

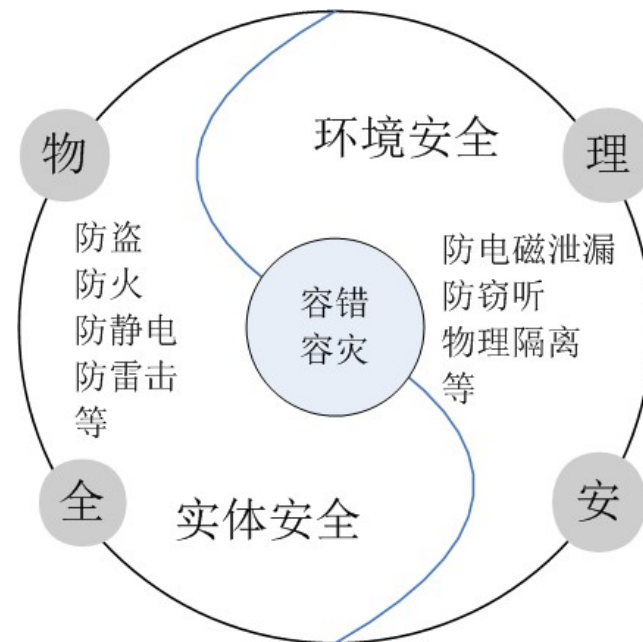


# 容灾(**Disaster Recovery**)

主讲教师：刘亚维

# 物理安全概述

- 物理安全:实体安全和环境安全
- 解决两个方面问题:
  - 对信息系统实体的保护;
  - 对可能造成信息泄漏的物理问题进行防范。
- 物理安全技术包括:
  - 防盗、防火、防静电、防雷击、防信息泄漏、物理隔离;
  - 基于物理环境的容灾技术和物理隔离技术也属于物理安全技术范畴。
- 物理安全是信息安全的必要前提
  - 如果不能保证信息系统的物理安全,其他一切安全内容均没有意义。



# 物理安全概述-容错

## 保证系统可靠性的三条途径

- 避错是完善设计和制造，试图构造一个不会发生故障的系统，但这是不太现实的
- 纠错做为避错的补充。一旦出现故障，可以通过检测、排除等方法来消除故障，再进行系统的恢复。
- 容错是第三条途径。其基本思想是即使出现了错误，系统也可以执行一组规定的程序；

# 容错系统分类

- ① **高可用度系统**：可用度用系统在某时刻可以运行的概率衡量。高可用度系统面向通用计算机系统，用于执行各种无法预测的用户程序，主要面向商业市场。
- ② **长寿命系统**：长寿命系统在其生命期中不能进行人工维修，常用于航天系统。
- ③ **延迟维修系统**：延迟维修系统也是一种容灾系统，用于航天、航空等领域，要求满足在一定阶段内不进行维修仍可保持运行。
- ④ **高性能系统**：高性能系统对于故障（瞬间或永久）都非常敏感，因此应当具有瞬间故障的自动恢复能力，并且增加平均无故障时间。
- ⑤ **关键任务系统**：关键任务系统出错可能危及人的生命或造成重大经济损失，要求处理正确无误，而且恢复故障时间要最短。

# 常用的数据容错技术

- ① **空闲设备**：也称双件热备，就是备份两套相同的部件。当正常运行的部件出现故障时，原来空闲的一台立即替补。
- ② **镜像**：镜像是把一份工作交给两个相同的部件同时执行，这样在一个部件出现故障时，另一个部件继续工作。
- ③ **复现**：复现也称延迟镜像，与镜像一样需要两个系统，但是它把一个系统称为原系统，另一个成为辅助系统。辅助系统从原系统中接收数据，与原系统中的数据相比，辅助系统接收数据存在着一定延迟。
- ④ **负载均衡**：负载均衡是指将一个任务分解成多个子任务，分配给不同的服务器执行，通过减少每个部件的工作量，增加系统的稳定性。

# 物理安全概述-容灾

- 容灾的含义是对偶然事故的预防和恢复。
- 解决方案有两类
  - 对服务的维护和恢复；
  - 保护或恢复丢失的、被破坏的或被删除的信息。
- 灾难恢复策略
  - (1) 做最坏的打算
  - (2) 充分利用现有资源
  - (3) 既重视灾后恢复，也注意灾前措施
- 数据和系统的备份和还原
  - 是事故恢复能力的重要组成，
  - 数据备份越新，系统备份越完整的机构部门就越容易实现灾难恢复操作。

# 主要内容

- 信息——企业的财富与麻烦
- 容灾概述
- 容灾的分级
- 国内外研究现状
- 关键技术



# 前言



- 1983年，个人电脑还处于萌芽期的时候
- 美国青年戴尔（**Dell**）成立了自己的个人电脑公司，主要销售**IBM**的旧电脑和自己组装的品牌电脑。
  - 彼时，行业的领导者们争相以引人注目的技术推出计算机，戴尔注意到了平凡的供应链。
  - 虽然没有傲视群雄的杰出技术，现在的戴尔公司却已成长为一个年销售额达**410**亿美金的企业。



# 前言



- 戴尔公司利用信息技术全面管理公司生产过程。
  - 通过互联网，戴尔公司及其上游的配件制造商能够对客户的订单迅速地做出反应：
    - 当订单传至戴尔的控制中心时，控制中心把订单分解为一个个子任务，并通过网络分派给各独立配件制造商进行生产。
    - 各制造商按照戴尔的电子订单进行生产组装，并按照戴尔控制中心的时间表来供货。
    - 戴尔所需要做的只是在成品车间完成组装和系统测试，剩下的就是客户服务中心的事情了。
    - “经过优化后，戴尔供应链每20秒钟汇集一次订单”，“平均库存时间仅有7小时”。

# 前言



- 对戴尔公司来说，市场信息的获取、物流信息的传递以及合作伙伴的信息交换，这些共同构成了拉动企业正常运转的信息链。
- 如果有一天，一场意外的事故导致供应链的崩裂，戴尔该如何面对客户恼怒的面容和企业直线下滑的利润？

# 前言



- 信息，作为企业宝贵的资源，其重要性已经得到了人们的充分认识。
  - 如何保护这一资源？
- 假设您就是某企业的一位高级管理人员，当您的企业遭遇以下事故时，您将如何去面对：
  1. 某一天，证券公司的交易数据因操作失误而损坏；
  2. 某一天，保险公司的所有保单数据因电源故障而丢失；
  3. 石油勘探公司辛苦一年获取的地质数据因人为的恶意操作而丢失；
  4. 医院保存的所有病历因为磁带的损坏而无法使用；

这样的例子还有很多很多。

# 前言




- 以上事故所带来的后果
  - 很难想象这个不幸的企业还能毫发无损的健康生存。
- 对于信息时代的企业而言，健全的信息往往是维持其运转所必须的基本条件。
- 如何保护企业的信息资源，如何使企业免遭信息灾难，已经成为企业所必须考虑的沉重问题。

# IT 大集中——把蛋都装进篮子里

- 在计算机应用的早期，大型主机一统天下。
  - 一种高度集中的信息应用模式。
  - 昂贵的计算机和存储设备躲藏在幽深的机房里，客户仅能依靠哑终端与主机进行交互，以完成自己的工作。

# IT 大集中



- 随着IT设备的降价和网络技术的发展，客户机 / 服务器体系结构和浏览器 / 服务器体系结构这样的信息应用模式应运而生。
  - 降低了用户进入计算机应用系统的门槛
  - 推进了计算机应用在现代社会的全面普及
  - 产生了计算机应用分布式存在和数据存储分布式存在的局面。

# IT 大集中




- 合久必分，分久必合。
  - 随着网络速度的进一步提高以及高速存储设备的降价，高速信息交换、大容量存储等困扰IT人员多年的问题基本得到了解决。
  - 过于分布的应用和数据所导致的日益昂贵的维护和运营费用，已经给大型企业的发展带来了束缚。

# IT 大集中

- 在银行信息化领域，数据大集中已经成为了一种必然趋势。
  - 在国内，中国工商银行在2000年就前瞻性地启动了数据大集中工程，并在2002年完成了全部工程的建设。
    - 现在，中国工商银行已经将分布在全国各地的四十多个数据中心整合为互相连接、互为备份的北京、上海两大数据中心，建成了全行统一的计算机系统平台。
  - 其它银行和大型证券公司也迎头赶上。
- 大集中已经成为包括银行、证券、保险等行业在内的整个金融信息化发展的大趋势。



# IT 大集中



- 鉴于信息资源对于企业的宝贵作用
  - 信息比作一枚枚金蛋
  - 信息基础设施就是用来装这些金蛋的篮子。
- 过去，不同的金蛋分布在不同地域的篮子里，而大集中所带来的信息基础设施整合则意味着我们将把越来越多的金蛋放进同一个篮子。
- 现在，一个不得不考虑的问题：
  - 如果这个篮子翻了，怎么办？覆巢之下，岂有完卵？

# 容灾，覆巢之下，亦有完卵

- 2001年9月11日，美国世贸中心双子大厦遭受了谁也无法预料的恐怖打击。
  - 灾难发生前，约有**350**家企业在世贸大厦中工作。
  - 事故发生一年后，重返世贸大厦的企业变成了**150**家，有**200**家企业由于重要信息系统的破坏，关键数据的丢失而永远的关闭、消失了。
    - 其中的一家公司称，自己要恢复到灾难前的状态需要**50**年的时间。

# 容灾



- 2003年，国内某电信运营商的计费存储系统仅发生了两个小时的故障，就造成**400**多万元的损失。
  - 尚不包括对公司声誉的影响所导致的无形资产流失。

# 容灾



- 据IDC的统计数字表明，美国在2000年以前的10年间发生过灾难的公司中，有**55%**当时倒闭。剩下的**45%**中，因为数据丢失，有**29%**也在两年之内倒闭，生存下来的仅占**16%**。
- 国际调查机构**Gartner Group**的数据表明，在由于经历大型灾难而导致系统停运的公司中，有**2 / 5**再也没有恢复运营，剩下的公司中也有**1 / 3**在两年内破产。

# 容灾



- 美国德克萨斯州大学的调查显示：“只有**6%**的公司可以在数据丢失后生存下来，**43%**的公司会彻底关门，**51%**的公司会在两年之内消失。”
- 另一份针对这一课题的研究报告也显示：在灾难之后，如果无法在**14**天内恢复信息作业，有**75%**的公司业务会完全停顿，**43%**的公司再也无法重新开业，**20%**的企业在两年之内被迫宣告破产。
- 美国明尼苏达大学的研究也表明，在遭遇灾难的同时又没有灾难恢复计划的企业中，将有超过**60%**在两到三年后退出市场。而随着企业对数据处理依赖程度的递增，此比例还有上升的趋势。

# 容灾



- 灾难的发生对企业的打击往往是致命的。
- 面对灾难，企业就真的不堪一击吗？

# 容灾



- “9·11”，世贸大厦倒塌后，在世贸大厦租有**25层**的金融界巨头摩根斯坦利公司最为世人所关注。
- 事发几个小时后，该公司宣布：全球营业部可以在第二天照常工作。
  - 该公司建立的数据备份和远程容灾系统，它们保护了公司的重要数据，在关键时刻挽救了摩根斯坦利，同时也在一定程度上挽救了全球的金融行业。
- 这一独特的例子说明了什么？
  - 拥有先知先觉的防范意识和充分的技术准备，即使是在突如其来的覆巢之灾下，亦有完卵，亦有企业的一线生机。

# 容灾



- 预防灾难的发生，充分考虑灾难发生后的快速恢复手段，成为现代企业的一门必修课。
  - 中国古代的智者早就提出了自己的观点：生于忧患，死于安乐。
  - 无论是对一个国家，还是一个企业，都是如此。



# 主要内容

- 信息——企业的财富与麻烦
- 容灾概述
- 容灾的分级
- 国内外研究现状
- 关键技术



# 容灾概述

- 哪些事件可以定义为灾难呢？

- 自然灾害

- 典型的灾难事件，如火灾、洪水、地震、飓风、龙卷风、台风等；

- 原先提供给业务运营所需的服务中断

- 设备故障、软件错误、电信网络中断和电力故障等等。

- 人为的因素往往也会酿成大祸

- 操作员错误、破坏、植入有害代码和恐怖袭击。

# 灾难与错误的区别

- 错误（容错领域中对于错误的定义）：
  - 逻辑上把硬件（软件）的实际输出与理论输出不一致称为错误，把导致错误的原因称为故障。
- 灾难：
  - 一旦发生系统便停止工作，不会产生输出，更不会有错误的输出。

现在的容错研究领域，已经涉及到了硬件损坏、断电等错误，他们称这类错误为**fail-stop**错误

这类错误出现后，系统便停止工作，而不是继续运行，产生错误的结果。从这个角度来看，所有的**fail-stop**错误都可以算做是灾难。

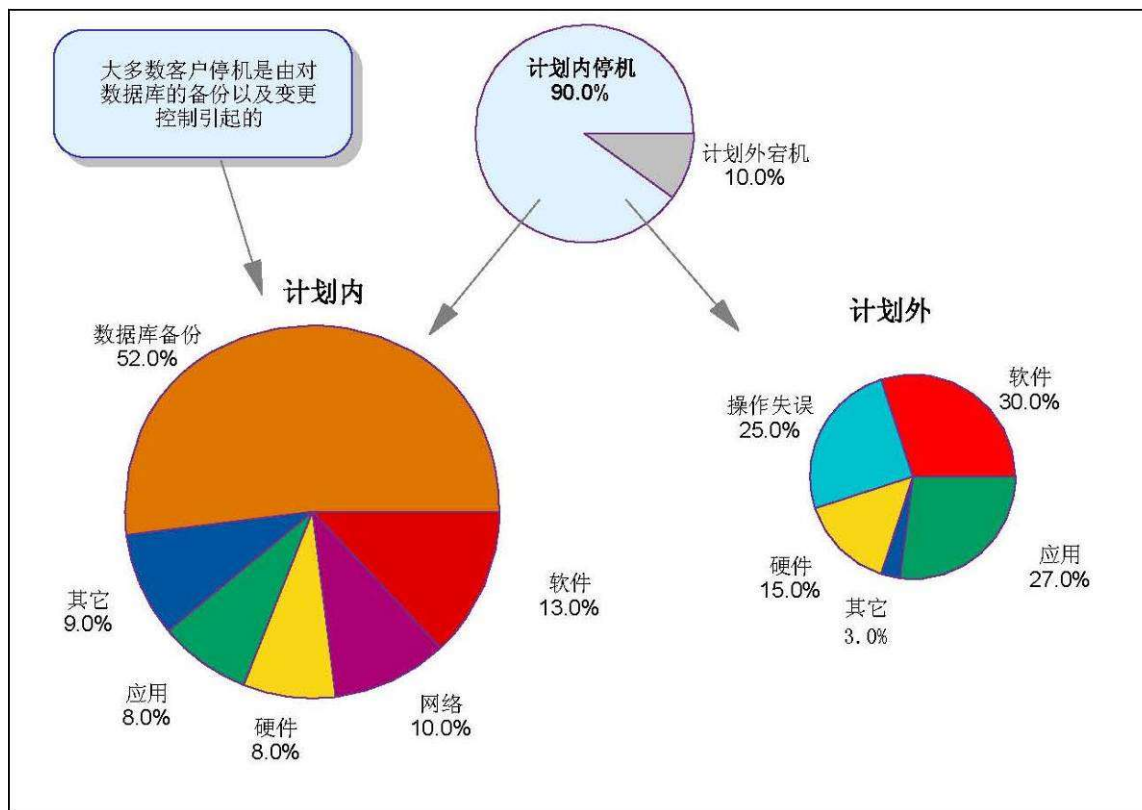
# 容灾概述

- 我国在这方面的损失屡见不鲜
  - 很多行业正处在高速发展的阶段
  - 很多生产流程和制度仍不完善
  - 缺乏经验。
    - 我国2003年遭遇的“非典”，某种意义上也是灾难。
- 对此，需要做到两点：
  1. 建立切实可行的应急机制  
这主要包含一套基于充分且清楚地将风险予以分类定义的业务持续计划
  2. 在危机突然降临时，此计划能被有效执行。

# 停机原因分析

- IT系统

- 除了上述的灾难之外，与系统相关的计划外宕机也可视作灾难



停机原因分析——北美

# 容灾概述



- “9·11”之后，全球各企业均认识到灾难防范保护的重要性。
  - 某些大型金融机构之所以能够在两天内恢复营业，其**主要原因**是它们不仅象一般公司那样在内部进行数据备份，而且在数英里外的数据备份中心也保留着数据备份。
    - 这些备份都是通过数据备份软件和数据复制软件进行的。
    - 采取了这种措施后，一旦工作现场发生意外，企业就可以立即使用另一套数据。

# 传统的数据系统的安全体系

- 备份系统:

- 提供应用系统的数据后援, 确保在任意情况下数据具有完整的恢复能力。

- 高可用系统:

- 确保本地应用系统在多机环境下具有抗御任何单点故障的能力, 一旦系统发生局部的意外(如操作系统故障、掉电、网络故障等), 高可用系统可以在最短的时间迅速确保系统的应用继续运行(热备份)。

# 容灾技术的原理



- 通过在异地建立和维护一个备份系统，利用地理上分散性来保证数据对于灾难性事件的抵御能力。



# 容灾技术的原理

- 容灾系统在实现中可分为两个层次：
  - **数据容灾**指建立一个异地的数据系统，作为本地关键应用数据的一个备份。
  - **应用容灾**是在数据容灾的基础上，在异地建立一套完整的与本地生产系统相当的备份应用系统(可以是互为备份)，在灾难情况下，远程系统迅速接管业务运行。
    - 应用容灾是更高层次的容灾系统。

# 容灾计划



- 华尔街的金融机构重新对灾难恢复的步骤做了评估，并认识到灾难恢复只是技术手段之一，它们开始强调**Business Continuity** (业务连续性)而不仅仅是**Disaster Recovery**(“灾难”恢复)。
  - 过去的“灾难”恢复计划并没有强调全局性及对整个市场的影响，而如何维持业务的连续运作将成为企业运营风险评估中至关重要的一环。

# 容灾计划

- 严格的说，容灾计划包括一系列应急计划，如

- 业务持续计划(BCP-Business Continuity Plan),
- 业务恢复计划(ERP-Business Recovery Plan),
- 运行连续性计划(COOP Continuity of Operations Plan),
- 事件响应计划(IRP Incident Response Plan),
- 场所紧急计划(OEP Occupant Emergency Plan),
- 危机通信计划(CCP Crisis Communication Plan),
- 灾难恢复计划(DRP-Disaster Recovery Plan)
- .....

# 容灾的实质 确保永不停顿的业务运营

- 一个真实的故事：
- Fred Alger基金管理公司的总部设在世贸中心北楼的93层。
  - 在上个世纪90年代，Fred Alger曾是美国业绩最好的一家基金管理公司。它旗下的“光谱共同基金”(Spectra mutual fund)的年均收益率曾达到让人惊羡的29%。然而，公司2000年的业绩大幅下滑，其前景不容乐观。
- 2001年9月11日上午发生恐怖袭击后，该公司正在上班的35人全部遇难，老板David Alger也在其中，这对Fred Alger公司来说无疑是灭顶之灾。

# 容灾的实质 确保永不停顿的业务运营



- 但所幸的是，该公司居安思危，在繁荣期建设的IT系统早早就考虑到容灾的需要，在**50**英里以外的新泽西中心区建有一个数据备份点。
- “9·11”过后的第三天，该公司幸存无几的人在那里发现，袭击之前所有的交易记录和所有的研究报告都有详细备份，并被完好无损地保留了下来。

# 容灾的实质 确保永不停顿的业务运营

- Fred Alger公司没有选择关张，而是决定重建。
- 几年前就已退休的Fred Alger，在弟弟David去世后立刻再度出山。
- 当整个市场在重新开市时，Fred Alger公司成了华尔街经纪公司中的股票大买家。
- 很快，Fred Alger公司的投资管理队伍也空前兴旺起来，并在第五大道的2层楼建立了新的总部。

# 容灾的IT实现



- 在技术层面上，容灾需要考虑
  - 数据版本保护
  - 实时数据保护
  - 应用系统恢复
  - 网络系统恢复
  - 容灾切换决策
  - 容灾切换过程
- 备份至关重要
  - 定期测试备份的可靠性

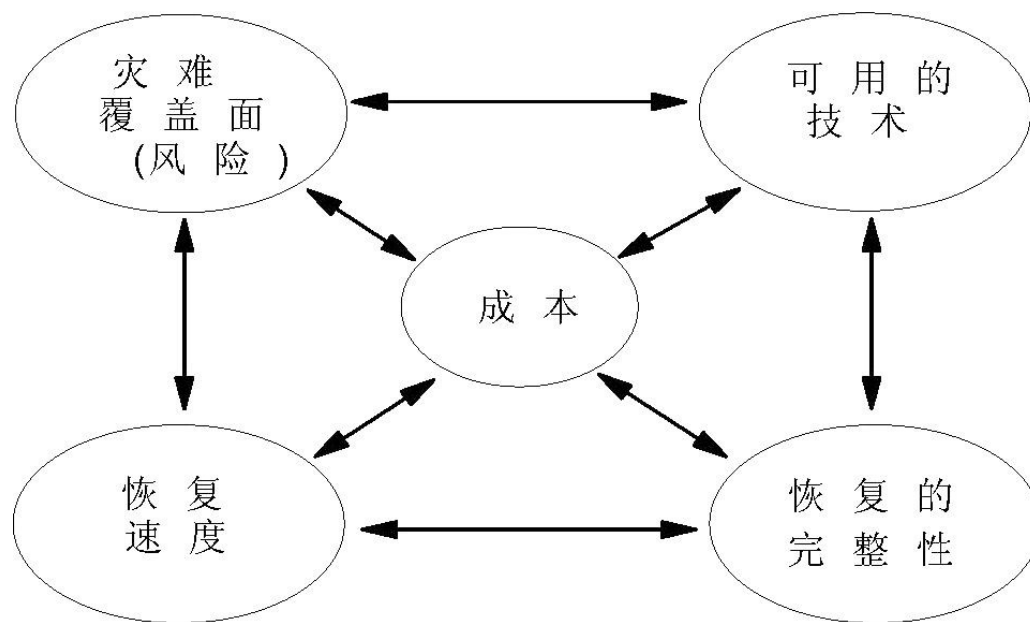
# 容灾的IT实现

- IT系统相关的灾难备份方案设计必须考虑以下五大因素
  - 1, 灾难类型
  - 2, 恢复速度
  - 3, 恢复程度
  - 4, 可用的技术
  - 5, 方案总体成本



# 容灾的IT实现


- 综合以上所述，可以如图所示：



灾难备份方案选择标准

# 主要内容



- 信息——企业的财富与麻烦
  - 容灾概述
  - 容灾的分级
  - 国内外研究现状
  - 关键技术
- 

# 容灾的7个层次



- 0层——没有异地数据 (No off-site Data)
- 1层——PTAM卡车运送访问方式(Pickup Truck Access Method)
- 2层——PTAM卡车运送访问方式+热备份中心(PTAM + Hot Center)
- 3层——电子链接(Electronic Vaulting)

# 容灾的7个层次



- 4层——活动状态的备份中心(Active Secondary Center)
- 5层——两个活动的数据中心，确保数据一致性的两阶段传输承诺(Two-Site Two-Phase Commit)
- 6层——0数据丢失(Zero Data Loss)，自动系统故障切换


# 容灾的业务恢复时间段

- IT系统的容灾指标，我们可以通过下列参数表示：
  - 以恢复点为目标(RPO Recovery Point Object)
    - 数据的完整性(无数据丢失)
    - 数据的一致性(数据正确且可用)
  - 以恢复时间为目标(RTO Recovery Time Object)
  - 以网络恢复为目标(NRO Network Recovery Object)
  - 以服务支持能力为目标(SDO Serviceability Degrade Object)

# 容灾所涉及的恢复技术

- DR (容灾Disaster Recovery)项目的实施中涉及到三类技术：
  - 应用恢复
  - 网络恢复
  - 数据恢复

# 恢复的7个层次



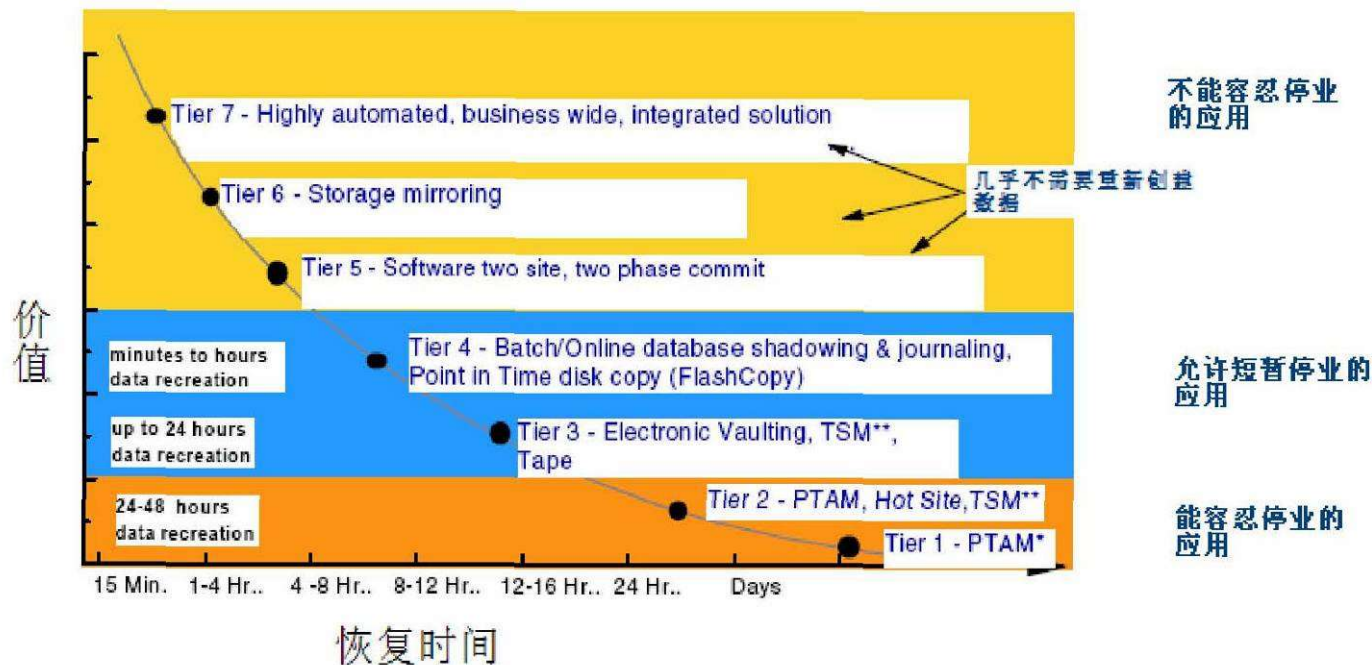
- 灾难恢复方案中的恢复时间与下列因素有关：

- 数据有效性的恢复
- IT基础设施的恢复
- 可操作流程的修复
- 关键业务的修复

# 灾难恢复的层次划分

## 灾难恢复的层次划分

最好的D/R解决方案是综合考虑不同层次的恢复方案,以最少的投资换取最大的收益.只使用一种方法,一种技术是不可能满足企业中所有应用的需要.

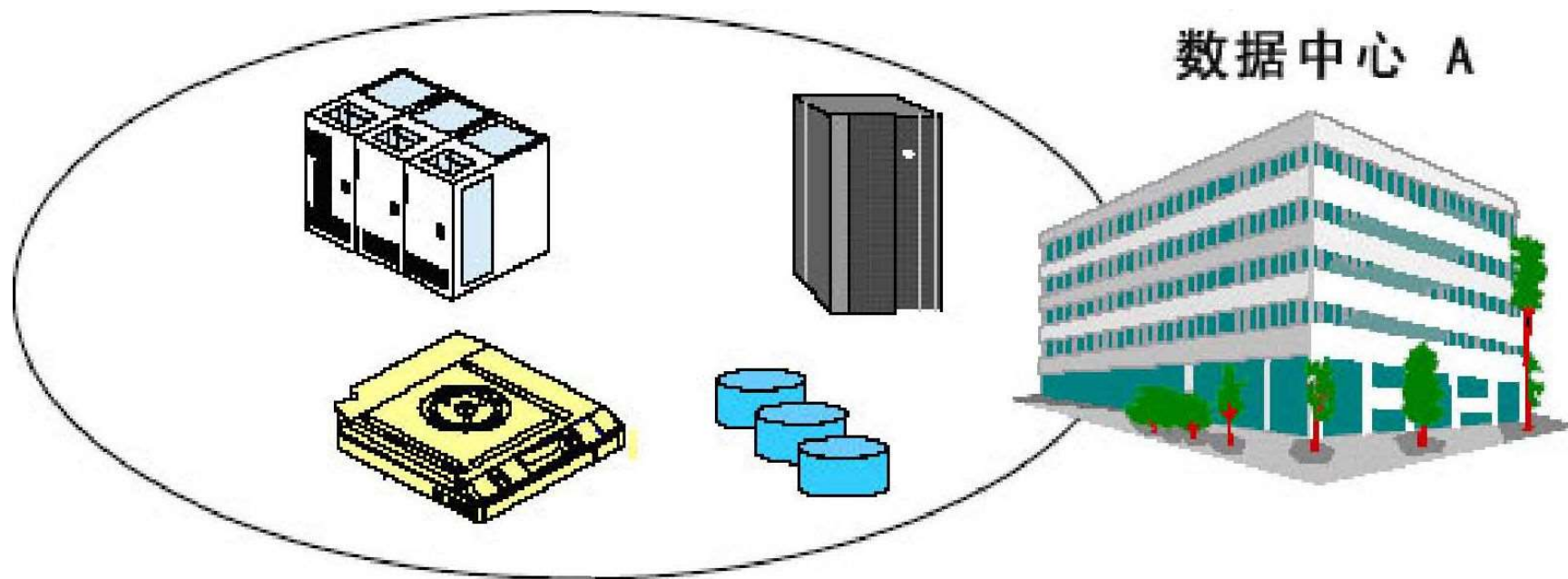




## 0层：无异地备份数据 (No off-site Data)

- 对于使用0层灾难恢复解决方案的业务，可称其为没有灾难恢复计划
  - 数据仅在本地进行备份恢复，没有任何数据信息和资料被送往异地，没有处理意外事故的计划。
  - 恢复时间：在此种情况下，恢复时间不可预测。事实上也不可能恢复。

# 0层：无异地备份数据



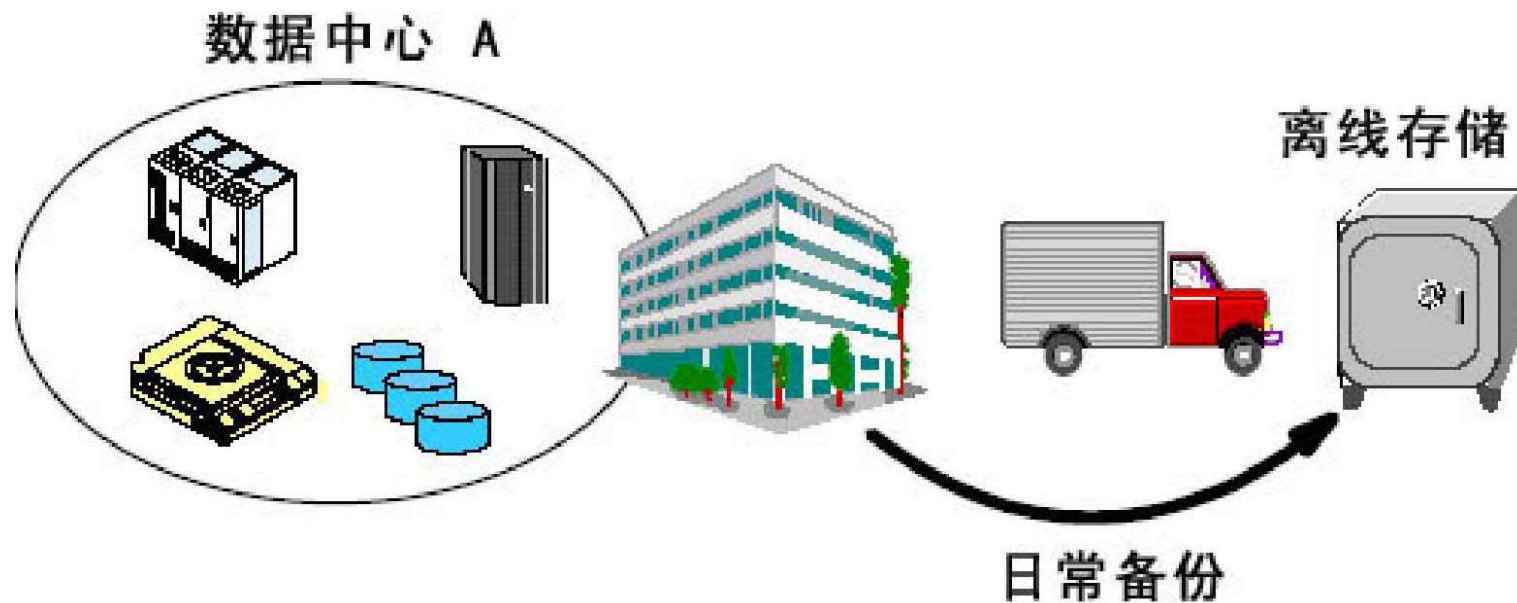
## 1层：有数据备份，无备用系统(Data Backup with No Hot Site)

- 使用1层灾难恢复解决方案的业务，通常将需要的数据备份到磁带上，然后将这些介质运送到其它较为安全的地方。
  - 但在那里缺乏能恢复数据的系统，若数据备份的频率很高，则在恢复时丢失的数据就会少些。
  - 此类业务应能忍受几天乃至几星期的数据丢失。

# 1层：有数据备份，无备用系统(Data Backup with No Hot Site)

- PTAM (Pickup Truck Access Method)是一种许多数据中心所采用的标准备份方式。
  - 在完成所需的数据备份后，用适当的运输工具将它们送到远离本地的地方，同时备有数据恢复的程序。
  - 灾难发生后，一整套系统安装需要在一台未开启的计算机上重新完成，系统和数据可以被恢复并重新与网络相连。
  - 这种灾难恢复方案相对来说成本较低(仅仅需要运输工具的消耗以及存储设备的消耗)。但恢复的时间长，且数据不够新。

# 1层：有数据备份，无备用系统(Data Backup with No Hot Site)



## 2层：有数据备份，有备用系统(Data Backup with Hot Site)

- 定期将数据备份到磁带上，并将其运到安全的地点。
- 在备份中心有备用的系统，当灾难发生时，可以使用这些数据备份磁带来恢复系统。
- 虽然还需要数小时或几天的时间来恢复数据以使业务可用，但不可预测的恢复时间减少了。

## 2层：有数据备份，有备用系统(Data Backup with Hot Site)

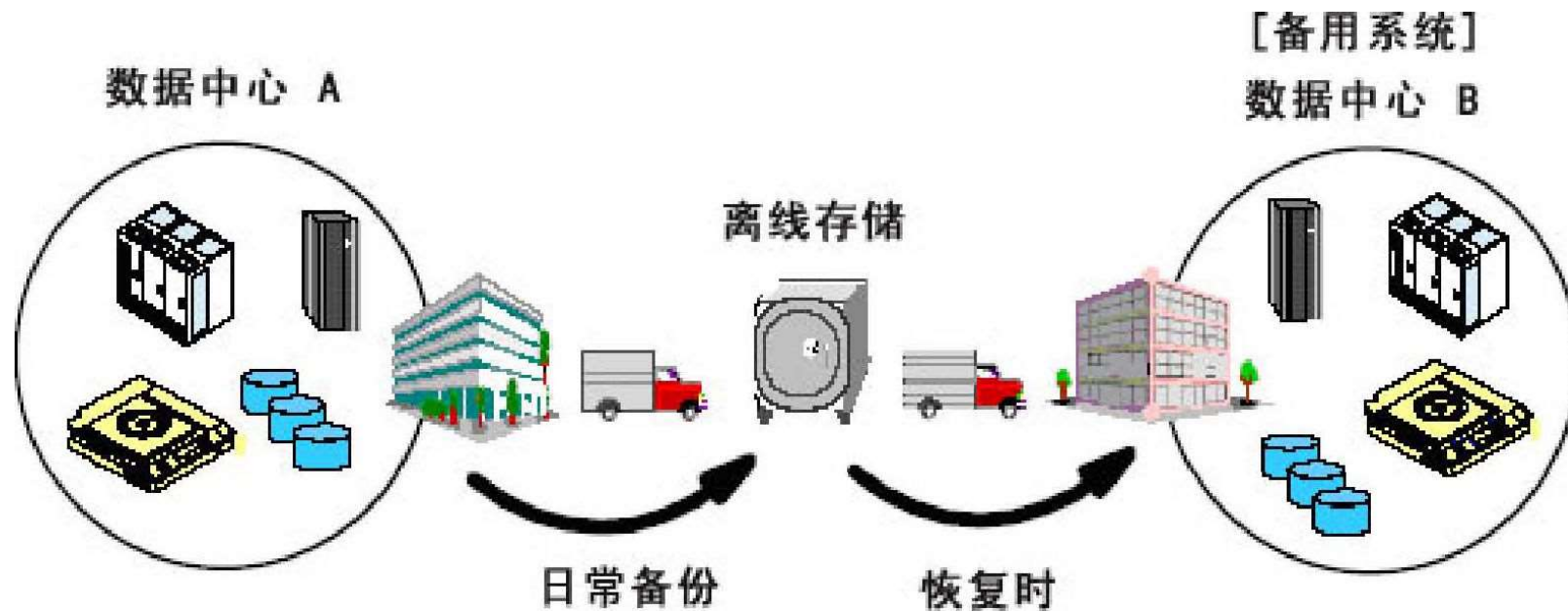
- 2层相当于在1层上增加了备份中心的灾难恢复。
- 备份中心拥有足够的硬件和网络设备来维持关键应用的安装需求
  - 这样的应用是十分的关键的，它必须在灾难发生的同时，在异地有正运行着的硬件提供支持。

## 2层：有数据备份，有备用系统(Data Backup with Hot Site)

- 这种灾难恢复的方式依赖于PTAM方法去将日常数据放入仓库，当灾难发生的时候，再将数据恢复到备份中心的系统上。
- 虽然备份中心的系统增加了成本，但明显降低了灾难恢复时间，系统可在几天内得以恢复。



## 2层：有数据备份，有备用系统(Data Backup with Hot Site)



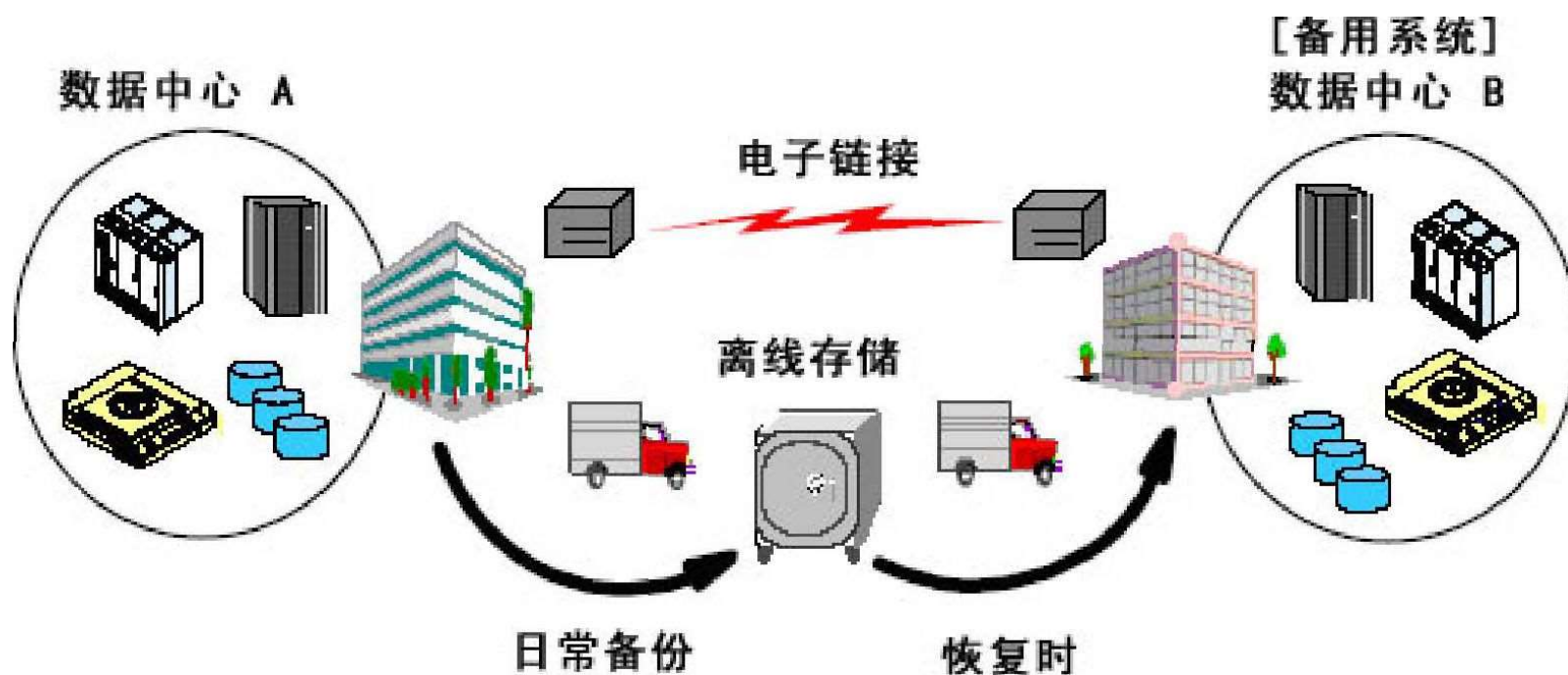
### 3层：电子链接 (Electronic Vaulting)

- 是在2层解决方案的基础上，又使用了对关键数据的电子链接技术。
  - 电子链接将磁带备份后更改的数据进行记录，并传到备用中心，使用此种方法会比使用传统的磁带备份更快地得到更新的数据。
- 恢复时间会缩短
  - 当灾难发生后，只有少量的数据需要重新恢复

### 3层：电子链接 (Electronic Vaulting)

- 增加了运营成本
  - 由于备用中心要保持持续运行，与生产中心间的通讯线路要保证畅通，
- 提高了灾难恢复速度。
  - 消除了对运输工具的依赖

### 3层：电子链接 (Electronic Vaulting)



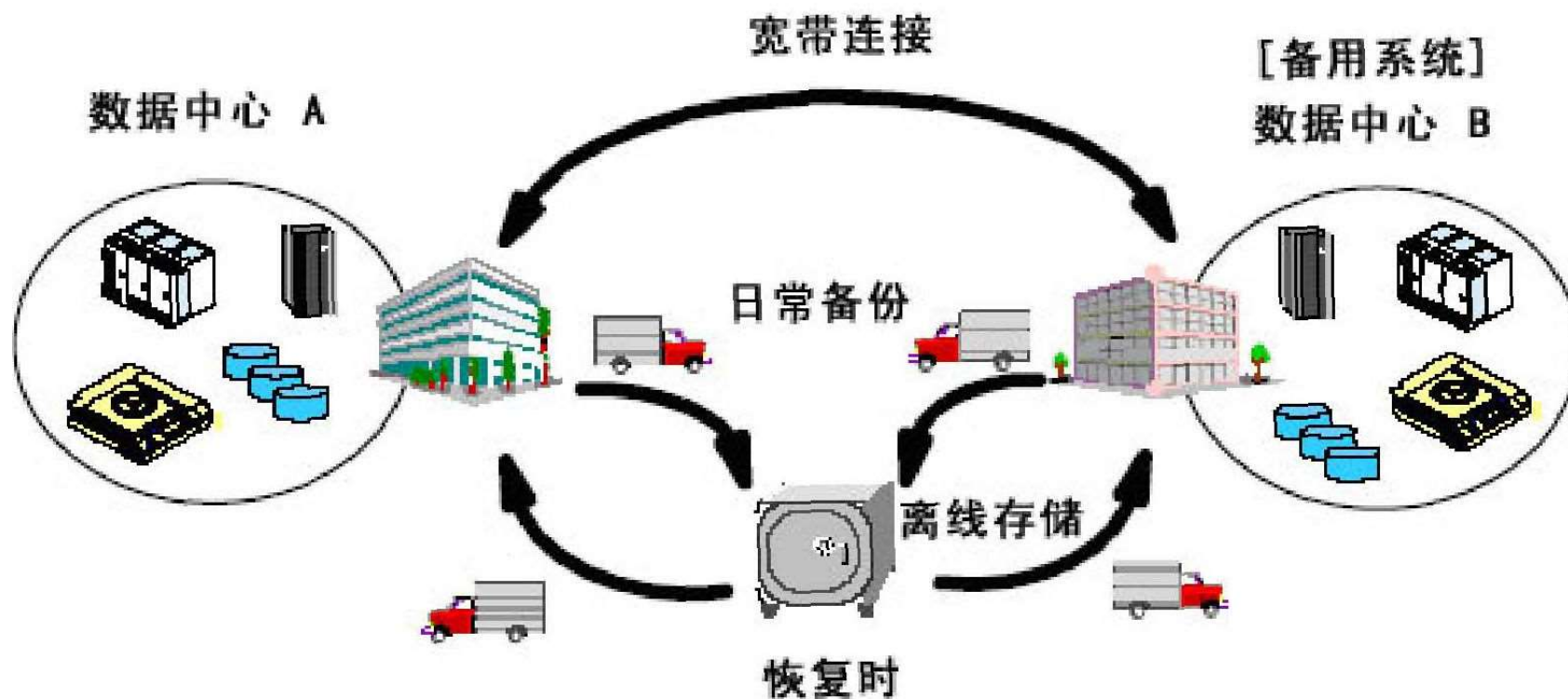
## 4层：使用快照技术拷贝数据(Point-in-time Copies)

- 对数据的实时性和快速恢复性要求更高些
- 开始使用基于磁盘的解决方案
  - 下3层的方案中较常使用磁带备份和传输
- 通过加快备份频率，使用最近时间点的快照拷贝恢复数据会更快
  - 仍然会出现几个小时的数据丢失
  - 同基于磁带的解决方案相比，系统可在一天内恢复。

## 4层：使用快照技术拷贝数据(Point-in-time Copies)

- 可有两个中心同时处于活动状态并管理彼此的备份数据，允许备份行动在任何一个方向发生。
- 接收方硬件必须保证与另一方平台在地理上分离
  - 工作负载可能在两个中心之间分享，中心1成为中心2的备份，反之亦然。
  - 在两个中心之间，彼此的在线关键数据的拷贝不停地相互传送着。
- 关键应用的恢复也可降低到小时级。
  - 在灾难发生时，需要的关键数据通过网络的切换，可迅速恢复

## 4层：使用快照技术拷贝数据(Point-in-time Copies)



## 5层：交易的完整性(Transaction Integrity)

- 要求保证生产中心和数据备份中心的数据的一致性。
- 在此层方案中只允许少量甚至是无数据丢失，但是该功能的实现完全依赖于所运行的应用。



## 5层：交易的完整性(Transaction Integrity)

- 除了使用4层的技术外，还要维护数据的状态，要保证在本地和远端数据库中都要更新数据。
  - 只有当两地的数据都更新完成后，才认为此次交易成功。
  - 生产中心和备用中心是由高速的宽带连接的，关键数据和应用同时运行在两个地点。
  - 当灾难发生时，只有正在进行的交易数据会丢失。
  - 由于恢复数据的减少，恢复时间也大大缩短。

# 5层：交易的完整性(Transaction Integrity)



## 6层：少量或无数据丢失(Zero or little data loss)

- 可以保证最高一级数据的实时性。
  - 适用于那些几乎不允许数据丢失并要求能快速将数据恢复到应用中的业务。
  - 提供数据的一致性，不依赖于应用而是靠大量的硬件技术和操作系统软件来实现的。
    - 这一级别的要求很高，一般需要整个系统应用程序层到硬件层均采取相应措施。



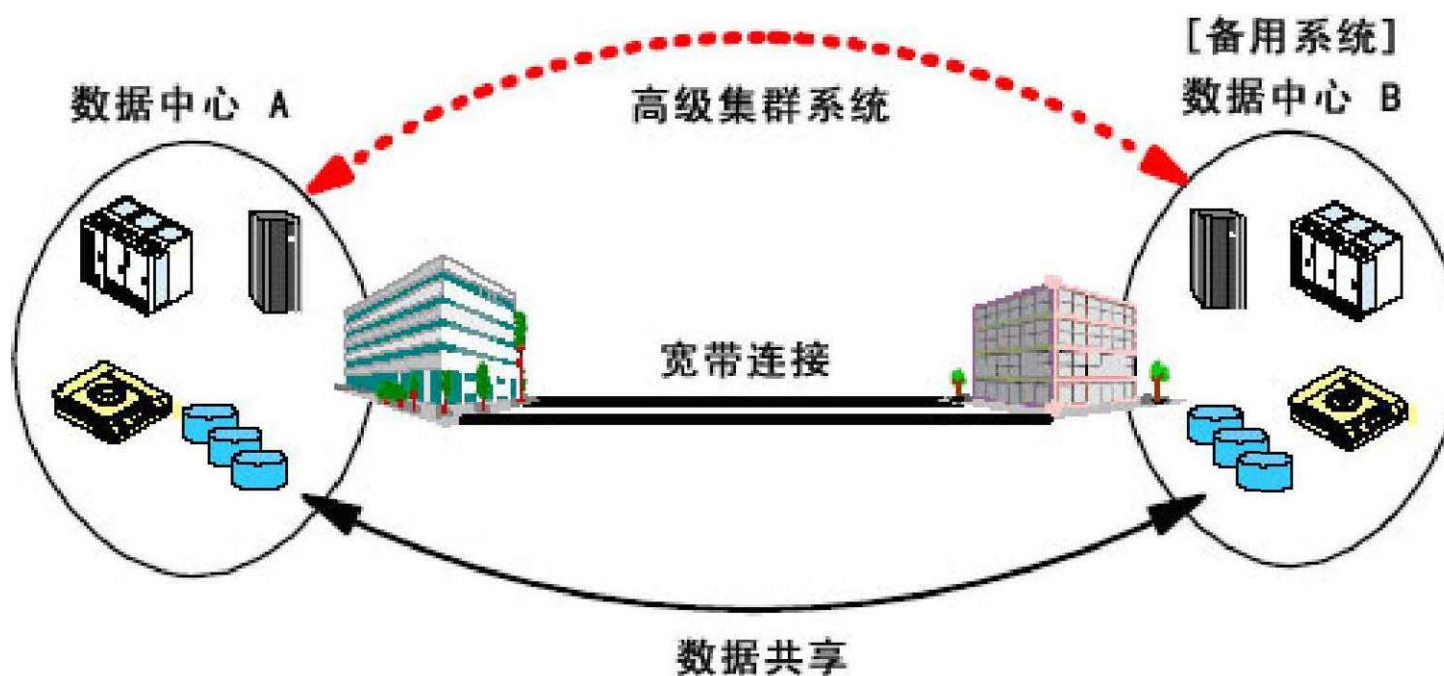
7层：解决方案与具体业务相结合，实现自主管理  
(Highly Automated, Bussiness Integrated Solution)

- 在第6层的基础上，集成了自主管理的功能。
  - 在保证数据一致性的同时，又增加了应用的自动恢复能力，使得系统和应用恢复的速度更快、更可靠
    - (按照灾难恢复流程，手工操作也可实现整个恢复过程)。

## 7层：解决方案与具体业务相结合，实现自主管理 (Highly Automated, Bussiness Integrated Solution)

- 可以实现0数据丢失率，同时保证数据立即自动地被传输到恢复中心。
- 被认为是灾难恢复的最高级别，在本地和远程的所有数据被更新的同时，利用了双重在线存储和完全的网络切换能力。
- 是灾难恢复中最昂贵的方式，但也是速度最快的恢复方式。
- 当一个工作中心发生灾难时，7层能够提供一定程度的跨站点动态负载平衡和自动系统故障切换功能。
  - 现在已经证明，为实现有效的灾难恢复，无需人工介入的自动站点故障切换功能需要一个应该纳入考虑范围的重要事项。

7层：解决方案与具体业务相结合，实现自主管理  
(Highly Automated, Bussiness Integrated Solution)



# 主要内容

- 信息——企业的财富与麻烦
- 容灾概述
- 容灾的分级
- 国内外研究现状
- 关键技术



# 国内外研究现状及相关产品

- 容灾概念提出来的时间不长，但容灾方面的研究很早就在进行
  - 实际上很多容错、高可靠性研究领域的很多东西都属于容灾的范畴
    - 例如：高可靠集群、分布式系统中关于容忍单点时效方面的研究都应该属于容灾的范畴。



# 国内外研究现状及相关产品

- 国内外的研究机构和商业公司也在该方面取得了很多研究成果，很多都已经进入广泛的应用阶段。
  - Hp开发了OpenVMS高可用集群系统
  - VERITAS提供了Volume Replicator
  - IBM给出了基于ESS企业存储服务器的PPRC(Peer to Peer Remote Copy)复制技术的数据容灾方案，以及基于IBM RS / 6000服务器的HAGEO(High Availability Geographic Cluster)异地群集技术的应用级容灾方案
  - 北京装甲兵工程学院、上海欣方智能网有限公司以及北京邮电大学计算机学院合作完成了基于主从异步复制技术的容灾系统
  - 浪潮软件公司实现了基于海量实时数据库的HLR容灾系统

# 主要内容

- 信息——企业的财富与麻烦
- 容灾概述
- 容灾的分级
- 国内外研究现状
- 关键技术



# 关键技术的分析与讨论

- 对于容灾系统来说，所包含的关键技术有五个方面
  - 数据存储管理
  - 数据复制
  - 灾难检测
  - 系统迁移
  - 系统恢复

# 数据存储管理



- 指对与计算机系统数据存储相关的一系列操作(如备份, 归档, 恢复, 近线等)进行的统一管理, 是计算机系统管理的一个重要组成部分, 也是建立一个容灾系统的重要组成部分。

# 数据存储管理

- 数据存储管理工作已经超出早期的数据备份工作的范畴，成为容灾系统管理的一个重要组成部分。
  - 数据备份
  - 数据恢复
  - 备份索引
  - 备份设备及媒体
  - 灾难恢复
  - .....

# 数据备份



- 是指为防止系统出现操作失误或系统故障导致数据丢失，而将全系统或部分数据集合从应用主机的硬盘或阵列复制到其它的存储介质的过程，数据备份是容灾的基石。

# 数据备份



- 分为在线备份和离线备份。

- 在线备份

- 对正在运行的数据库或应用进行备份，通常对打开的数据库和应用是禁止备份的，因此要求数据存储管理软件能够对在线的数据库和应用进行备份。

- 离线备份

- 指在数据库**Shutdown**或应用关闭后对其数据进行备份，离线备份通常只采用全备份。

# 数据归档



- 是将硬盘数据复制到可移动媒体上。
- 与数据备份不同的是，数据归档在完成复制工作后将原始数据从硬盘上删除，释放硬盘空间。



# 数据归档



- 层次化的存储方案


- 计算机硬盘-光盘库-磁带库构成三级模式

- 当上一级媒体的数据量达到规定的上限，便将访问率很低的数据迁移到光盘库中，而当上一级媒体的数据量降低到下限，便将下一级媒体的数据迁移回来

- 这种层次化的存储技术增加了存储的灵活性，又降低了存储代价，通过层次化存储管理软件可以实现迁移的自动化


- 这种方案也存在一定的缺陷，如对数据库的迁移和因策略制定不完善造成的抖动(数据频繁迁移)问题。

# 数据复制



- 容灾系统的核心技术
- 将一个地点的数据拷贝到另外一个不同的物理点上的过程。
- 数据复制一般分为同步数据复制和异步数据复制。

# 数据复制



- 同步数据复制

- 通过将本地生产数据以完全同步的方式复制到异地，每一本地I/O交易均需等待远程复制的完成方予以释放。这种复制方式基本可以做到零数据丢失。

- 异步数据复制

- 指将本地生产数据以后台同步的方式复制到异地，每一本地I/O交易均正常释放，无需等待远程复制的完成。这种复制方式，在灾难发生时，会有少量数据丢失，这与网络带宽、网络延迟、I/O吞吐量相关。

# 数据复制

- 实现数据的异地复制，有软件方式和硬件方式两种途径。

- 软件方式

- 是通过主机端软件来实现。即在主系统和容灾系统的主机上，安装专用的数据复制软件。这种方式的特点是与硬件无关，而且成本较低。但是由于效率较低和可管理性也较差。

- 硬件方式


- 就是数据直接在存储设备之间传输，并不依赖主机的管理。这种方式要求在主系统和容灾系统配置上支持这种功能的专用存储设备，所以成本较高。

# 数据复制



- 根据复制数据的层次进行细化，可以分为以下四种类型：
  - 硬件级的数据复制
  - 操作系统级的复制
  - 数据库级的复制
  - 业务数据流级复制

# 数据复制



- 硬件级的数据复制

- 主要是在磁盘级别对数据进行复制，包括磁盘镜像、卷复制等
- 这种类型的复制方法可以独立于应用，并且复制速度也较快，对生产系统的性能影响也较小，但是开销比较大。


# 数据复制



- 操作系统级的复制

- 主要是在操作系统层次，对各种文件的复制
- 这种类型的复制受到了具体操作系统的限制。

# 数据复制



- 数据库级的复制

- 是在数据库级别将对数据库的更新操作以及其它事务操作以消息的形式复制到异地数据库
- 这种复制方式的系统开销也很大，并且与具体数据库相关。



# 数据复制



- 业务数据流级复制

- 就是业务数据流的复制，就是将业务数据流复制到异地备用系统，经过系统处理后，产生对异地系统的更新操作，从而达到同步。
- 这种方式，也可以独立于具体应用，但是可控性较差。
  - 现在利用这种方式来实现容灾系统的例子还很少。

# 灾难检测

- 有些灾难依靠人很难及时察觉
  - 对于火灾、地震等大规模灾难，当然可以依靠人为确定
  - 对于停电、硬件毁坏等很难觉察到的灾难就不能仅仅依靠人去发现。
- 现在对灾难的发现方法
  - 心跳技术
  - 检查点技术

# 灾难检测

## ● 心跳技术（拉技术）

- 每隔一段时间都要向外广播自身的状态(通常为“存活”状态)，在进行心跳检测时，心跳检测的时间和时间间隔是关键问题
- 如果心跳检测的太频繁，将会影响系统的正常运行，占用系统资源；
- 如果间隔时间太长，则检测就比较迟钝，影响检测的及时性。

# 灾难检测

- 检查点技术（主动检测）

- 每隔一段时间周期，就会对被检测对象进行一次检测，如果在给定的时间内，被检测对象没有相应，则认为检测对象失效。

- 与心跳技术相同，检测点技术也受到检测周期的影响
    - 如果检测周期太短，虽然能够及时发现故障，但是给系统造成很大的开销；如果检测周期太长，则无法及时的发现故障。

# 系统迁移



- 能够利用备用系统透明的代替生产系统。
  - 在发生灾难时，为了保证业务的连续性，必须实现能够实现系统透明的迁移，

# 系统迁移



- 通过**DNS**或者**IP**地址的改变来实现系统迁移
  - 实时性要求不高的容灾系统
- 进程迁移算法
  - 对于可靠性、实时性要求较高的系统
  - 进程迁移算法的好坏对于系统迁移的速度有很大影响
  - 现在该算法在分布式系统和集群中得到了广泛的运用，并发挥着重大作用，也有很多研究对该算法的性能进行了改进。

