CENTER OF EXPERTISE FOR ENERGY EFFICIENCY IN DATA CENTERS

BERKELEY LAB

# Guidance on Identifying Zombie Servers in Data Centers

**June 14, 2024**

This work sponsored by FEMP

**Magnus Herrlin, Ph.D.**

Energy Technology Researcher

Lawrence Berkeley National Laboratory

One Cyclotron Road

Berkeley, CA 94720

**Federal Energy Management Program (FEMP)**

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Federal Energy Management Program, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

**Table of Contents** **Page**

## 1. Introduction

Data centers are an important part of modern society, and most computer servers in these facilities are active and productive in storing, processing, and transmitting data. However, some are “zombies”. There are quite a few definitions of zombie servers, but a common thread is a server that consumes electricity but serves no useful purpose. That seems to encapsulate with a few words what a zombie does and does not do, although it still leaves room for interpretation. A zombie server, also known as a comatose or a ghost server, is often, but not always, a server that runs at very low utilization. Not only physical servers can be zombies, but also virtual servers. This document guides the important task of identifying zombie servers.

## 2. Defining Zombie Servers and Utilization

**How do zombie servers come into existence?**

Zombie servers come into existence for many reasons.

- Excess capacity is installed to guard against future needs
- Excess capacity is installed for redundancy purposes
- Decommissioning of older servers may not have high priority
- A server may no longer be needed for its original purpose but is still powered on
- When workloads are moved to the Cloud, old servers are left in place
- A server may be left running with a single application that is difficult to migrate
- Servers may be left idling because of poor inventory and maintenance practices

Zombie servers are a big problem in data centers since they drain many resources. Multiple studies show that up to 30% of all servers in any given data center may be zombies. This is true for both physical and virtual servers. Physical servers are tangible, standalone hardware that runs an operating system and host applications. Virtual servers, on the other hand, are software-based instances of servers running on a physical host machine.

**Why is it important to run servers at high utilization?**

Utilization is, in general, the act of using something effectively. More specifically, utilization can be defined as actual performance divided by max performance. Note that utilization can mean *power* utilization or *computational* utilization. The rationale for running servers highly utilized stems from the fact that energy efficiency degrades at lower utilization levels. There is a minimum power draw in idle mode (= zero throughput). For modern servers, this is around 10% to 30% of full power.

The figure below shows a compilation of data from SpecPower (2023) incorporated into the U.S. DOE IT Equipment Simulation Tool (LBNL, 2024). We analyzed the comprehensive SpecPower server data and created three generalized server power performance curves based on commonly found 2U servers in data centers. The computational utilization is here shown as a function of power utilization.

The dotted line shows the reduction in energy efficiency at lower power utilization levels for the mid-range power performance servers (not particularly bad, not particularly good). The solid thin line illustrates a perfect server for comparison purposes. In this example, a server at 10% computational utilization still consumes 35% of maximum power. Since the server is not doing much, if anything, productive work at that level, this energy consumption is wasted. The server also continues to be a load on the cooling system adding to the energy waste.

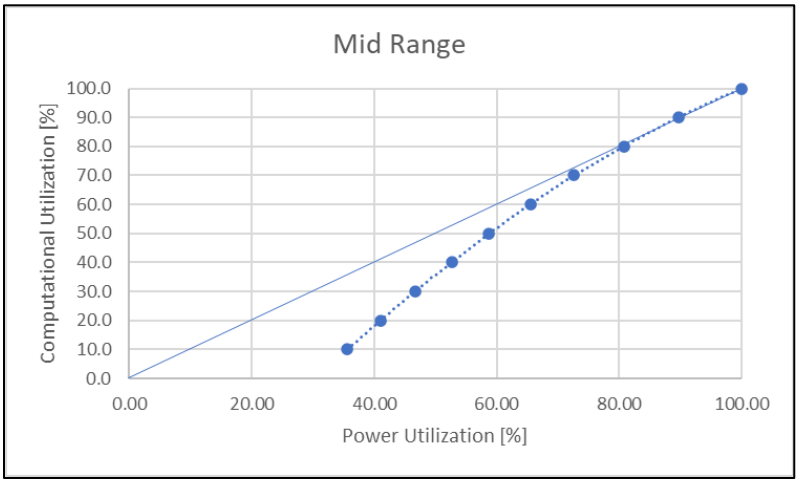


Figure 1. Server power performance curves.

Furthermore, identifying and decommissioning zombie servers saves resources other than the energy to run the servers and associated cooling equipment. Examples include reduced carbon footprint, fewer security risks, freeing up data center and rack space, and less need for new servers. Zombie servers are vulnerable in many dimensions since they receive less attention. In summary, zombies lead to unnecessary costs for the data center. Savings due to the elimination of zombies can add up quickly.

**How can we ensure high utilization?**

One powerful technology to ensure high server utilization is virtualization, which automatically consolidates applications onto fewer highly utilized servers. The remaining "empty" servers can then be decommissioned or reassigned. Strictly speaking, this applies only to under-utilized servers (doing little) rather than zombie servers (doing nothing). However, there is a significant gray zone between the two types. Manual utilization optimization efforts are only fruitful in small data centers.

Before we can decommission idling servers, however, we need to identify the zombies.

## 3. Identifying Zombie Servers

Identifying zombie servers is not simple since the server inventory is often inadequate. Adding to that challenge is that zombie servers can be forgotten or abandoned. As usual, the devil is in the details. Zombie servers are surprisingly hard to find since the utilization does not tell the entire story.

So, it is not always as simple as looking at server utilization. Some servers can be used for light processes that result in low utilization. On the other hand, servers executing maintenance processes like virus scans or software updates, often stress the CPU and memory. Such servers appear active but serve no useful purpose, and are therefore, per definition, zombies.

We will now address two important scenarios.

A) <u>Low utilization does not necessarily mean a zombie server.</u>
   Servers with low utilization could be kept at low loads for a particular purpose. A zombie or underutilized *physical* server can be decommissioned or reassigned. Identifying these zombies is relatively easy since their power consumption is constant (a flat line) over time. But zombie *virtual* servers are trickier. These server systems usually require a group of idle servers standing by for new workloads. Finding zombie servers is no longer as simple as finding servers that consume below a certain threshold power. A virtual server becomes a zombie when it's not utilized for its intended purpose over an extended period. That will occur if it is rarely or never activated, meaning that the server is not doing what it was supposed to do (no useful purpose, that is).

B) High utilization could mean a zombie server.
Here is another twist. A server might have high utilization but is not doing productive work or useful tasks contributing to the organization's bottom line. There are usually many overhead processes in servers like virus scans, operating system (O/S) updates, backups, security scans, O/S rot, stalled processes, unused applications, etc. One can argue that some of these processes are necessary. However, we did not invest in the server, maintenance, license fees, and data center resources to support virus scans. These overhead processes were not for the business purpose of acquiring the equipment.

We need to look at utilization and determine how much of the utilization is business-purposeful (useful) or just overhead (non-useful). The point is that we are not trying to maximize computational utilization, per se. We want to maximize the useful utilization which can be defined differently in different organizations.

Even simple computer systems have built-in tools for analyzing what makes up the computational utilization. Have you ever wondered why your laptop gets slower the older it gets? Take a peek at the utilization makeup and you may find the answer.

Server underutilization or utilization caused by overhead processes drives IT waste. Servers may run at a fraction of their processing capability, data storage may be oversized and infrequently accessed, and network traffic may be well below max transfer rates. Monitoring power and computational utilization is the first step toward identifying zombie servers, keeping in mind the purpose of the server. Is it intended as a physical server or a virtual server? As we have seen, physical servers should have high utilization made up of useful work. Virtual servers may not be zombie servers even at very low utilization levels.

The remaining part of this guide outlines a few ways of identifying zombie servers. We will cover the following steps:

- Keeping an inventory
- Collecting relevant data
- Using analytics

## 4. Keeping an Inventory

An inventory of all servers in the data center is a prerequisite for knowing the purpose of each server (intended use), which will help determine whether a server can be classified as a zombie. A thorough inventory may also include server age, IP address, name, operating system, hardware configuration, applications installed, and ownership. Ideally, the inventory should also

include information about usage patterns for each server. The difficulty in creating the inventory depends on the size and complexity of the data center. Small data centers may also have challenges tracking the inventory due to a lack of human resources. Establishing and keeping an inventory up to date can be a time-consuming activity with low priority.

Besides aiding the process of identifying zombie servers, an inventory serves other important purposes in data centers. It provides critical information for managing these critical environments, including the life cycles of electronic equipment.

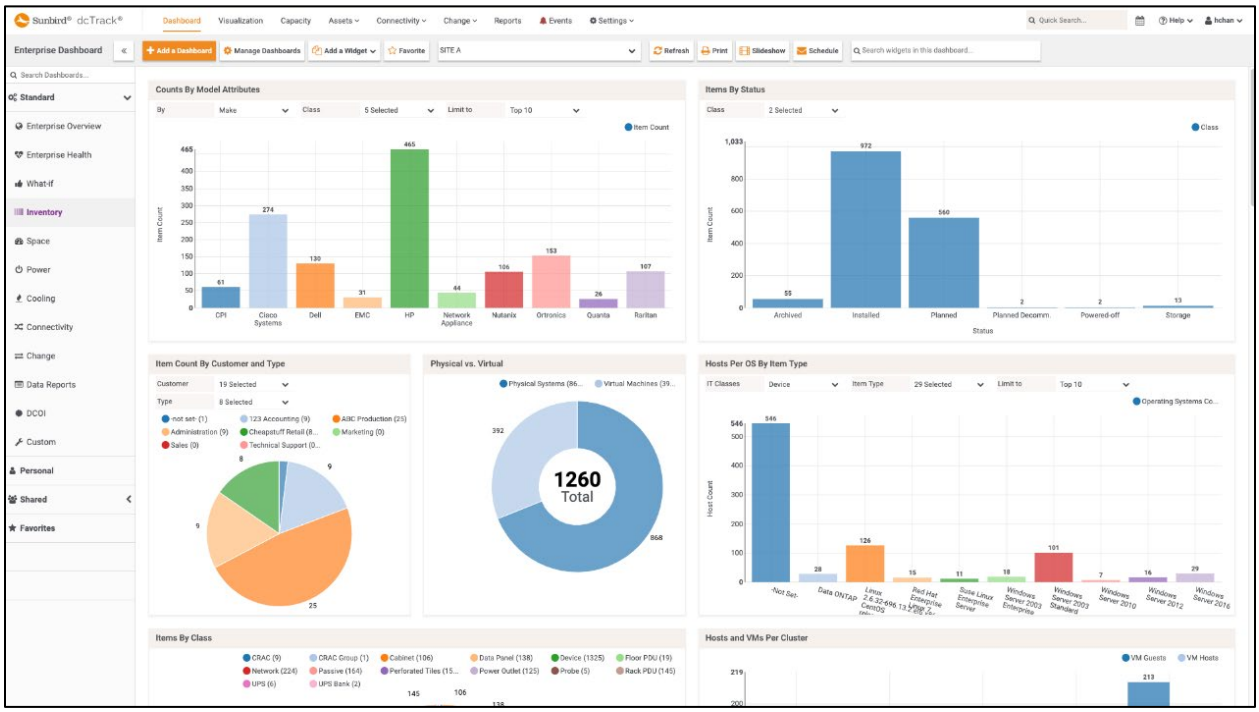


Figure 2. Inventory screen in Sunbird Data Center Infrastructure Management (DCIM) software.

## 5. Collecting Relevant Data

There are three main ways of accessing physical data from a server.

- Power cord measurements, e.g., intelligent power strips
- Onboard sensors through the server's Operating System (OS)
- Onboard sensors through the server's Baseboard Management Controller (BMC)

Power cord measurements are external to the server. These measurements may produce accurate power data but it could require site labor unless automation is used, such as a DCIM system.

The placement of the measuring device in the power chain is of importance. The closer we get to the server the more accurate readings, which also allow measurements of each server. In-rack measurements are generally the best. Intelligent power strips (also called Power Distribution Units or PDUs) are often installed vertically in the server rack and each server is connected to its outlet. If we instead measure the power at a certain point up the electrical power chain, maybe at the UPS, the power data quality will decline.


Figure 3. Intelligent power strip (PDU).

Server onboard sensor data is internal to the server. All modern servers allow viewing power in real-time using their onboard, built-in sensors. This is typically the best option for collecting accurate data.

The onboard server sensors provide support for the server's operation. A list of sensors for energy efficiency work in data centers includes input power, processor (CPU) utilization, and intake air temperatures. We are primarily interested in the two first parameters. One significant benefit of moving away from using external sensors to internal sensors is the addition of processor utilization readings.

The LBNL document “Accessing Onboard Server Sensors for Energy Efficiency in Data Centers” provides a roadmap to accessing server onboard physical data (LBNL, 2021). In summary, a Baseboard Management Controller (BMC) is included in all servers for accessing physical server data. The BMC is accessed through a separate port on the back of the server. There are many benefits to accessing the data through the BMC rather than through the operating system (OS).

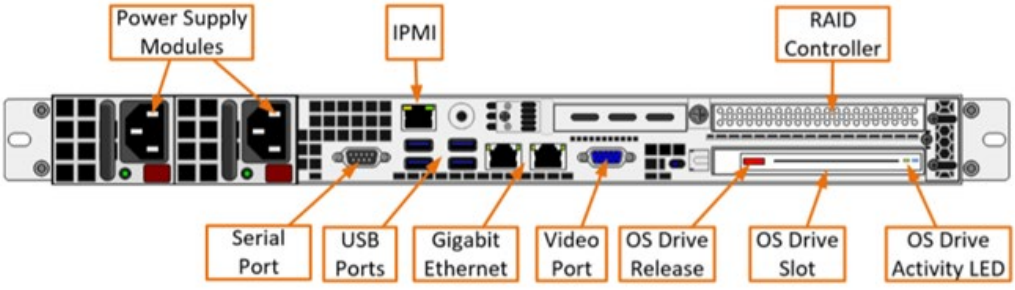


Figure 4. BMC (here labeled IPMI) dedicated management port.

After knowing the purpose of each server (physical or virtual), one effective way to find zombie servers is to use software that scans for power and computational utilization. Unfortunately, accessing the onboard sensors is often not understood (especially among smaller data centers) although many commercial data center management tools are readily available. Common management tools can be divided into:

- Data Center Infrastructure Management (DCIM) tools, focusing on the support infrastructure, such as cooling and electrical systems
- Data Center Networking tools, focusing on the IT equipment and IT connectivity
- Hybrid software tools may collect and analyze the physical health of many devices in data centers.

Over time, the line between these tools has become blurred. For example, many DCIM systems are now capable of receiving and analyzing IT equipment data besides the traditional infrastructure data.

Data Center Infrastructure Management (DCIM) tools collect, store, analyze, trend, and report real-time measured data from the physical infrastructure. DCIM tools can help identify zombie servers by monitoring power and CPU utilization. Some tools even have built-in zombie server reports. You can then quickly determine which of your servers draw less than a certain wattage over time. Comparing that with the minimum, maximum, and average power allows you to start identifying zombie servers. Your facilities group may already have this tool, especially for medium and large data centers.

Data Center Networking tools focus on IT equipment and IT connectivity. Network monitoring tools can help identify zombie servers by monitoring network traffic and identifying servers not sending or receiving data. Your IT department usually has this type of tool. This category also includes virtualization monitoring tools. If your data center uses virtualization, these tools can scan the virtual environment and detect virtual machines (VMs) that are zombies, i.e., servers powered on but not used for an extended period.

Hybrid software tools are positioned between the two main tool categories discussed above and often have a more targeted (limited) scope. These software solutions collect and analyze the real-time health, power, and thermals of many devices in data centers helping you improve the overall efficiency and uptime.

## 6. Using Analytics

We need analytics to determine if a server is truly a zombie. A relatively simple way is to look at power consumption, which is significantly easier to measure and monitor than CPU utilization. Many server racks have intelligent rack PDUs, which provide power data at the server level. An idling computer server often draws only a quarter of its maximum power. That in itself, however, does not guarantee a zombie server.

We can determine whether we are dealing with a true zombie by examining power consumption patterns. For example, knowing it is a physical server with a constant low power consumption indicates that we are dealing with a zombie. And, if we have a virtual server with low or no utilization over an extended time, it is likely to be a virtual zombie. In both cases, we can find zombie servers using automated tools - as outlined above - that scan the network for activity or non-activity.

To gain a complete and more accurate picture of zombie servers, it's sometimes necessary to use CPU utilization data. It will provide information about whether the utilization makeup is useful or not. A server serving no useful purpose over an extended period is a zombie although it does use significant power. This could be considered another "type" of zombie without the characteristic low power consumption.

Some suggest that Artificial Intelligence (AI) could help identify zombie servers, arguing that conventional human-crafted rules often don't work. AI is a complementary tool to such rules-based approaches. AI can be trained to become experienced to determine whether a server is a zombie. It can detect complex patterns. However, such discussion is well beyond the scope of this paper. The reference list includes information for those who are interested in this subject.

Zombie virtual servers have some typical characteristics that differ from their physical counterparts. These servers are powered on but are not used for a prolonged period. Of course, they may be inactive for shorter times since that is the nature of some virtual machines. Therefore, it is essential to study the server utilization over an extended time to determine if it could be a true zombie.

You don't want to turn off a virtual server thinking it is a zombie and then have the orchestration software looking for it unsuccessfully. This software (e.g., VMware) assigns workloads across servers. Besides using the built-in server sensors, you can also access computational utilization directly from the orchestration layer. For the software to operate successfully, a pool of idle servers must be available to draw upon. Therefore, declaring a server

a zombie requires more effort and involves analysis of both idle time and understanding why the orchestration software is not used. Data analytics can point to a zombie server.

Lastly, you also want to look out for accidentally creating zombies. It's important to track servers over time to prevent creating zombie servers. You should record the server purpose, status, and activity in addition to information about who deployed, owns, and uses the server.

Although finding zombies is not a simple task, these servers are an opportunity to boost your organization's bottom line. What to do with the newly found zombies to avoid decommissioning? One way is to give the servers more tasks to use the extra capacity. Or, you can repurpose the capacity elsewhere in your organization. A last example, although not attractive since it does not resolve the root cause, is to use power-saving modes to make the idle server consume less energy.

## 7. Creating an Action Plan

Drawing from the previous discussion, we can now assemble a step-by-step action plan in the hunt for zombie servers. It starts with creating *and* maintaining an inventory of all servers, it continues with studying the power utilization over time, and it wraps up with studying computational utilization over time and the makeup of that utilization.

- Step 1: Create and maintain an inventory.
  An inventory of all servers in the data center is a prerequisite for the remaining steps in our action plan. Without an inventory, it will be difficult to hunt for zombie servers. For example, to be able to distinguish physical servers from virtual servers is key. Not just that, an inventory is important for running the data center efficiently in general.

- Step 2: Study power utilization over time.

  A relatively simple way to find zombie servers is to look at power consumption and how it varies over time. We need to understand patterns of power consumption. A flatliner is a good indication of a physical zombie but not necessarily for a virtual zombie. An idling computer server often draws only a fraction of its maximum power. However, that does not guarantee a zombie server but is a first step in the right direction.

- Step 3: Study computational (CPU) utilization over time.

  To better understand if a physical server is a zombie server, it is sometimes necessary to use an analytical tool to explore the CPU utilization data over time and the type of

utilization. It provides information about whether the computational utilization makeup is useful or not. A server serving no useful purpose over an extended period is a zombie although it may consume significant power.

## Summary

Most computer servers in data centers are active and productive. Yet, some are of an entirely different breed called zombie servers. Not only physical servers can be zombies, but also virtual servers. They consume electricity but serve no useful purpose. They usually have a very low utilization, but not always since they can work on overhead processes. Furthermore, some servers appear to be zombies but low utilization could be by design, e.g., some virtual servers.

In the hunt for zombie servers, an inventory of all servers is a prerequisite for knowing the purpose of each server, which, in turn, will help us determine if they are true zombies. The next step is to access physical data from the equipment. It can be accomplished by power cord measurements or built-in server sensors for power and utilization. The data can then be funneled into various analytical tools to make the determination more effective. These tools examine patterns of different parameters to gain valuable information on whether we are dealing with a zombie.

Although identifying and eliminating zombie servers can be complicated and require specialized expertise, the extra effort is generally greatly rewarded as an improved bottom line for your organization.

**References**


Finding Zombie Servers - An AI-Driven Cloud Cost Optimization Solution, IBM

https://community.ibm.com/community/user/aiops/blogs/khalid-ahmed/2021/06/29/finding-zombie-servers

How To Attack Zombie Servers In The Data Center, Horizon Technology

https://horizontechnology.com/news/attacking-zombie-servers-in-the-data-center/

How to Find (Power) Zombies in Your Data Center, DELL

https://www.dell.com/support/contents/en-ph/videos/videoplayer/how-to-find-power-zombies-in-your-data-center/6293551522001

How to Find the Ghost Servers Haunting Your Data Center, Sunbird

https://www.sunbirddcim.com/blog/how-find-ghost-servers-haunting-your-data-center

LBNL, 2021. Accessing Onboard Server Sensors for Energy Efficiency in Data Centers

http://datacenters.lbl.gov/resources/accessing-onboard-server-sensors-energy

LBNL, 2024. The IT Equipment Tool

http://datacenters.lbl.gov/tools

SpecPower, 2023

http://www.spec.org/power_ssj2008/results/power_ssj2008.html

Three Ways To Locate Zombie Servers and Cables in Data Centers, powerwhips.com

https://powerwhips.com/3-ways-to-locate-zombie-servers-and-cables-in-data-centers/

Zombie Servers in Data Centers: What They Are and How to Find Them, LinkedIn

https://www.linkedin.com/pulse/zombie-servers-data-centers-what-how-find-them-kitty-chen/