THE JOURNEY OF A BI-STACK TO THE CLOUD

Pieter Van Puymbroeck
• Dad
• Oracle DBA
• Sailor
• Musician
• working for Exitas Oracle Platinum Partner in BELGIUM

http://vanpupi.stepi.net

@vanpupi
About Exitas

Quality - Passion - Personality

- Team of 25 dba’s
- Consultancy, managed services,…
- http://www.exitas.be
500+ Technical Experts Helping Peers Globally

3 Membership Tiers
- Oracle ACE Director
- Oracle ACE
- Oracle ACE Associate

bit.ly/OracleACEProgram

Connect:
- oracle-ace_ww@oracle.com
- Facebook.com/oracleaces
- @oracleace

Nominate yourself or someone you know: acenomination.oracle.com
Agenda

• The project
• Cloud semantics
• The design
• Tests
• Provisioning
• Initial load
• Lessons learned
• Q/A
Safe Harbour statement

The project dates from January 2017
(The new content is added, where applicable)
And the message is still valid ;-)

Little red ridinghood and the big bad GDPR: how does it impact you
The project
The project
The journey of a bi-stack to the cloud
How it started …

• Request for assistance for a move to the cloud?
• It’s about 55TB prod data
  • Prddb1: 34TB
  • Prddb2: 16TB
  • Prddb3: 5TB
• Agile: “iterative incremental process”
• Oh by the way … non-prod is ~ 360TB … we think …
How it started …
Fact check

The journey of a bi-stack to the cloud
Regions

Location: Amsterdam

The journey of a bi-stack to the cloud
It’s all about avoiding limits
Disks

This is what was available during the project:

**Standard Disks (HDD)**

<table>
<thead>
<tr>
<th>Storage disk type</th>
<th>Standard Disk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk size</td>
<td>1024 GB</td>
</tr>
<tr>
<td>IOPS per disk</td>
<td>500</td>
</tr>
<tr>
<td>Throughput per disk</td>
<td>~60 MB per second</td>
</tr>
</tbody>
</table>

**Premium Disks (SSD)**

<table>
<thead>
<tr>
<th>Storage disk type</th>
<th>P10</th>
<th>P20</th>
<th>P30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk size</td>
<td>128 GB</td>
<td>512 GB</td>
<td>1024 GB</td>
</tr>
<tr>
<td>IOPS per disk</td>
<td>500</td>
<td>2300</td>
<td>5000</td>
</tr>
<tr>
<td>Throughput per disk</td>
<td>100 MB per second</td>
<td>150 MB per second</td>
<td>200 MB per second</td>
</tr>
</tbody>
</table>

**Performance Expectations**

**Performance is not provisioned**

**Consistency of provisioned performance**
Disks

And this is available nowadays:

<table>
<thead>
<tr>
<th>Premium Disks Type</th>
<th>P4</th>
<th>P6</th>
<th>P10</th>
<th>P20</th>
<th>P30</th>
<th>P40</th>
<th>P50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk size</td>
<td>32 GB</td>
<td>64 GB</td>
<td>128 GB</td>
<td>512 GB</td>
<td>1024 GB (1 TB)</td>
<td>2048 GB (2 TB)</td>
<td>4095 GB (4 TB)</td>
</tr>
<tr>
<td>IOPS per disk</td>
<td>120</td>
<td>240</td>
<td>500</td>
<td>2300</td>
<td>5000</td>
<td>7500</td>
<td>7500</td>
</tr>
<tr>
<td>Throughput per disk</td>
<td>25 MB per second</td>
<td>50 MB per second</td>
<td>100 MB per second</td>
<td>150 MB per second</td>
<td>200 MB per second</td>
<td>250 MB per second</td>
<td>250 MB per second</td>
</tr>
</tbody>
</table>
Networking

Accelerated networking (ACN)

Accelerated networking:
• General available
• Recommended

Availability sets


- VMs deployed on different tracks (= Fault Domain (FD)) and different physical servers (= Update Domains (UD))

- Cfr VMware Anti-Affinity Rules
The journey of a bi-stack to the cloud
The journey of a bi-stack to the cloud

<table>
<thead>
<tr>
<th>Type</th>
<th>Sizes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>General purpose</td>
<td>Dv3, Dv3, D5v2, Dv2, D5, D, Av2, A0-7</td>
<td>Balanced CPU-to-memory ratio. Ideal for testing and development, small to medium databases, and low to medium traffic web servers.</td>
</tr>
<tr>
<td>Compute optimized</td>
<td>Fs, F</td>
<td>High CPU-to-memory ratio. Good for medium traffic web servers, network appliances, batch processes, and application servers.</td>
</tr>
<tr>
<td>Memory optimized</td>
<td>Esv3, Ev3, M, GS, G, D5v2, D5, Dv2, D</td>
<td>High memory-to-CPU ratio. Great for relational database servers, medium to large caches, and in-memory analytics.</td>
</tr>
<tr>
<td>Storage optimized</td>
<td>Ls</td>
<td>High disk throughput and I/O. Ideal for Big Data, SQL, and NoSQL databases.</td>
</tr>
<tr>
<td>GPU</td>
<td>NV, NC</td>
<td>Specialized virtual machines targeted for heavy graphic rendering and video editing. Available with single or multiple GPUs.</td>
</tr>
<tr>
<td>High performance</td>
<td>H, A8-11</td>
<td>Our fastest and most powerful CPU virtual machines with optional high-throughput network interfaces (RDMA).</td>
</tr>
</tbody>
</table>

The current vm’s:

**Linux:**
https://docs.microsoft.com/en-us/azure/virtual-machines/linux/sizes

**Windows:**
https://docs.microsoft.com/en-us/azure/virtual-machines/windows/sizes
Choosing a VM

- MS defined categories
- Own requirements
- VM = predefined set of resources
  - # vCPUs
  - # vRAM
  - standard vs premium storage os/data disks
  - Maximum disk throughput
  - # data disks
  - Max disk throughput
  - Network bandwidth

The journey of a bi-stack to the cloud
Choosing a VM

1. Workload
   - Cpu requirements
   - Memory requirements
   - Disk requirements

2. Storage
   - S: fast storage
   - VMs can use premium storage
   - VMs use solid state drives (SSD)

3. VM Size
   - CPU
   - RAM
   - # data disks
   - Disk throughput

Standard_DS15_v2
Choosing a VM

Requirements:
- High workload
- High #disks
- ACN needed
- High throughput

-> DS15_V2

-> Hard coupling between memory and cpu
The journey of a bi-stack to the cloud

Topologies:

1. **SAP BI BO**
   - **DEDICATED**
   - DC1/DC2
   - VMs spread across DCs

2. **BI DB**
   - **DEDICATED**
   - DC1/DC2
   - VMs spread across DCs
   - ESX Cluster PROD/ESX Cluster NON-PROD
   - Dataguard

3. **BACKUP**
   - **SHARED**
   - DC1/DC2
   - VMs spread across DCs

**RHS** – 8 nodes spread over 2 DCs
Topology - To be: Requirements

- High network throughput ~ 825 MB/s
  → RMAN Backup – CIFS share (Mos-note: 444809.1)
- High # number of disks
  → Oracle database requires 34 TB
- Accelerated networking only available on DS15_v2
  → Backup Server
- Currently DS15_v2 only available and West Europe
  → with accelerated networking
Testing

Network tests
  • Simple ping + TCP tests
Local disk performance - 2 tests:
  • 1 storage account + 34 disks
  • 34 storage accounts + 1 disk
Layout:
  • 1 volume group
  • XFS striped lvm
Testing

Network tests - normal network

2.77

0.75
Testing

Network tests - ACN
Testing

Test case 1: Windows 2 windows

SMB Server Write MB/sec

Network Adapter Received MB/sec

The journey of a bi-stack to the cloud
Testing

Test case 2a: Linux 2 windows - Non Accelerated Networking
Testing

Test case 2b: Linux 2 windows - Accelerated Networking

![Graph showing SMB Server Write MB/sec with and without Accelerated Networking for Linux](image)

<table>
<thead>
<tr>
<th>Counter Sample Number</th>
<th>No Accelerated Networking for Linux</th>
<th>With Accelerated Networking for Linux</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>2</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>3</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>4</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>5</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>6</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>7</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>8</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>9</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>10</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>11</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>12</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>13</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>14</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>15</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>16</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
<tr>
<td>17</td>
<td>0 MB/sec</td>
<td>0 MB/sec</td>
</tr>
</tbody>
</table>

The journey of a bi-stack to the cloud
Testing

The journey of a bi-stack to the cloud

DATA

PRODUCTION/STANDBY

VM1

VM2

VM3

AZURE West-Europe

DATAGUARD

DB VM

DB VM

DB PRD1

DB PRD2

DB PRD3

DB PRD4

DB PRD5

DB CM

DB VM

Snapshots

REFRESH (Non-Prod)

ACCDB1

DEVDB1

UATDB1

TESTDB1

ACCDB2

DEVDB2

UATDB2

TESTDB2

ACCDB3

DEVDB3

UATDB3

TESTDB3

NFS Server Archive Data

AZURE West-Europe

Folder per DB
The journey of a bi-stack to the cloud

Physical I/O - Read MB/s

upd 0%
upd 10%
upd 20%
upd 30%
Physical I/O - Read MB/s
Testing

Backups

- Prod db
  - Protected by Dataguard
  - Bi-weekly full backup
    - Week 1 -> Big PRDDB1 backup
    - Week 2 -> Medium PRDDB2 / Small PRDDB3 backup
  - Retention
    - 1st week is on premium storage available via share
    - Week 2 -> 4: cfr azure offload
Testing

- 3 backup servers in AS + ACN, but in Separate AS than the db servers
- 1 share per server
- rman
  - 4 channels per share
  - tuning parameters needed:
    - sql 'alter system set "_backup_disk_bufcnt"=20 scope=memory';
    - sql 'alter system set "_backup_disk_bufsz"=1048576 scope=memory';
    - sql 'alter system set "_backup_file_bufcnt"=20 scope=memory';
    - sql 'alter system set "_backup_file_bufsz"=1048576 scope=memory';
    - CONFIGURE DEVICE TYPE DISK PARALLELISM 4;
    - CONFIGURE CHANNEL DEVICE TYPE DISK MAXPIECESIZE 20G MAXOPENFILES=40;
- Maximum backup speed achieved was the vm limit (450MB/s) , so throttling will be needed.
- Restore speed is about 400MB/s per oracle server
Expect the unexpected

Managed disks
Expect the unexpected

The journey of a bi-stack to the cloud
Expect the unexpected

- Managed disks are now globally available.
- Simple management
- Snapshots possible

→ Recommended!

Treat them as cattle, not as pets
## Provisioning

### Full provisioning Scenario

<table>
<thead>
<tr>
<th>Component</th>
<th>Source State</th>
<th>Target State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New: target VM</td>
<td></td>
</tr>
<tr>
<td>Storage</td>
<td>Exists: source storage</td>
<td>Exists: target storage</td>
</tr>
<tr>
<td></td>
<td>New: target storage new VM</td>
<td>New: target storage target VM</td>
</tr>
<tr>
<td>Data</td>
<td>Exists: source db</td>
<td>Exists: source db</td>
</tr>
<tr>
<td></td>
<td>New: target db</td>
<td>New: target db</td>
</tr>
</tbody>
</table>

### Refresh Scenario

<table>
<thead>
<tr>
<th>Component</th>
<th>Source State</th>
<th>Target State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Machine</td>
<td>Exists: source VM + db (standby)</td>
<td>Exists: target VM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storage</td>
<td>Exists: source storage</td>
<td>Exists: target storage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New: target storage target VM</td>
</tr>
<tr>
<td>Data</td>
<td>Exists: source db</td>
<td>Exists: source db</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New: target db</td>
</tr>
</tbody>
</table>
Provisioning

1. STDBY DB COMPONENTS
2. ACC DB COMPONENTS
3. ACC DB PRE-REFRESH SCRIPTS
4. SHUTDOWN ACC DB
5. UNMOUNT ACC FILESYSTEMS
6. SHUTDOWN STDBY DB
7. SNAPSHOT STDBY DISKS
8. DETACH ACC AZURE DISKS
9. SNAPSHOT -> AZURE DISK
10. ATTACH ACC AZURE DISKS
11. MOUNT ACC FILESYSTEMS
12. STARTU/RECOVER ACC DB
13. ACC DB POST-REFRESH SCRIPTS

The journey of a bi-stack to the cloud
Initial Dataload

The journey of a bi-stack to the cloud
Initial data load

• AZ copy

• express route

• No snowball / truck /… (yet)

And we needed it …
Initial data load

So we created it ourselves
Initial data load

• Lowest Lead Time for production cut-over Oracle Databases
• Phased approach to spread workload
• Tested and validated through dry-runs to assure production move
• One-off approach : good enough solution/cost
Initial data load

Our solution? Dataguard!
Initial data load

- Full Production backup – **Physical Transfer**
  When ? Migration Day – 6
  What ? 50 TB of three databases
  How ? Disk copy and Courier Transfer

- Incremental Production backup – **Physical Transfer**
  When ? Migration Day-2
  What ? 15 TB of incremental changes of 5 days
  How ? Disk copy and Courier Transfer

- Incremental Production backup – **Network Transfer**
  When ? Migration Day
  What ? 6 TB of incremental changes of 5 days
  How ? ExpressRoute Network Transfer
Initial data load
It’s all about avoiding limits
LIMITS:
• Azure BLOBS have a throughput limit of 60 Mbytes / sec
• Azure Storage Accounts : # : 250 per subscription
• Max ingress per storage account (EU) : 5 Gbps (GRS/ZRS), 10 Gbps (LRS)
• Max egress per storage account (EU) : 10 Gbps (GRS/ZRS), 15 Gbps (LRS)
• Standard Storage Account : 20.000 IOPS, 800MB/sec
• Premium Storage disk P20 : 2300 IOPS, 150 MB/sec
• Premium Storage disk P30 : 5000 IOPS, 200 MB/sec
• VM DS15_V2 : Disk : 64.000 IOPS, 640 MB/sec , Network : 15 Gbps
• VM GS5 : 80.000 IOPS / 2000 MB/sec, , Network : 32 Gbps
Lessons learned

• Get people out of their comfortzones!  
  —> old procedures, lots of manual work.

• People get uncomfortable by change.  
  —> No decent service catalog so every exceptional 
      situation is executed.

• Trust but verify the facts
Lessons learned

• Check! What?! Double check
• Fail early and correct.
• Involve the vendor
• Ask for previews, document and Test them!
Lessons learned

Expect the unexpected
• project manager got fired
• oracle changed the licensing rules

!!! But we delivered !!!
Would I do it again?

The journey of a bi-stack to the cloud
Mail:  pieter.vanpuyymbroeck@exitas.be
Twitter:  @vanpupi