Education

Exchange Performance 101

Ray Lucchesi, Silverton Consulting, Inc.
http://www.SilvertonConsulting.com

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Abstract

- Exchange Performance 101
  - Exchange storage performance is constantly evolving. Tuning storage to support Exchange services for the best performance is a complex and never ending task. As new storage becomes available tuning for email services also becomes more complex.
  - We will discuss some techniques to tune storage for email services. Also, we will look at some storage product's exchange performance to reveal what was done to support the service levels reported.

Exchange architecture

- User mailboxes
- System mailbox databases
- Storage groups
  - Databases
  - Log files
- Exchange servers/hosts
- Exchange replication options

User mailboxes

- Size – 100MB to 2GB
  - 1 to 2GB typical
  - Exchange 2010 to10GB
- IOPs/user mailbox key storage factor
- Mailboxes aggregated into one or more mailbox databases
- Mailboxes served via one or more Exchange servers
- Mailboxes can be moved
  - From one mailbox database to another
  - From one server to another

Mailbox databases

- Database is unit of backup & reliability
  - Database problems impact all mailboxes in database
- Database size based on many factors - Microsoft recommends
  - Maximum of 100GB when not using continuous replication
  - Maximum of 200GB with continuous replication
  - Exchange 2010 maximum database size increased to 2TB
- Database size mostly dictated by backup & recovery SLAs
- Databases are aggregated into one or more Storage Groups

Mailbox databases (cont.)

- Database to LUN layout
  - Database cannot span more than one LUN
  - More than one database can reside on a single LUN
  - Databases cannot span multiple LUNs
- Database I/O is generally random
  - More mailboxes hosted on a database the more IO there is to the database

Storage groups

- One or more databases can be assigned to a single storage group
- One log per storage group maintaining transaction playback
- Maximum of 50 storage groups per Exchange server
- Continuous replication requires at most one database per storage group
- Exchange 2010 is removing storage group concept

Storage group log

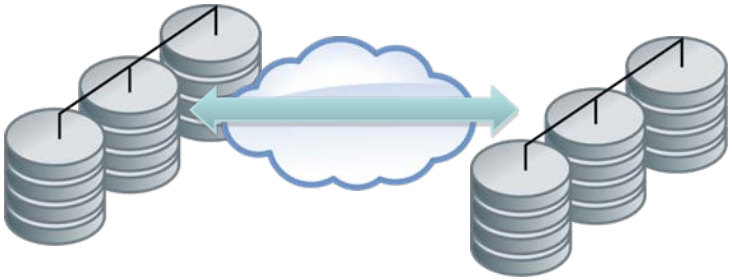
- One set of log files per storage group
  - A single Log file cannot span more than one LUN
  - Multiple storage group logs can reside one LUN
- Log files record all transactions to databases in storage group
  - Data written to logs first and then lazily written to database(s)
  - Log files are 1MB in Exchange 2007
- Backed up log files = incremental backup for the database(s) in the storage group
- Sequential I/O

Exchange server

- Exchange 2007 supports
  - 64 bit O/S supports much more memory
  - Maximum of 50 storage groups per server
- Database caching is an important consideration
  - Changes R:W ratio from 2:1 to 1:1 by caching random database reads
  - More memory the better up to recommended 5MB of cache per user

Exchange replication options

- Local continuous replication (LCR)
- Cluster continuous replication (CCR)
- Disk replication
  - Synchronous
  - Asynchronous

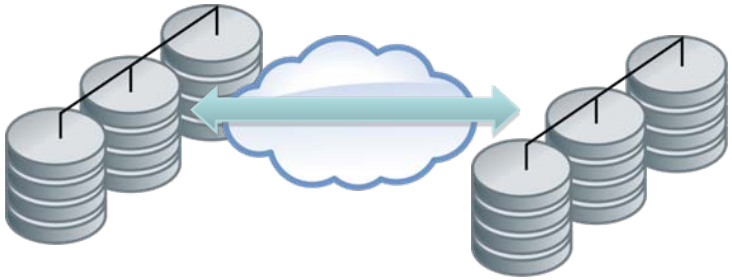


LCR & CCR

- Passive database(s) and shipping (or copying) log files
  - Log files replayed against the passive database
  - For LCR passive database and log copies are on same Exchange server
  - For CCR passive database and log copies on different Exchange server
    - CCR also uses Microsoft Cluster services for failover
- For CCR and LCR
  - Only one database per storage group
  - Standby continuous replication supports multiple recovery points

Disk replication or mirroring

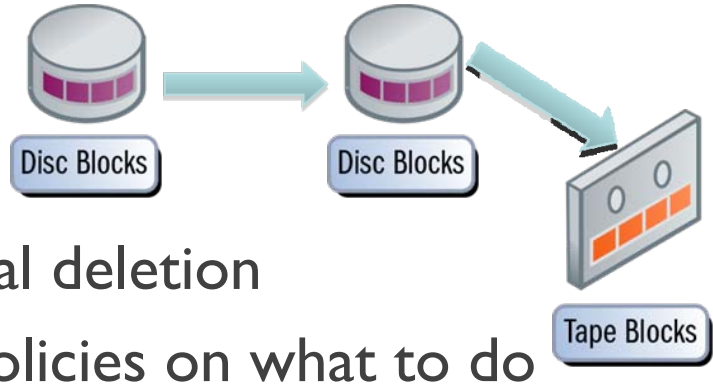
- Storage group needs to be replicated in total
  - Database LUN(s)
  - Log LUN(s)



- Synchronous replication
  - Current database & logs mirrored at remote site
- Asynchronous replication
  - Log files need to be replayed to bring database current
  - Recovered database(s) may be downlevel
- Possible to asynchronously replicate database while synchronously replicate log files

Exchange IO operations

- Transactional IO
  - Mail user read, write, send emails
  - Impacted by Exchange server database caching
    - In Cached Exchange mode transactional IO generally 1:1 read:write
- Non-transactional IO
  - Database backup and restores
  - Online maintenance – defrag and real deletion
  - Messaging records management – policies on what to do with email



Transactional IO - database

- Database reads for 50KB avg message
  - (Msgs rcvd*0.0048)*(Cache per user **(-0.65)
- Database writes for 50KB avg message
  - Msgs snt*0.00152

User	Sent/received per day	DB cache size per user	Estimated IOPS
Light	5snt/20rcvd	2MB	0.11
Average	10snt/40rcvd	3.5MB	0.18
Heavy	20snt/80rcvd	5MB	0.32
Very heavy	30snt/120rcvd	5MB	0.48
Extra heavy	40snt/160rcvd	5MB	0.64

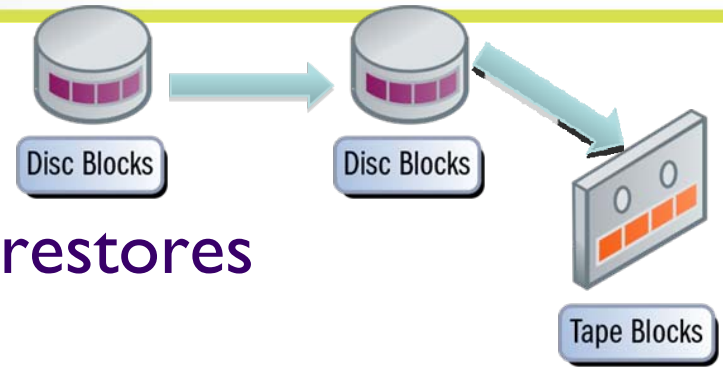
Transactional IO (cont.)

- Log file IO = 75% of database IO per user
  - Recommend add 20% safety factor to account for busier than normal periods
- Online client mode increases database read by 1.5X
  - And mailbox size increases database reads, e.g. every multiple of 250MB will double the database readcount again

Non-transactional IO

- Online maintenance, happens every nite
  - Lot's of IO but hopefully not competing with user IO
  - IO amount proportional to database size
  - Defrag database
  - Physically delete emails
- Message records management (MRM)
  - MRM is scheduled, can be down during idle periods
  - MRM IO amount depends on the number of items to act upon

Non-transactional IO



Backup and restore sequential I/O

- VSS to create backups, perform restores
- Continuous replication backups
  - Done against the passive databases and logs
- Can be staggered
  - 1/7 of databases full backup each night
  - Other 6/7 incrementally backed up
- Recommend separate LUNs and physical disks between source and target for backup-restore

Storage considerations – part 1

- Database transactional IO random and 1:1 R:W
  - Probably best to use RAID1 or 10 due to the heavy random write
- Spread LUNs over many physical disks
  - Random IO benefits from more spindles
- Try to segregate database LUN physical spindles from Log LUNs
  - For availability in case of spindle failure
  - For better performance, not mixing random and sequential IO on same disk drives

Storage considerations – part 2

- Having multiple databases on one LUN is ok
  - Consider how your backup IO impacts database IO
- Disk replication may require that only one DB per LUN
  - CCR and LCR require this already
- Separate primary database LUN from backup database LUN
  - Both at the LUN level and at the physical disk level

Storage considerations – part 3

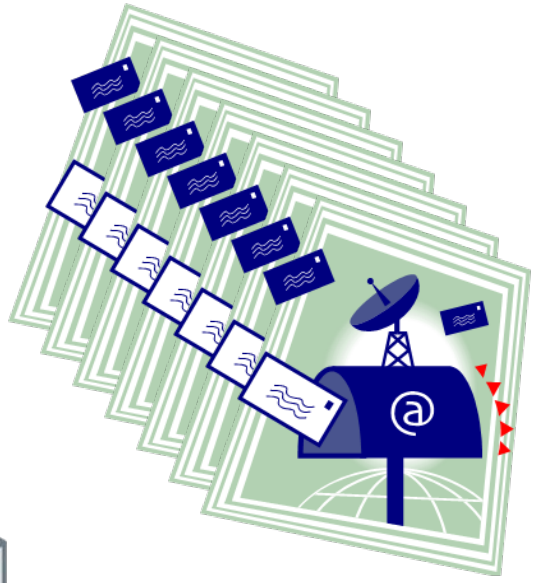
- Storage read caching probably not that effective
  - But depends on sophistication of algorithms
  - Heavy write workload may require more % of cache devoted to write data
- Recommend a 2 LUN SG
  - LUN for Database(s)
  - LUN for Log files

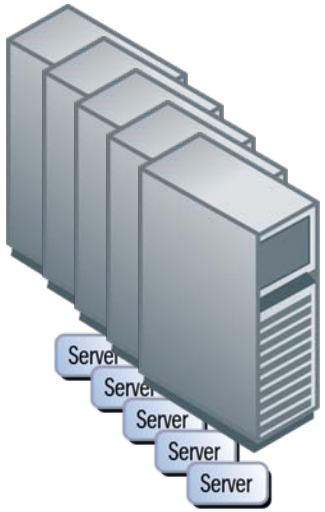
Tools for Exchange

- Jetstress – simplified storage validator
  - Set up parms, and storage configuration and jetstress simulates user mail activity against the storage
- Exchange Load Gen – fully functional Exchange validator
  - Set up exchange servers, storage config, user mailboxes
  - Will simulate various Email services MAPI, OWA, POP against your Exchange environment

Jetstress parameters

- Mailbox count
- IOPs/mailbox – user level of activity
- Number of Exchange servers
- Database size
- Log size
- Replication option





Jetstress tests

- Simulates user mailbox activity against the storage configuration
- Database backups
- Log playback time

Jetstress measures

- Database transfers per second
  - Read, write, log writes
- Database transfer latency
  - In msecs for read, write, log write
- Database backup throughput
  - In MB/sec per storage group, per database
- Log playback time
  - In seconds for 1MB logfile

Exchange performance results

Education SNIA

Exchange Solution Review Program (ESRP)

- Available at http://technet.microsoft.com/en-us/exchange/bb412164.aspx
- Uses Jetstress to validate vendor storage configurations
  - Results split into <= 1000 mailboxes, 1000..5000 mailboxes, >5000 mailboxes
  - Valid results must have response time <20msec.
  - Supports FC, iSCSI, and SAS storage

Load generator

- Simulates active Exchange Clients
  - MAPI, OWA, IMAP, POP and SMTP
- Full end-to-end simulator of Exchange sever-storage configuration
- Loadgen runs on client computers

Best practices storage config's

- Latency
- Database backup
- Database and log latency
- Database design
- Log design
- Jetstress results
- Jetstress parameters

Best practice - latency

- Database latency is a key end-user factor
  - Latency can affect all end-users as they access Exchange
- Focus on speeding up random access for database
  - Fast disks
  - Mailstore/database to LUN to physical disk layout
  - Storage RAID level
- Focus on sequential access for log file
  - Fast disks
  - Storage RAID level

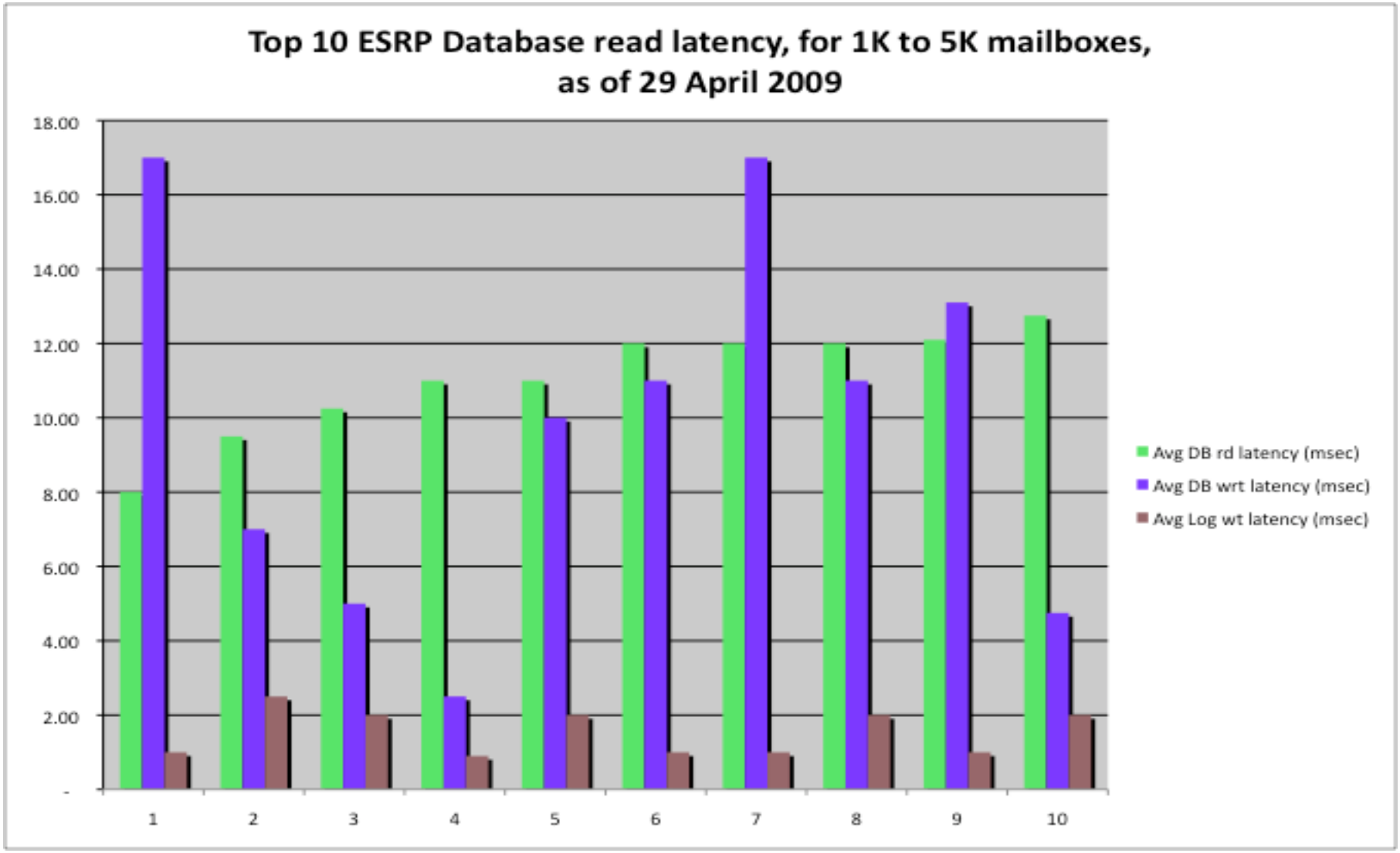
Database design for best latency

- Supported 1800 user mailboxes
  - 360-mailboxes/mailbox database,
  - 5-databases in1-storage group
- Uses high performance,15Krpm disks
  - Small capacity - 146GB
  - Many disks - 10 disks
- 10 disks configured as database LUNs
  - RAID 10 to support random access
  - As 5 90GB LUNs M:..Q: for Exchange
- Not a lot of controller cache

Log design for best latency

- Used fast disks, same as database design
  - Log I/O mainly sequential writes/reads
- Used RAID 1
- Configured as single 130GB Log LUN

Database latency top 10 – 1..5K



Jetstress parameters

Simulated profile

- 0.5 IOPs/sec/mailbox (very heavy workload)
- 1800 user mailboxes supported
- 200MB mailbox size
- No replication
- 1 Exchange server
- Small to medium sized Exchange environment
- iSCSI SAN storage

Best practice - backup

- Focus on sequential/streaming performance
  - Mailstore/database to LUN to physical disks
  - Spread the database and storage group over many disks
    - More disks the better
  - High sequential throughput disk drives
- Often what the datacenter is concerned about
  - Done often
  - Lot's of data to backup

Database design - backup

- Uses high performance,15Krpm disks
  - Small capacity - 146GB
  - Many disks – 448 disks
- 21 LUNs configured as 1-database
  - RAID 1+(0) (2D+2D) to support random access
  - Each RAID group 4 disks supports 3 ~90GB LUNs for Exchange database access
    - 7 RAID 1+(0) groups/database
    - 38 physical disks per database

Database design part 2

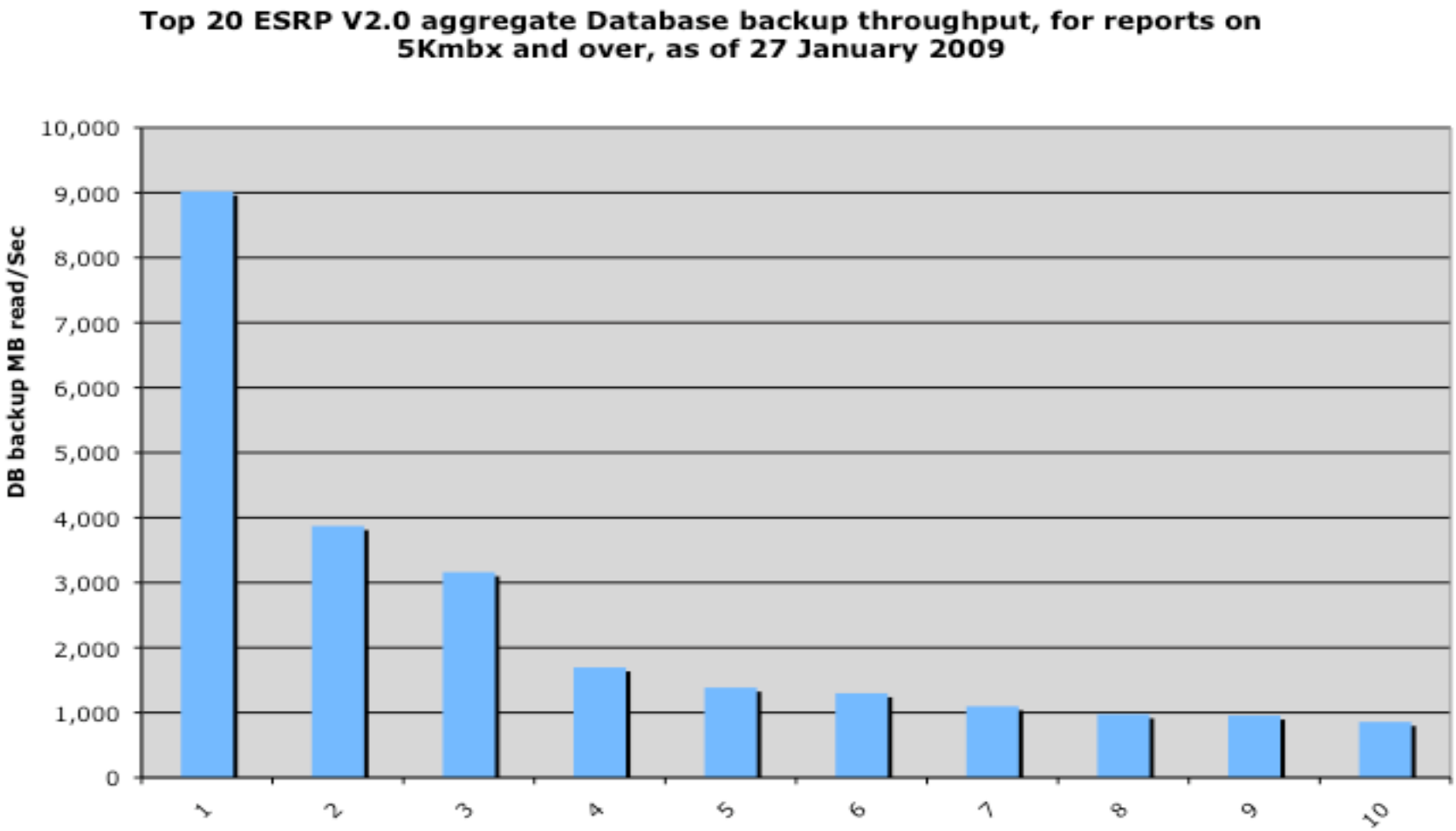
Supported 100,000 user mailboxes

- 298-mailboxes/mailbox database,
- 1-database/storage groups
- 21-storage group per exchange server
- 336 storage groups for all mailboxes
  - 448 146GB 15Krpm drives for mailstore database
- 25TB mailstore databases

Log design for best backup

- Used fast disks, same as database design
  - Log I/O mainly sequential writes/reads
- Used RAID 1+(0)
  - RAID 1+(0) supported 8 Log LUNs
  - Each LOG RAID 1+0 (2D+2D) used 4 disk drives
    - RAID 5 would have been OK with more disks
- Configured as 8 12.75GB LUNs

Database backup top 10 - >5k



Jetstress parameters

Simulated profile

- 0.384 IOs/sec/mailbox (heavy workload targeted)
  - 0.67 IOs/sec/mailbox achieved
- 100,000 user mailboxes supported
- 250MB mailbox size
- No replication
- 16 Exchange servers
  - 6250 mailboxes/server
- Large enterprise Exchange environment
- FC SAN storage

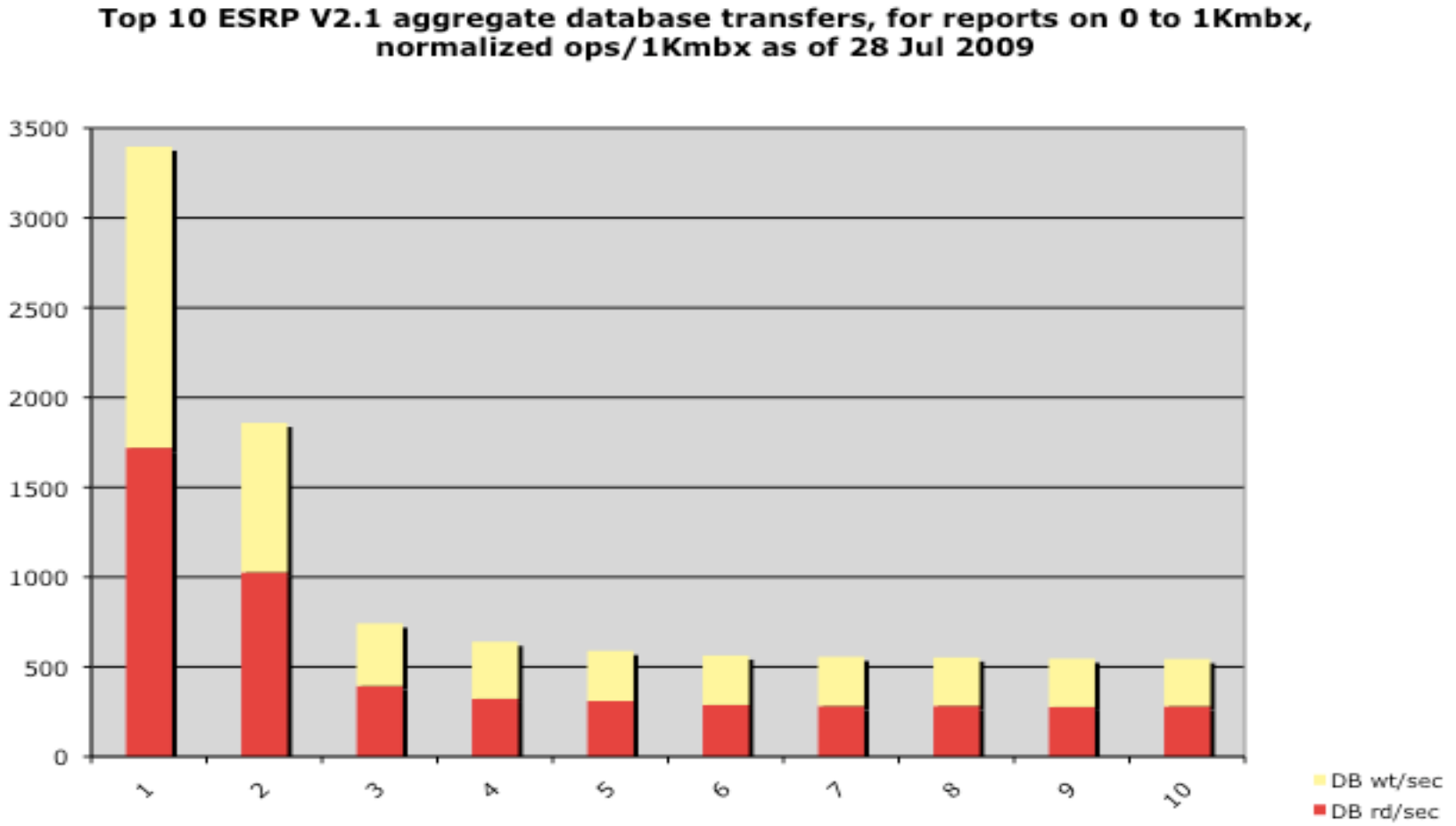
Best practice - database transfers

- Improve random access
  - Fast disks
  - Better database layout
  - More spindles
  - RAID level
  - Speedy replication options

Combined database&log design

- Uses medium performance, **10Krp**m disks
  - Small capacity - 146GB
  - Many disks – 32+ disks
- 1 LUN/database, 1database/storage group, 12 storage groups for configuration
- 2 RAID 10 groups
  - Each RAID group supports 6 database and 6 Log LUNs together
  - Each RAID group had 16 physical disks

Database xfers/sec top 10 – 0..1K



Jetstress parameters

Simulated profile

- 0.576 IOs/sec/mailbox (heavy workload targeted)
  - 3.4 IOs/sec/mailbox achieved
- 1000 user mailboxes supported
- 1000MB(1GB) mailbox size
- No replication
- 1 Exchange server
- Recovery storage group
- Small to medium sized Exchange environment
- FC SAN storage

For more information

- ESRP http://technet.microsoft.com/en-us/exchange/bb412164.aspx
- Exchange planning storage configurations http://technet.microsoft.com/en-us/library/bb124518.aspx
- Silverton Consulting

Q&A / Feedback

- Please send any questions or comments on this presentation to SNIA: trackstoragemgmt@snia.org

Many thanks to the following individuals for their contributions to this tutorial.

- SNIA Education Committee

Ray Lucchesi