Market Share


Paul Maguranis  Peter Rutten

IDC MARKET SHARE FIGURE

FIGURE 1

Worldwide AL4 Server 2019 Share Snapshot

Note: 2019 Share (%), Revenue ($M), and Growth (%)

Source: IDC, 2020
IN THIS EXCERPT

The content for this excerpt was taken directly from Worldwide AL4 Server Market Shares, 2019: Fault-Tolerant Systems Become Digital Transformation Platforms (Doc #US46640020). All or parts of the following sections are included in this excerpt: Executive Summary, Advice for Technology Suppliers, Situation Overview, Market Share, Who Shaped the Year, Market Context, and Market Definition sections that relate specifically to Oracle, and any figures and or tables relevant to Oracle.

EXECUTIVE SUMMARY

In 2019, the worldwide AL4 server market declined 5.6% compared with 2018 from $3.68 billion to $3.5 billion in large part because of lower average selling prices (ASP) for the platforms included. IBM dominates the market with 78.5% share, Oracle has 7.8%, and Hewlett Packard Enterprise (HPE) has 7.5%. IBM's share has declined somewhat; HPE, on the other hand, has picked up share; and Oracle has remained relatively steady.

Since the last published AL4 server market share, which covered the 2015 AL4 market, the worldwide market has declined from $5.2 billion in 2015 to $3.5 billion in 2019. In 2015, the $5.2 billion market was considered a "new baseline," registering 57.0% growth compared with 2014 because of the addition of three new platforms – IBM Power Systems, HPE Superdome, and Oracle SuperCluster. For 2019, Oracle SuperCluster is no longer included as Oracle ceased selling the platform when it stopped designing new SPARC processors. Also no longer included is Unisys Clearpath, which today is sold as firmware under the brand ClearPath Forward.

The computing platforms that are included in the AL4 market are:

- IBM Z
- IBM Power Systems E980
- Oracle Exadata Database Machine X8M
- HPE Integrity NonStop
- HPE Superdome family
- NEC ACOS-2 and ACOS-4
- NEC Express5800/FT
- Fujitsu BS2000
- Stratus Computer ftServer

AL stands for "Availability Level." There are four availability levels for servers – the higher the number, the higher the availability:

- **AL1**: Servers that have been designed without specific availability features
- **AL2**: Servers that ship with virtualization and workload balancing solutions to achieve availability
- **AL3**: Servers that ship with clustering software to facilitate failover to another node in a cluster (Some servers achieve AL4 availability with a combination of clustering software plus other technologies; they are included in the AL4 category.)
- **AL4**: Servers, often labeled as "fault tolerant" servers, that guarantee continuous processing under any circumstances with an extensive set of hardware RAS features and redundancy throughout the system
The purpose of AL4 systems is to address the following requirements, which are becoming increasingly important for businesses around the world (see the Market Context section for more detail):

- Support an increasing number of mission- and business-critical workloads
- Support increasing business transaction volumes and speeds
- Fend off all varieties of cyberattacks
- Provide a single source of truth (SSOT) for all data
- Provide round-the-clock data processing uninterrupted by scheduled or unscheduled downtime
- Prevent application downtime and its related monetary and reputational costs for the business
- Prevent or circumvent outages, delays, data loss, and data corruption
- Provide complete compliance with national and international regulations

Today, the percentage of all systems that need to be highly available is growing. IDC data (see Managing Server and Storage Availability in Enterprise Infrastructure, April 2019, IDC #US44988119, April 2019) shows that, in the financial industry, 23.5% of systems need to be highly available; in the retail and wholesale industry, this is 17.7%; in the healthcare industry, 16.5%; and in manufacturing, 16%. Across all industries, more than 60% of businesses have 21-30% of all their servers in the highest availability tier.

AL4 systems, which are also sometimes referred to as "legacy" systems, have proven themselves to be indispensable in businesses' digital transformation (DX). They have largely been de-siloed and modernized, allowing several of them to operate as open systems that run as a hybrid cloud leveraging open source software.

Furthermore, the homogeneous datacenter of the past, in which they stood out, is changing rapidly, with purpose-built platforms for various workloads now becoming the norm. As a result, AL4 systems are no longer considered the "sore thumb" in an otherwise homogeneous datacenter but are rather seen as one of those purpose-built platforms – the one where the most valuable data is securely processed with extremely high availability (HA) as part of an organization's on-premises/cloud infrastructure. These diverse platforms built for various workloads increasingly share a common, open source plane for application mobility and data sharing.

Please note the following:

- IDC categorizes a server as AL4 based on the platform's vendor-reported characteristics, not based on availability tests.
- IDC determines a system's market share based on the number of systems that have shipped in pairs to enable remote failover or that have shipped as a single unit that has been connected to an existing unit using vendor-designed software for failover.
- IDC no longer sizes or forecasts the market for servers that ship as AL1, AL2, or AL3.

Refer to the Market Context section for more detailed definitions of AL1-AL4.

This IDC study discusses the market for AL4 platforms, which remains strong at $3.5 billion worldwide. The siloed AL4 platforms of the past have been opened up and can now be fully leveraged for digital DX while the rest of the datacenter has been diversified and become more heterogeneous with purpose-built platforms for specific workloads, including those AL4 platforms.
"Developments of the past few years have been a perfect storm for the continuing importance of AL4 platforms. The platforms have changed, and the environment around them has changed, allowing high-value data on fault-tolerant systems to be fully leveraged for digital transformation," said Peter Rutten, research director at IDC’s Infrastructure Systems, Platforms, and Technologies group.

ADVICE FOR TECHNOLOGY SUPPLIERS

Remain Focused on Core Characteristics

The worse thing an AL4 vendor can do is throw the baby out with the bathwater, meaning compromising an AL4 system’s HA features to modernize the platform. AL4 vendors must continue to differentiate their platforms with technologies that ensure that high-volume transaction processing continues securely regardless of breakdowns, disasters, corruptions, cyberattacks, and so forth. To achieve AL4, these platforms are designed very differently than AL1–AL3 servers and they should be. At the same time, they are being augmented with many new technologies to transform them from previously siloed systems into true team players with the rest of the datacenter and the outside world. They need to support open source container-based cloud technology (e.g., enable modern programming languages, be fully API-enabled, support mobile and web). The latter is critical for future success, but this important transformation should not expose a platform to new liabilities. This makes modernizing more challenging but can at the same time result in exciting innovation.

Make Sure That Data Can Be Made Available

Making the data available to new customer-facing applications is what transforming AL4 systems is all about. The data that AL4 systems process is of the highest value, often with real-time aspects, SSOT implications, and in-flight security and data privacy considerations. AL4 systems are very good at processing that data type without interruption but have traditionally not been very good at securely making the data available for new, customer-facing applications, which today are the lifeblood of any enterprise. Some vendors in the AL4 space have made tremendous progress on this front, but not all, and there is still more to be done. DevOps with AL4 systems included is evolving but not optimal yet. Agile development is still slow compared with AL1–AL3 systems. The number of new apps and new feature releases are growing, but there is still an enormous amount of untapped potential. AL4 server vendors should focus on unlocking that potential so that customers can leverage the data that their platforms process in many more revenue-generating ways.

Automate with AI

AI automation for AL4 systems should be a top priority. AL4 requires operational complexity and the unique skill sets to manage that complexity are thinning out. AL4 vendors have been addressing this problem in myriad ways – from changing the GUIs of their platforms to training new generations of staff to reducing complexity through automation. The latter should get more attention. There is no reason why an AL4 platform should not be able to run nearly autonomously in 5–10 years using AI technologies for system analysis, prediction, decision making, and even physical maintenance through robotics.

MARKET SHARE

This IDC study provides an overview of the worldwide AL4 server market during 2019. Vendor revenue in the worldwide AL4 server market declined 5.6% year over year (y/y) to about $3.5 billion. Notably, 2018 was a historic year for the broader server market and makes for a difficult comparison in 2019. Shipment
volume has remained relatively steady while average price points have fallen as fault-tolerant systems have become more reasonably priced, which is largely reflected in the decline seen in 2019.

In 2019, IBM declined by 8.4% y/y to about $2.7 billion but still took the lion’s share of the market at 78.5%. The decline is largely due to IBM Power Systems underperforming in the market. Another factor is System Z – IBM launched the z15 platform in September 2019, providing a strong boost, with some potential to spill over into 2020. Another factor is that the new Z systems are becoming more affordable than in past years. Despite showing a decline at first glance, System Z performed well in MEA, Latin America, and the United States in the second half of 2019.

Oracle declined as well in 2019 by 8.2% y/y, with a market share of 7.8% totaling to $270.1 million. Oracle had a strong 2018 with Exadata but experienced a 6.9% y/y decline in 2019. Oracle’s SuperCluster has reached end of life and declined 25.3% y/y. 2019 revenue for SuperCluster consists mostly of support and maintenance for the enterprises that continue to rely on these systems for their mission-critical workloads.

HPE grew 21.8% y/y to $260.7 million, largely because of a shift in product mix. As businesses continue to shift away from the proprietary Itanium-based Integrity NonStop to the more affordable x86-based NonStop X, ASPs have declined significantly, resulting in savings for their customers and allowing HPE to capture some market share. HPE Superdome revenue, which has shifted away from the Superdome X model to the new Superdome Flex series, grew by 41.3%, while related shipments increased drastically by 111.8%. This is indicative that the new platform is much more affordable and becoming a strong driver of DX growth for their consumer base.

NEC and Fujitsu were able to grow revenue by 5.5% and an impressive 26.3% y/y respectively, while both tied for a market share of 2.5%. NEC saw declining revenues for their ACOS systems by 13.7% y/y, but this was offset by growth in fault-tolerant 5800 systems, which achieved 16.5% growth y/y. As for Fujitsu, its BS2000/OSD and GS lines grew by 1.6% and 46.2%, respectively. Stratus declined 4.8% y/y and attained a market share of 1.2% with the ftServer as Stratus is focusing its strategy on edge servers rather than AL4 datacenter products. Unisys has effectively exited the server market as its ClearPath Forward technology is now sold as firmware on any x86-based systems (see Tables 1 and 2).

### TABLE 1

**Worldwide AL4 Server Revenue by Vendor, 2017-2019 ($M)**

<table>
<thead>
<tr>
<th>Vendor</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Revenue ($M)</td>
<td>Share (%)</td>
<td>Revenue ($M)</td>
</tr>
<tr>
<td>IBM</td>
<td>3,404.6</td>
<td>82.6</td>
<td>2,978.7</td>
</tr>
<tr>
<td>Oracle</td>
<td>298.1</td>
<td>7.2</td>
<td>294.3</td>
</tr>
<tr>
<td>Hewlett Packard Enterprise</td>
<td>184.3</td>
<td>4.5</td>
<td>214.1</td>
</tr>
<tr>
<td>Other</td>
<td>235</td>
<td>5.7</td>
<td>194.8</td>
</tr>
<tr>
<td>Total</td>
<td>4,122.0</td>
<td>100.0</td>
<td>3,681.9</td>
</tr>
</tbody>
</table>
TABLE 1

Worldwide AL4 Server Revenue by Vendor, 2017-2019 ($M)

<table>
<thead>
<tr>
<th>Vendor</th>
<th>2017 Revenue ($M)</th>
<th>2017 Share (%)</th>
<th>2018 Revenue ($M)</th>
<th>2018 Share (%)</th>
<th>2019 Revenue ($M)</th>
<th>2019 Share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source: IDC, 2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 2

Worldwide AL4 Server Shipments by Vendor, 2017-2019

<table>
<thead>
<tr>
<th>Vendor</th>
<th>2017 Shipment</th>
<th>2017 Share (%)</th>
<th>2018 Shipment</th>
<th>2018 Share (%)</th>
<th>2019 Shipment</th>
<th>2019 Share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM</td>
<td>6,896</td>
<td>66.4</td>
<td>5,063</td>
<td>57.0</td>
<td>4,336</td>
<td>47.3</td>
</tr>
<tr>
<td>Hewlett Packard Enterprise</td>
<td>619</td>
<td>6.0</td>
<td>1,021</td>
<td>11.5</td>
<td>1,906</td>
<td>20.8</td>
</tr>
<tr>
<td>NEC</td>
<td>880</td>
<td>8.5</td>
<td>835</td>
<td>9.4</td>
<td>1,268</td>
<td>13.8</td>
</tr>
<tr>
<td>Stratus Computer</td>
<td>1,089</td>
<td>10.5</td>
<td>994</td>
<td>11.2</td>
<td>909</td>
<td>9.9</td>
</tr>
<tr>
<td>Oracle</td>
<td>852</td>
<td>8.2</td>
<td>937</td>
<td>10.6</td>
<td>712</td>
<td>7.8</td>
</tr>
<tr>
<td>Other</td>
<td>53</td>
<td>0.5</td>
<td>30</td>
<td>0.3</td>
<td>39</td>
<td>0.4</td>
</tr>
<tr>
<td>Total</td>
<td>10,389</td>
<td>100.0</td>
<td>8,880</td>
<td>100.0</td>
<td>9,170</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Source: IDC, 2020

WHO SHAPED THE YEAR?

In 2019, IBM and Oracle both introduced impressive new generations and HPE succeeded in gaining market share thanks to its Superdome Flex. IBM launched the IBM Z15, and Oracle launched the Oracle Exadata Database Machine X8M. IBM Power Systems had already launched a new enterprise-class product, the E980, in 2018, which contributed to IBM’s AL4 market share in 2019. This section provides a brief overview of these top systems in the AL4 market and some of their characteristics. The HA features listed here are extremely incomplete by necessity as all these platforms have a massive set of unique HA features, both in hardware and software.

Oracle

Oracle Exadata Database Machine X8M

The Exadata Database Machine is a combined compute and storage system. It is an engineered system designed to host the Oracle Database for numerous applications with very high performance...
and fault tolerance. Oracle launched the Oracle Exadata Database Machine X8M in 2019, with various new HA features.

Select HA Features

- **Oracle Exadata Database Machine X8M-8** and the X8M-2 have redundant networking, redundant power distribution units, redundant power supplies, and redundant database and storage servers.
- **Oracle Real Application Clusters (RAC)** is used to protect against database server failure while Oracle ASM provides data mirroring to protect against disk or storage server failures.
- **Oracle RMAN** is used for backups to Oracle's Zero Data Loss Recovery Appliance, disk, or tape.
- Oracle’s Flashback Technology is a feature that allows the system to roll back human inputs; essentially, this is a "go back" feature for databases that can function even at the row level of a database.
- Within an Oracle Maximum Availability Architecture (MAA) set up, Oracle Active Data Guard can be used to create a completely redundant, remote second site that will maintain a copy of the environment in real time.
- Exadata hardware and software integration allows for detection of failures within the hardware itself. The X8M uses RDMA-based detection, which Oracle claims can detect node failures in "subseconds," with the aim of unperceivable downtime.
- In the case of slow or faulty I/O read and write operations, Exadata System Software uses machine learning techniques to predict components susceptible to failure and takes proactive action to gracefully take such components out of service. Oracle Exadata System software will redirect the workloads to a healthy mirrored copy of the data.
- When disks or other hardware do fail, Oracle ASM will rebalance the workload and data from the faulty disks to other healthy hardware and, in some cases, even to a different storage server.
- Replacement components are hot swappable for disks, fans, power supplies, and PCIe flash cards. All monitoring and alerts of hardware performance is automated out of the box.

MARKET CONTEXT

The Transformation of AL4 Platforms

Today's dominant AL4 platforms have made major strides toward becoming fully integrated platforms in the datacenter that not only participate in an organization's digital transformation (DX) but in fact drive it. The most important reason is that these systems process many enterprises’ most critical and most valuable data, often with greater volumes than any other data types, and businesses have decided to unlock this data and leverage it for their digital transformation.

Until the mid-2010s, unlocking the data in an AL4 system without spending significant time and resources on moving it off-platform — and potentially compromising the data in the process — was nearly impossible. Businesses were unable to easily perform real-time analytics on their core transactional data; securely make that data available to applications outside the system with APIs, let alone in the cloud; safely combine that high-value data with other types of data from external sources; enable mobile and web access to the AL4 platform; or even run open source software, which had begun revolutionizing the software ecosystem. Basically, AL4 systems were siloed transaction monsters that would continue processing data during floods, earthquakes, or power failures, but that could not easily be used for running a cool consumer application.
Those days have passed. AL4 systems are still disaster proof, withstanding anything from tiny data corruptions to volcanic eruptions, but most of them are now fully DX friendly. In the last half decade, vendors of AL4 systems have spent billions of R&D dollars on securely de-siloing and transforming their products, making them hybrid cloud enabled, web enabled, mobile enabled, API enabled, open source enabled, and container enabled, just to name a few of the technologies that have opened these systems up to the rest of the datacenter and to the outside world. IBM Z, for example, sometimes still referred to as "the mainframe," runs OpenShift today. The emphasis here is on securely because technologies were required to solve the paradox of opening up the platforms while safeguarding the data, otherwise these systems would have lost their fundamental purpose.

Increasing Need for AL4

The fundamental purpose of AL4 systems is to address the following problems that organizations are going through on their DX journey, which ultimately focuses on the end user's entire digital experience as a fundamental measure of business success:

- IDC research shows that the top 3 causes of application downtime are failure in the network (16.2%), a failure in the servers (15.5%), and malware (10.3%). In other words, the most important aspects of IT infrastructure to improve are the network, the server platform, and security.
- The cost of downtime is increasing as businesses become more and more dependent on their infrastructure for daily operations. For 20.7% of organizations, the cost of downtime is $5,000-10,000 per hour; for 18.4%, it is $10,000-25,000 per hour; for 17%, it is $25,000-100,000 per hour; and for some businesses (1.4%), it is $500,000.
- Mission- and business-critical workloads are growing, and business support functions that previously could be run on a low availability tier are increasingly deemed business critical.
- Transaction volumes are increasing dramatically, and businesses need ever faster transaction speeds to satisfy their customers.
- Security characteristics of a system are critical as the number, sophistication, and severity of attacks on systems across datacenters and clouds has grown significantly.
- Single source of truth (SSOT) practices are becoming more important as DX drives businesses toward ever vaster and faster data processing, requiring data to be maintained in one place with only references to it to avoid duplication and incorrect representation of the data.
- The end of "regular business hours," with businesses' applications required to be available to customers at all times, has put tremendous pressure on the infrastructure that supports those applications, allowing for little if any scheduled or unscheduled downtime.
- Tolerance for outages, delays, data loss, and data corruption is zero – both from businesses and consumers – and any breaches or errors can have catastrophic consequences for an organization's reputation.
- As businesses engage digitally with consumers or citizens and with other businesses more often and in many more diverse ways, compliance with national and international regulations on data availability, security, and privacy is of paramount importance.
- Even as availability and security in the public cloud have greatly improved, true fault tolerance continues to be seen as an on-premises or hybrid cloud capability, not as a public cloud capability. IDC research shows that 38.5% of businesses hosts the highest availability tier on on-premises infrastructure, whereas only 2% of businesses host this tier in a public cloud.
Today, the percentage of all systems that need to be highly available is growing as a result. IDC data shows that, in the financial industry, 23.5% of systems need to be highly available; in the retail and wholesale industry, 17.7%; in the healthcare industry, 16.5%; and in manufacturing, 16%. Across all industries, more than 60% of businesses have 21-30% of all their servers in the highest availability tier.

The Rise of the Heterogeneous Datacenter

At the same time, another trend has started changing the datacenter, this one just in the past few years. Data is no longer considered to be a homogeneous concept. It is generated in many different ways, managed with a widening set of new and existing technologies, and has led to a remarkable diversity of application that leverage that data. This data heterogeneity has prompted a “purpose built” approach to the platforms that process the data, rather than the homogeneous infrastructure strategies from years past. IT is increasingly inclined to match a specific type of data with the optimal combination of hardware and software technologies. Training a large AI model, for example, is done on a cluster of tightly connected nodes, each with multiple GPUs; running S/4HANA is done on a scale-up system with extensive memory for the in-memory database. Some data types can be processed in a public cloud; others require on-premises processing.

IDC has seen infrastructure vendors — processor manufacturers as well as storage and server OEMs — embrace this “purpose built” approach for various workloads and their specific data processing and management requirements. This is relevant in the context of the AL4 server market because AL4 platforms foreshadowed this diversity of purpose-built, heterogenous data processing. For many years, they were exactly that: built for a specific data processing purpose, namely mission-critical Online Analytical Processing (OLAP) and Online Transaction Processing (OLTP) on relational database management systems (RDBMS). This emerging diversification in the datacenter is not threatening the mobility of applications though, thanks to today’s open source layers that provide a shared plane regardless of the underlying hardware, such as Kubernetes, OpenShift, or Gardener. These shared planes are also enabling AL4 systems to operate as a hybrid cloud.

Significant Market Developments

The most significant developments in the market in 2019 were:

- IBM launched the IBM z15 in 2019 with various new modernization characteristic and features for high availability.
- Oracle launched the Oracle Exadata Database Machine X8M in 2019 with various new HA features.
- Various cloud service providers have started targeting the customers in the AL4 market, trying to make a case that they can move some or all their mission-critical workloads to the public cloud. Cloud service providers consider this segment as the last on-premises holdout and as a significant competitive differentiation opportunity.

METHODOLOGY

The tables and figures in this study are generated from a proprietary IDC database and analytical tools. Our census process researches enterprise storage information on a product-, vendor-, and geography-specific basis. IDC is tracking server vendor results on a quarterly basis. IDC’s methodology provides our customers with a nearly unlimited ability to analyze the server market from
many perspectives. IDC’s continuous information services (CISs) and consulting services can provide additional insights beyond the scope of this study using the database supporting it.

Note: All numbers in this document may not be exact due to rounding.

MARKET DEFINITION

Servers Shipped as Highly Available

High availability can be achieved in a number of ways, including:

- **Hardware redundancy:** This approach takes the form of redundancy in select hardware components (processors and memory), input/output (I/O) paths, or totally redundant hardware systems (with redundancy for all hardware components) such as fault-tolerant systems that don’t stop, even in the event of hardware failure.

- **Analysis engine:** This feature reduces human error through predictive fault handling. It monitors resources continuously, predicts hardware faults, and initiates self-repair without operator assistance.

- **Use of reliability, availability, and scalability (RAS) features:** The use of RAS features, which are implemented in high-availability server hardware, means that instructions can be retried, even in the event of transient errors. For example, instructions can be retried with the processor, I/O can travel along an alternate path (multipathing), or memory can be mirrored in the case that one segment of memory needs to be used because another segment is corrupted.

- **Software support for HA failover of applications and data through the use of clustering software:** This approach involves the use of software that spans multiple servers, all of them linked by interconnects with equal access to the same shared data stores. These storage resources can be direct-attached storage, or they can be data arrays, storage area networks (SANs), or other types of shared data such as network file system (NFS)-based, network attached storage (NAS) filers.

- **Use of workload-balancing software:** This approach to achieving high availability has been used for many years, but it has accelerated with the wide adoption of virtualization in the x86 server world. These servers achieve high availability by shifting workloads from one server to another and allowing new requests to be redirected to alternate servers, if needed. Large arrays of small servers are often deployed in this way, achieving availability of applications and business services via redundancy rather than relying on built-in RAS features and resilience alone to improve availability.

- **True fault tolerance through redundancy in hardware and software:** This approach provides continuous data processing, even in the event of the failure of one hardware or software component. Because all data processing (i.e., support for a given application) occurs in a side by-side fashion, the failure of any single component in the system will not affect other sets of computing resources that also support that same application. As a result, computing can continue uninterrupted. However, in the event of a natural disaster or widespread power outage, the workload would still need to be replicated on alternate servers outside the campus, city, or geographic region that experienced the outage.

**IDC Availability Spectrum: AL1-AL4**

Table 3 outlines the four levels of availability as defined by IDC.
### TABLE 3

**IDC’s Availability Spectrum**

<table>
<thead>
<tr>
<th>Availability Level</th>
<th>Characterization</th>
<th>Impact of Component Failure</th>
<th>System Protection Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability level 1 (AL1)</td>
<td>Not shipped as highly available (no special provisions)</td>
<td>There is a need to switch to redundant resources before processing resumes.</td>
<td>No special protection for availability</td>
</tr>
<tr>
<td>Availability level 2 (AL2)</td>
<td>Workload balancing</td>
<td>Balancing may not be perceptible to end users because of retry.</td>
<td>User request redirected to alternate resources</td>
</tr>
<tr>
<td>Availability level 3 (AL3)</td>
<td>Clustered server</td>
<td>Short outage is needed for failover to take place.</td>
<td>User workload fails over to alternate resources</td>
</tr>
<tr>
<td>Availability level 4 (AL4)</td>
<td>Fault-tolerant server</td>
<td>The switch to alternate resources is not perceptible to end users.</td>
<td>100% component and functional resiliency</td>
</tr>
</tbody>
</table>

Source: IDC, 2020

In more detail:

- **AL1 (no special provisions):** The server is not designated as a highly available server when it is shipped by the vendor. Users always have the option to add availability software, following initial deployment and installation, which would convert an AL1 system into a highly available server.

- **AL2 (workload-balancing servers).** Workload-balancing servers are often deployed as arrays or grids of servers. As previously mentioned, these servers leverage redundancy in the form of multiple servers, any of which can be “tapped” to take over a workload, if needed. Examples include servers that are able to redirect tasks to other servers and servers that leverage the migration of virtual machines (VMs) — whether it is live migration or migration as part of disaster recovery after initial deployment. There is no formal “failover” involved in the redirection of work; rather, new requests are redirected to alternate resources on other servers as planned before any outage takes place.

- **AL3 (clustered servers).** This applies to all servers shipped with clustering software, which is available from a number of ISVs and system vendors. Examples include Microsoft Windows Failover Cluster (WFC), Veritas Cluster Server, HPE Serviceguard, IBM PowerHA, and Oracle Real Application Clusters (RAC).

- **AL4 (fault-tolerant servers).** At this level, the combination of multiple hardware and software components allows a near-instantaneous failover to alternate hardware/software resources so that business processing continues as before without interruption. The software components must be created by the original hardware OEM and not a third party. There must be one single point of contact for all service needs, hardware, software, and networking.

**RELATED RESEARCH**
▪ At THINK 2020, IBM Demonstrates How Red Hat Helps Drive Its Hybrid Cloud Strategy (IDC #US46348220, June 2020)
▪ Managing Server and Storage Availability in Enterprise Infrastructure (IDC #US44988119, April 2019)
About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com

Copyright Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or web rights.

Copyright 2020 IDC. Reproduction is forbidden unless authorized. All rights reserved.