KDD：

1）数据截取：爬取郑州的 7 天内的天气

2）数据处理：应用过滤器过滤除掉天气以外所有杂讯

3）数据转换：将处理好的数据分离出来成表格存储

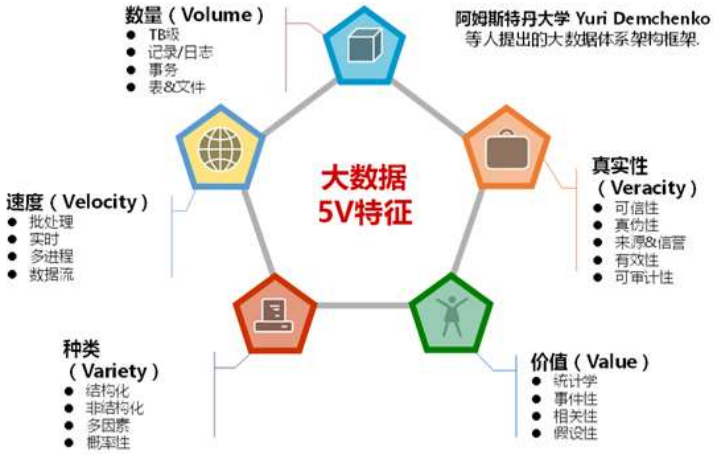
4）数据建模：通过观察提取转好的数据建立模型，如高低温差分析，并且分别提取做温差减法

5）数据解读：利用建立好模型的数据进行图表的生成，报告的撰写

满足数据存储的三个调节：1. 两个或以上的状态 2.状态可以识别 3. 状态可以改变

海量数据：TB 或 PB 级以上大量數據集合

大数据特点：



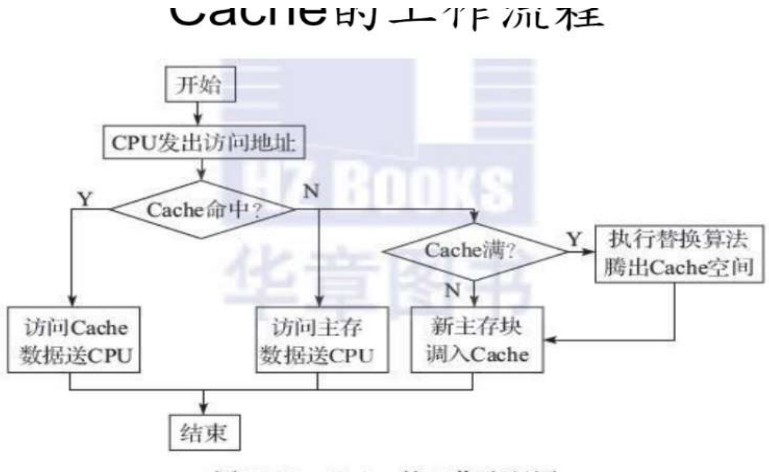
海量数据要求什么：存储容量，实时性，数据安全，存储成本

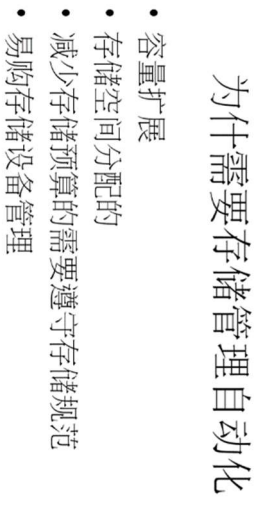
纵向升级 Scale Up：不可以超越系统限制。单单在一个系统升级更多硬件，传统模式。

横向升级 Scale Out：通过网络将多个系统连立起来，不仅扩充容量，还扩充其他系统资源。

存储设备分类：存储介质：光(DISK)，电(SSD)，磁性(HDD)，光电 根据时间：易失和非易

Cache：加速，有效性基于时间局部性和空间局部性





- 主要的替换策略有四种：
  - 先进先出（FIFO）算法、
  - 最不经常使用（LFU）算法、
  - 近期最少使用（LRU）算法和
  - 随机替换。

成员磁盘上的分区可以进一步细分为更小的段，这些更小的段即单个I/O操作的对象，称为分块。

分条：同一磁盘上的分块组成一个分区，各个磁盘的分区构成一个阵列。阵列以分条为单位进行读写。

条带：磁盘中单个或者多个连续的扇区构成一个条带。它是组成分条的元素。

→ 全相联映射

主存中的任一块可以被放置到Cache中的任意一个位置。其特点是空间利用率高，冲突概率最低，实现最复杂。

→ 直接映射

主存中的每一块只能被放置到Cache中的唯一的一个位置。其特点是空间利用率低，冲突概率最高，实现最简单。

→ 组相联映射

主存中的每一块可以被放置到Cache中的唯一的一个组中的任何一个位置。组相联映射是直接映射和全相联映射的一种折中。

- 写回法
  - CPU写Cache命中时，只修改Cache的内容，而不立即写入主存；只有当此行被换出时才写回主存。这种方法减少了访问主存的次数，但是存在不一致性的隐患。实现这种方法时，每个Cache行必须配置一个修改位，以反映此行是否被CPU修改过。
- 全写法
  - 当CPU写Cache命中时，Cache与主存同时发生写修改，因而较好地维护了Cache与主存内容的一致性。当CPU写Cache未命中时，直接向主存进行写入。Cache中每行无需设置一个修改位以及相应的判断逻辑。缺点是降低了Cache的功效。
- 写一次法
  - 基于写回法并结合全写法的写策略，写命中与写未命中的处理方法与写回法基本相同，只是第一次写命中时要同时写入主存（全写法）。这便于维护系统全部Cache的一致性。

CPU在执行一段程序时，Cache完成存取的次数为1900次，主存完成存取的次数为100次。假设Cache存取周期为50ns，主存存取周期为250ns，求Cache/主存系统的访问效率和平均访问时间。

命中率：

$h = N_c / (N_c + N_m) = 1900 / (1900 + 100) = 0.95$

访问主存与访问Cache的时间比：

$r = t_m / t_c = 250ns / 50ns = 5$

访问效率：

$e = 1 / (h + (1 - h) r) = 1 / (0.95 + (1 - 0.95)\times5) = 83.3\%$

平均访问时间：

$t_a = t_c / e = 50ns / 0.833 = 60ns$ 或

$t_a = ht_c + (1 - h) t_m = 0.95\times50ns + (1 - 0.95)\times250ns = 60ns$

存储设备自动化工作

- 存储安装与初始化配置。
- 存储监控
- 存储优化
- 存储日常分配
  - 利用脚本来创建LUN、CIFS/NFS
- 应用的存储自动化分配
- 存储的回收与迁移
- 与主机设备集成。

RAID 特点：高性能、高可用、可扩展、易使用、系统监测和故障报警、支持多种管理

NAS 连接：文件级别(File-level)的访问协议，不同于块级别(Block-level)

云储存系统结构模型

多级结构存储服从原则：一致性，包含性

描述		
网络速度		
网络架构	单独建设光纤网络和HBA卡	使用现有IP网络
传输距离	受到光纤传输距离的限制	理论上没有距离限制
管理、维护	技术和管理较复杂	与IP设备一样操作简单
兼容性	兼容性差	与所有IP网络设备都兼容
性能	非常高的传输和读写性能	目前主流1Gb，占用主机CPU资源
成本	购买（光纤交换机、HBA卡、光纤磁盘阵列等）、维护（培训人员、系统设置与监测等）成本高	与FC－SAN相比，购买与维护成本都较低，有更高的投资收益比例
容灾	容灾的硬件、软件成本高	本身可以实现本地和异地容灾，且成本低
安全性	较高	较低、容易丢包、截取

的
具
个物
个逻
。
储信
息的地方。就像通常的分区一样，在逻辑卷上可以创建文件系统。

有效性局部性原则：时间局部性，恐惧局部性

一致性与效能机制

- 档案一致性机制
  - 删除档案、新增写入档案、读取档案是Namenode负责
- 巨量空间及效能机制
  - 以Block为单位： 64M为单位
  - 在HDFS上得档案有可能大过一颗磁盘
  - 大区块可提高存取效率
  - 区块均匀散布各节点以分散读取流量
- 数据完整性
  - checked with CRC32
  - 用副本取代出错资料
- Heartbeat
  - Datanode 定期向Namenode heartbeat
- Metadata
  - FSImage、Editlog为核心文件
  - 多份储存，当NameNode坏掉可以手动复原

存储设备自动化实现

- 存储设备来自同一厂商。
- 购买第三方存储自动化软件。
- 自行开发：
  - Command Line Interface
  - API
  - REST API
  - SMI-S (storage management initiative specification, SNIA 订定标准管理规范)

如何进行虚拟化

- 带内 In-Band
  - 存储系统、网络、主机、文件系统上实现。
- 带外 Out-of-Band

CDP與快照、傳統備份比較

技術類別	CDP	複製	快照	傳統備份
啟動機制	I/O存取事件	I/O存取事件或手動啟動	預設時間點或手動啟動	預設時間點或手動啟動
持續追蹤系統資料異動	可	可	無	無
備份窗口	無	無	數秒到數分鐘	數小時
多還原點選擇功能	可	無	可	可
還原點選擇彈性	完全無限制，可回復到過去任意時間點	無，與來源端磁碟最新狀態同步，不能回到更早狀態	依快照頻率而定，只能回復到快照啟動的時間點	只能回復到備份啟動的時間點

資料來源：ITbome整理，2008年10月

备份粒度

- 完全备份：每次备份都做完整的备份，备份周期长，存储量大
- 增量备份：先做完全备份，之后每天只做与前一天有差异的部分，对恢复的要求大，速度快，恢复较慢。
- 差异性(累计)备份：先做完全备份，之后每天只备份和第一天有差异的部分，恢复快。

连续数据保护（Continuous Data Protection，CDP）技术

- 自动持续捕捉源数据卷数据块的变化，并**连续完整地记录**这些数据块版本。
- **每一次数据块**变化都会被记录，生成瞬间快照，这与其他快照技术在快照时间点上创建快照是不同的。
  - **写操作都被记录保存**下来，因此能够动态地访问任意一个时间点的数据状态，提供了细粒度的数据恢复。
  - 可以实现**瞬间和实时的恢复**，有效拉近恢复点目标。
- 数据块级的持续数据保护技术的优点是与应用的**藕合比较松**，**性能和效率比较高**，系统连续不间断运行，**不存在快照窗口**问题。
- 它的缺点是对**存储空间的要求比较高**，这也是限制数据块级持续数据保护技术广泛应用的根本原因。

SNIA存储虚拟化

- 通过对存储(子)系统或存储服务的内部功能进行**抽象、隐藏或隔离**，使存储或数据的管理与应用、服务器、网络资源的管理分离，从而实现**应用和网络**的**独立管理**。
- 对**存储服务和设备**进行虚拟化，能够在对下一层存储资源进行扩展时进行资源合并、降低实现的**复杂度**。
- 存储虚拟化可以在系统的多个层面实现。

存储优化技术

- **存储空间利用率**
  - 自动精简技术
- **存储性能**
  - 分层存储
  - Cache
- **数据可用性**
  - 快照
  - 克隆
  - 远程复制
  - LUN拷贝

- 基于主机/服务器的虚拟化
  - 服务器的存储空间可以跨越多个**异构的磁盘阵列**，常用于在不同磁盘阵列之间做**数据镜像保护**。
  - 操作系统下的**逻辑卷管理**软件完成（安装客户端软件），不同操作系统的逻辑卷管理软件也不相同。
  - **LVM，Veritas Volume Manager**等
- 基于网络的虚拟化
  - 异构存储系统整合和统一数据管理。
  - 通过在存储域网（SAN）中添加虚拟化引擎实现。
- 存储子系统虚拟化
  - 在磁盘阵列内部实现灵活调用存储资源；**块级虚拟化，精简配置**等。
- 存储管理
  - 将来自服务器本地的闪存盘、机械盘，存储数组，JBOD等存储资源，通过存储管理协议（如SMI-S等），进行特性描述和虚拟化，构建出存储资源池。
- 数据服务
  - 存储资源池化后，数据服务即可按照用户对存储服务级别（如金银铜）的要求提供。
  - 数据服务包含：空间部署、数据保护、数据可用性、性能、数据安全性。
- 数据请求
  - 存储资源的用户，如软件开发人员通过数据管理接口（如CDMI），向SDS发起数据请求。
  - SDS开放了丰富的API供调用，SDS能够满足用户的数据请求，按照服务级别，提供相应的存储资源。

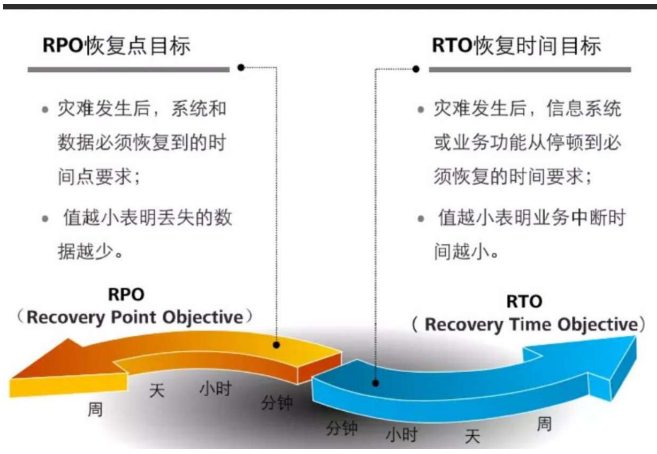
备份软件(客户端/服务器架构)

- 服务器端：备份服务器
  - 管理备份操作、
  - 维护包括备份过程、
  - 备份元数据信息的相关目录，
  - 依靠客户端收集将要备份的数据、
  - 储存节点负责向备份设备写数据。
- 客户端：
  - 客户端收集要备份的数据。

备份方式

- 热备份：应用处在运行状态，备份不影响用户获取信息。
  - 缺点：有些文件始终处于打开状态，始终在变化；热备份会引起系统性能下降。
- 冷备份：应用处在停止状态，备份时用户无法获取信息。
  - 缺点：备份时无法使用系统

备份网络拓扑结构：直接连接，基于局域网的备份，基于SAN 的备份



CDP 3 类部署架构

- 不同的复制机制，也就构成了3类不同架构的产品。

1) 主机端（Host-Based）
2) 网络端（Network-Based）
3) 储存端（Storage-Based）

连续数据保护（Continuous Data Protection，CDP）技术特性

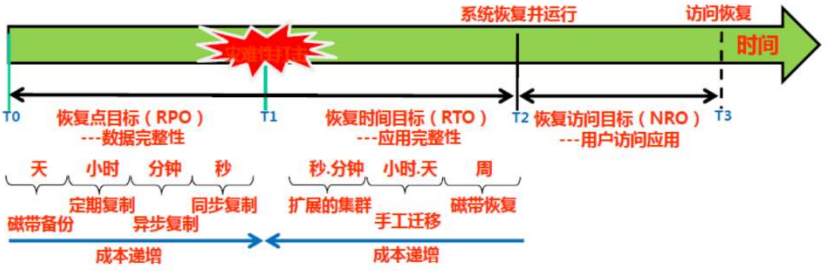
储存网络工业协会（SNIA）的定义，CDP具备3个特性：

1) 数据的更动必须连续的被记录与追踪。
2) 所有数据的变化历程都被保存在与主储存地点不同的独立地点。
3) 资料还原点（Recovery point objectives，RPO）是任意的。

容灾等级	RTO	RPO
第1级	2天以上	1天至7天
第2级	24小时以上	1天至7天
第3级	12小时以上	数小时至1天
第4级	数小时至2天	数小时至1天
第5级	数分钟至2天	0至30分钟
第6级	数分钟	0

数据大集中初期建设目标

最终目标



重复数据删除(De-duplication)技术

- 实际的利益：
  - 满足ROI(投资回报率，Return On Investment)/TCO(总持有成本，Total Cost of Ownership)需求；
  - 可以有效控制数据的急剧增长；
  - 增加有效存储空间，提高存储效率；
  - 节省存储总成本和管理成本；
  - 节省数据传输的网络带宽；
  - 节省空间、电力供应、冷却等运维成本。

Dedupe的衡量维度

- 重复数据删除率(deduplocation ratios)
- 数据自身的特征和应用模式所决定
- 性能

数据库检索：动态数组、数据库、RB/B/B+/B*树、Hashtable等

数据块指纹计算：哈希函数

数据压缩与重复数据删除对比分析

- 消除冗余范围
- 发现冗余方法
- 冗余粒度
- 性能瓶颈
- 数据安全
- 应用角度

- 重复数据删除率(影响因素

存储优化技术
•存储空间利用率：自动精简技术
•存储性能：1 分层存储 2 Cache
•数据可用性：1 快照 2 克隆 3 远程复制 4LUN 拷贝

高重复数据删除率	低重复数据删除率
数据由用户创建	数据从自然世界获取
数据低变化率	数据高变化率
引用数据、非活动数据	活动数据
低数据变化率应用	高数据变化率应用
完全数据备份	增量数据备份
数据长期保存	数据短期保存
大范围数据应用	小范围数据应用
持续数据业务处理	普通数据业务处理
小数据分块	大数据分块
变长数据分块	定长数据分块
数据内容可感知	数据内容不可知
时间数据消重	空间数据消重

重复数据关键技术

- 文件数据块切分
  - 文件级的dedupe技术也称为单一实例存储(SIS, Single Instance Store)
  - 数据块级的重复数据删除其消重粒度更小，可以达到4-24KB之间。
- 数据分块算法主要有三种，即定长切分(fixed-size partition)、CDC切分(content-defined chunking)和滑动块(sliding block)切分。