数据传输机制。

第三点：提高主节点对集群运行状态的存储和展示效率

大数据集群通过主监控节点，向集群资源调度模块或集群管理员提供对整个集群运行状态的监控数据。在大规模的大数据集群中，工作节点的数量巨大（几百台至上千台），如何在主节点上高效的存储整个集群的监控数据，用来反映集群当前的资源使用情况，是本发明关注的问题。此外，如何向集群管理员提供当前集群运行状态监控的概览和细致数据，也是本发明要解决的问题。

**4、 本发明的优势**

1.设计了大规模集群中工作节点监控系统的架构，设计了工作节点日志存储和去冗余的方式，在降低了长期日志存储的存储空间开销的同时，确保了重要监控数据的完好保存。

2.设计了从工作节点日志中提取并单独存储重要监控信息的工作方式，通过程序接口将重要监控数据提供给主节点，集群中的主节点可调用工作节点的数据获取接口，获取到指定工作节点的重要监控数据。

3.为应对集群规模较大时，主节点存储大量监控数据的压力，设计了一种多粒度的循环存储监控数据的方法，可显著提高大规模集群监控数据存储、展示和检索的效率。