# 大数据容灾备份技术挑战和增量备份解决方案

罗圣美[1,2]，李 明[1]，叶郁文[1]
1. 中兴通讯股份有限公司 南京 210012；2. 清华大学计算机科学与技术系 北京 100084

**摘要**
大数据已成为当前社会各界关注的焦点，是当前世界各国新一轮的科技竞争和综合国力较量的重点，必须做好大数据的容灾备份工作。为此，在分析大数据容灾备份现状的基础上，结合行业对大数据容灾备份需求，讨论了几种典型的技术解决方案及其优缺点，提出了一种基于HDFS的增量数据备份恢复方案，具备分钟级RPO的系统远程备份特性，可以较好地解决目前大数据容灾备份项目建设规划面临的实际需求。

**关键词**
大数据；备份；恢复；业务连续性

doi: 10.11959/j.issn.2096-0271.2015033

# *Challenge and Solution of Big Data Backup and Recovery*

**Luo Shengmei[1,2], Li Ming[1], Ye Yuwen[1]**
1. ZTE Corporation, Nanjing 210012, China;
2. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

***Abstract***
Big data has become the focus of the social attention, it will raise a new competition in science and technology and comprehensive national strength, especially in the disaster recovery and backup data. Therefore, based on the analysis of the current industrial status and the requirements about big data disaster recovery and backup, the advantage and disadvantage of several typical technology solutions were discussed, then a better incremental backup data recovery solution was proposed. This solution can support minute RPO, and meet current requirements about the disaster recovery and backup data.

***Key words***
big data, backup, recovery, service continuity

# 1 引言

随着移动互联网信息技术的演进和社会的发展，人类在生产和生活中会产生更多、更复杂的数据。据 IDC 报告显示，2020 年全球企业数据总量将从目前的 1.2 ZB 增加到 35 ZB。从一般意义上讲，这种量级和复杂度的数据，业界称为大数据。大数据是人类社会重要的信息资产，在科技发展和生活生产中起到非常重要的作用，已成为当前社会各界关注的焦点，是当前世界各国新一轮的科技竞争和综合国力较量的重要体现。

近些年来飓风、地震、海啸、火灾等自然灾害频发，电脑病毒泛滥、黑客攻击猖獗等日益严重的互联网危机，已让无数企业遭受了数据丢失所带来的沉重打击。多个研究机构提供的数据表明，公司数据丢失将对公司带来严重影响，甚至导致公司倒闭。来自IDC的统计数据显示，1990-2000年发生灾难导致数据丢失的公司生存下来的仅有16%，美国明尼苏达大学研究报告显示，发生数据丢失的公司如果无法快速恢复数据，近3/4的公司业务将停顿，近2/5的公司将倒闭[1]。

由此可见，数据是企业最宝贵的资产，是企业生存的基础，也是企业核心竞争力的重要组成部分，一旦丢失，其产生的后果可能是灾难性的，甚至会引发社会性问题，所以大数据的安全、备份和容灾就显得尤为重要。

# 2 大数据容灾备份

## 2.1 容灾备份现状

容灾备份系统的目的在于保证系统数据可靠和服务的在线性，即当主用系统发生故障时，仍能提供数据和服务，保证系统业务不受影响。

灾备领域国际和国内都制定了相关标准，国际标准中SHARE78具有较大影响力，针对灾难恢复定义了Tier-0至Tier6/7共7个层次。

我国的国家标准GB20988-2007-T《信息安全技术信息系统灾难恢复规范》对容灾备份进行了标准化[2]，与SHARE78的7个层次具备对应关系，并进一步细化了具体要求。

在设计容灾系统时，容灾要达到什么样的目标和层次，需要用一些定量的指标来衡量，这就是灾难恢复能力指标，具体介绍如下。

RTO（recovery time object，恢复时间目标）：指信息系统从灾难状态恢复到可运行状态所需要的时间，用来衡量容灾系统的业务恢复能力。

RPO（recovery point time，恢复点时间）：指业务系统所允许的在灾难过程中的最大数据量丢失，用来衡量容灾系统的数据冗余备份能力。

NRO（network recovery object，网络恢复时间目标）：指在灾难发生后网络恢复或切换到灾备中心的时间，通常网络要先于应用恢复才有意义，但应用恢复后才能提供业务访问。

某行业灾难恢复等级对应的能力指标见表1。

数据的备份策略一般分为全量备份（full backup）、差异备份（differential backup）和增量备份（incremental backup）。

全量备份：间隔一段时间就对整个系统进行全面备份，包括系统和数据。

差异备份：针对前一次完全备份后发生变化的所有信息进行备份。

增量备份：针对前一次备份后所有发生变化的信息进行备份，增量备份方式备份的数据量最小，但恢复时要利用全备份的数据，并叠加以前的增量备份，数据恢复时间也最长。

无论哪种备份策略，在一个备份周期内都首先要进行一次完全备份，然后再选择进行增量备份或者累计备份。通常在数据更新不太频繁且数据量不太大的情况下，可以选用累计备份的方式。若数据量更新很频繁，更新量又很大，那么备份周期后几次的累计备份数据量就很大，这时使用累计备份就不太经济，可以考虑增量备份或者增量备份和累计备份相结合的方式，也可以考虑缩短备份周期。

表1 RTO/RPO 与灾难恢复能力等级的关系 [2]

| 灾难恢复能力等级 | RTO | RPO |
|---|---|---|
| 1 | 2天以上 | 1天至7天 |
| 2 | 24 h以上 | 1天至7天 |
| 3 | 12 h以上 | 数小时至1天 |
| 4 | 数小时至2天 | 数小时至1天 |
| 5 | 数分钟至2天 | 0~30 min |
| 6 | 数分钟 | 0 |

备份系统结构一般包括DAS-base备份、LAN-base备份、LAN-free备份和Server-free备份，如图1所示。

（1）DAS-base备份是利用服务器自带的磁带机或备份硬盘手工进行数据备份。优点是维护简单，数据传输速度快；缺点是可管理的存储设备较少，对大数据量备份场景或者实时数据备份场景不适用。

（2）LAN-base备份是专门使用一台服务器作为备份管理服务器，通过备份管理服务器实施系统的专用备份操作。优点

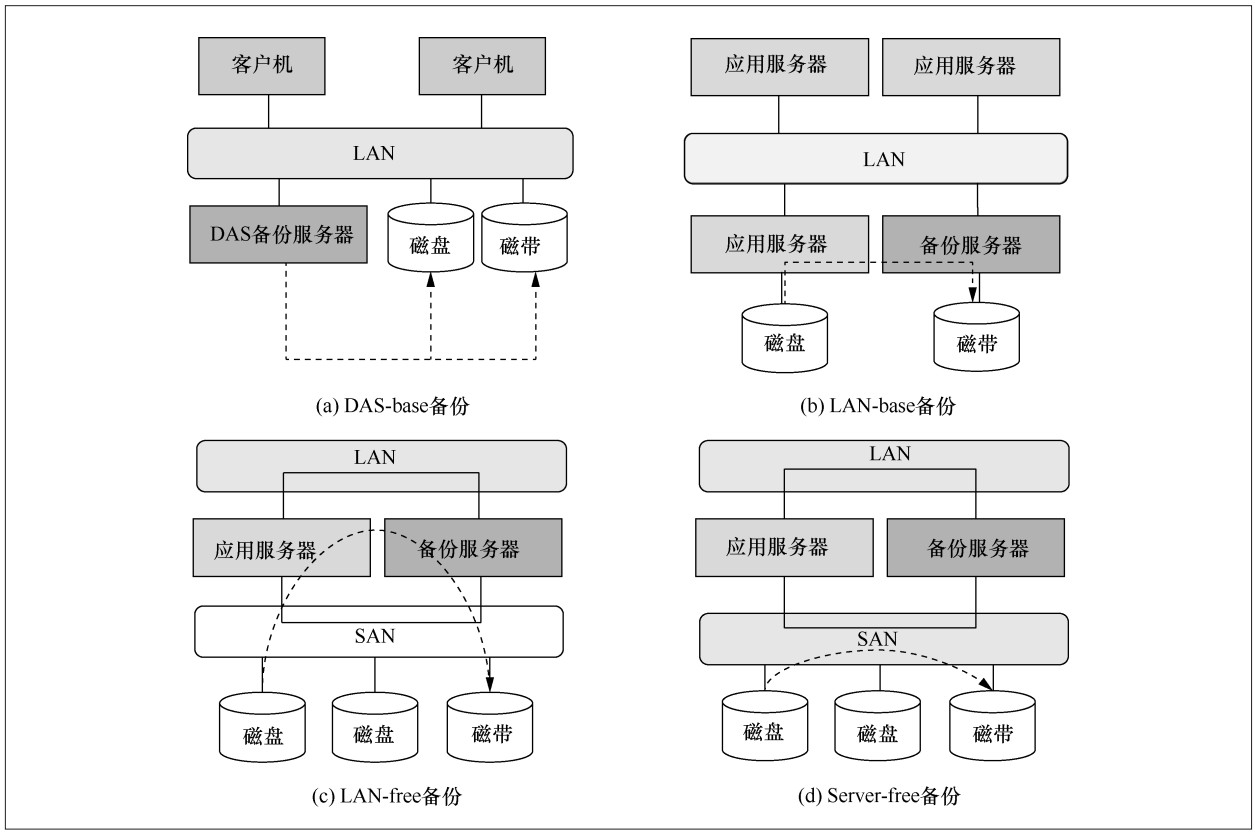


图1 常见备份系统架构

是投资经济、磁带库共享、集中备份管理；缺点是网络传输压力大，对备份数据量大/备份频率高的场景不适用。

（3）LAN-free备份是将备份设备连接到SAN（storage area network，存储区域网络）上，数据无须通过局域网而直接进行备份，局域网只承担各服务器之间的通信任务，实现了控制流与数据流分离。优点是数据备份统一管理、备份速度快、网络传输压力小、磁带库资源共享；缺点是恢复操作繁琐、实施复杂、投资较高，不适用于少量文件备份场景。

（4）Server-free备份不需要在服务器中缓存数据，显著减少对主机CPU占用，采用NDMP（网络数据管理协议）应用是实现方式之一。优点是数据备份和恢复时间短、网络传输压力小、便于统一管理和备份资源共享，但由于需要特定的备份应用软件进行管理，需要考虑厂商的兼容性问题，实施起来比较复杂，成本也较高。

## 2.2 大数据容灾备份特征与挑战

大数据同过去的海量数据有所区别，其4V特征（volume、variety、value、velocity）体现了量大、多样、密度低、速度快的特点，采用磁带复制方式不现实。

传统备份产品不适用于大数据领域。传统备份产品大多基于主机、网络或者磁阵，都是单机备份系统，而云存储现有的数据多副本和EC技术应用都只能保证单数据中心内的数据可靠性，不提供数据备份能力。

HDFS的数据节点（DataNode）采用多副本、纠删码（erasure coding，EC）等高可用（high availability，HA）技术，可以提供数据高可靠性。HDFS的主从和单个元数据节点（NameNode）设计方式使得元数据节点成为单点故障，HDFS的HA发展经历了3个过程。

（1）借助分布式块复制设备（distributed replicated block device，DRBD）、心跳服务（heartbeat）HA组件实现主备切换。使用DRBD实现两台物理机器之间块设备的同步，即通过网络实现raid1，辅以heartbeat HA实现两台机器动态角色切换，对外使用虚拟IP地址来统一配置。

（2）主元数据节点（Primary NameNode）与备元数据节点（Standby NameNode）之间通过网络文件系统（network file system，NFS）来共享FsEdits、FsImage文件，这样主备NameNode之间就拥有了一致的目录树和block信息。DataNode向两个NameNode上报块信息，辅以虚拟IP地址，可以较好地达到主备NameNode快速热切换的目的。

（3）基于Paxos算法实现的HDFS HA方案。基本原理是用2$N$+1台JournalNode存储EditLog，每次写数据操作有大多数（$\geq N$+1）返回成功时，即认为该次写成功，数据不会丢失。

虽然基于HDFS的大数据系统解决了HA问题，但是异地备份和容灾仍存在问题。所以，需要有一种容量更大、安全性更高、数据存储及恢复更快的容灾备份解决方案来满足实际应用需求。

## 2.3 大数据容灾备份解决方案

系统容灾备份是一项系统工程，一般要从灾备中心基础设施、数据备份系统、备用数据处理系统、备用网络系统、灾难恢复预案、运维管理能力以及技术支持能力7个要素进行统筹考虑。

本文主要从数据备份系统以及备用数据处理系统角度描述一种可行的灾备方案，该容灾备份方案是基于HDFS的增量

数据备份恢复方案，是具备分钟级RPO的HDFS远程备份系统，提供数据一致性备份以及高效备份恢复机制，如图2所示。

为了实现大数据的增量备份与恢复，系统可配置备份服务器地址和备份时间间隔，可在任意时刻输入备份命令开始进行备份，并可在指定的时间间隔进行快速增量备份，数据块备份过程中能够通过重复数据删除减少网络传输，支持对任意版本的数据快速恢复，可以指定目录备份，并支持多源端备份。

大数据容灾备份总体架构如图3所示。

HDFS上存储的数据包括元数据和业务数据，分别存储在NameNode和DataNode上。为实现基于HDFS的增量备份与恢复，生产系统即源端的NameNode和DataNode需要针对备份与恢复进行改造。详细结构如图4所示。

NameNode包括对元数据的内容检测、备份存储、备份过滤、备份传输以及备份恢复。

DataNode负责发送和接收要备份/恢复的块数据，DataNode包括如下模块。

- 备份传输：根据NameNode指令将块数据传送到备份服务器。
- 备份删冗：计算块数据的散列值。
- 传输加密：对需要传输的数据进行

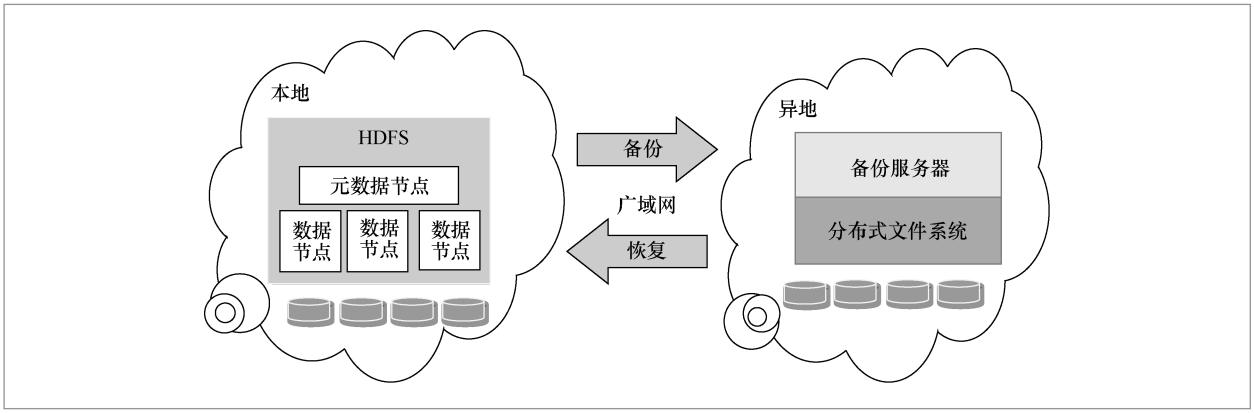


图 2 基于 HDFS 的远程容灾备份系统

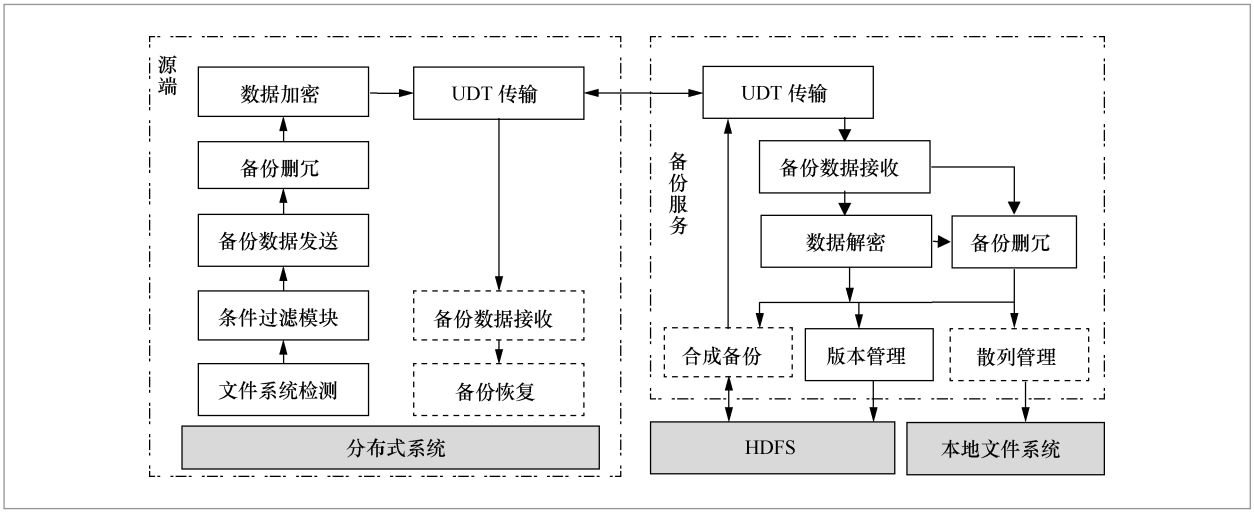


图 3 大数据容灾备份总体架构

加密，加密算法为DES算法。

● 备份恢复：从备份端恢复数据。

备份端对备份数据进行删冗处理，减少传输过程的数据量，提升备份性能，传输前进行数据加密，保证数据传输安全。备份端模块架构如图5所示。

备份服务器模块组成功能如下。

● 备份接收：接收源端发送的备份数据。

● 传输解密：对备份数据进行DES解密。

● 备份删冗：建立备份端块数据散列池，并比对源端发送的散列值是否存在于此散列值，若是，指示源端不必发送数据。

● 版本管理：对源端的多次备份区分不同版本号，并加以时间戳标志。

● 备份存储：将备份历史数据存入文件系统。

● 合成备份：合成最新备份版本。

● 备份恢复：向恢复端发送指定版本的元数据和数据。

备份数据传输前，传输删冗大幅降低冗余备份数据传输量，DataNode切分数据块并计算每个块的散列值，备份端保存历史散列池。同时，采用数据加密保证数据传输安全。详细流程实现如图6所示。

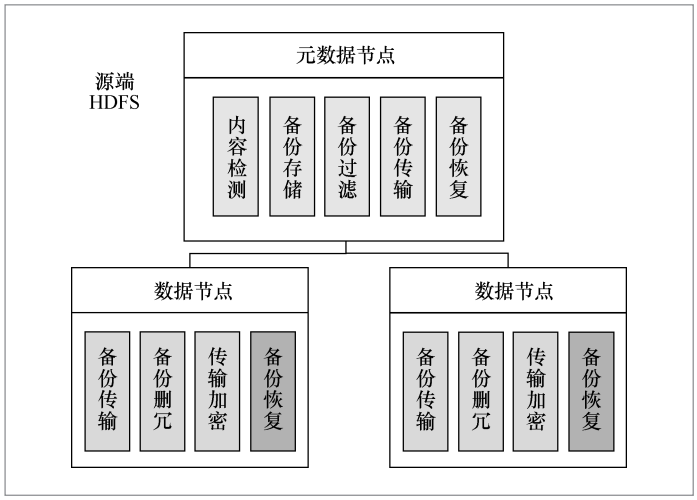


图 4 源端模块组成

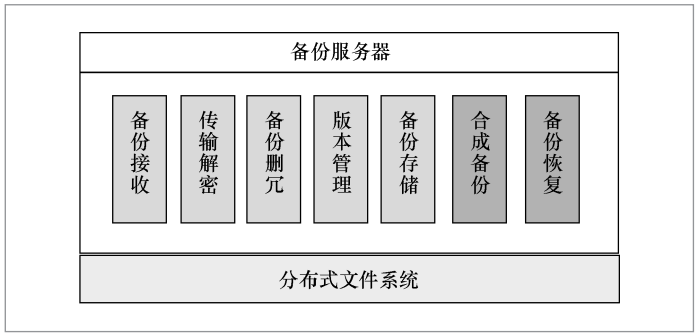


图 5 备份端模块组成

## 3 结束语

在大数据时代，数据是企业最重要的资产，确保数据在环境异常情况下的可靠

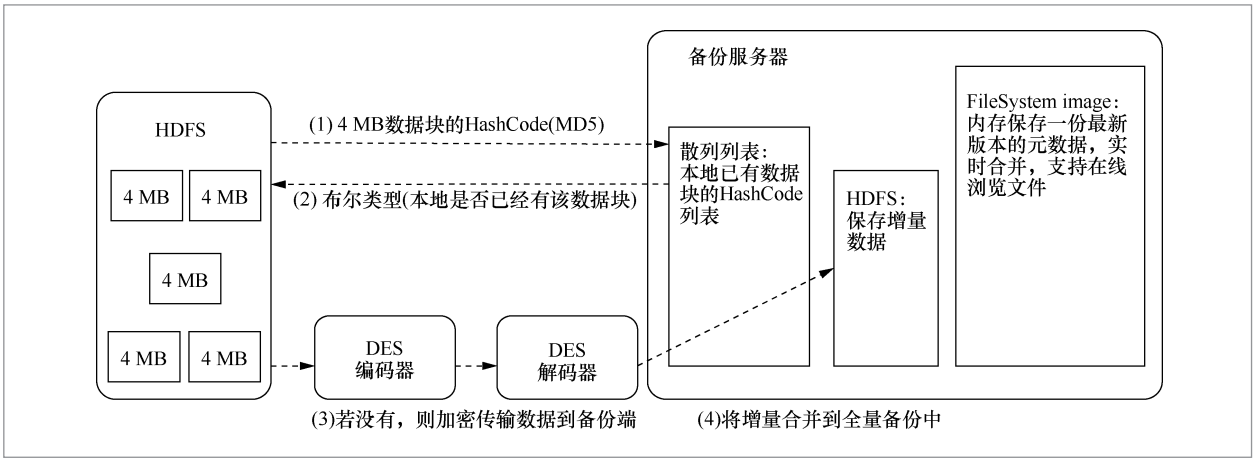


图 6 数据备份流程

性和可用性是目前技术研究和产业应用的重点方向。本文通过大数据容灾备份现状分析，结合行业对大数据容灾备份需求，讨论了几种典型的技术解决方案及其优缺点，提出了一种基于HDFS的增量数据备份恢复方案，具备分钟级RPO的系统远程备份特性，可以较好地解决目前大数据容灾备份项目建设规划面临的实际需求。随着大数据技术的发展以及应用需求的不断变化，对大数据容灾备份的方案也会提出更高的要求，需要持续性地针对业务需求、应用场景和运维成本进行针对性的设计和实现。

## 参考文献

[1] 姚文斌，伍淳华. 中国灾备标准和产业发展现状. 中兴通讯技术，2010，16(5)
Yao W B，Wu C H. Development of standards and industry of disaster backup and recovery in China. ZTE Communications，2010，16(5)

[2] GBT 20988-2007. 信息安全技术信息系统灾难恢复规范，2007
GBT 20988-2007. Information Security Technology Disaster Recovery Specifications for Information Systems，2007

作者简介

**罗圣美**，男，中兴通讯股份有限公司首席架构师，目前主要从事云计算和大数据技术研究工作。担任2项“863”计划和国家科技重大专项课题组长，荣获省部级科学技术进步奖励3项，拥有20多项发明专利，并在国内外核心期刊发表30多篇学术论文。

**李明**，男，中兴通讯股份有限公司产品经理，目前主要从事云存储、大数据和分布式数据库的研究和管理工作。参加1项国家科技重大项目，申请发明专利5项，在多媒体应用和大数据领域发表论文2篇。

**叶郁文**，男，中兴通讯股份有限公司产品规划部长。目前主要从事大数据存储和应用研究，荣获国家科技发明奖1项，参与2项国家科技重大专项课题研究，申请发明专利8项，发表论文4篇。

**收稿日期：** 2015-08-24
**基金项目：** 国家科技重大专项基金资助项目（No.2013ZX03002004）
**Foundation Item:** The National Science and Technology Major Project (No.2013ZX03002004)
**论文引用格式：** 罗圣美，李明，叶郁文. 大数据容灾备份技术挑战和增量备份解决方案. 大数据，2015033
Luo S M，Li M，Ye Y W. Challenge and solution of big data backup and recovery. Big Data Research，2015033