(19) 中华人民共和国国家知识产权局



(12) 发明专利申请



(10)申请公布号 CN 106445403 A (43)申请公布日 2017.02.22

- (21)申请号 201510489865.3
- (22)申请日 2015.08.11
- (71) 申请人 张一凡 地址 271100 山东省莱芜市莱城区风城西大 街 191 号
- (72) 发明人 张一凡
- (51) Int. CI.

G06F 3/06(2006.01) *G06F 17/30*(2006.01)

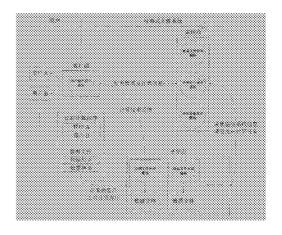
权利要求书2页 说明书6页 附图4页

(54) 发明名称

针对海量数据成对存储的分布式存储方法和 系统

(57) 摘要

本发明涉及一种针对海量数据成对存储的分布式文件存储系统,包括基于主从式分布式集群的存储系统设计,主结点功能模块,子结点功能模块及存储及计算流程设计。该分布式存储系统旨在应用分布式文件存储系统高效处理满足数据集文件比对模式的计算问题。本发明通过将文件存储及文件处理协同考虑,从数据存储初始阶段即满足后续比对计算的需要,实现了对于同一计算任务,一次存储,无需移动的数据文件存储及处理解决方案;此分式文件系统同时兼顾降低存储空间,保证数据文件可靠性的存储需要。此外,本发明提出的布式文件存储系统与硬件无关,可以方便的在不同分布式机群环境中实现。



CN 106445403 A

- 1. 一种针对海量数据成对存储的分布式文件存储系统,其特征在于,该分布式文件系统包括基于主从式分布式集群的存储系统设计,主结点功能模块,子结点功能模块及存储及计算流程设计。该分布式存储系统旨在应用分布式文件存储系统高效处理满足数据集文件比对模式的计算问题。本发明通过将文件存储及文件处理协同考虑,从数据存储初始阶段即满足后续比对计算的需要,实现了对于同一计算任务,一次存储,无需移动的数据文件存储及处理解决方案;此分布式文件系统同时兼顾降低存储空间,保证数据文件可靠性的存储需要;同时此分布式文件系统支持通用的分布式集群结构,具备很高的扩展性。
- 2. 根据权利要求 1 所述的一种针对海量数据集比对计算的分布式文件存储系统,其特征在于,所述的分布式文件存储系统包括以下三个独立组成部分:用户客户端、系统主结点及系统子结点。所述每个组成部分都具有不同的功能:所述客户端为用户提供基本操作接口,包括数据文件管理,计算程序管理及计算结果采集等。所述系统主结点负责收集用户提供的数据集信息及相应计算任务信息,确定数据文件部署方案。所述系统主结点同时采集系统各子结点信息,确定相应计算任务调度。所述系统子结点同时具备数据文件存储及数据对比处理计算功能。依据主结点的调度信息,所述的分布式文件系统各子结点以数据对为单位存储各相应的数据文件,并进行数据文件的比对计算任务,最终计算任务全部结果在所述主结点的管理下汇总至所述用户客户端。
- 3. 根据权利要求 1 所述的一种针对海量数据集比对计算的分布式文件存储系统,其特征在于,所述主结点及子结点中主要包括以下五种功能模块:数据文件管理模块,计算任务调度模块,数据文件处理模块,系统资源监控模块和用户操作接口模块:

其中数据文件管理模块在用户在提交所有数据文件后,数据文件管理模块根据相应数据处理任务信息,分布式文件系统中可用子结点的数量及空间资源信息,生成数据成对存储的具体部署方案。随后每个对应子结点存储相应数据文件对,同时主结点保存所有数据文件在整个文件系统中的存储情况(原数据信息)。

其中计算任务调度模块运在用户在提交比对计算任务后,计算任务调度模块计算整个 计算完成需要的子任务数量及及每个计算任务需要的基本计算资源。随后计算任务调度 模块将计算任务信息传递至数据文件管理模块,协助数据管理模块确实数据文件的存储方 案。计算任务调度模块随后根据系统中各子结点的空闲计算资源信息,各子结点存储的数 据文件对信息,计算任务所对应计算资源要求信息等生成计算任务队列,刚不同的计算任 务调度至不同的子结点进行处理。

其中数据文件处理模块在子结点接受主结点的调度信息后,从子结点本地存储空间中 读取相应数据文件对,进行数据文件比对计算,并将计算任务状态实时反馈回主结点,最终 计算结果按用户计算程序中的要求返回至用户客户端。

其中系统资源监控模块负责收集分布式文件系统中各子结点的运行信息,监控子结点存储及计算资源使用情况,提供基本的结点通信,数据传输,故障处理,数据备份功能。

其中用户操作接口模块提供用户相应的数据操作接口,包括数据上传,下载,修改,查询等基本功能。其次,用户操作接口模块提供用户接口负责上传数据比对计算任务相应的计算程序及计算信息。此外,用户操作接口模块提供用户读取和操作最终数据比对计算结果的功能。

4. 一种利用如权利要求 1 所述分布式文件存储系统进行数据成对存储和数据比对计

算的方法,其特征该方法包括步骤如下:

- (1) 用户根据具体数据比对计算问题,准备以下信息:数据文件集,比对计算程序,计算资源需求信息。62
- (2) 数据对比计算任务分析阶段:在此阶段,所述分布式文件系统根据用户提供的相关信息,确定完成比对计算所需要的全部子计算任务数量,每个子计算任务对应的具体数据对及每个子计算任务需要的硬件资源。
- (3)数据成对存储阶段:在此阶段,所述分布式文件系统根据第二步中产生的信息,以计算任务为分配依据,考虑不同结点的处理能力,将全部计算任务预分配至不同结点;同时将每个计算任务对应的数据文件对预分配至相应结点,从而保证了每个计算任务所需要处理的数据均能够存储在相应子结点。所述分布式文件系统根据预分配产生的数据文件及计算任务分配方案,生成生成每个子结点需要存储的数据文件信息及每个子结点需要处理的比对计算任务信息。随后所述分布式文件系统将全部数据文件依数据分配方案转输分配至不同的子结点。
- (4)数据比对处理阶段:在此阶段,所述分布式文件系统根据阶段3中生成的各子结点处理比对计算任务的具体信息,考虑每个子任务需要的具体计算资源,对应每个不同子结点生成各自的任务队列。随后所述分布式文件系统中的每个子结点依据各自计算任务队列执行比对计算任务,并将计算任务执行情况实时传递给主结点。
- (5)数据输出阶段:在此阶段,每个计算任务产生的比对计算结果在主结点的调度下汇总至用户客户端,输出最终结果文件供用户进一步分析需要。

针对海量数据成对存储的分布式存储方法和系统

技术领域

[0001] 本发明涉及一种针对海量数据文件成对存储的分布式文件存储系统和方法,属于分布式存储技术论域。

技术背景

[0002] 随着信息技术的迅速发展,科研,商业及工业领域对于海量数据的存储及分析规模不断扩大。分式式文件系统由于自身的高可靠性,高可扩展性及相对低廉的使用成本获得了越来越多的关注。对于目前的分布文件存储系统,尽管具备不同的数据存储及恢复策略,支持不同的数据类型及兼容不同的操作系统平台,这些分布式文件存储系统在设计之初就将数据存储与数据处理分割成相互分离的两个阶段。这种对于数据处理场景的滞后考虑使得现有分布式文件系统在处理某些特定模式的大数据问题时存在严重的性能问题。

[0003] 文件比对作为一类特殊的计算模式(任一比对计算均需要处理两个不同数据文件),在生物科学,数据挖掘,图像处理等领域应用广泛。比如在基因识别,基因表达及物种进化分析的过程中,为了揭示数据所蕴含的生物学意义,数据集中的任意两个不同基因序列文件或核酸序列文件均需要进行两两比对计算。在图像识别领域,图书、商品、人脸、视频识别等均包含比对计算这一重要的处理过程。因此,优化这一计算模式的大数据处理问题具有很强烈的现实需求。

[0004] 目前的解决方案主要分为两类,一类方案针对小规模数据集,采取单机或在分布式计算系统中的所有结点统一部署全部数据的方法,此类方案虽然可以确保任一比对计算任务均可在本地存储中读取全部所需数据,此类解决方案的使用场景严重受限于单个结点的存储能力;另一类方案采用现有通用分布式文件系统,每个结点只存储部分数据。虽然可以解决海量数据存储问题,但由于在数据存储时并未考虑文件比对计算的特点,导致大量比对计算任务需要远程读取所需数据,数据的频繁传输严重的影响了文件比对计算的效率。

[0005] 文件比对计算属上笛卡尔乘积问题的一类情况。考虑数据集合 A, 比对计算 C 和表示结果的相似度矩阵 M, 多文件比对问题如图 1 所示:

[0006] 在图 1 所示的文件比对计算过程中,全部所需完成的比对计算任务如图中相似度矩阵 M 所示,其中矩阵中每一项 M[i,j] = C(A[i],A[j]),对集合 $A = \{A_1,A_2\cdots A_n\}$,整个集合的任意两个不同元素之间都要进行比对运算,因此整个过程需要进行 $\frac{n\times(n-1)}{2}$ 次比对运算。

[0007] 在对于大数据计算问题的一般解决方案中,在计算任务开始之前全部待处理数据需要提前部署在计算集群的分布式文件系统中。目前广泛使用的分布式文件系统如(HDFS,GFS等)在数据存储的具体策略上多以单个文件或文件块为单位,在数据存储的位置选择上以提高文件可靠性,平衡系统数据存储空间占用等为主要考虑,并未考虑不同数据之间的依赖关系,而对于文件比对类问题,计算任务以成对数据为基本输入,传统分布式文件系

统的存储策略必然导致了大量数据在计算过程中需要重新移动位置,大量的数据移动和硬盘读写严重的影响了文件比对的计算效率。此外,在计算过程中的数据重新移动和副本生成也严重的抵消了最初数据存储策略带来的存储空间节省及存储速速率提高等优点。因此,迫切需要研究一种针对文件比对计算模式的分布式数据存储方法及系统,即满足海量数据文件的存储需要,又可以避免文件比对过程中的远程数据传输,提高大规模文件比对的计算效率及存储效率。

发明内容

[0008] 针对现有技术的局限与不足,本发明提供一种针对海量数据文件成对存储的分布式数据存储系统,属于分布式存储技术领域

[0009] 本发明还提供一种利用上述分布式计算系统存储数据的方法。

[0010] 本发明的技术方案在于:

[0011] 一种针对海量数据文件成对存储的分布式文件存储系统,包括基于主从式分布式集群的存储系统设计,主结点功能模块,子结点功能模块及存储及计算流程设计。该分布式存储系统旨在应用分布式文件存储系统高效处理满足数据集文件比对模式的计算问题。本发明通过将文件存储及文件处理协同考虑,从数据存储初始阶段即满足后续比对计算的需要,实现了对于同一计算任务,一次存储,无需移动的数据文件存储及处理解决方案;此分式文件系统同时兼顾降低存储空间,保证数据文件可靠性的存储需要,并通过物理存储与逻辑存储的分离设计保证了系统对于多用户多应用场景的支持。

[0012] 根据本发明优选的,所述的分布式文件存储系统包括以下三个独立组成部分:用户客户端、系统主结点及系统子结点。基于数据成对存储的高性能分布式文件系统组成如图 2 所示:

[0013] 所述每个组成部分都具有不同的功能:所述客户端为用户提供基本操作接口,包括数据文件管理,计算程序管理及计算结果采集等。所述系统主结点负责收集用户提供的数据集信息及相应计算任务信息,确定数据文件部署方案。所述系统主结点同时采集系统各子结点信息,确定相应计算任务调度。所述系统子结点同时具备数据文件存储及数据对比处理计算功能。依据主结点的调度信息,所述的分布式文件系统各子结点以数据 对为单位存储各相应的数据文件,并进行数据文件的比对计算任务,最终计算任务全部结果在所述主结点的管理下汇总至所述用户客户端。在本发明所述分布式文件系统中主结点的控制下,不同子结点协同运行,构成一个具体的数据文件成对存储及后续比对计算过程。

[0014] 根据本发明优先的,所述基于分布式集群的分布式文件系统中主要包括以下五种功能模块:数据文件管理模块,计算任务调度模块,数据文件处理模块,系统资源监控模块和用户操作接口模块。

[0015] 本发明针对海量数据文件比对计算数据文件规模庞大,分布式集群硬件配置不同,计算资源动态变化的特点,提出了高性能及高扩展性的以数据文件成对存储为基础的分布式文件存储系统所述分布式文件系统主要负责数据文件的存储管理,基于数据文件对的计算处理工作。各种功能模块在此分布式文件系统的组成如图 3 所示:

[0016] 其中数据文件管理模块运行于所述分布式文件系统主结点(Master 结点)。用户在提交所有数据文件后,数据文件管理系统根据相应数据处理任务信息,分布式文件系统

中可用子结点的数量及空间资源信息,生成数据成对存储的具体部署方案。随后每个对应子结点存储相应数据文件对,同时主结点保存所有数据文件在整个文件系统中的存储情况(元数据信息)。

[0017] 其中计算任务调度模块运行于所述分布式文件系统主结点(Master 结点)。用户在提交比对计算任务后,计算任务调度模块计算整个计算完成需要的子任务数量及及每个计算任务需要的基本计算资源。随后计算任务调度模块将计算任务信息传递至数据文件管理模块,协助数据管理模块确实数据文件的存储方案。计算任务调度模块随后根据系统中各子结点的空闲计算资源信息,各子结点存储的数据文件对信息,计算任务所对应计算资源要求信息等生成计算任务队列,刚不同的计算任务调度至不同的子结点进行处理。

[0018] 其中数据文件处理模块运行于分布式文件系统子结点(Slaves 结点)。子结点在接受主结点的调度信息后,从子结点本地存储空间中读取相应数据文件对,进行数据文件比对计算,并将计算任务状态实时反馈回主结点,最终计算结果按用户计算程序中的要求返回至用户客户端。

[0019] 其中系统资源监控模块运行于分布式文件系统主结点(Master 结点)。资源监控模块负责收集分布式文件系统中各子结点的运行信息,监控子结点存储及计算资源使用情况,提供基本的结点通信,数据传输,故障处理,数据备份功能。

[0020] 基中用户操作接口模块运行于分布式文件系统的用户客户端(Client 结点)。用户操作接口模块提供用户相应的数据操作接口,包括数据上传,下载,修改,查询等基本功能其次,用户操作接口模块提供用户接口负责上传数据比对计算任务相应的计算程序及计算信息。此外,用户操作接口模块提供用户读取和操作最终数据比对计算结果的功能。

[0021] 一种利用上述分布式文件存储系统进行数据成对存储和数据比对计算的方法,包括步骤如下:

[0022] (1) 用户根据具体数据比对计算问题,准备以下信息:数据文件集,比对计算程序,计算资源需求信息。

[0023] (2)数据对比计算任务分析阶段:在此阶段,所述分布式文件系统根据用户提供的相关信息,确定完成比对计算所需要的全部子计算任务数量,每个子计算任务对应的具体数据对及每个子计算任务需要的硬件资源。

[0024] (3)数据成对存储阶段:在此阶段,所述分布式文件系统根据第二步中产生的信息,以计算任务为分配依据,考虑不同结点的处理能力,将全部计算任务预分配至不同结点;同时将每个计算任务对应的数据文件对预分配至相应结点,从而保证了每个计算任务所需要处理的数据均能够存储在相应子结点。所述分布式文件系统根据预分配产生的数据文件及计算任务分配方案,生成生成每个子结点需要存储的数据文件信息及每个子结点需要处理的比对计算任务信息。随后所述分布式文件系统将全部数据文件依数据分配方案转输分配至不同的子结点。

[0025] (4)数据比对处理阶段:在此阶段,所述分布式文件系统根据阶段3中生成的各子结点处理比对计算任务的具体信息,考虑每个子任务需要的具体计算资源,对应每个不同子结点生成各自的任务队列。随后所述分布式文件系统中的每个子结点依据各自计算任务队列执行比对计算任务,并将计算任务执行情况实时传递给主结点。

[0026] (5) 数据输出阶段:在此阶段,每个计算任务产生的比对计算结果在主结点的调

度下汇总至用户客户端,输出最终结果文件供用户进一步分析需要。

[0027] 本发明的优势在于:

[0028] 本发明针对海量数据文件成对存储的分布式文件系统针对但不限于基于 Linux 操作系统的异构分布式集群。在异构分布式集群环境中,不同结点具备不同的数据存储及数据处理能力,各结点之间由高速网络相连,在主控制结点的管理下协同运行。本文所述分布式文件系统的处理能力可随着集群中结点数量的增加而扩充,满足了数据存储规模扩大下的高扩展性要求。

[0029] 本发明基于主从式分布式集群结构,易于构建。用户可以通过自主选择主控制结点和子结点部署所述的分布式文件存储系统,达到海量数据文件成对存储和数据文件比对计算的任务要求

[0030] 本发明针对大规模数据文件比对运算的计算特点,以计算任务的分配为指导,使数据文件的分配不再简单以增加副本,提高存储可靠性为唯一指标。以比对计算任务为目标的分配方式使得所述的分布式文件系统对于每一比对计算问题可以达到数据文件一次存储,无需移动的优点,从而避免了在数据处理过程中产生的大量数据移动和数据副本生成,极大的提高了比对计算的计算性能。此外,本发明所提出的基本数据成对存储的分布式文件系统与硬件无关,可以方便的在不同的分布式集群环境中部署使用。

附图说明

[0031] 图 1 为海量数据文件比对计算问题的图解描述;

[0032] 图 2 为本发明所述的针对海量数据文件成对存储的分布式存储系统架构图;

[0033] 图 3 为本发明所述的针对海量数据文件成对存储的分布式存储系统功能模块图;

[0034] 图 4 为利用上述分布式文件存储系统进行数据文件比对计算的方法流程框图;

[0035] 图 5 为利用上述分布式文件存储系统进行数据文件比对计算的数据存储策略图。

[0036] 具体实施方法:

[0037] 下面结合实施例和说明书附图对本发明做详细的说明,但不限于此。

[0038] 实施例 1、

[0039] 如图 2-3 所示。

[0040] 一种针对海量数据成对存储的分布式文件存储系统,包括基于分布式集群的文件存储主结点,子结点及各功能模块。此分布式文件存储系统旨在应用分布式存储和计算技术高效处理海量数据文件比对的计算问题,在节省存储空间的同时提高比对计算的计算效率。本发明通过以计算任务分配做为数据存储的指导,避免了传统分布式文件系统数据存储阶段只考虑存储可靠性及存储空间,计算阶段无法避免大量数据复制和传输的问题;将数据文件成对存储过程与数据文件比对计算过程结合,保证用户在处理比对计算问题时计算程序只需读取本地文件资源,避免了用户考虑数据远程调用,网络传输等性能问题;基于分布式集群环境,提供了很好的系统存储和计算扩展能力。

[0041] 所述的基于数据文件成对存储的分布式文件系统由三个部分组成:系统主结点,系统子结点及客户端结点。在分布式文件系统的工作过程中,用户通过客户端上传全部数据文件,并上传相应数据文件比对程序:数据存储阶段数据文件及比对程序在系统控制下被部署在相应的结点;数据处理阶段各子结点依据所分配的对比任务各自处理存储在本地

的数据文件:最终处理结果在主结点调度下反回给客户端结点供用户进一步分析。

[0042] 所述的分布式文件系统中主要包括以下五种功能模块:数据文件管理模块,计算任务调度模块,数据文件处理模块,系统资源监控模块和用户操作接口模块。

[0043] 其中数据文件管理模块负责依据分布式文件系统中可用子结点数量,存储资源信息,相应数据比对任务信息等生成数据对存储方案,并且管理数据文件在所有结点的存储信息。此模块以数据对为基本存储单位,在存储阶段提前考虑了不同数据文件之间的关联性,避免了计算过程中数据文件的二次移动。

[0044] 所述计算任务调度模块负责计算完成整个数据文件比对处理所需要的任务数量,并依据用户提供的任务信息,确实每个计算计算所需要的具体计算资源。此外,所述计算任务调度模块与数据文件管理模块协同操作,确实不同结点需要处理的不同计算任务并进行具体的任务调度工作。计算任务调度模块确保了所有的数据比对计算任务均在包含有完整数据对的子结点上执行,充分利用了数据存储阶段的数据部署方案,避免了计算过程中数据的远程调用和复制。

[0045] 所述数据文件处理模块负责进行具体的数据比对计算工作。每个子结点的数据处理模块在接受主结点调度的子任务后,分配相应的计算资源处理相应数据文件对并将计算状态实时反馈给主结点。此外全部计算结果在主结点的管理下返回至客户端结点供用户进一步分析。数据文件处理模块避免了用户考虑复杂的远程数据传输,结点间交互等问题,简化了数据对比程序的开发工作。

[0046] 所述系统资源监控模块负责整个分布式文件系统的运行控制。系统中所有子节点的运行信息,存储及计算资源使用情况等均被主结点收集分析,以提供基本的结点通信,数据传输,故障处理,数据备份等功能。

[0047] 所述用户操作接口模块负责提供简单的操作接口供用户进行数据文件和计算程序的管理工作。用户使用此模块提供计算所需要的相应信息,并接受最终计算结果。所述操作接口模块使用户可以专注单机操作,隐藏了分布式文件系统复杂的运行细节和系统结构,降低了用户的使用成本。

[0048] 实施例 2、

[0049] 一种利用如实施例 1 所述分布式文件系统处理数据比对问题的方法,包括步骤如下:

[0050] (1)用户根据具体数据比对计算问题,准备相应数据及计算程序信息。

[0051] (2)数据对比计算任务分析阶段:在此阶段,系统计算对比计算所需要的全部子计算任务并确实每个子计算对应的数据对及计算资源。

[0052] (3)数据成对存储阶段:在此阶段,所述分布式文件系统根据第二步中产生的信息,同时收集系统中可用子结点数量信息及全部待处理数据文件数量信息,以计算任务为分配依据,生成每个子结点需要存储的数据文件信息及每个子结点需要处理的比对计算任务信息。所有数据文件在此阶段被传输及存储在不同的系统子结点。

[0053] (4)数据比对处理阶段:在此阶段,所述分布式文件系统根据阶段3中生成的各子结点处理比对计算任务的具体信息,考虑每个子任务需要的具体计算资源,对应每个不同子结点生成各自的任务队列。随后每个子结点接受相应的任务序列执行比对计算任务,任务状态在过程中时实反馈给主结点进行管理。

[0054] (5) 数据输出阶段:在此阶段,每个计算任务产生的比对计算结果在主结点的调度下汇总至用户客户端,输出最终结果文件供用户进行进一步分析。

[0055] 下面结合生物计算领域一个具体的海量 RNA 序列文件对比计算问题对本发明做进一步说明,但不限于此。此具体实例依照图 5 所示处理过程,包含此海量生物数据文件的成对存储及基因序列文件比对计算两部分。

[0056] 针对海量 RNA 序列文件对比计算问题的数据存储及计算方法,包括步骤如下:

[0057] (1) 分析 RNA 序列文件对比计算问题,依据本发明提出的针对海量数据成对存储的分布式系统的需要,提供以下信息:

[0058] a. RNA 基因序列文件信息,包括文件格式,文件本地路径,分布式文件系统中存储的相对位置

[0059] b. 比对程序相关信息,包括比对计算程序,所对应需要处理基因序列文件集,每个比对计算所需要硬件资源,包括 CPU 数量,内存数量。

[0060] (2)应用本系统提供的用户操作接口,将以上信息传递至系统主结点点,在数据文件管理模块及计算任务调度模块的处理下,进行数据对比计算任务分析,确定比对任务总数,每节点需要执行的比对计算任务数量等信息。

[0061] (3) 根据分析产生的结果,将数据文件以数据对的方式存储在不同的子结点,如图 5 所示。在图 5 实例中,系统分析产生了全部比对计算任务如图 5 中矩阵所示。随后,不同计算任务被分配至相应子结点,每个计算任务所涉及的数据文件存储位置也随之确实。根据此分配方法,每个子结点均含有不同的序列文件并且每个子结点只存储部分数据文件,有效节省了空间。同时,每个子结点所需要处理的计算任务也在此过程中被确定,具中图 5 各子结点中星号代表此子结点需要执行的比对计算任务。

[0062] (4) 所述分布式文件系统根据阶段 3 中生成的各子结点对应的处理比对计算任务,考虑第一步中用户提供的每个子任务需要的具体计算资源,对应每个不同子结点生成各自的任务队列如图 5 所示。每个子结点随后依据各自己的任务队列,依次独立的执行不同比对计算任务,并将计算信息反馈给主结点。从图 5 示例可见,所有计算结点均包括均衡数理的计算任务并且所有计算任务仅需要处理结点本存储的数据文件。

[0063] (5) 计算完成后,所有计算结果按用户定义返回至客户端结点,如下所示,供用户分析使用,完成整个过程。

[0064] 最终得到计算结果如下所示,输出结果格式如下所示:

[0065] Key: [AcMNPV. faa, ASFV. faa] Value: 0.0016810058966743133

[0066] Key: [AsGV. faa, AdhoNPV. faa] Value: 0.011683080339163805.

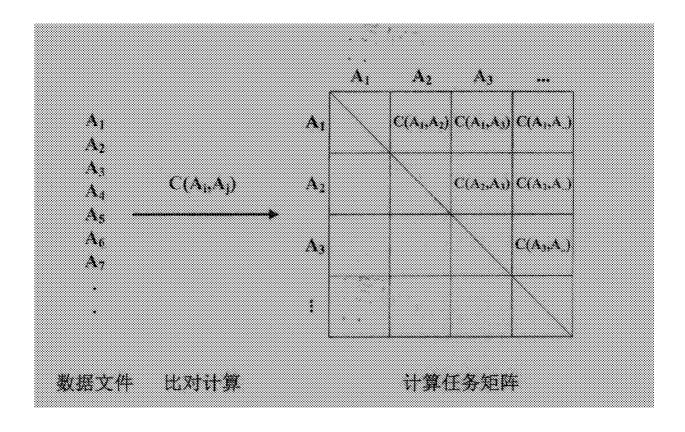


图 1

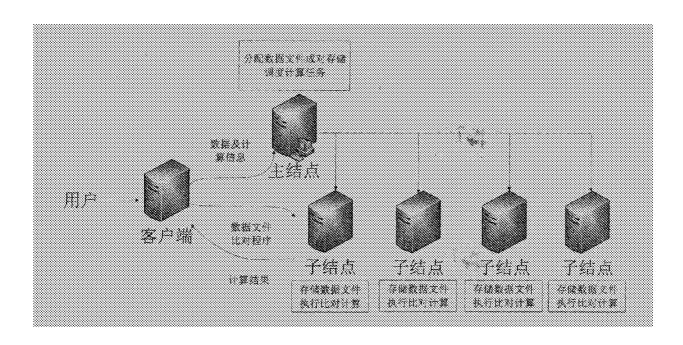


图 2

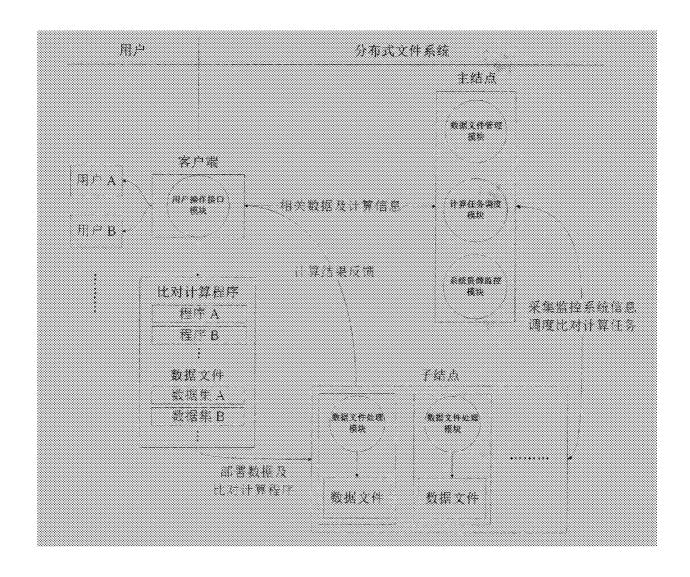


图 3

分析具体比对计算问题

确定相应数据文件及比对计算程序

提交主结点相关信息,系统确实子任务数量及相应资源信息

系统收集子结点相关信息,确定数据分配策略及各子 结点需处理的任务信息

各子结点依任务队列处理比对任务,并反馈任务信息 到主结点

计算完成, 计算结果返回至客户端结点

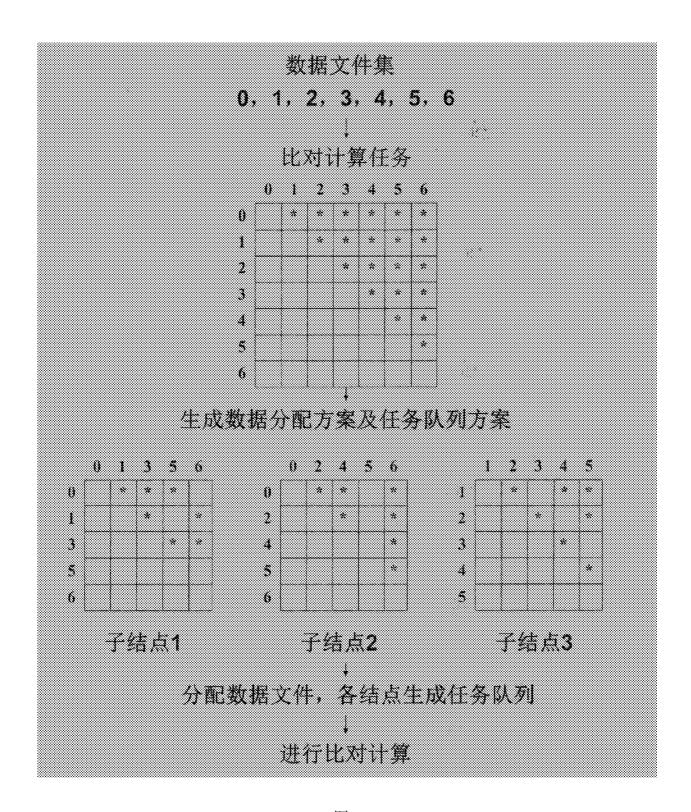


图 5