WANDISCO Fusion POC Test


WanDisco Overview
产品简介
Google | google.com/datacenters

WANDISCO FUSION 适用场景

WANDISCO FUSION FOR REAL-TIME ANALYTICS

- 实时接收和复制来自多个来源和位置的数据
- 确保您遵守数据主权要求和隐私条例
- 简单的设置可以让你在几分钟内启动并运行
- 直观的管理控制台，用于监控，调度和审计
- 使用标准云供应商程序进行安装和部署
- 完全支持云供应商的功能，Hadoop集群按需启动和关闭
- 支持在任何本地和云平台环境的Fusion之间移动数据

WANDISCO FUSION 适用场景

WANDISCO FUSION

- 可以实现数据PB级的数据备份，在遇到设备故障时实现最小的RTO和RPO保障数据无丢失
- 数据始终准确：在任何情况下数据发生更改都会被记录
- 数据始终可用：数据备份到私有云和公有云，实现RTO和RPO几乎为零
- 降低成本：消除其他解决方案所需的专用硬件备份的投入费用。

WANDISCO Fusion 与 DistCp和HDFS快照 功能与性能对比

WANDISCO FUSION 性能卓越

DistCp 或 HDFS快照

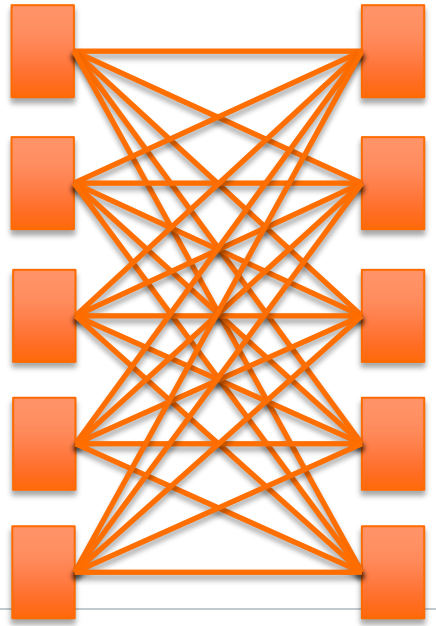
- 需要专用资源进行备份，而无法在这些集群上提供数据服务，资源浪费
- 只能手动备份 或 定时备份
- 无高可用，在网络/服务器故障之后，无法自动续传
- 无法实现带宽控制，避免网络拥塞
- 无法制订同步规则，确定文件的同步策略
- 无法集中化管理，需要手动管理多个集群
- 只能单向数据同步
- 性能差，同步数据慢

WANDISCO Fusion

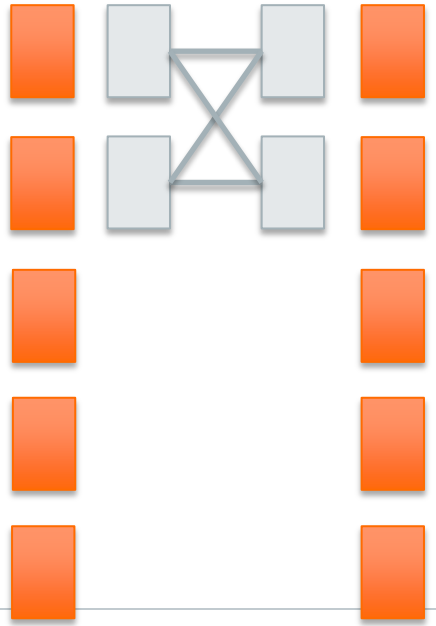
- 数据同步的时候，参与同步的数据库可以同时提供数据服务，最大利用服务器资源
- 实时同步，无需设置同步时间表
- 高可用，在网络/服务器故障之后，可自动续传
- 可以带宽控制，避免网络拥塞
- 可制定同步规则，确定文件同步策略
- 集中化管理，一个界面即可实现多个集群同步
- 双活 或 单向数据同步
- 性能好，无论服务器负载如何，Fusion基本上可以只用DistCp一半的时间完成数据同步。（Accenture | 2017）

WANDISCO Fusion 与DistCp和HDFS快照 安全性对比

DistCp 安全性差



Fusion 安全性好



wanDISCO®

WANDISCO Fusion 产品优势

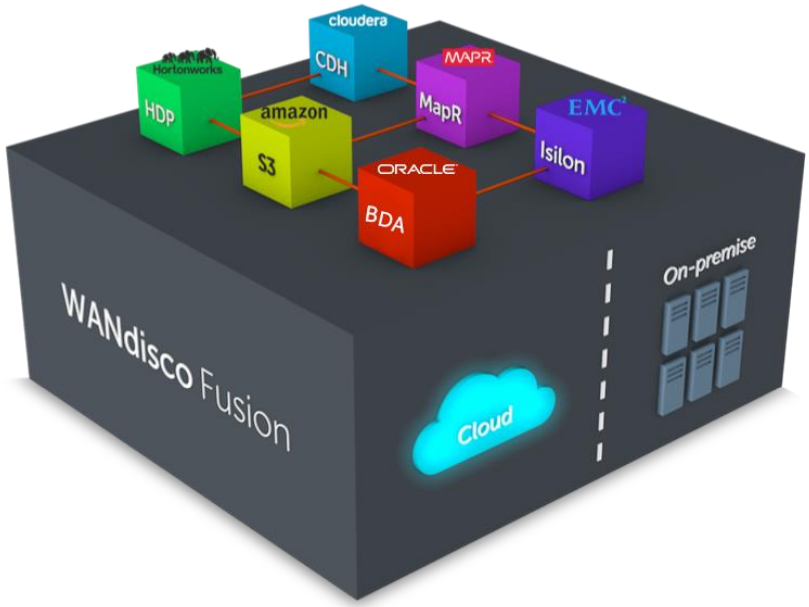
WANDISCO FUSION

- 高性能：Hadoop 集群中 进行高效的 双活（Active-Active）热数据 实时同步
- 使用灵活：自定义同步规则、带宽控制、单向双向、自动续传、定时同步
- 技术领先：在数据同步领域拥有独一无二的国际专利
- 数据安全：Hadoop集群间只需要2台Fusion进行数据同步，极大简化架构，有效提高数据同步的安全性

WANDISCO FUSION 的核心价值

WANDISCO FUSION FOR CLOUD MIGRATION

- 构建企业数据湖，打造全方位的灵活、高效的企业内部数据实时流动的解决方案。

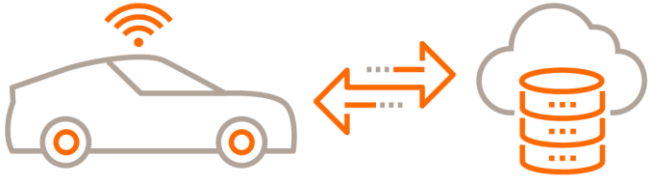


我们的客户

wanDISCO

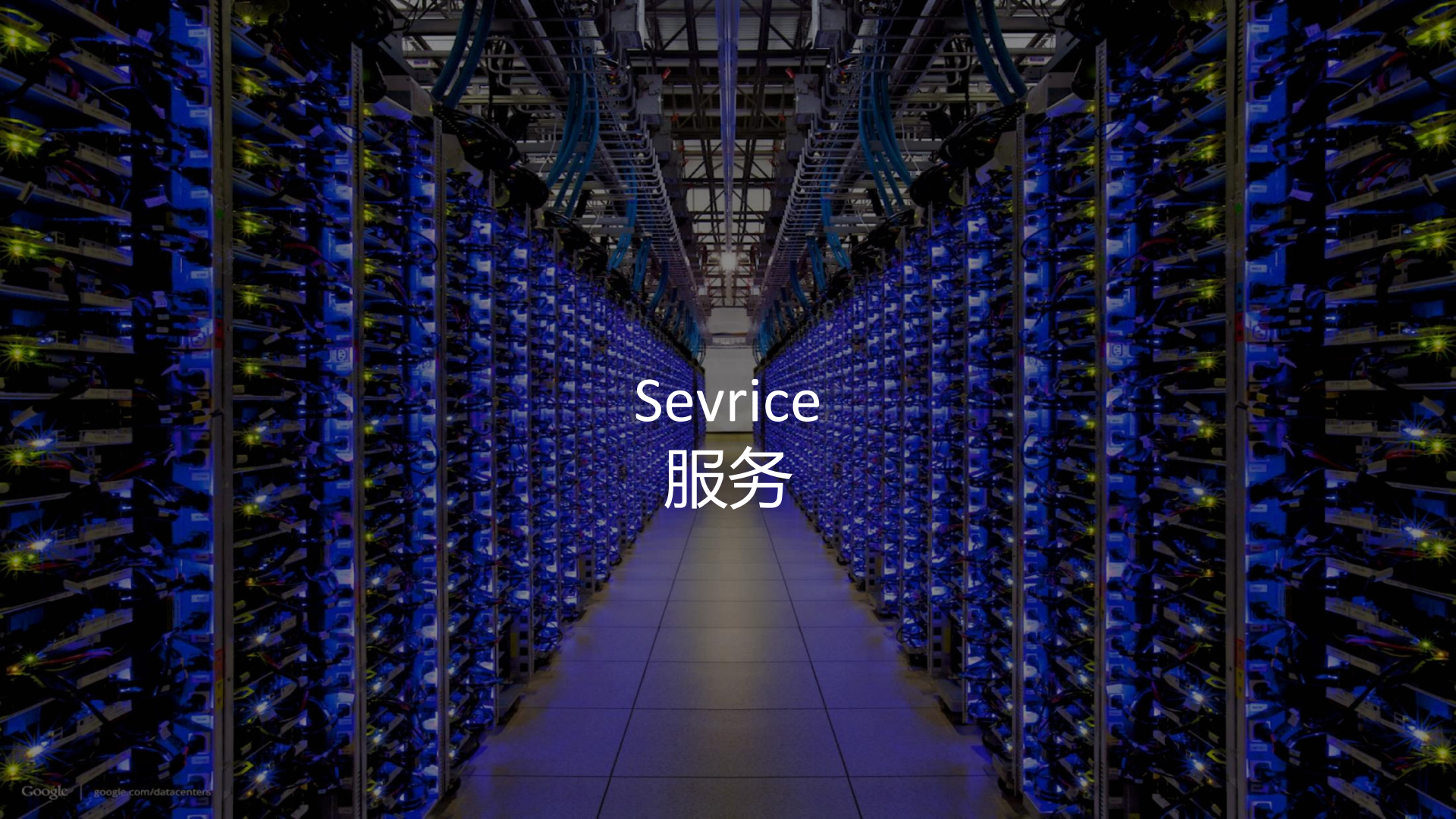
重点客户案例—全球著名汽车制造商

WANDISCO FUSION FOR CLOUD MIGRATION



- 机遇与挑战
  - 自动驾驶汽车项目每天产生超过200 TB的快速流数据。这些数据被放入2个集群以平衡负载分析工作需要两活的多数据中心摄取来保持集群的同步和数据的高可用性
    - 用户希望每个集群都能实时的为另一个集群提供备份和恢复
    - 在传统模式下数据管理工作繁重，并无法处理在每个集群上独立创建的数据，无法保证方案的可靠性

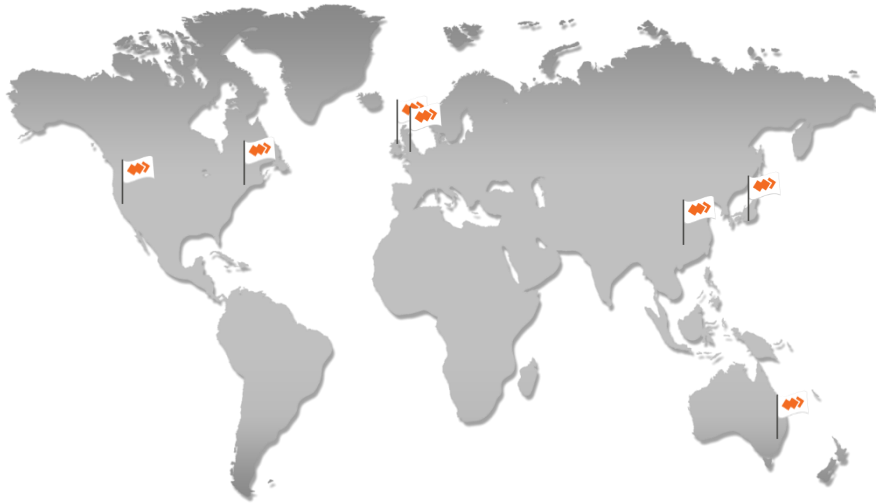
- 解决方案
  - 用户使用了 WANdisco Fusion 的方案:
    - 持续自动同步，在高负载的情况下，有效的保证集群中的数据的一致性
    - 对两活集群的全主动读/写访问，在集群之间实现持续的高可用性数据备份和自动恢复。
    - 支持随集群的规模增长而扩容，可提供超过400 PB的可扩展能力。


Sevrice
服务
Google | google.com/datacenters

WANDISCO FUSION 的服务

WANDISCO FUSION FOR CLOUD MIGRATION

- 全球化的服务管理体系
- 7 X 24 小时服务支持及快速服务响应
- 中国成都服务中心
- 本地化的服务提供



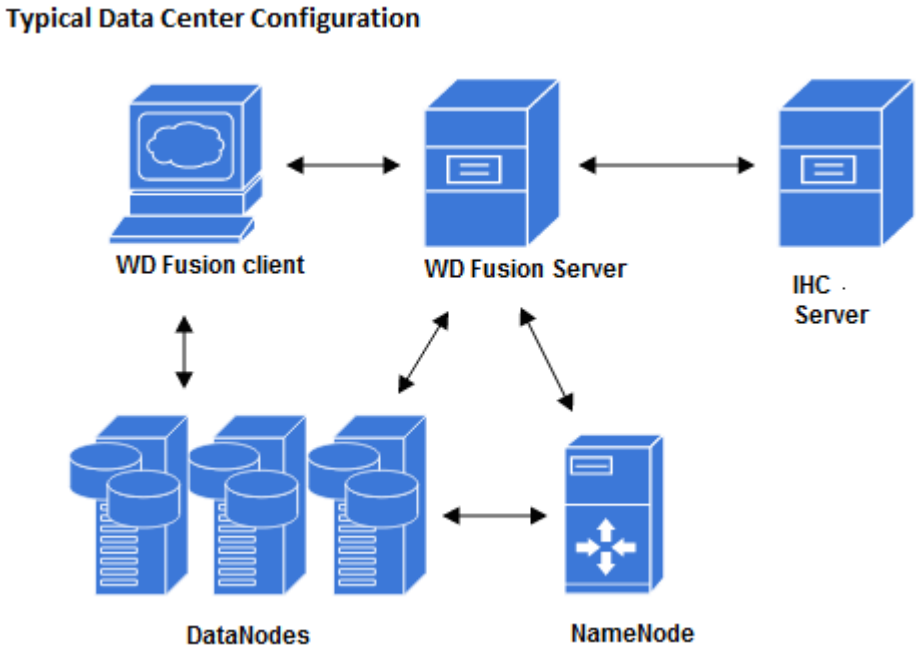

POC Testing Overview
测试案例汇总
Google | google.com/datacenters

Fusion 标准部署

WANDISCO FUSION FOR REAL-TIME ANALYTICS

简单部署模式

- Fusion Server- installed on Server
- IHC Server- installed on with Fusion Server
- Fusion Client-installed on Data Node



测试案例汇总

WANDISCO FUSION FOR CLOUD MIGRATION

- 测试案例1/2（通过） ： Fusion服务安装
- 测试案例3（通过） ： 设置同步规则
- 测试案例4（通过） ：检查并修复数据的‘不一致’
- 测试案例5（通过） ：日志查询
- 测试案例6（通过） ： API调用
- 测试案例7（通过） ：**实时同步，灵活配置同步规则**
- 测试案例 8 （通过） ：在已有文件中进行新数据的实时同步
- 测试案例 9（通过）：HDFS checksum 实时同步
- 测试案例 10（通过）：HDFS文件权限的实时同步

测试案例汇总

WANDISCO FUSION FOR CLOUD MIGRATION

- 测试案例 11 （通过） ：高可用，在网络/服务器故障之后，自动续传
- 测试案例 12/13 （通过） ：高可用，HDFS 服务故障之后，自动续传
- 测试案例 14/15（通过）：高性能，同步速度远远优于其他产品例如DistCp
- 测试案例 16/17（通过）：HDFS文件属性的实时同步
- 测试案例 18（通过）：数据同步时，带宽控制，避免网络拥塞
- 测试案例 19（通过）：Hive元数据及数据内容的手工修复
- 测试案例 20（通过）：Hive元数据及数据内容的实时同步
- 测试案例 21/22（通过）：高可用，Hive 服务故障之后，Hive元数据及数据内容的实时同步
- 测试案例 23（通过）：Hive 数据库的各类操作实时同步
- 测试案例 24（通过）：HDFS 控制列表(ACL)的实时同步

测试案例 1 / 2要点（Fusion服务安装）

WANDISCO FUSION FOR CLOUD MIGRATION

- 测试步骤:
  - 安装Fusion服务和各个部件
  - 启动各个Fusion部件
- 测试结果（ 通过 ）:
  - Fusion服务成功安装
  - Fusion各个服务成功启动

测试案例 3要点（设置同步规则）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 按HCFS文件系统文件夹，在集群间同步数据
- 功能：
  - Fusion服务器支持用户灵活配置需要同步的文件夹 / 文件
- 测试步骤：
  - 在Fusion管理界面中，使用“Replication”页面
  - 创建基于HCFS的，选定文件夹的同步规则
  - 设定Cluster1（重要的）为高优先级集群
- 测试结果（ 通过 ）：
  - 两个集群中的选定文件夹下，任何一方创建的文件都可以同步到另一集群去

测试案例 4要点（修复数据的‘不一致’）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 某些特殊因素，导致集群间数据‘不一致’，需要同步数据使其‘一致’
- 功能：
  - Fusion可以检查各个集群之间的文件不一致性，并加以修复，使所有文件一致
- 测试步骤：
  - 在Cluster1中/repl-00文件夹下，创建数据文件
  - 设置Cluster1到2的同步规则
  - 使用该规则检查“一致性”（文件夹中已经存在数据，所以会不一致）
  - 使用该规则，解决“不一致”
- 测试结果（ 通过 ）：
  - 系统侦测到集群中数据的“不一致”
  - 两个集群中的数据“不一致”被迅速解决，且同步过程通过管理界面可观测

测试案例 5要点（日志查询）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 系统管理员需要了解数据同步的历史记录，以便排查技术问题
- 功能：
  - Fusion提供接口，为日志查询提供便利
- 测试步骤：
  - 通过Fusion提供的http接口访问各个数据节点
  - 打开通过页面打开“fusion-server.log”
  - 通过文件系统访问日志 /var/kig/fusion/server(ui or ihc)
- 测试结果（ 通过 ）：
  - 各个服务的日志通过界面成功访问
  - 各个服务的日志通过文件系统成功访问

测试案例 6要点（API调用）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 系统管理员对数据同步需要通过代码进行管理
- 功能：
  - Fusion提供API接口，为二次开发提供便利
- 测试步骤：
  - 远程访问Fusion节点
  - 通过curl命令调取各个节点的API接口（memberships, nodes, locations, path等）
- 测试结果（ 通过 ）：
  - 各个服务Restful API接口被成功调取

测试案例 7要点（实时同步/灵活配置）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 高效运作的数据库/文件系统，需要各个集群之间的数据实时同步
- 功能：
  - Fusion服务器允许灵活的数据同步配置，各个集群间数据按需求实时同步（单向、双向）
- 测试步骤：
  - 配置同步规则
  - 在Cluster1中创建数据文件（删除/增加文件）
  - 在Cluster2中创建数据文件（删除/增加文件）
- 测试结果（ 通过 ）：
  - 两个集群中的任何数据文件操作都被实时的同步到按规则配置的另一个集群中

测试案例 8要点（在已有文件中进行新数据的实时同步）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 高效运作的数据库/文件系统，需要各个集群之间的数据实时同步
- 功能：
  - Fusion服务器允许灵活的数据同步配置，各个集群间数据按需求实时同步（单向、双向）
- 测试步骤：
  - 在已有文件中添加新内容
- 测试结果（ 通过 ）：
  - 两个集群中的任何新增数据都被实时的同步到按规则配置的另一个集群中

测试案例 9要点（HDFS checksum 实时同步）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 高效运作的数据库/文件系统，需要各个集群之间的数据实时同步
- 功能：
  - Fusion服务器实时同步HDFS文件的checksum
- 测试步骤：
  - 配置同步规则
  - 在Cluster1中创建数据文件（删除/增加文件）
  - 对比Cluster1和Cluster2 上该文件的checksum
- 测试结果（ 通过 ）：
  - 两个集群中数据文件checksum都被实时的同步到按规则配置的另一个集群中

测试案例 10要点（HDFS文件权限的实时同步）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 高效运作的数据库/文件系统，需要各个集群之间的数据实时同步，包括相关权限
- 功能：
  - Fusion服务器实时同步HDFS文件的权限
- 测试步骤：
  - 配置同步规则
  - 在Cluster1中创建数据文件
  - 在Cluster1上修改数据文件权限
- 测试结果（ 通过 ）：
  - 两个集群中的任何数据文件权限修改被实时的同步到按规则配置的另一个集群中

测试案例 11 要点（高可用，在网络/服务器故障之后，自动续传）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 数据同步过程中，可能因为网络或其他原因，导致无法同步。在网络故障或其他原因排除后，系统可以自动同步不一致的数据
- 功能：
  - Fusion服务器自动侦测各个集群的可用性，在网络出现问题的时候，停止同步，在网络恢复后，自动同步不一致数据
- 测试步骤：
  - 暂停Cluster2的Fusion服务（模拟网络故障或服务下线 5～10分钟）
  - 在Cluster1导入新的数据文件，启动Cluster2的Fusion服务（模拟故障排除）
- 测试结果（ 通过 ）：
  - Cluster2的数据在Fusion服务恢复后，第一时间自动完成同步，数据保持一致

测试案例 12/13要点（高可用，在HDFS故障之后，自动续传）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 数据同步过程中，可能因为hdfs服务故障，导致无法同步。在故障排除后，系统可以自动同步不一致的数据
- 功能：
  - Fusion服务器自动侦测各个集群的可用性，在hdfs出现问题的时候，停止同步，在hdfs恢复后，自动同步不一致数据
- 测试步骤：
  - 暂停Cluster2的hdfs服务（模拟服务下线）
  - 在Cluster1导入新的数据文件，启动Cluster2的hdfs服务（模拟故障排除）
- 测试结果（ 通过 ）：
  - Cluster2的数据在hdfs服务恢复后，第一时间自动完成同步，数据保持一致

测试案例 14/15要点（高性能同步）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 不同地区之间的数据，需要实时快速同步
- 功能：
  - Fusion在文件层提供实时快速同步，效果优于DistCp
- 测试步骤：
  - 在Cluster1上生成10（待确认）个10GB的数据文件
  - 使用DistCp进行数据同步，观测所需要时间
  - 使用Fusion进行数据同步，观测所需要时间
- 测试结果（通过）：
  - 在同步Hadoop数据文件时，相同环境下，Fusion的数据同步远优于DistCp

测试案例 16/17要点（HDFS文件属性的实时同步）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 高效运作的数据库/文件系统，需要各个集群之间的数据文件属性实时同步
- 功能：
  - Fusion服务器允许灵活的数据同步配置，各个集群间数据属性实时同步
- 测试步骤：
  - 在Cluster1上生成1个数据文件
  - 查看Cluster1和Cluster2上的文件属性
- 测试结果（ 通过 ）：
  - Cluster1和Cluster2上文件属性一直

测试案例 18 要点（带宽控制）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 跨地域进行数据同步时，带宽占用过多，导致地区和北京之间其它应用无法链接
- 功能：
  - 在数据同步时，Fusion Server可控制带宽占用，避免网路拥塞
- 测试步骤（简略）：
  - 配置传输速率（100MB/秒），启动传输，并观察传输进度和预估时间
  - 增加传输速率（200MB/秒），观察传输进度和预估时间
- 测试结果（通过）：
  - 带宽资源的使用，根据设定而调整，立即生效，可以根据业务需求灵活调整

测试案例 19 要点（Hive元数据及数据内容的手工修复）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 因为故障或特殊原因，导致Hive元数据出现不一致时，需要修复Hive 元数据
- 功能：
  - 在Hive元数据不同步时，Fusion Server可手工修复Hive元数据
- 测试步骤（简略）：
  - 在Cluster1创建Hive数据库并写入数据
  - 创建同步规则并手工修复元数据
- 测试结果（通过）：
  - Hive元数据被手工同步至Cluster2

测试案例 20 要点（Hive元数据及数据内容的实时同步）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 高效运作的数据库/文件系统，需要各个集群之间的数据实时同步
- 功能：
  - Hive元数据及数据内容的实时同步
- 测试步骤（简略）：
  - 创建Hive同步规则
  - 在Cluster1创建Hive数据库并写入数据
- 测试结果（通过）：
  - 两个集群中的任何Hive元数据及数据内容都被实时的同步到按规则配置的另一个集群中

测试案例 21/22 要点（高可用，Hive 服务故障之后，Hive元数据及数据内容的实时同步）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 数据同步过程中，可能因为Hive服务故障，导致Hive元数据无法同步。在故障排除后，系统可以自动同步不一致的数据
- 功能：
  - Fusion服务器自动侦测各个集群的可用性，在Hive出现问题的时候，停止同步Hive元数据，在Hive恢复后，自动同步不一致数据
- 测试步骤（简略）：
  - 创建Hive同步规则
  - 停掉Cluster1 Hive 服务
  - 在Cluster2创建Hive数据库并写入数据
  - 启动Cluster1 Hive 服务
- 测试结果（通过）：
  - Cluster1的数据在Hive服务恢复后，第一时间自动完成同步，元数据及数据内容保持一致

测试案例 23 要点（Hive 数据库的各类操作实时同步）

- 场景：
  - 高效运作的数据库/文件系统，需要各个集群之间的数据实时同步
- 功能：
  - Hive数据库各类操作的实时同步
- 测试步骤（简略）：
  - 创建Hive同步规则
  - 在Cluster1执行各类Hive数据库操作
- 测试结果（通过）：
  - 两个集群中的任何Hive数据库操作都被实时的同步到按规则配置的另一个集群中

测试案例 24 要点（HDFS 控制列表-ACL的实时同步）

WANDISCO FUSION FOR CLOUD MIGRATION

- 场景：
  - 高效运作的数据库/文件系统，需要各个集群之间的数据实时同步，包括其相关权限
- 功能：
  - HDFS控制列表的的实时同步
- 测试步骤（简略）：
  - 创建Hive同步规则
  - 在Cluster1创建文件并修改文件ACL
- 测试结果（通过）：
  - 两个集群中的文件控制列表ACL都被实时的同步到按规则配置的另一个集群中

感谢！