

Server Team Process Right Sizing Virtual Machines



July 2, 2015

Table of Contents

1	Right Sizing Virtual Machines	3
1.1	Action or Event that Requires the Procedure	3
1.2	Goals of the Procedures	3
1.3	Roles that Participate in the Procedures:	3
2	Right Size VMs: Procedure Flow and Narrative	4
2.1	Right Size VMs Procedure	4
2.2	Right Size VMs Narrative	5
2.3	Right sizing compute standards	7
2.4	Right sizing for priority 1 & priority 2 disk space alerts	9
2.5	Right size for priority 1 & priority 2 disk space alerts narrative	9
2.6	Communication	10

1 Right Sizing Virtual Machines

The procedures in this document cover Right Sizing Virtual Machines (VMs).

1.1 Action or Event that Requires the Procedure

The Right Sizing procedures are implemented on a monthly basis. This is done to provide agency partners an opportunity to control virtual machine (VM) operating costs in which they are charged for.

This document includes the following procedures:

- Right Sizing VMs

1.2 Goals of the Procedures

The goals of the procedures are to:

Control compute (CPU, memory, disk) resources across the State Private Cloud to prevent excessive costs added to the overall service.

- Alert application owners of either compute excess or compute shortages based on trended VM usage.
- Gain agreement to adjust compute resources.
- Adjust compute resources.
- Verify that the compute resources were adjusted correctly without incident to the application.

Success is defined as:

- Adjust compute resources on a regular basis for the different VMs residing in the State Private Cloud.
- Server Management Team completes quarterly audit of overall process to validate effectiveness, efficiencies, accuracy, and adherence.

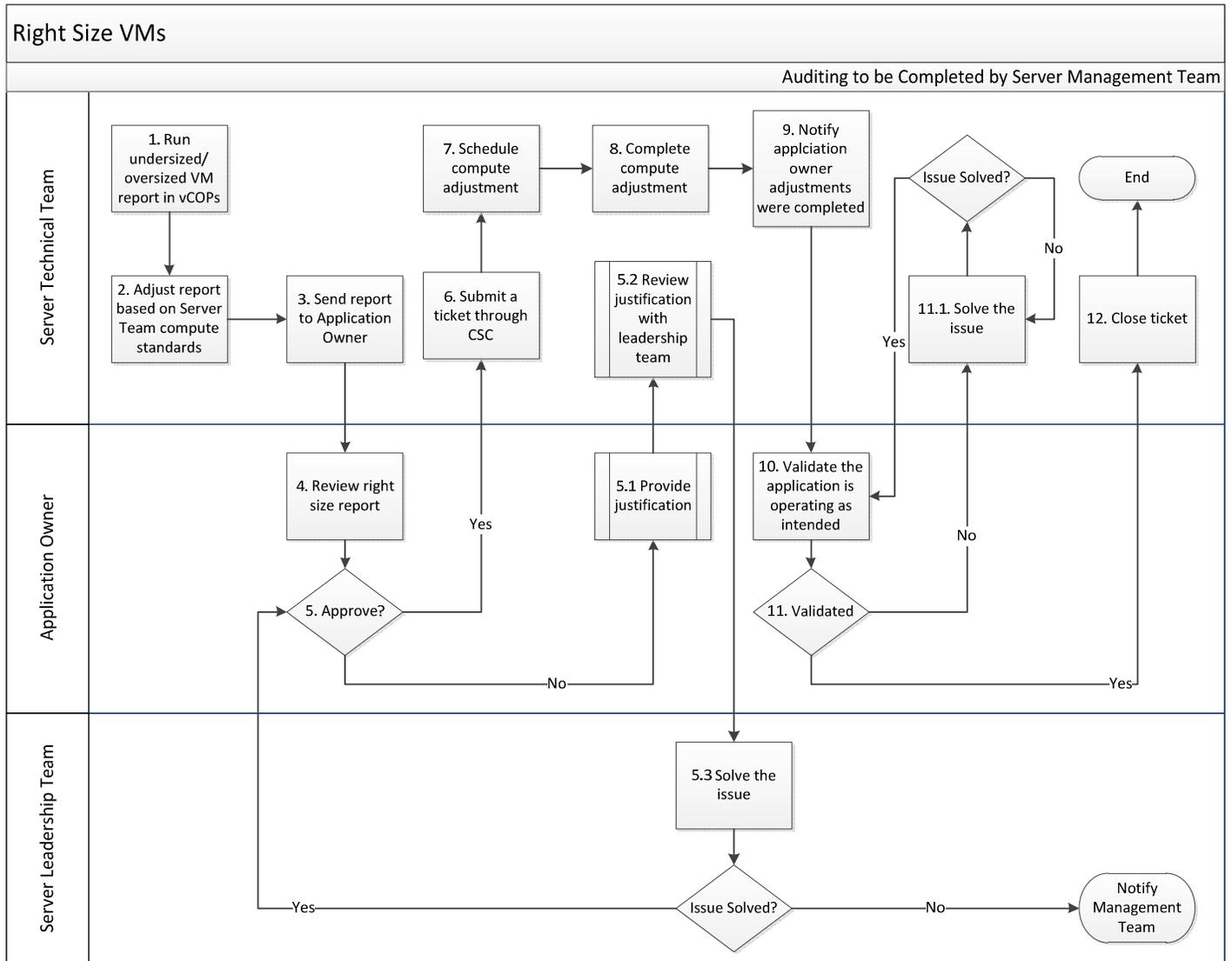
1.3 Roles that Participate in the Procedures:

- The Server Leadership Team
- Server Technical Team
- Agency application owners

2 Right Size VMs: Procedure Flow and Narrative

2.1 Right Size VMs Procedure

The following flow illustrates the interactions between the server team to Right Size VMs.



2.2 Right Size VMs Narrative

The following table describes the steps for Right Size VMs.

Table 1: Right Size VMs

Role	Step
Server Technical Team	1. Last Friday of the month, run undersized/oversized VM report in vCOPs. Create undersized/oversized VM report from vCenter Operations Manager, based on agency. Note: We will need to work with the Application Owner on right sizing their database servers separately.
	2. Adjust report based on Server Team compute standards. Consolidate reports using the Server Team template and apply appropriate standards and costs.
	3. First Tuesday of the month, send report to Application Owner. Use applicable distribution group.
Application Owner	4. Review right size report. Review information about recommended compute adjustments, cost impact, service impact/disruption.
	5. Approve? The application owner may approve or reject the recommendations. Note: Application Owners must provide a justification for rejected recommendations. 5.1 Provide justification to Server Technical Team. 5.2 Review justification with Server Leadership Team. 5.3 Resolve the issue with original recommendations, if unable to do so notify Server Management Team.

Right Sizing Virtual Machines

Role	Step
Server Technical Team	<p>6. Schedule resource adjustment Adjustments are scheduled based upon agreed upon maintenance window for the particular virtual resource. The application owner identifies contact information for their staff who will be available to verify after the change.</p> <p>Note: If adjustment poses no service disruption the compute adjustment may be completed outside of the normal maintenance window.</p>
	<p>7. Submit a change ticket in service now. Upon approval of recommendations Right Sizing Team is to submit a standard change ticket indicating approval and time for change.</p>
	<p>8. Complete compute adjustment.</p>
	<p>9. Notify Application Owner that adjustments were completed. Notify the Application Owner that the approved adjustments were completed as instructed via phone call.</p>
Application Owner	<p>10. Validate the application is operating as intended. Complete regression testing as necessary to validate compute adjustments pose no negative impact to application.</p>
	<p>11. Validated? Is the change validated as not impacting the application?</p> <ul style="list-style-type: none"> a. Yes, go to step 12 b. No, go to step 11.1 11.1 Technical Team works with application owner to solve the issue, escalates if needed.
	<p>12. Close ticket. Server Technical Team closes ticket upon confirmation from Application Owner.</p>

2.3 Right sizing compute standards

The compute standards form a baseline of minimum resources required for an operating system and do not take into account application needs. The right sizing effort will walk the resources down slowly to minimize the impact on the applications. The following table describes the current compute standards for right sizing.

Memory compute standards

Configured Memory	Recommended Memory (VROPS)	DAS/OIT Memory Right Sizing
2 GB	1 GB	Leave at 2 GB

Beyond 2 GB of memory, recommendations will be rounded up to the nearest whole number. (i.e. 3 GB, 4 GB, etc.). Please note for excessively large amounts of provisioned memory discussions may be held due to a need to align memory allocations with NUMA performance standards.

CPU compute standards

Configured vCPU	Recommended vCPU	DAS/OIT CPU Right Sizing
2 vCPU	1 vCPU	1 vCPU

Beyond 2 vCPU, recommendations will be rounded up to the nearest whole even number (i.e. 4 vCPU, 6 vCPU, etc.).

Right Sizing Virtual Machines

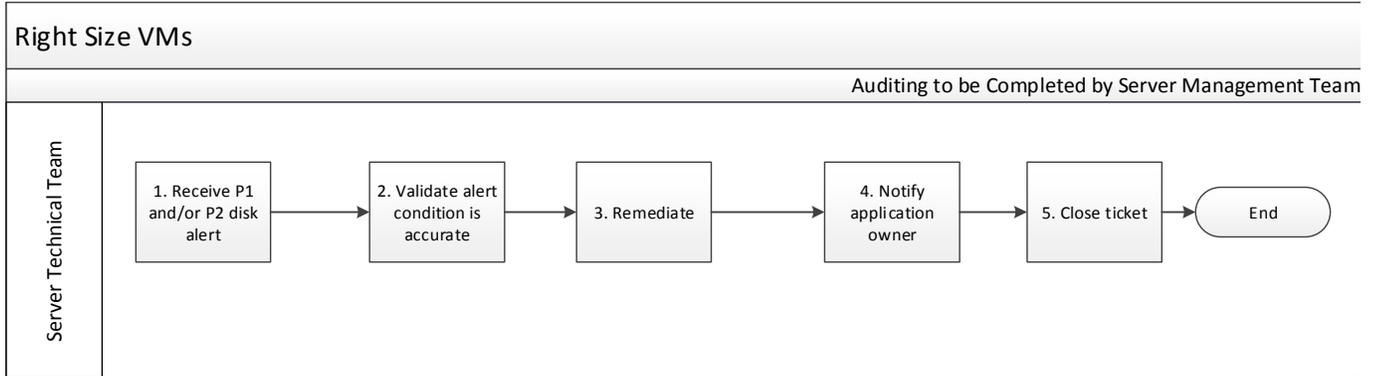
Windows Disk space standards

Warning	Critical	Fatal
90% used and less than or equal to 95% used and less than 5GB free	Greater than 95% AND less than or equal to 98% used AND less than 5GB free	Greater than 95% AND less than or equal to 98% used AND less than 5GB free
Priority 4 Incident	Priority 2 Incident	Priority 1 Incident
Work with customer to correct	Resolve to under warning levels and notify customer of action taken	Resolve to under warning levels and notify customer of action taken

Unix/Linux Disk space standards

Warning	Critical	Fatal
/: Greater than or equal to 90% AND less than 95% used /home: Greater than or equal to 90% AND less than 95% used /opt: Greater than or equal to 90% AND less than 95% used /tmp: Greater than or equal to 99% used /usr: Greater than or equal to 99% used /var: Greater than or equal to 90% AND less than 95% used	NOT "/var" or "/" AND Greater than or equal to 90% used	/: Greater than or equal to 90% used /var: Greater than or equal to 90% used
Priority 4 Incident	Priority 2 Incident	Priority 1 Incident
Work with customer to correct	Resolve to under warning levels and notify customer of action taken	Resolve to under warning levels and notify customer of action taken

2.4 Right sizing for priority 1 & priority 2 disk space alerts



2.5 Right size for priority 1 disk space alerts narrative

Role	Step
Server Technical Team	1. Receive P1 and/or P2 disk alert Priority 1 and/or P2 incident opened in ServiceNow via server monitoring tools.
	2. Validate alert condition is accurate Server team to respond to incident and verify condition is valid and requires action.
	3. Remediate Grow disk space on server to place server below warning levels. Verify server operating system sees the new disk space.
	4. Notify application owner Send notification to application owner explaining condition and corrective actions that occurred. Inform that if further remediation action is required to notify server team for additional assistance.

Right Sizing Virtual Machines

Role	Step
	5. Close ticket. Server Technical Team closes ticket upon confirmation from Application Owner.

2.6 Communication

First Tuesday of the month the server technical team will contact the application team with right sizing recommendations. These recommendations should be reviewed by the agency application team and business unit for approval and validation.