银行虚拟化部署设计方案

一. 引言

随着银行开发项目的数量和规模日益增长，对项目质量控制的要求也越来越高，目前采用的目前采用CDC传统的数据中心解决方案已经满足不了目前的应用需求，运维带来很多不便，存在着成本、管理、故障（容灾）等问题，服务器利用率低，缺乏灵活性和管理型，达不到高可用性要求。

针对大作业要求，本文细致分析了存储系统的用户需求；提出了存储技术与网络拓扑整体设计方案；具体分析了存储方案与成本，以及网络拓扑结构与带宽设计；再分析了容灾备份方案与单点故障问题的应对措施，以及数据安全保护机制；并分析了云计算存储的实现。

二. 问题描述

2.1 背景

某银行现阶段规划X个分区，分别为生产业务区、综合管理区、网银在线区、产品测试区、运维基础区。每个业务区承载着不同数量的虚拟机，并且随着业务发展不断增加。随着新业务不断增长要求，未来将逐步把小型机服务器上应用系统迁移到虚拟化平台，最终实现数据中心x86服务器的全部虚拟化。

根据我们所学习存储信息管理知识，设计一套虚拟化部署方案，考虑到业务的迁移的复杂性，默认不在使用原有设备(降低设计难度)，因为是银行业务需要100%的冗余方案，尽量考虑容灾，存储部分设计按照基础数据为100TB设计，年增长40%,设计一个满足3年的方案。

2.2 设计要求

考虑以下指标点:

- 主机配置:（计算和存储）
- 用户、组、权限和角色
- 共享存储容量规划
- 虚拟机命名规划
- 地址池管理规范化

三. 需求分析

对银行X86物理设备进行虚拟化改造，搭建虚拟化平台。建设桌面虚拟化系统，以保障数据安全同时快速提高开发测试环境。

实施虚拟化后，所有数据集中存储在存储设备上，为保障数据安全，防止误操作引起的数据丢失问题，建设数据备份系统。

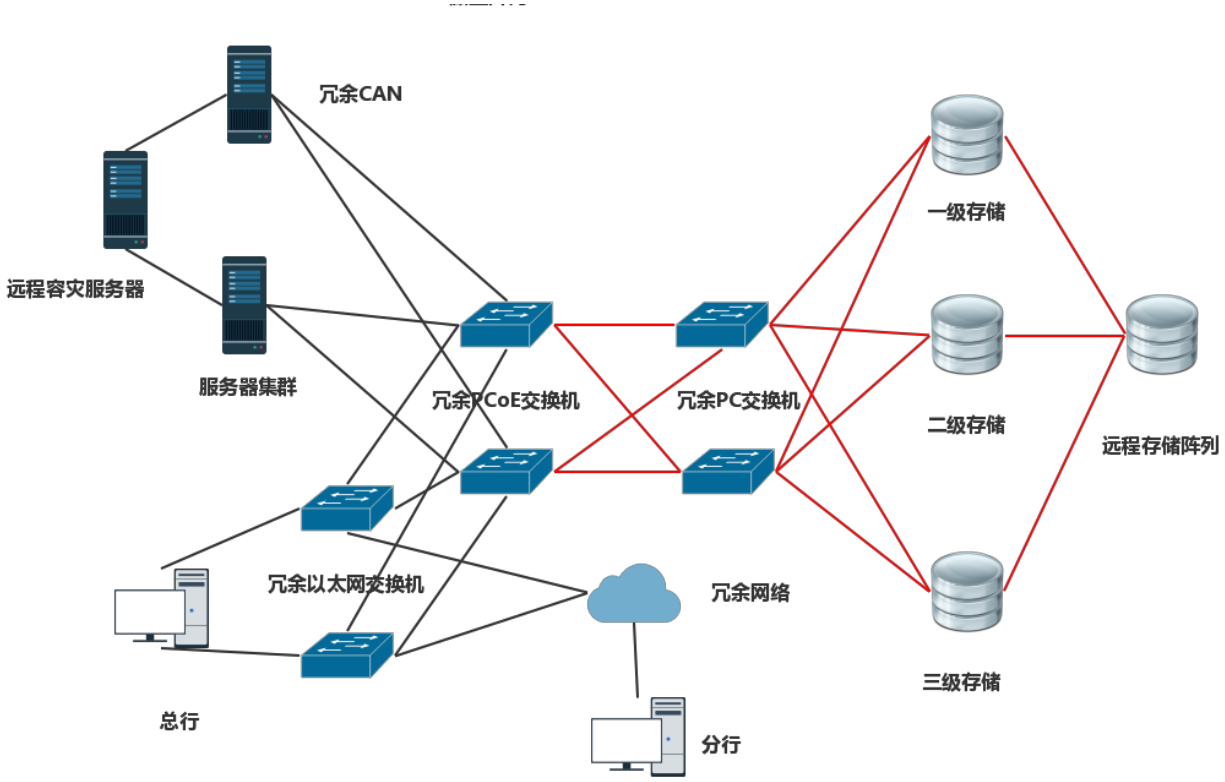
项目实施应充分考虑监管要求和行内要求。需满足银行虚拟化平台建设项目的安全性、操作性、兼容性、稳定性、冗余性及未来系统建设扩展性的相关要求。

四. 设计方案

4.1 整体概述

采用分布式存储方案，因为银行每天的数据量相当大，需要的磁盘等存储设备数量也很多。如果使用集中式存储建设周期太长，架构复杂，故障排查难度大，而且扩展性不够好，所以采用分级存储的三级架构。

本系统以FCOE网络为核心，通过FC SAN链接存储，通过以太网设备与总行、分行终端相连。同时在核心节点设置了冗余级备份设备或链路，但又充分考虑到了综合成本问题，使得在保证安全性、可用性的前提下成本尽量低。



4.2 存储系统

本系统采用分级存储方案，共分为三级。

根据以往银行业务办理经验，我们认为业务办理的第一个月数据是最活跃的，将会产生大量数据，且这些数据会被频繁调用，所以第一级存储将存放首月的数据，采用读写性能都相对出色，且具备镜像保护的RAID10阵列。考虑到此级存储中的数据活跃性、重要性等方面的因素，在灾备方面将采用零PRO同步远程镜像的方

式。由于受到此级存储总数据量的约束，即使在使用了诸多高成本的设备和方案的情况下，也不会产生太高的费用。

一级在线数据存储空间需求：

$$100TB * 0.4 * \frac{1}{12} = 3.4TB$$

配置冗余100%的存储空间后的总存储空间需求为：

$$3.4TB * 2 = 6.8TB$$

配置500G的SSD磁盘，单盘成本为2000元，RAID10磁盘利用率为50%，需要28块磁盘，56000元。

首月后，数据利用率将显著下降，但仍有一定频率，对磁盘阵列的读写性能级数据保护的要求略低于前者，耳机存储将存放第二个月到第六个月的数据，使用RAID5阵列。由于此级存储中的数据也是经常被调用的，仅仅是频率相对第一级较低，所以使用分钟级PRO的异步远程镜像方式，这样既满足了基本要求，又节约了部分成本。

二级在线数据存储空间需求：

$$100TB * 0.4 * \frac{6}{12} - 3.4TB = 16.7TB$$

配置冗余100%的存储空间后的总存储空间需求为：

$$16.7TB * 2 = 33.4TB$$

配置1T SAS 15000转磁盘,单盘成本为2000元。RAID5磁盘阵列，磁盘利用率为$((n-1))/n$，
$n \times 1TB \times ((n-1))/n \geq 33.4TB, n \in N$解得，$n \geq 35$，所以需要35块磁盘，70000元。

而六个月后的数据调用频率就很低了，但此级存储的数据量非常大，所以磁盘和阵列的性能并不是要考虑的首要因素，我们仍然使用RAID5阵列，并且每天进行异步远程备份。

三级在线数据存储空间需求：

$$100TB * (1+0.4)^3 - 16.7TB - 3.4TB = 254.4TB$$

配置冗余100%的存储空间后的总存储空间需求为：

$$254.4TB * 2 = 508.8TB$$

配置1T SAS 15000转磁盘,单盘成本为2000元。RAID5磁盘阵列，磁盘利用率为$((n-1))/n$，
$n \times 1TB \times ((n-1))/n \geq 508.8TB, n \in N$解得，$n \geq 510$，所以需要510块磁盘，1020000元。

综上所述，总成本为 `56000+70000+1020000=1146000`

五. 容灾方案设计

5.1 容灾备份概述

灾难恢复可以分为六个等级，由低至高依次为基本支持、备用场地支持、电子传输和部分设备支持、电子传输及完整设备支持、数据零丢失和远程集群支持。每个等级都有数据备份系统、备用数据处理系统等七个要素，分别有对应的容灾方案，一般来说数据越重要，容灾等级就应当越高。

在考虑容灾方案时，应当综合多方因素，而不是简单地以最高等级为目标，应当寻找最合适的总体投入（TCO）和投资汇报（ROI），用户应根据数据的重要性确定相应的容灾等级。根据对系统的保护程度来分，容灾系统可以分为数据容灾和应用容灾。

数据容灾，就是至少在异地保存一份可用的关键业务数据，该数据可以是与本地生产数据的完全实时复制，也可以比本地数据稍微落后，但一定是可用的。

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。这样一套系统相对比较复杂，除了数据复制还包括网络、主机、应用、甚至IP等资源，以及各资源之间良好协调。

对于银行的数据，需要进行容灾方案设计以应对突发灾难。其中，由于在存储方案设计过程中已经充分考虑了服务器的存储备份机制，本地的容灾问题通常可以被及时解决，不会导致严重的后果。因此，在容灾方案的设计中，我们需要重点考虑遇到重大灾害，导致本地的系统和数据全部无法及时恢复时，启用在异地设置容灾中心的远程容灾方案。

在远程的容灾系统中，要实现完整的应用容灾，既要包含本地系统的安全机制、远程的数据复制机制，还应具有广域网范围的远程故障切换能力和故障诊断能力。也就是说，一旦故障发生，系统要有强大的故障诊断和切换策略制定机制，确保快速的反应和迅速的业务接管。

5.2 容灾方案

容灾功能基本要求是在异地建立独立的容灾备份中心，保证在发生灾难时对数据保护。对于银行而言，RPO的需求不需要过高，但仍需保证一定的容灾能力。因此可以采用异步复制模式将本地数据保存到远程的数据中心，即一个写操作提交给源存储阵列后就立即通知主机，数据保持在源存储阵列并稍后传送给远程站点。

总体工作流程如下：正常运行时，应用程序在修改、保存等操作时将数据写入应用数据中心磁盘，同时也将数据发送到灾备系统并存储，若灾备系统检测到数据中心的数据变化，也会将数据中心的数据传输到灾备系统中，并存储在远程的备份磁盘内。

当源设备损坏时，将由远程主机替代原设备进行工作负载。需要注意的是，异步复制模式不能保证与原数据一致，且性能会有所下降，因此不建议长时间使用远程的系统。另外，若存在价值特别高的数据，需要保证能够备份成功的，可以给数据设立优先级机制，优先级高的数据可以采用同步复制，而优先级低的数据采用异步复制。

相比于同步复制，异步复制可能会出现由于网络中断导致数据丢失的问题。为应对这种情况，一方面需要指定灾害恢复计划，在网络重新连接后对远程备份的数据进行校验和补充；另一方面，需要在数据传输过程中

以某种方法保证其时序一致性，

为保持生产库和灾备库数据一致，用户可以自主设置备份间隔时间以及定时备份、灾备级别, 并指定哪些一作业复制到灾备数据数据库灾备系统可以实现为期三个月、半年、一年和用户自定义时间范围的图像文件自动转储, 转储过后删除灾备系统中相应的数据, 从而节约磁盘空间。由于灾备系统绝大多数时间处于闲置状态，从经济性考虑，灾备系统可以在源系统的基础上适当降低性能，如不使用SSD磁盘而统一用SAS磁盘存储，这样虽然会造成一些性能损失，但能够在很大程度上避免资源的浪费。

灾备系统还具有数据校验的功能, 可以验证生产数据库和灾备数据库中数据的一致性,从而确定数据是否被正确备份。用户可以实时监控数据存储情况、数据增长情况、数据文件使用率、数据移动、数据有效性等。

同时, 灾备系统完全实现了模块化, 通过与底层平台的无缝组合, 在模块有新功能或新技术更新后只需升级此模块即可, 同时随时可装配更多模块。灾备系统的操作界面简单, 提供实时报警与预警机制, 当灾备系统出现一些日常信息、警告、错误时, 会通过提示、邮件告知使用者。

六. 用户、组、权限和角色

用户	用户组	角色	权限
业务员	用户	用户	读写本人负责业务
部门经理	部门管理员	管理员	读写部门所有业务
总经理	总管理员	管理员	读所有部门业务
系统开发及运维人员	Admin	系统管理员	读写所有数据

七. 虚拟机命名规划

虚拟机由运维部门统一进行管理，按需创建，每台虚拟机与特定业务相绑定，做到专机专用。关键业务虚拟机、一般业务虚拟机和测试用虚拟机应分开，停用的虚拟机在规定时限内应及时删除。

创建后的虚拟机及时登记并纳入服务器统一管理，每开通一个虚拟机要及时把虚拟机配置，用户名，密码，主机地址纳入管理库。

虚拟机命名：使用的格式，如“1.42*SCYW*SYJS_WIN2003”即为生产业务区上用于收益计算的WIN2003虚拟机

八.地址池管理规范化

使用全局流量管理提供地址池的方式来简化应用服务的地址管理。一个全局流量管理实例，可以配置多个地址池，便于实现不同地区的用户访问不同的地址池，实现就近接入。同时当地址池整体不可用时，可以做备份切换。

一个地址池内可能会存在多个IP地址，在通过健康检查对IP地址监控时，会实时统计地址池内健康IP地址的数量，并自动隔离故障IP。假如地址池内，健康IP地址的数量小于您设置的最小可用地址数量，系统会将自动将地址池设置为不可用。

在全局配置中均衡策略选择加权轮询时，地址池内添加IP地址会提供权重配置，可以对每个IP地址设置不同的权重，实现访问流量按照权重分配到每个IP地址。

在地址池内添加IP地址，可以对IP地址设置不同的工作模式。默认对IP地址启用智能返回的工作模式，工作方式介绍如下：

- 智能返回：根据健康检查状态，对IP地址进行动态选择，即IP地址健康检查正常时，DNS解析向用户返回IP地址，IP地址异常时，DNS解析把异常的IP地址暂时删除；
- 永远在线：系统认为该IP地址永远处于正常状态，DNS解析始终向用户返回该IP地址，健康检查不对永远在线的IP地址进行监控；
- 永远离线：系统认为该IP地址永远处于异常状态，DNS解析不会向用户返回该IP地址，该IP地址只存在系统配置中，等待以后某个时间启用，健康检查不对永远离线的IP地址进行监控；