第五次作业

19030100332 徐浩东

1. (1)块存储系统：

块存储器（有时称为“块级存储器”）是一种用于在存储区域网络 (SAN) 或基于云的存储环境中存储数据文件的技术。块存储器可用于需要快速、高效和可靠地传输数据的计算场景。

块存储器将数据分解成块，然后将这些块存储为单独的部分，而每个部分都具有唯一标识。SAN 将这些数据块放在能实现最高效率的位置。这意味着可以将这些块存储在不同系统中，并且每个块都可以配置（或分区）为使用不同的操作系统。

块存储器还将数据与用户环境分离，允许将数据分布在多个环境中。这样就会创建多个数据路径，让用户能够快速检索到数据。当用户或应用程序从块存储系统请求数据时，底层存储系统将重新组装数据块并将数据提供给用户或应用程序。

(2) 对象存储系统：

对象存储器（也称为“基于对象的存储器”）将数据文件分解成多个部分（称为“对象”）。 然后，它将这些对象存储在单个存储库中，该存储库可以分布在多个联网系统中。

在实践中，应用程序可以管理所有对象，而不需要使用传统文件系统。 每个对象都会收到一个唯一的 ID，应用程序将使用该 ID 来识别对象。 每个对象都会存储元数据，即存储在对象中的文件的相关信息。

对象存储器与块存储器之间的一个重要区别是它们处理元数据的方式。在对象存储器中，可以定制元数据以包含有关存储在对象中的数据文件的其他详细信息。例如，可以定制视频文件附带的元数据，以说明该视频的制作地点、用于拍摄的摄像机类型，甚至每一帧中拍摄的主体。在块存储器中，元数据仅限于基本文件属性。

对应的，块存储器最适合存储不经常更改的静态文件，因为对文件所做的任何更改都会导致创建新对象。

(3) 文件存储系统：

文件存储器（也称为“文件级存储器”或“基于文件的存储器”）通常与网络连接存储 (NAS) 技术有关。NAS 使用与传统网络文件系统相同的机制向用户和应用程序提供存储器。用户或应用程序通过目录树、文件夹和单个文件接收数据。这与本地硬盘驱动器的功能类似。但是，NAS 或网络操作系统 (NOS) 可以处理访问权限、文件共享、文件锁定和其他控件。

文件存储器的配置过程非常简单，但数据访问受到单一数据路径的限制，与块存储器或对象存储器相比，这会影响性能。文件存储器同样只使用常见的文件级协议，例如用于 Windows 的新技术文件系统 (NTFS) 或用于 Linux 的网络文件系统 (NFS)。 这可能会限制在不同系统中的可用性。

	块存储	文件存储	对象存储
速度不同	低延迟 10ms,热点突出	不同技术不同	100ms-1s,冷数据
可分布性不同	异地不现实	内可分布式，但有瓶颈	分布并发能力强
文件大小不同	大小都可以，热点突出	适合大文件	适合各种大小
接口不同	Driver,kernel module	POSIX	Restful API
典型技术不同	SAN	HDFS,GFS	Swift,Amazon
适合场景不同	银行等	数据中心	网络媒体文件存储

2.1 客户端读取 HDFS 系统中指定文件指定偏移量处的数据时，工作流程是什么？

当一个应用程序读取一个文件时，HDFS client 首先向 NameNode 询问请求托管文件块副本的数据节点列表。该列表按与客户端之间的网络拓扑距离排序。NameNode 服务器会将文件包含地 DataNode 服务器的 IP 地址进行返回。
客户端根据指定的文件偏移量选取相应数据块编号的 DataNode 服务器，并以数据流的方式访问 DataNode 服务器获取数据.

2.2 客户端向 HDFS 系统中指定文件追加写入数据的工作流程是什么？

应用程序通过创建新文件并将数据写入 HDFS 来向 HDFS 添加数据。关闭文件后，无法更改或删除写入的字节，除非可以通过重新打开要追加的文件将新数据添加到文件中。

打开文件进行写入的 HDFS 客户端被授予该文件的租约;没有其他客户端可以写入该文件。写入客户端通过向 NameNode 发送检测信号来定期续订租约。当文件关闭时，租约将被吊销。租约期限受软限制和硬限制的约束。在软限制到期之前，编写器确定对文件的独占访问权限。如果软限制过期，并且客户端无法关闭文件或续订租约，则另一个客户端可以抢占租约。如果在硬限制到期（一小时）后，客户端未能续订租约，HDFS 将假定客户端已退出，并代表编写器自动关闭文件，并恢复租约。
当数据写入 HDFS 文件之后，HDFS 可以显示地调用æ ush 操作确保应用程序可见性地保证，然后将数据包推送到管道上，æ ush 操作将等待管道中的所有数据节点确认数据包的成功传输。

2.3 新增一个数据块时，HDFS 如何选择存储该数据块的物理节点？

创建新块时，HDFS 会将第一个副本放在 writer 所在的节点上。
第二个和第三个副本放置在不同机架中的两个不同节点上。其余的放置在随机节

点上，但有限制，即在任何一个节点上放置的副本不超过一个，并且如果可能，在同一机架中放置的副本不超过两个。选择将第二个和第三个副本放在不同的机架上，可以更好地在整个群集中分发单个文件的块副本。如果前两个副本放在同一个机架上，对于任何文件，其三分之二的块副本将位于同一机架上。

2.4HDFS 采用了哪些举措应对数据块损坏或丢失的问题？

checkpoint 节点定期组合现有的检查点和日志，以创建新的检查点和空日志，定期创建 checkpoint 是保护文件系统元数据的一种方法。

在 HDFS 中创建快照，便于 HDFS 系统中数据出现错误或丢失后的回滚；

复制管理，将节点分布在多个机架上，数据冗余保证数据不会完全丢失或损坏。

HDFS 为 HDFS 文件的每个数据块生成和存储校验和，以帮助检测由客户端、数据节点或网络引起的任何损坏。

每个数据节点运行一个块扫描程序，定期扫描其块副本，并验证存储的校验和是否与块数据匹配。

2.5HDFS 采用了什么举措应对主节点失效问题？

心跳机制，DataNodes 向 NameNode 发送心跳信号，确保节点间的数据正常交互。
数据冗余，主节点的数据有多个副本，以防主节点失效后，数据不会丢失。
安全模式，NameNode 在启动的时候会先经过一个安全模式节点，增加主节点或数据的稳定性。
checkpoint 创建新的检查点与日志，保护系统元数据.

2.6NameNode 维护的“数据块-物理节点对应表”需不需要在硬盘中备份，为什么？

不需要
HDFS 为了提高耐用性，通常将存储在 NameNode 的本地本机文件系统中的图像的持久记录的检查点和日志的冗余副本存储在多个独立的本地卷和远程 NFS 服务器上。
复制管理：NameNode 确保每一个块之中都具有与预期数量的副本，会自动复制和删除机架上的块来确保数量，块副本的位置不属于持久性检查点.