Growing Demand for Warehouse-Scale Computing

Chetan Arvind Patil
Contributing Writer | EPDT

Warehouse-scale computing (WSC) describes an approach to building computing infrastructure. Here, the data centre is designed and operated as a single large-scale computer. The main unit of design is not one server, but a coordinated fleet of silicon-powered devices. These devices work together to deliver performance, reliability and predictable operation at scale, often spanning from 10,000 to over 100,000 servers.

AI processing requirements have further intensified demand for WSC by increasing both the scale and density of computation. Training large models and serving inference at production scale requires massive parallel processing, high-bandwidth memory access, plus tightly synchronised communication between nodes. As a result, WSC platforms are increasingly built around accelerator-rich systems and high-speed interconnects that can sustain continuous data movement.

As WSC systems scale, the platform becomes increasingly shaped by silicon constraints. Power delivery, thermal limits, signal integrity and high-speed data movement define the real performance envelope. This is why WSC has become a silicon-driven system problem, rather than a simple server scaling problem. At a warehouse scale, operational realities also change. Hardware failures are routine across large fleets, and compute density can concentrate rapidly within a single rack. The system must therefore maintain stable behaviour across compute, memory and networking silicon - operating simultaneously under sustained load, where power transients, thermal gradients and tight interconnect margins become decisive factors in performance and reliability.

WSC from a silicon point of view

From a silicon perspective, WSC is best understood as a multi-chip platform, not a single-processor story. Modern systems combine compute, memory, storage and networking silicon that must operate together as one product. Each block has a clear job to do, but performance becomes predictable only when the interfaces between them are solid, and the system behaves consistently under real workloads. In other words, the platform is only as strong as its weakest interaction point.

What makes WSC silicon especially interesting is how tightly coupled everything becomes at scale. Accelerators can deliver GBytes to TBytes of throughput, but only if memory keeps up and the network can move data fast enough to prevent stalls. At the same time, failure modes are rarely isolated. A marginal power condition on a computer device can appear to be a memory issue. A weak high-speed link can appear as a software instability. This is why WSC is not just about adding more chips, but about making diverse silicon blocks function as a coordinated machine.

From a silicon engineering standpoint, the biggest requirement for WSC is that each silicon category has its own definition of quality. Some issues show up as clean functional failures and get screened early. Others only reveal themselves as intermittent behaviour, link instability, or performance degradation during sustained load. Those are the failures that hurt warehouse platforms the most because they are difficult to reproduce and expensive to debug at scale.

This is why the WSC silicon qualification requires a layered mindset. It is not enough to prove that each chip meets its standalone specifications. The goal is to ensure the platform remains stable when these components interact at full speed across thousands of nodes for extended periods. In practice, a good WSC test strategy is the bridge between silicon capability and system-level reliability, reducing the risk that small margins can turn into large operational problems once the platform is deployed.

Silicon Building Block	Role in WSC Platforms	Test Engineering Focus
CPU	General compute/control.	Platform stability and correlation across configurations.
Accelerator (GPU or AI ASIC)	High throughput parallel compute.	Sustained correctness under load and sensitivity to power and thermal conditions.
HBM or DDR memory	Bandwidth and data supply.	Data integrity and interface margin stability.
NIC or DPU	Node connectivity and offload.	Link stability and interoperability under high speed IO conditions.
Switch ASIC	Fabric scaling.	Signal integrity behaviour and stable bandwidth operation.
Retimers and related high speed components	Signal conditioning.	Robust link margin and resistance to intermittent failures.
Storage controller	Data persistence and performance.	Integrity and predictable behaviour during long running operation.

Figure 1: The PAN9019 certified radio module containing dual-band WiFi 6 and Bluetooth 5.4 sub-systems

5V Operation in Any Condition

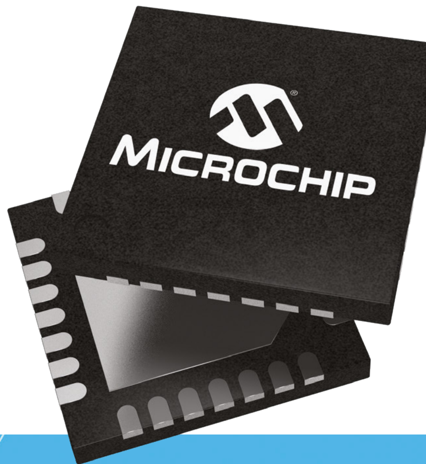
Efficient Performance Without Power Penalties

The PIC32CM PL10 microcontrollers redefines what's possible for engineers seeking an ideal balance between simplicity and performance. Built on the ARM Cortex-M0+ core, these microcontrollers deliver true 5V operation, rare among 32-bit MCUs, ensuring exceptional noise immunity for industrial, appliance and automotive applications. With advanced touch sensing, ultra-low power consumption and seamless support for familiar development tools, the PIC32CM PL10 bridges simplicity and performance with robust capacitive touch capability and reliable 5V support.

Key Features

- True 5V Operation: Robust performance in noisy environments
- Advanced Touch Sensing: Peripheral Touch Controller supports high channel count and resists interference
- Ultra-Low Power Modes: Sleepwalking and low standby current consumption extend battery life
- Easy Migration: Designed for 8-bit users to upgrade with no growing pains
- Familiar Development Tools: Compatible with Microsoft® Visual Studio Code (VS Code®) and MPLAB® Code Configurator, and supported by partner toolchains like IAR Embedded Workbench, Keil and Segger.
- Competitive Price Point: High-end features without premium cost

Upgrade your next design with the PIC32CM PL10 and experience 32-bit performance without the complexity.



MICROCHIP

microchip.com/pic32cm-pl10

The silicon testing aspect of WSC components

WSC changes silicon testing because the chips do not operate in isolation. Each component has its own specifications and manufacturing test requirements, but the system value comes from how these components interact. A platform can include devices that are fully functional on their own yet still deliver unstable results after integration if the interfaces are marginal, operating conditions shift, or behaviour changes under real workload stress.

Traditional semiconductor production test focuses on confirming that devices meet functional and parametric requirements at the chip level. This remains essential because it screens manufacturing defects, enforces specifications and supports high-volume shipments. In WSC platforms, however, passing the production test is only the 1st step, since system-level reliability depends on silicon behaviour under sustained utilisation and high concurrency.

The reason is practical. WSC environments push silicon to the limits of real operation and many failure mechanisms manifest as intermittent behaviour rather than clean pass-fail defects. Marginal high-speed links may train successfully in most conditions - but fail under specific temperature, voltage, or traffic combinations, while compute devices that pass standard patterns can still show instability when memory bandwidth is saturated, and power transitions occur rapidly. At a warehouse scale, low-probability issues stop being rare because the fleet is large enough for these edge cases to surface repeatedly across the system.

A practical WSC testing strategy proves quality in layers. Firstly, devices must pass production screening to enforce device-level compliance. Secondly, they must behave predictably after board and system

Figure 1: Silicon testing and WSC computing

integration, where margins shift. Thirdly, the platform must remain stable under sustained operating patterns that mimic deployment, because warehouse-scale reliability is ultimately defined by long-duration stability, low intermittent fault rates, and consistent performance.

How silicon testing for WSC will evolve

WSC platforms are becoming more heterogeneous, meaning more silicon types must come together into a single product and a single qualification flow. As CPUs, XPUs, HBM stacks, NICs, switch ASICs and high-speed retimers are integrated into a single platform, the qualification scope expands rapidly. Each silicon class has different specifications, operating limits and failure signatures, which means test engineering cannot rely on a one-size-fits-all approach.

One major challenge is correlation across test stages and environments. Production ATE screening verifies device-level specifications, but the platform still needs board-level validation and system-level qualification to confirm stability after integration. If ATE results, hardware behaviour and system validation outcomes do not align, teams either ship risk into deployment or spend excessive time compensating with redundant tests, repeated characterisation and conservative guard bands that reduce yield and throughput.

Another challenge is intermittent behaviour that sits at the boundary of specification and real-world operation. At a warehouse scale, low-rate issues become operationally significant, especially when they occur only under specific combinations of temperature, voltage, traffic and workload. These are not always caught by standard pass-fail criteria, so screening strategies must evolve to include stronger stress coverage, tighter monitoring of marginal patterns and validation that reflects realistic concurrency and utilisation profiles.

WSC will also continue to increase performance density and I/O complexity, pushing hardware closer to system limits. Higher bandwidth links, tighter timing margins and higher power density make platforms more sensitive to signal integrity, power integrity, thermal gradients and workload-driven power transients.
As a result, test engineering will place greater emphasis on proving robustness under these conditions, rather than merely meeting a static specification at a single operating point.

In the coming years, WSC silicon testing will increasingly be treated as a productization discipline that bridges specifications, validation and real application behaviour. The goal will extend beyond device-level pass criteria on ATE to proving that compute, memory and networking silicon behave reliably together in the integrated hardware stack, under sustained workloads that resemble deployment conditions.