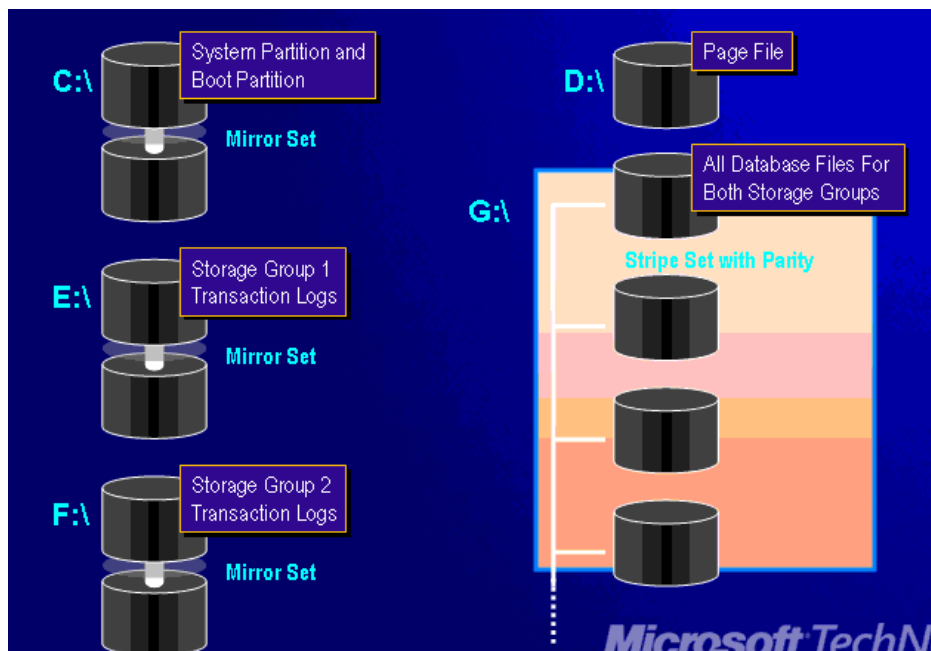


第七节：管理exchange数据库下

硬件规划



数据文件的读写是频繁的随机读写I/O

日志文件时顺序的连续的I/O的操作

有条件的情况下，推荐RAID10，无条件RAID5也可以。

RAID10保证了数据最快，而事务日志文件要求快速写入，所以建议RAID10，日志文件只会读一次，写一次，不会频繁读写的。

而RAID5可以保证完整写入，有数据校验，数据库文件有大量的读写操作。

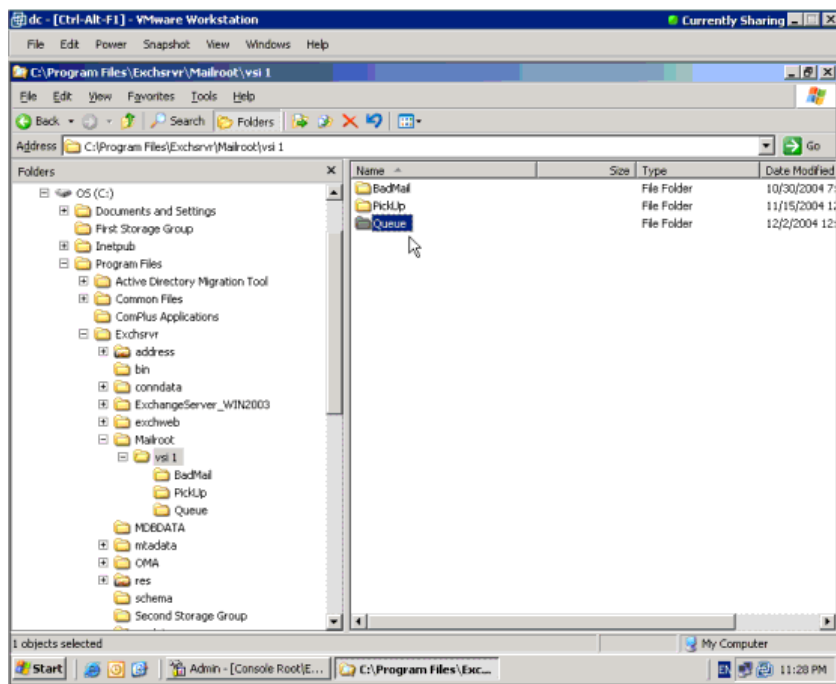
现在的磁盘阵列卡性能已经很高了。

exchange其他的优化选项：

- ◆ 有条件的情况下，尽可能选择高速的**SCSI硬盘(15K RPM)**和带有大容量**cache**和后备电池支持的磁盘阵列卡
- ◆ **SMTP Queue**文件夹移动
- ◆ 页面文件的位置
- ◆ 尽可能的提高**C盘**的速度
- ◆ **IIS Log**文件和**Messaging Tracking Log**文件移动
- ◆ 对存放**Log**的阵列，应该使用**diskpar**工具进行初始化，不要使用**Windows**自带的磁盘管理工具创建分区
 - **Diskpar**可以强制性的指定分区在磁盘磁道上的开始扇区
 - **Exchange**的数据页面是以**4KB**为单位的，使用**diskpar**可以避免在写入一个**4KB**页面的时候跨越两个扇区(造成性能问题)
 - 参考：
http://www.microsoft.com/technet/prodtechnol/exchange/guides/E2k3Perf_ScalGuide/0e24eb22-fbd5-4536-9cb4-2bd8e98806e7.msp

exchange server 2003队列目录

队列目录的读写快慢，决定了邮件收发的快慢，磁盘的I/O情况。当然邮件收发还取决于网络性能，DNS解析等。



页面文件的位置一般是在C盘，所以需要尽可能提高C盘的速度。

可以把页面文件移走，放在独立的磁盘。

4K扇区对齐，可以提高20%的性能。

在一个LUN上划分多个磁盘，对性能提高没有太大帮助，可能对减少磁盘碎片有帮助。

在线碎片整理。

◆ 了解在线碎片整理

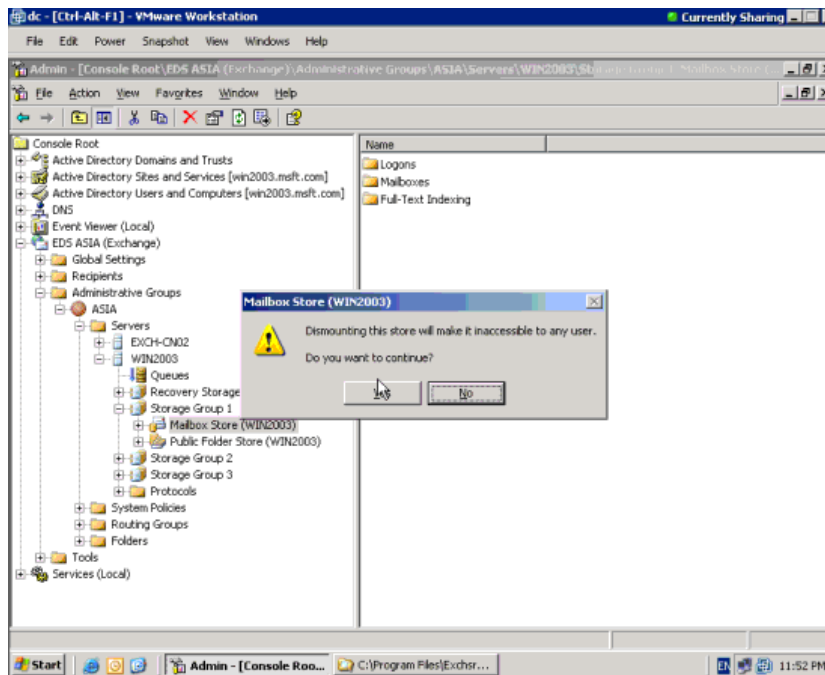
- 通过查询活动目录来确定Store中是否有被删除的邮箱。
- 物理的删除所有超过保留时间的邮件和邮箱。
- 执行在线碎片整理。

Event: 1221^v
 Source: MSeXchangeIS Private^v
 Type: Information^v
 Category: General^v
 Description: The database has nnn megabytes of free space after online defragmentation has terminated. ^v

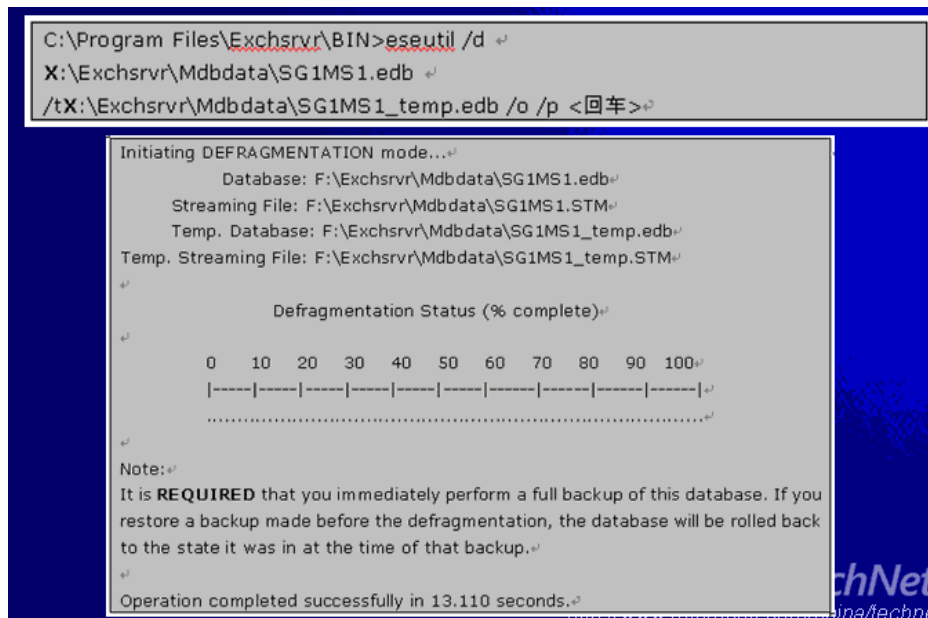
◆ 什么是离线碎片整理，何时应该进行碎片整理

当数据库撑得很大，而且磁盘没有剩余空间的时候，可以做数据库的离线整理。

数据库离线碎片整理需要downtime时间。



每小时可以做5-7GB左右。

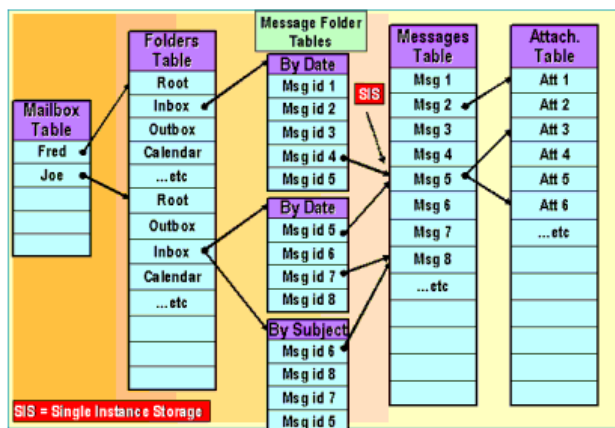


离线整理后，会提示我们对整理后的数据库做完全备份。

如果在线碎片整理和备份冲突，那么在线碎片整理会停止。

single instance单一实例存储。 在同一个数据库里面，如果给多个收件人发了一封邮件，那么邮件只会在数据库里面存放一个副本，每个用户都会有一个到这个副本的引用关系，这就是单一实例存储的概念。single instance 是为了降低数据库的容量的。

Information Store Structure



deleted retention是增大数据库的因素。用户虽然删除邮件，减少了用户的邮箱大小，但是已删除的邮件还是在数据库上，所以数据库的大小并不会变小。

第八课：管理exchange消息路由

邮件传输的四个阶段：提交、分类、路由、投递。

邮件提交的方式：

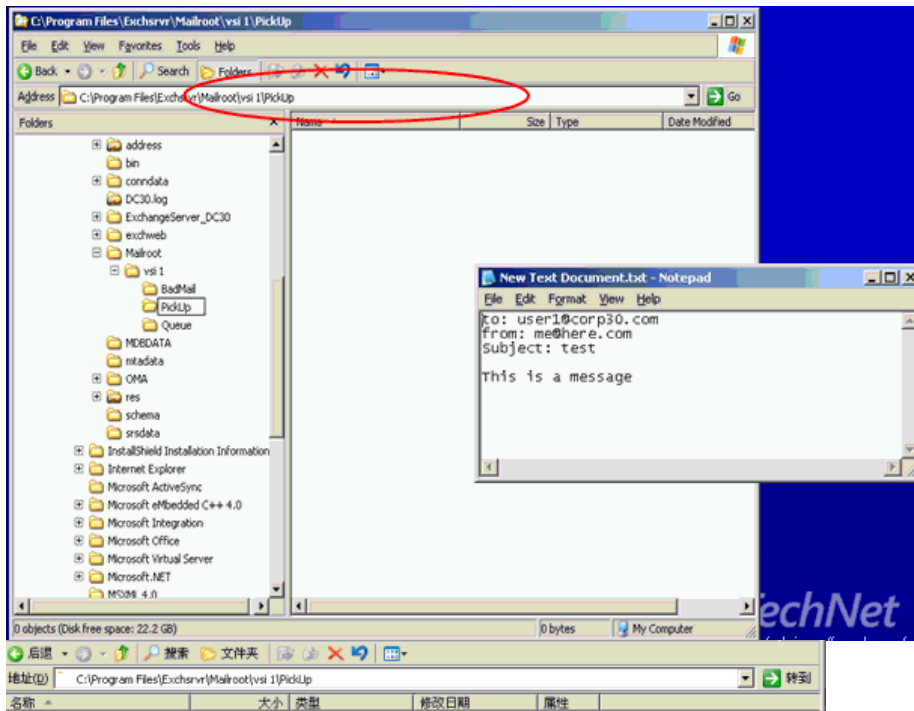
- ◆ 使用**MAPI** 客户端 (Outlook)
- ◆ 使用**SMTP** 客户端 (或者 MTA/X.400)
- ◆ 使用**pickup** 目录
- ◆ 使用邮件网关
- ◆ 邮件是否传输到服务器？
 - 使用 **Telnet**
 - 使用 **pickup directory**
 - 使用排错日志
 - 使用协议日志

pickup目录：可以直接向服务器提交邮件，如果把邮件直接拷贝到pickup，那么pickup目录会直接提交邮件。

一些第三方的OA系统，是直接往exchange的pickup目录投递邮件的。

下面是exchange的pickup目录，我们可以在pickup目录里面随便建一个txt文档，会自动送出去。

如果是IIS的pickup目录，位置不一样。



以上列举的只是2003时代常见的一些方式。

用户的邮件可能根本没有提交到服务器：比如exchange的服务器磁盘比较繁忙，就会做暂缓提交的动作。对于MAPI客户端和OWA客户端，如果邮件因为繁忙无法提交，exchange会把邮件放在草稿箱。还有可能因为一些情况导致exchange服务器拒绝连接。

如果通过CMD测试邮件收发？

先telnet上来。

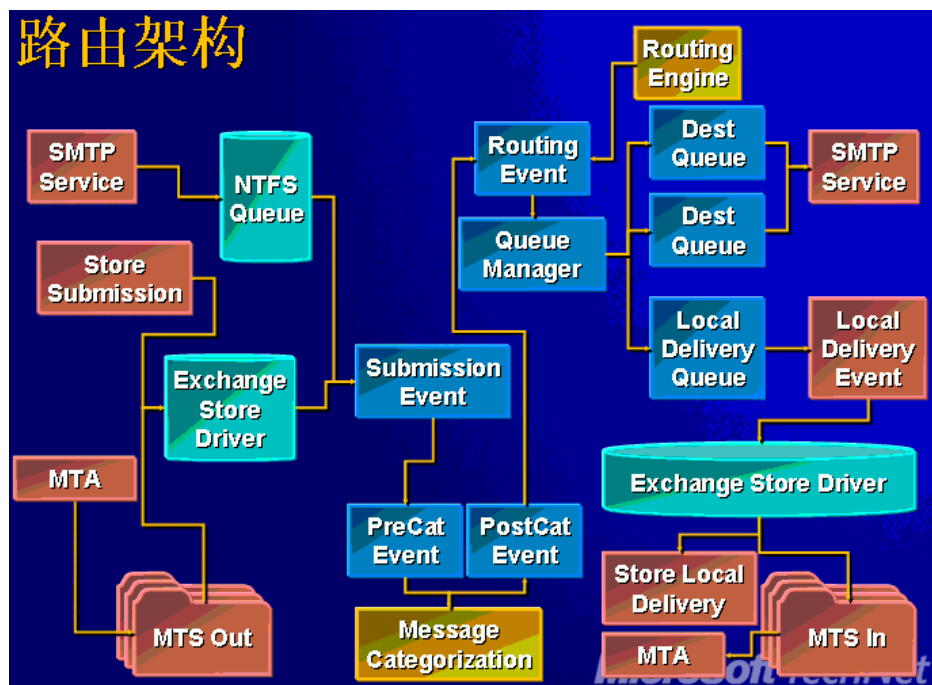
```

C:\WINDOWS\system32\cmd.exe
220 London.resource.msft Microsoft ESMTMP MAIL Service, Version: 6.0.3790
ready at Wed, 29 Jun 2005 14:23:08 +0800
helo
250 London.resource.msft Hello [10.100.10.62]
mail from
501 5.5.4 Unrecognized parameter
mail fo
501 5.5.4 Unrecognized parameter fo
mail from : <>
250 2.1.0 <>...Sender OK
rept to : <administrator@contoso.msft>
550 5.7.1 Unable to relay for administrator@contoso.msft
rept to : <adminis
501 5.5.4 Invalid Address
rept to : administrator@Resource.msft
250 2.1.5 administrator@Resource.msft
data
354 Start mail input: end with <CRLF>.<CRLF>
from:testuser<tu@micosot.com>
to:admin<administrator@esou.msft>
subject:ts mail
cads

```

可以在exchange的系统管理器里面打开exchange的诊断日志，如果有相关的错误，则会记录到相关事件日志当中。exchange会对IIS里面很多组件做扩充，比如IIS的metadata等。如果重装IIS，那exchange的一些东西就不存在了，比如分类器，消息队列。

路由架构

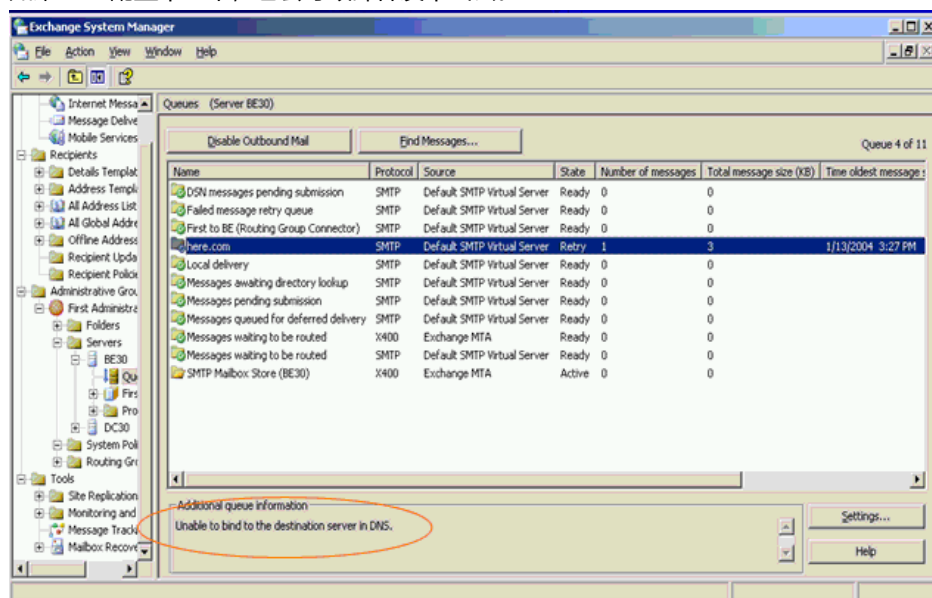


有了event 之后，就可以改变SMTP处理邮件的方式。在分类器阶段执行各种event。

四个过程是由exchange的高级队列引擎控制的。

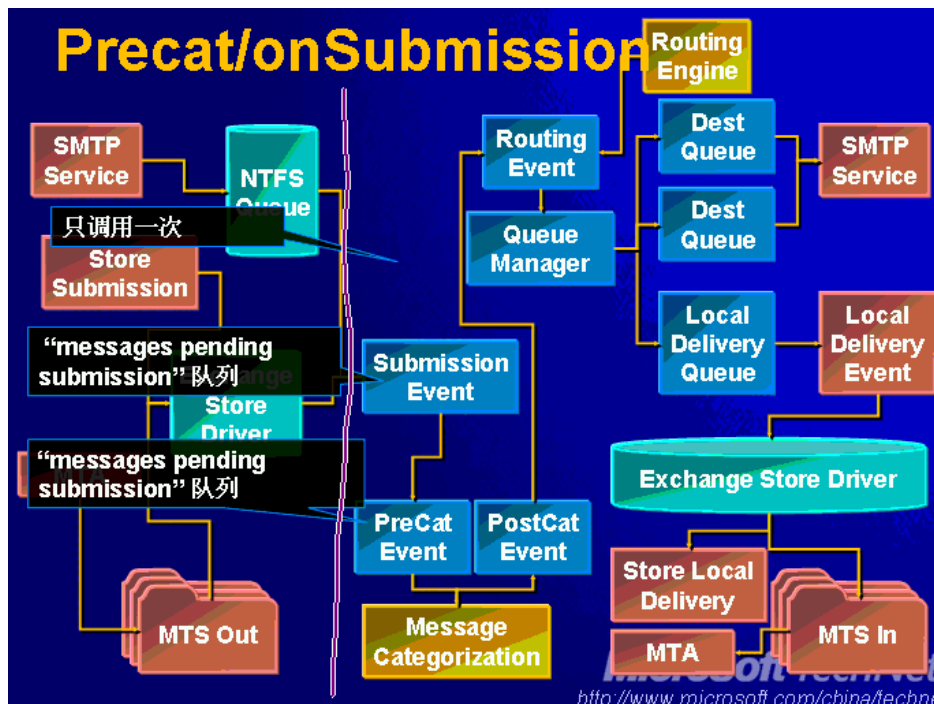
下图是exchangeserver的队列查看器。

如果DNS配置不正常，也会导致邮件发不出去。



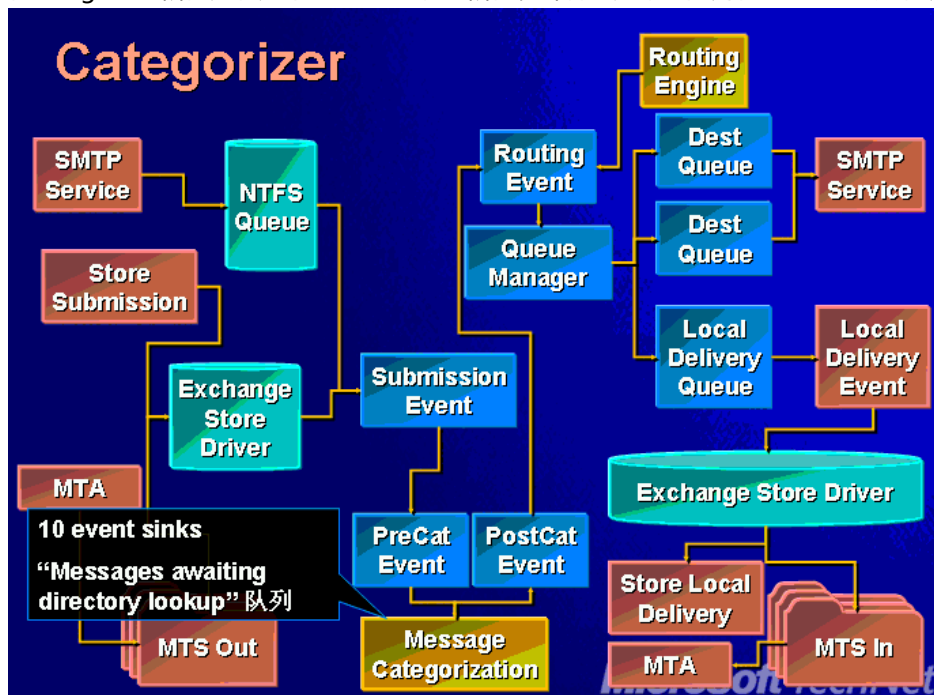
第一个submission阶段详情。

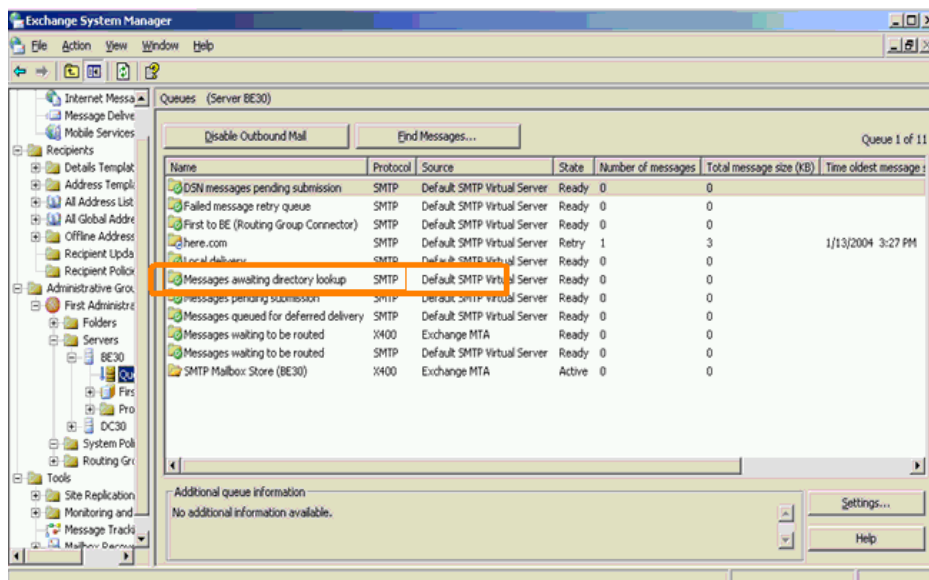
提交到message pending submission队列



exchange的SMTP和IIS的SMTP差别就在于categorization阶段。

在categorizer阶段是有是个event的。在此阶段，邮件由原来的队列转到目录查找的队列。





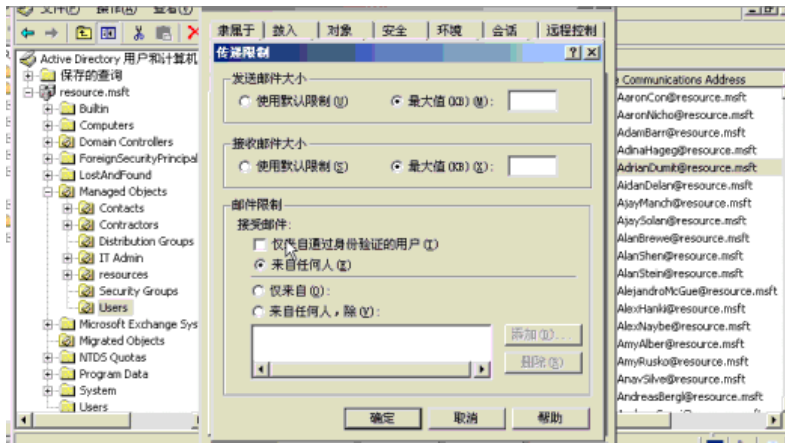
邮件列表仅仅是一个列表，没有邮箱。

比如邮件发往一个mail group，这封邮件会转到组的所有成员里面，这就是邮件列表扩展的实例。

Categorizer

1. 解析发送人和收件人地址 (DSAccess)
 - 基于活动目录对象的属性对邮件进行操作 (applies default proxy addresses, stamps SMTP address on message)
2. 添加收件人
 - 邮件列表扩展, 邮件转发, 内容转换
3. 实施收件人限制
4. 拆分邮件
5. 检查是否产生循环
 - 防止分发列表循环
 - 防止转发循环
6. 执行邮件归档 (journaling)
7. 转发所有未解析地址的邮件
8. 在邮件头标记路由地址

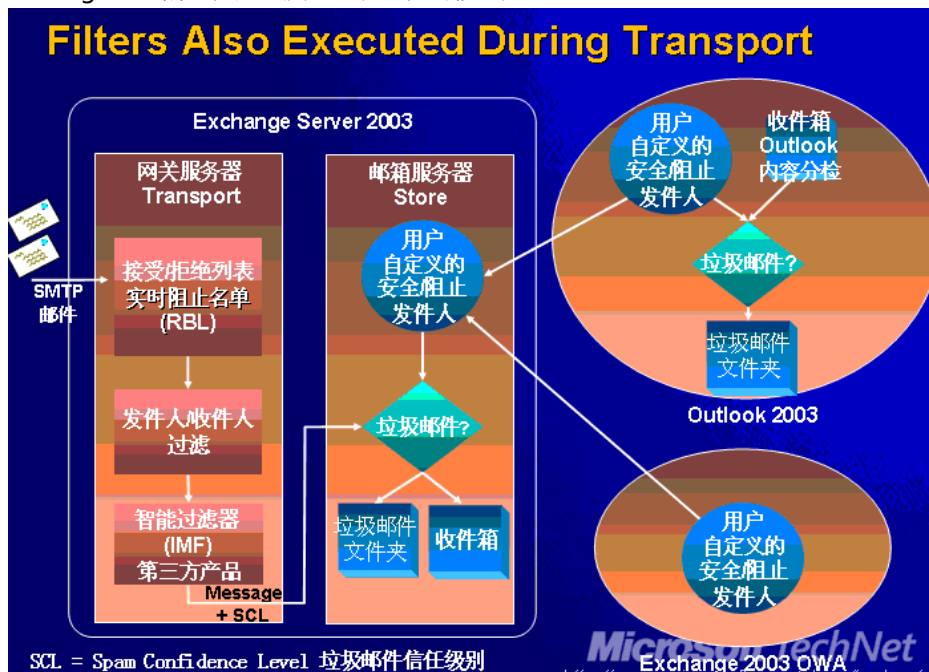
在categorizer阶段实施收件人限制。主要处理收件人限制、传递限制。



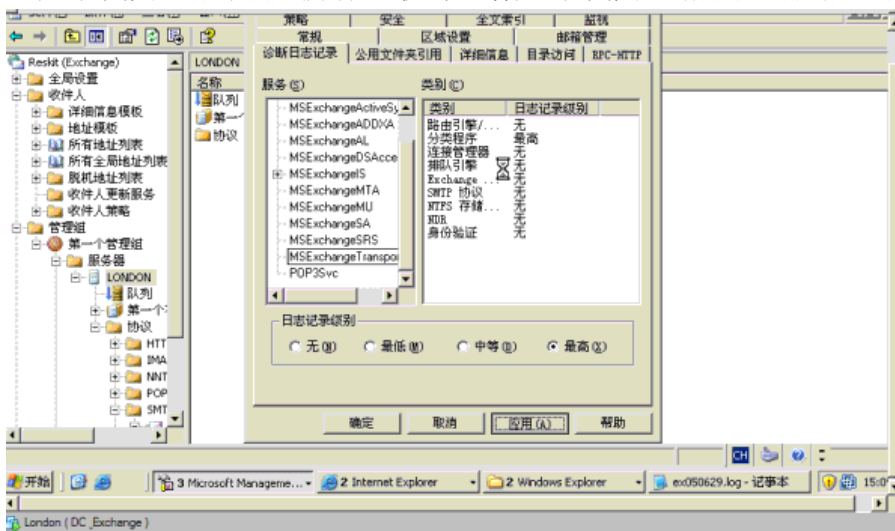
设置邮件转发的时候，要特别注意，不要让两个用户互相转发，会产生转发循环。

要防止分发列表循环，防止转发循环。

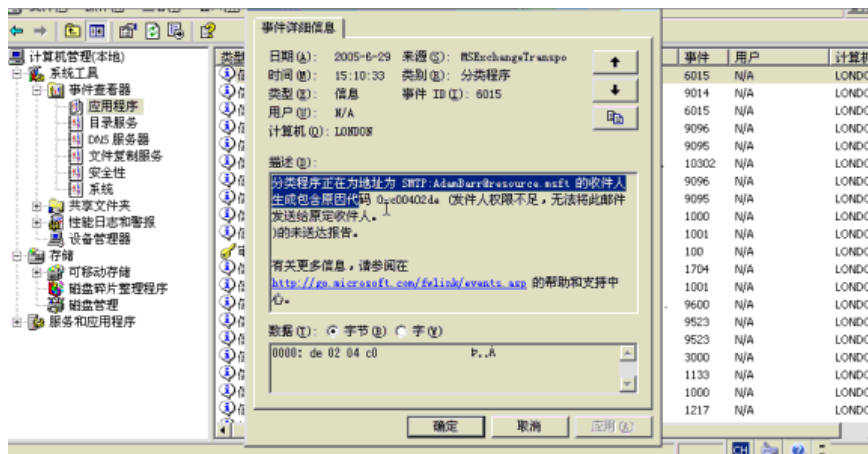
在categorizer阶段，处理反垃圾设置，传输服务器。



可以把分类程序的协议日志级别调到最高，这样关于分类程序的所有日志都会记录下来。

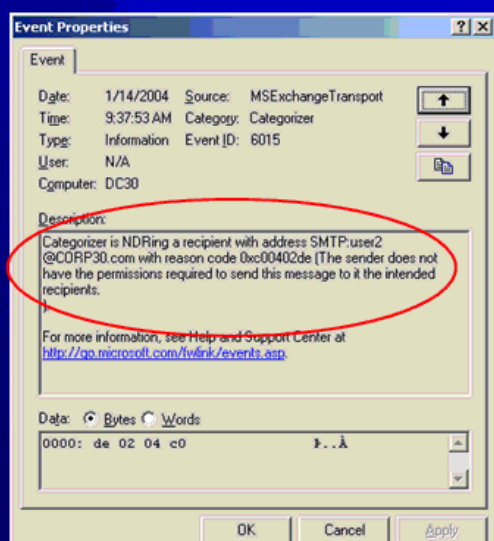


例子：如果一个用户设置了只接受经过身份验证的人发的邮件，那么匿名发往该用户的邮件是会被接收的，关于这方面的信息就会被记录到事件查看器里面。事件类型为分类程序。



退信报告，也是由分类程序处理。

Generation of non-delivery report (NDR)

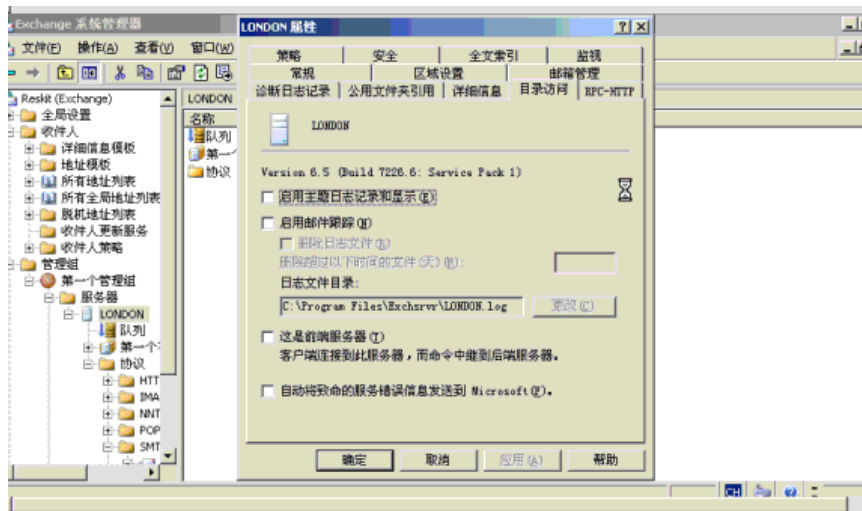


分类程序的排错。

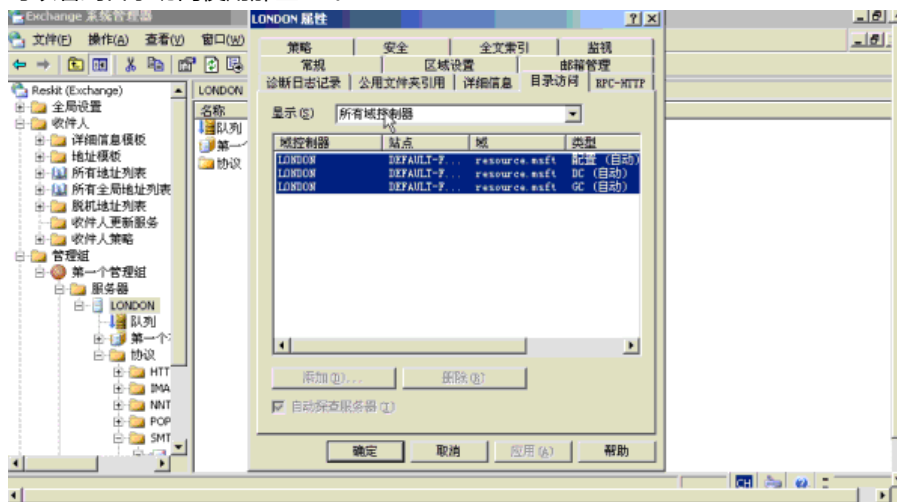
◆ Categorizer

- Directory Service Access (DSAccess) 决定使用哪一个GC
- KB article 284204 – 投递状态通知 delivery status notifications (DSNs)
- 将Categorizer日志级别设置为排错 (注册表)
 - 设置为最大不会产生帮助
- 邮件跟踪
 - Shows messages stuck in categorizer
- 使用性能监视器 (KB article 231734)

下图的目录访问就是DSAccess。



可以看到目录访问使用那些DC。



如果邮件发不出去，要看DSAccess是否正常。

其他排错工具。

◆ Categorizer, other tools

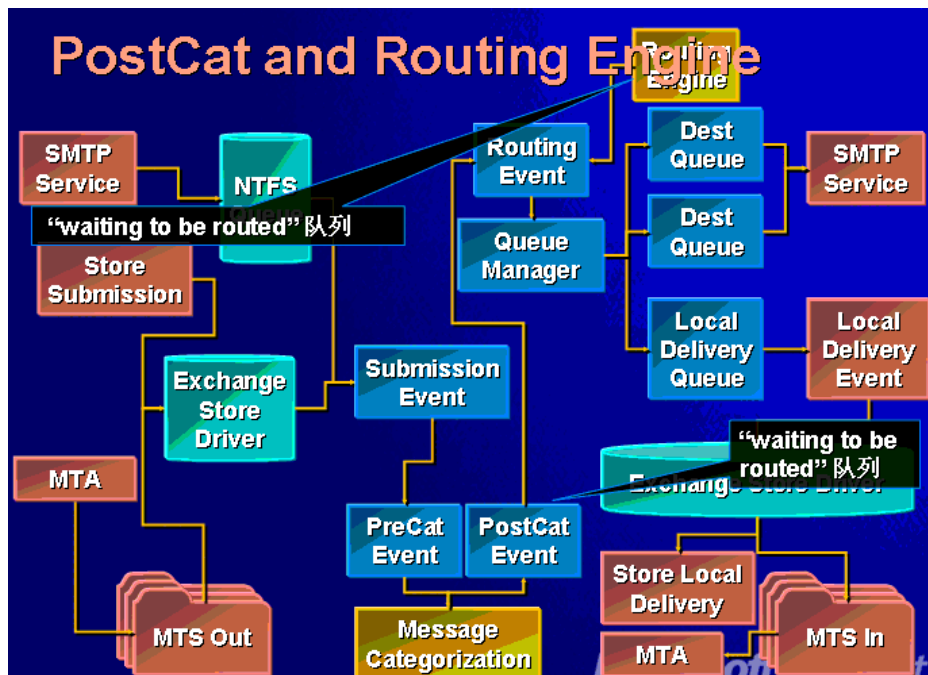
- ADSIEDIT, LDIFDE, LDP, LDSU reports
 - Determine if AD attributes are set properly
- DUMPRP.vbs (Recipient policy dump)
 - RecipPolicies pushed from Active Directory > Metabase but not the other way. Be sure they are correct.
- Metabase edit tool
- Netmon trace (check that global catalog queries are properly formed and responses are as expected)
- Regtrace (KB article 238614)
- Archive sink
 - Archive messages on both OnSubmission and PostCat and compare how messages are changed (KB article 307798)

Microsoft TechNet
<http://www.microsoft.com/china/technet>

archive sink是微软归档的小工具。

regtrace工具，是作为最后的排错工具，不到万不得已，一般不用。

第三个阶段：路由阶段



下图的message waiting to be routed代表邮件还没有到路由引擎，但是已经过了分类程序了。

Name	Protocol	Source	State	Number of messages	Total message size (KB)	Time oldest message
DSN messages pending submission	SMTP	Default SMTP Virtual Server	Ready	0	0	
Failed message retry queue	SMTP	Default SMTP Virtual Server	Ready	0	0	
First to BE (Routing Group Connector)	SMTP	Default SMTP Virtual Server	Ready	0	0	
here.com	SMTP	Default SMTP Virtual Server	Retry	1	3	1/13/2004 3:27 PM
Local delivery	SMTP	Default SMTP Virtual Server	Ready	0	0	
Messages awaiting directory lookup	SMTP	Default SMTP Virtual Server	Ready	0	0	
Messages pending submission	SMTP	Default SMTP Virtual Server	Ready	0	0	
Messages queued for deferred delivery	SMTP	Default SMTP Virtual Server	Ready	0	0	
Messages waiting to be routed	X400	Exchange MTA	Ready	0	0	
Messages waiting to be routed	SMTP	Default SMTP Virtual Server	Ready	0	0	
SMTP maximum store (legacy)	X400	Exchange MTA	Active	0	0	

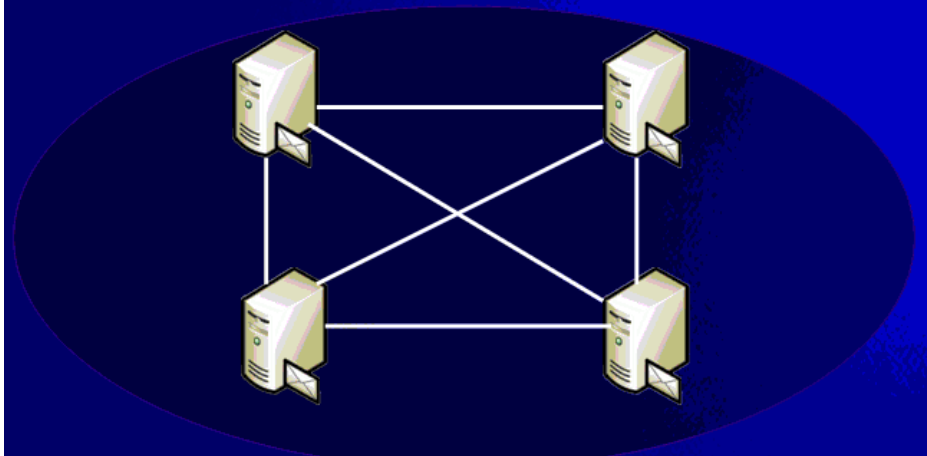
邮件路由路径——路由引擎

可以通过邮件跟踪看到。

- ◆ 决定邮件传递路径
 - 本地投递
 - 如果路由地址符合本地HomeMDB地址，则跳过路由
 - 通过 SMTP 或 X.400 远程路由
 - 使用路由决定下一站
 - 网关
- ◆ 邮件传递路径使用链路状态数据表决定

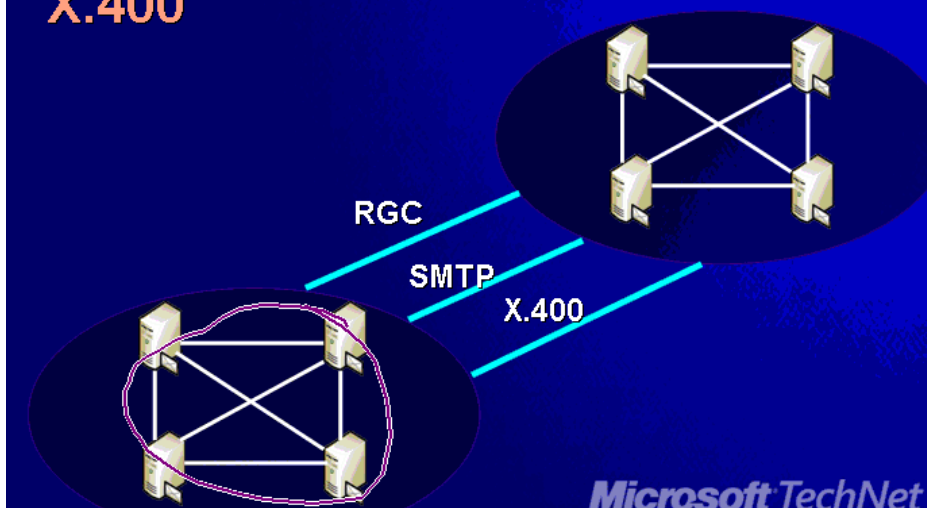
同一个路由组内，直接连接25端口，如图。

Point-to-point, SMTP, Port 25



如果不在同一个路由组。一定要通过connectors连接器来传递邮件。

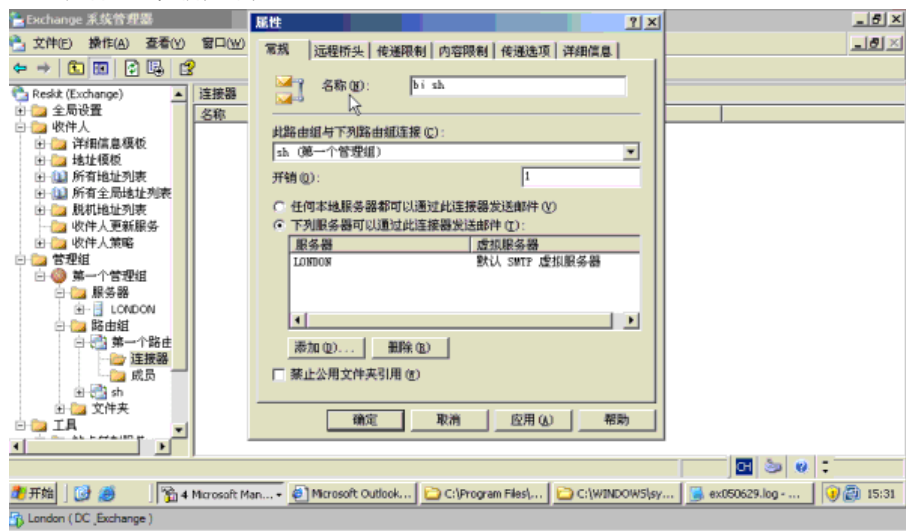
Routing group connector, SMTP, X.400



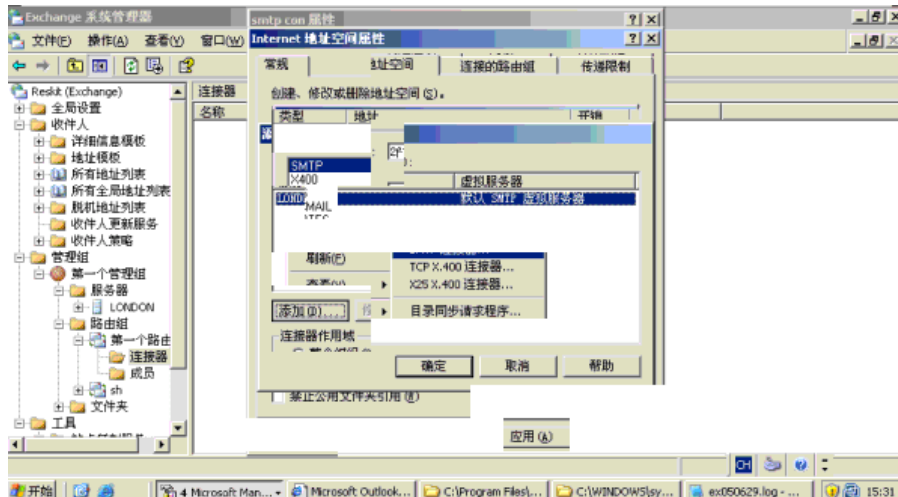
路由组设计的时候，一般都是按照链路状态来设计。路由组之间可以创建连接器，本地路由组和远程路由组可以进行连接。

下面是路由组连接器的属性。

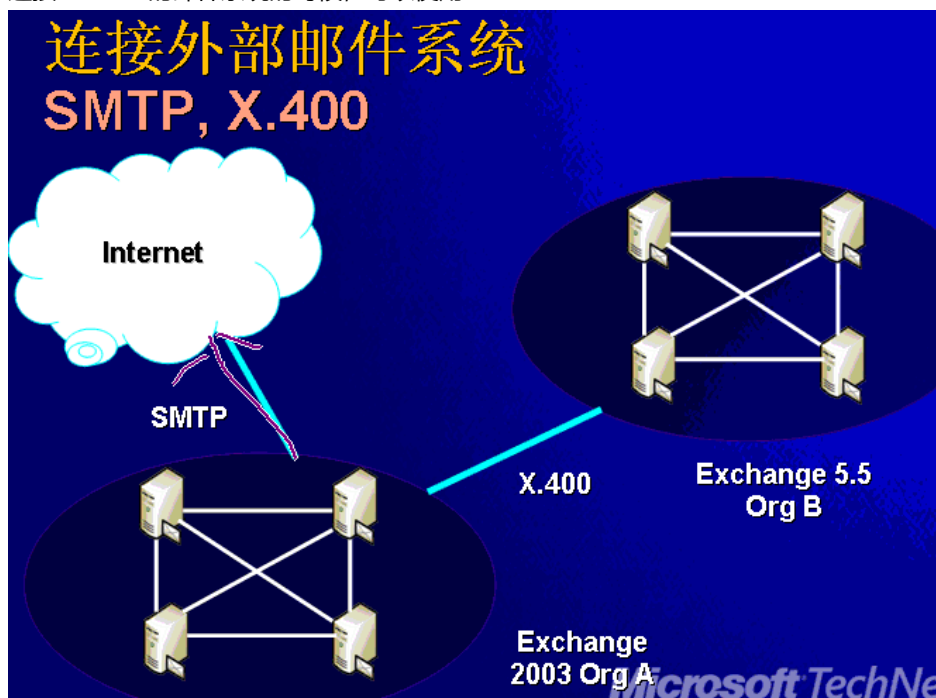
还可以指定远程桥头服务器。



也可以创建SMTP的连接器，连接到特定的地址空间。



连接Internet的邮件系统的时候，可以使用SMTP connector



基本就三种情况：

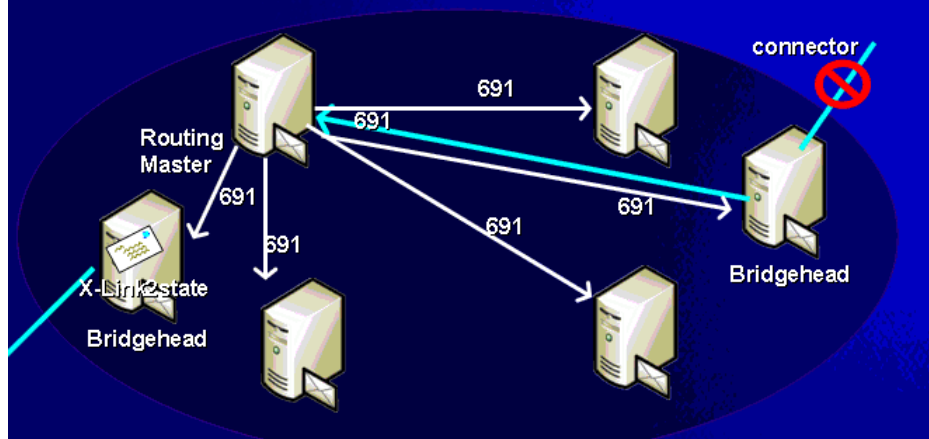
4. 路由组之间
5. 邮件系统与外部
6. 路由组内部

如果路由组之间有问题，可以在桥头堡服务器去抓netmon的数据包，来看连接请求情况。

路由组的主服务器和路由组的桥头堡服务器不一定是同一台服务器，是不同的概念。

路由表和链路状态的信息是由路由组的master来维护的，而桥头堡服务器是不会维护这些信息的。其实这和活动目录站点的概念特别类似。

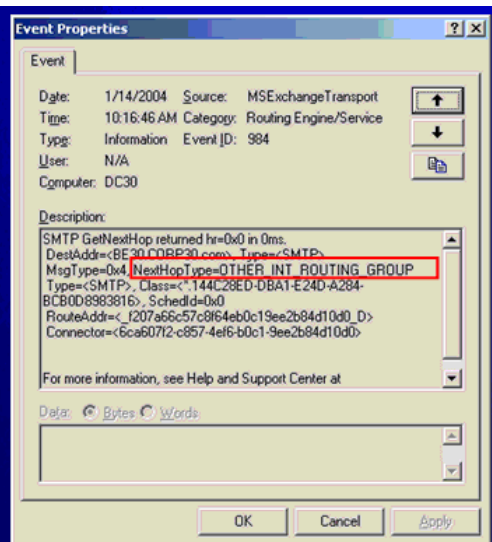
路由组主服务器: Link state table updates



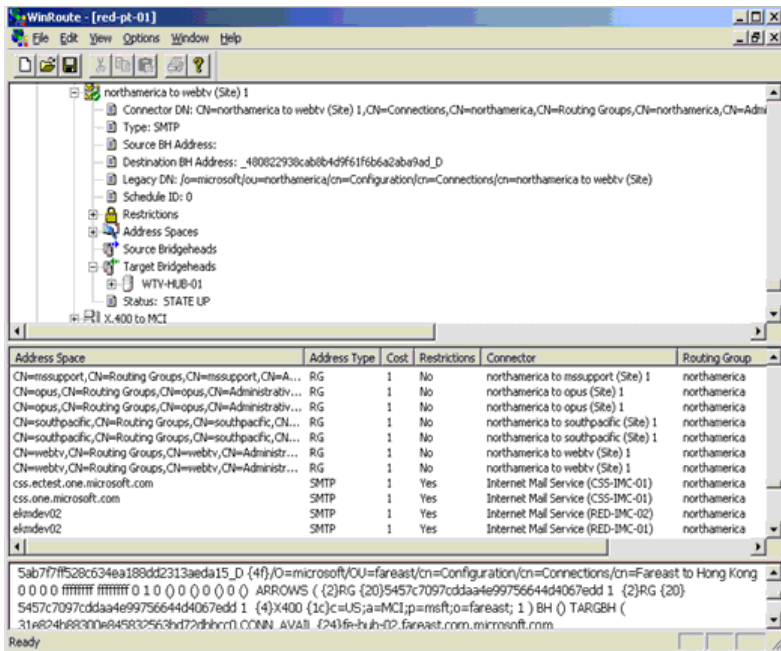
在AD站点中桥头服务器和ISTG服务器也是不同的角色。

可以查看路由的日志信息。

- ◆ 邮件将会被路由至何处
 - 内部还是外部?



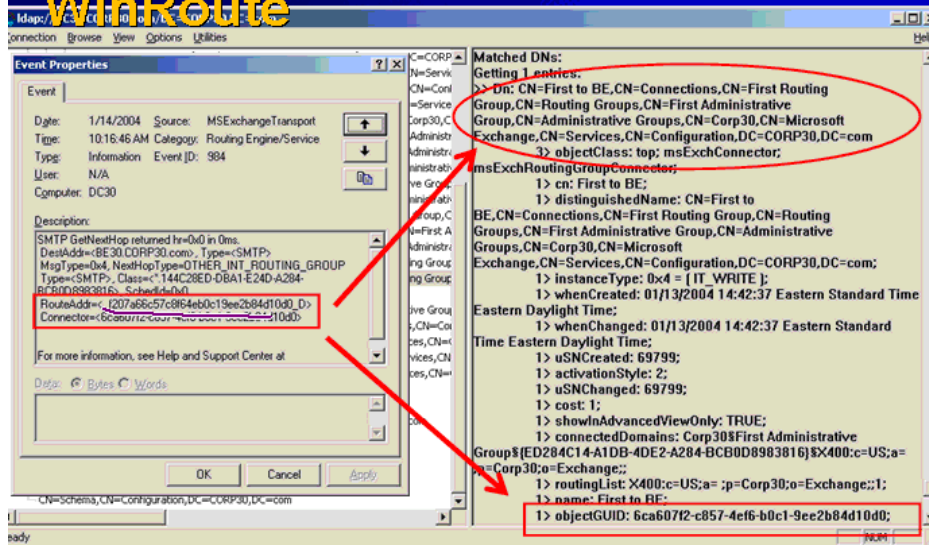
通过winroute来查看路由表的情况。



通过一周或两周的时间会生产一个完整的链路状态数据库。

事件日志、AD、winroute三者的信息是可以匹配上的。

Match Between Event Log and Information in Active Directory or WinRoute



远程投递。

- ◆ 队列和链路
 - 相同目的地
 - Connector
- ◆ 连接到目的地IP
 - 每个连接每次发送20封邮件
 - 投递状态通知DSNs
 - 400 level – 可重试
 - 尝试三次Glitch retry (60 seconds), 然后尝试 true retry
 - See <http://blogs.technet.com/exchange/archive/2005/04/04/403297.aspx>
 - 500 level – 永久错误
 - 生成NDR
 - 不可投递 NDRs => badmail 文件夹
 - See KB article 284204

共用文件夹是通过邮件进行复制的，也是基于邮件传递的一种复制。

入站邮件需要允许匿名，所以入站邮件必须有反垃圾邮件过滤；而出站邮件可以不允许匿名。

对于邮件归档软件来说，在inbound和outbound都需要部署。

第九节：exchange性能调优上

关键性能影响因素

性能计数器指标

性能分析和排错工具

- ◆ 了解Exchange系统中对性能有显著影响的因素
 - CPU
 - RAM
 - Network
 - Disk I/O
 - DsAccess (AD/DC)
- ◆ Exchange中需要注意的性能计数器指标
- ◆ 常见的性能分析和排错工具
 - Exchange BPA
 - Perfmon
 - Netmon
 - Systemal Tools (Regmon/FileMon/Process Explorer) <http://www.sysinternals.com>
 - Task Manager
 - Call MS PSS

当CPU从四个加到8个的时候，性能不升反降，对于前端服务器建议最高不超过4颗CPU。

如果四颗CPU的前端负载很高，可以通过增加多台前端服务器的方式来分散负载。

更快的CPU比更多的CPU更重要。

还要看操作系统能够支持的最大CPU数量。

超线程芯片显示的情况，比如实际处理器数量是2个，但是在操作系统显示4个。

CPU其实并不会产生很严重的瓶颈。

exchange server里面哪些模块对CPU要求高？1、加密模块，HTTPS；2、本身information store对CPU压力不是很高，但是当部署防病毒软件的时候，且邮件发送比较密集，也会要求CPU占用高；3、IIS里面处理的所有的HTTPS的加密运

算，所以IIS的进程消耗CPU也很高；4、机器响应慢，发现某一个进程占用CPU很高，在不正常的使用CPU，可能这个程序已经挂起了，程序代码还在运行，但是出于死循环的状态，比如早期的exchange的information store占用了很高CPU；

◆ **CPU -- Scalability**

- 推荐使用Xeon和超线程芯片(Pentium IV XEON MP)
- 使用你能够承担得起的最快的芯片，faster CPU is better than more CPU
- 不同角色的Exchange Server,对CPU要求也不同
 - Mailbox Store Server
 - 2-8 CPU, 不要超过8个。
 - Front-end Server (前端服务器, OWA/邮件网关)
 - Up to 4CPU
- Link:
 - <http://www.microsoft.com/windows2000/server/evaluation/performance/reports/hyperthread.asp>
 - <http://www.microsoft.com/exchange/techinfo/planning/2000/PerfScal.asp>

下图的三个进程会消耗大量的进程。

information store，即store.exe会占用大量的内存

IIS的模块，即inetinfo.exe

MAD.exe是exchange的核心模块之一，承担对AD的访问操作

下图的公式是一个算法，动态缓冲区的分配，根据内存的使用趋势来分配内存占用的空间，动态调整。

◆ **RAM**

- **Exchange会消耗大量的RAM**
 - Store.exe → 1.7GB
 - InetInfo.exe → 500MB
 - MAD.exe → 60MB
- **DBA: Dynamic Buffer Allocation**

$$\frac{\partial C}{\partial t} = \frac{A}{T} \times \frac{\partial B}{\partial t} - \frac{C}{T} \times \frac{\partial P}{\partial t}$$

C= Cache Memory A= Available Memory
T= Total Memory B= Buffer Allocations
P= OS Page Allocations

LRU-K replacement algorithm (Partial Derivative) [最近最少使用]

除了exchange本身以外，防病毒，操作系统，备份软件，也会占用一定的内存的。

下图是网络对性能的影响。

OUTLOOK RPC协议对网络要求比较高。

网卡的对端全双工模式匹配，网卡速率要匹配。

网卡模式不匹配的话会导致outlook连接特别慢。

如果DC的网卡不匹配，也会导致连接DC的时候特别慢。

◆ Network

- 100M Full Duplex
- 如果用 IPsec, 应该选用针对IPsec优化的网卡以减轻对CPU的负载
- 在前端服务器上使用快速的网卡

◆ Disk I/O

- RAID 1+0 is much better than RAID 5
- Faster RAID card

◆ DsAccess (AD/DC)

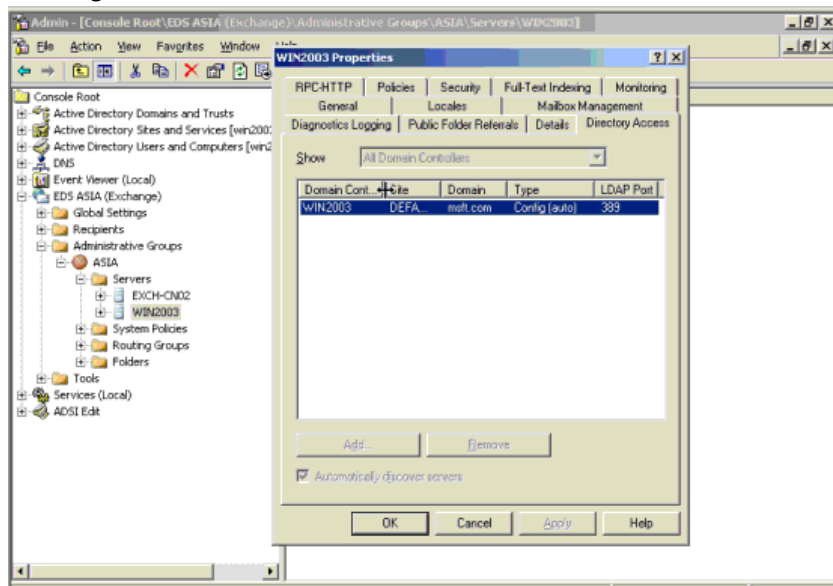
- 了解Exchange查找DC的规律, 确保Exchange使用最近范围内的DC

Microsoft TechNet

exchange的用户信息, 配置信息都是放在DC的。用户邮箱是放在exchange的。

exchange server查找DC的顺序, 使用的是LDAP协议。找DC的顺序:

exchange存储组的DA选项卡:



与DC建议三个连接: 1、取配置信息; 2、取用户信息; 3、联系GC, 拿所有的全球地址簿;

取信息的时候依赖DNS, 通过SRV记录定位信息, 发现在同一个站点有DC存在的时候, 优先找站点内部的DC, 如果本地站点没有DC, 找远端的DC, 所以exchange的DNS要指向本地内网的DC; 离exchange最近的DC服务器。

dasdiag命令, 会模拟exchange查找DC的过程。

找到DC之后, 再通过LDAP进行查询。

网卡建议手动设置双工, 不建议设置auto。

内网exchange server不使用MX记录, 内网exchange使用DNS时为了找到DC信息。

exchange server需要关注的性能计数器指标

Database (Information Store)↵	↵ PhysicalDisk↵
- Database Cache Size↵	- % Disk Time↵
- Log Record Stalls/sec↵	- Avg. Disk sec/Read↵
- Log Threads Waiting↵	- Avg. Disk sec/Transfer↵
- Log Writes/sec ↵	- Avg. Disk sec/Write↵
- Table Opens/sec↵	- Current Disk Queue Length ↵
	- Disk Transfers/sec↵
LogicalDisk ↵	↵ Process (select all instances)↵
- % Disk time (only after diskperf -y)↵	- % Processor Time↵
- % Free Space↵	- % User Time↵
- Avg. Disk Queue Length ↵	- Elapsed Time↵
- Avg. Disk sec/Read↵	- Handles↵
- Avg. Disk sec/Write↵	- Page Faults/sec ↵
- Avg. Disk sec/Transfer↵	- Page File Bytes↵
- Current Disk Queue Length↵	- Pool Nonpaged Bytes↵
- Free Megabytes↵	- Private Bytes ↵
	- Virtual Bytes ↵
	- Working Set ↵

日志文件写得越多，服务器越繁忙。

当前磁盘队列长度，一般要小于物理磁盘的磁头数。特别需要看C盘。比如有四块物理磁盘，那么队列不要超过4或者2。

如果页面错误比较高，要考虑增加物理内存。

下面的指标非常重要。

PRC average latency非常重要，直接反应了服务器的状态，健康状况。一般情况下要小于50毫秒。

MSExchangeIS↵	↵ Server↵
- Active Connection Count ↵	- Bytes Total/sec↵
- Active User Count ↵	- Pool Nonpaged Bytes↵
- Connection Count↵	- Pool Nonpaged Failures↵
- RPC Average Latency ↵	- Work Item Shortages↵
- RPC Operations/sec↵	Server Work Queues↵
- User Count↵	- Active Threads↵
- Virus Scan Queue Length ↵	- Bytes Sent/sec↵
- VM Largest Block Size↵	- Queue Length↵
- VM Total 16MB Free Blocks ↵	- Read Bytes/sec↵
- VM Total Free Blocks↵	- Write Bytes/sec↵
- VM Total Large Free Block Bytes↵	- Write Operations/sec↵

outlook把邮件投递的exchange所用的时间，可以通过average delivery time来体现。

MSExchangeIS Mailbox↵	↵ SMTP Server↵
- Active Client Logons ↵	- Categorizer queue Length ↵
- Average Delivery Time ↵	
- Average Local Delivery Time↵	System↵
- Message Opens/sec↵	- Processor Queue Length↵
- Received Queue Size ↵	- System Up Time↵
- Send Queue Size ↵	↵
	↵ TCP↵
MSExchangeIS Public↵	- Segments Received/sec↵
- Average Delivery Time↵	- Segments Retransmitted/sec↵
- Average Local Delivery Time↵	
- Folders Open/sec↵	Thread (select all instances)↵
- Message Opens/sec↵	- % Processor Time↵
- Received Queue Size↵	- ID Thread ↵
- Send Queue Size↵	- Thread Wait State↵
	- Thread Wait Reason↵
Network Interface (select all instances)↵	↵ Paging File↵
- Bytes Received/sec↵	- % Usage↵
- Bytes Sent/sec↵	
- Bytes Total/sec↵	
- Output Queue Length ↵	

exchange BPA会自动分析exchange服务器的配置是否正确，然后给出修改或者优化的建议。

=====

