GRC

THE IMMERSION COOLING AUTHORITY®

WHITE PAPER

Dealing with the Density Dilemma

How To Prepare Your Data Center Now for Next-Gen Application Architectures in AI, Microservices, and IoT.

11525 Stonehollow Drive, Suite A-135 Austin, TX 78758
+1.512.692.8003 • ContactUs@grcooling.com • grcooling.com

Data Center

Density Requirements Are **Changing**

Dramatically

A rapid, significant shift in data center infrastructure needs, like none other seen before, is being driven by the advancement of next-generation application architectures in AI, IoT, AR as well as microservices.[1] As a result, many hyperscale, high-performance computing (HPC), and bimodal data centers are already combining additional compute-dense co-processors with their current servers to boost mission-critical workloads.[2]

The electrical load of these additional co-processors plays a key role in accelerating the thermal design power (TDP) trend shown in Figure 1.

As kilowatt-hungry, next-gen application architecture continues to grow — and the thermal design power trend progresses, how can your data center framework keep up without significant expenditures in time and money? **This paper will help answer that question.**

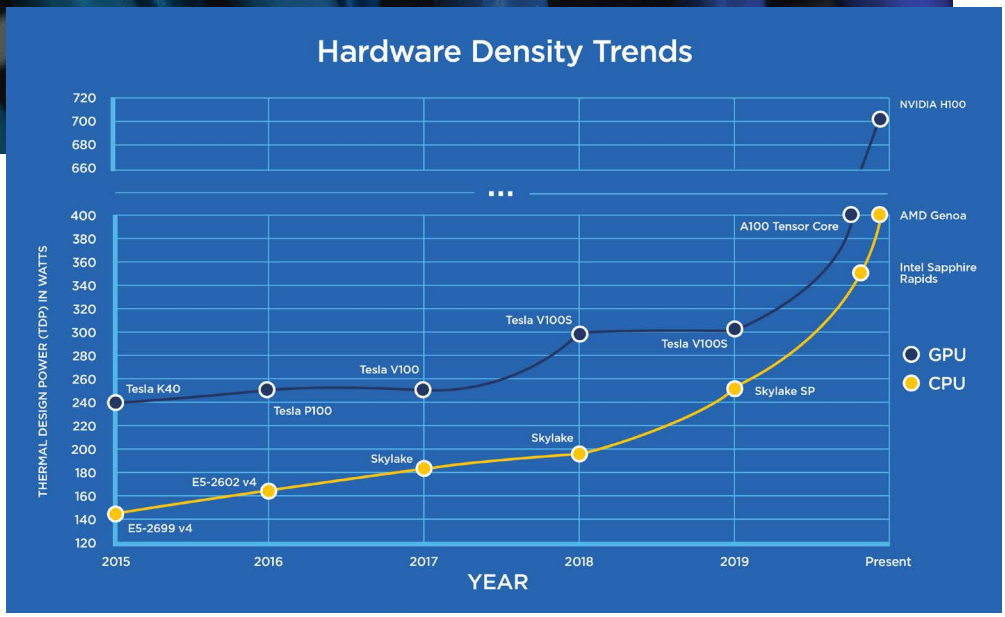


Figure 1.

Source: GRC Internal Research

[1] Gartner (April 10, 2018) 2018 Strategic Roadmap for Compute Infrastructure
[2] Gartner (October 30, 2017) IT Market Clock for Bimodal Compute Infrastructure, 2017.

The Density Dilemma Explained

Businesses are now seeking to "weaponize" their IT infrastructure via compute-intensive applications such as artificial intelligence, Internet of Things (IoT), machine learning, virtual reality, and faster streaming platforms for video and gaming. Running these new applications without latency requires ever more computing power. Often combining the raw power of GPUs with today's more powerful CPUs, a single 1U GPU-centric server can in fact drive 1500-2000 watts by itself.

But that's just part of the challenge. The denser the processors the more electrical power is required to operate them — and the more heat they generate. This can be problematic. Why? Data centers are typically designed on a basis of watts-per-square-foot, servers-per-square-foot, or square-feet-per-rack. Until now, floor space arrangements and electrical requirements for these designs have dictated cooling the processors with fans and additional air conditioners to keep the equipment running at an optimal temperature.

> Since more computing power means more heat, and more heat means more cooling, how do you integrate next-gen applications without disrupting your current infrastructure? Answers range from enhancing your infrastructure to building an entirely new facility — we'll share more about this later.

This increasing complexity of next-gen applications requires additional co-processors working side-by side with a computer's main CPU to accelerate particular crucial application algorithms.

The question before data center operators — both on-premises and co-location: "How do I facilitate this evolution in computing easily, at a low cost, and with no disruptions?"

Cold Hard Fact: Traditional Data Centers Can't Take the Heat

The necessity for more densely packed processors and co-processors presents a dual challenge for data center operators: these new compute requirements consume more power and generate more heat than most existing centers were designed to handle.

While it's true that operators must anticipate continued application advances, the scale and complexity of the newer compute-dense environment has, in many cases, far exceeded the 5-10 kW/rack threshold originally provisioned for traditional centers, in which heat is dissipated with inefficient and expensive air-based cooling technologies. Even newer cold plate solutions and rear-door heat exchangers (RDHx) are not up to the challenge.[3]

[3] Gartner (August 3, 2018) Integrate Cloud Into Your HPC Strategy with This Decision Framework.

This Brings Us to the Real Density Dilemma:

Since more computing power means more heat, and more heat means more cooling —

- How do you integrate next-gen applications without disrupting your current infrastructure?
- What are the pros and cons of various options?
- How quickly do you need to react?

Gartner suggests there will be a next higher "phase" of expected data center operations level in about two years.[4] Whether you're already ramping up, or struggling to get systems up to speed, you should balance agility with risk in calculating where and how dense your compute is going to be, as well plan for unlimited capacity plus density and location agnosticism. Most important, you should definitely be taking action now.

> High-heat, compute-dense processors, and co-processors quickly reduce the density dilemma to what is essentially a cooling issue.

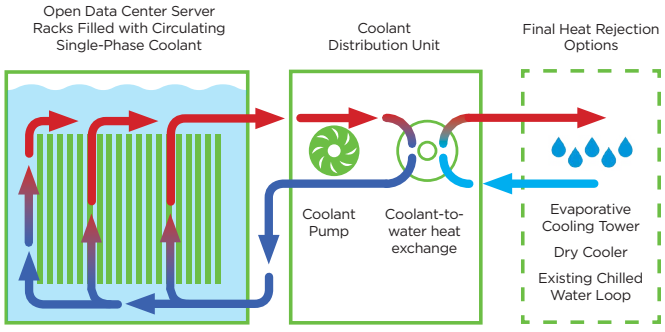
Fortunately, proven technology exists to solve the density dilemma while offering further optimizations to your current data center. That technology is called immersion cooling.

An Overview of Immersion Cooling Technology

With immersion cooling technology, servers are installed in horizontal racks filled with a dielectric coolant that is an excellent conductor of heat but not electricity. Coolant is circulated between the racks and a coolant distribution unit (CDU) connected to a warm water loop, which may use a cooling tower, dry cooler, or a chiller return line as the final form of heat removal.

The standalone system doesn't require chillers, CRACs, CRAHs, or raised floors, but can easily integrate with legacy infrastructure to cool over 100 kW per rack while delivering a PUE <1.03 in any climate. In addition to a reduction in cooling power for ITE of up to 90%, it also reduces critical (IT) load by 11%, through the removal of server fans and superior cooling performance.

How Liquid Immersion Cooling Works



Heated coolant exits top of rack. Coolant returns to rack from heat exchanger at user-specified temperature.

[4] Gartner (October 30, 2017) IT Market Clock for Bimodal Compute Infrastructure, 2017.

> "Within five years, ten on the outside, there will be no alternative to immersion cooling."
>
> – Dr. Satoshi Matsuoka, Tokyo Institute of Technology

Immersion cooling frees you to more easily accommodate HPC apps, saving you time, money, and essentially future-proofing your data center against the density dilemma.

Placing servers in liquid has also been shown to address many issues pertaining to data center growth, rack density, cooling capacity, data center design, hardware reliability, and location options. For this reason and many others, experts are calling immersion the next generation in data center cooling.

In fact, when asked about high-density environments, Dr. Satoshi Matsuoka of the Tokyo Institute of Technology has asserted that, *"Within five years, ten on the outside, there will be no alternative to immersion cooling."*

Immersion Cooling Is the Next Generation

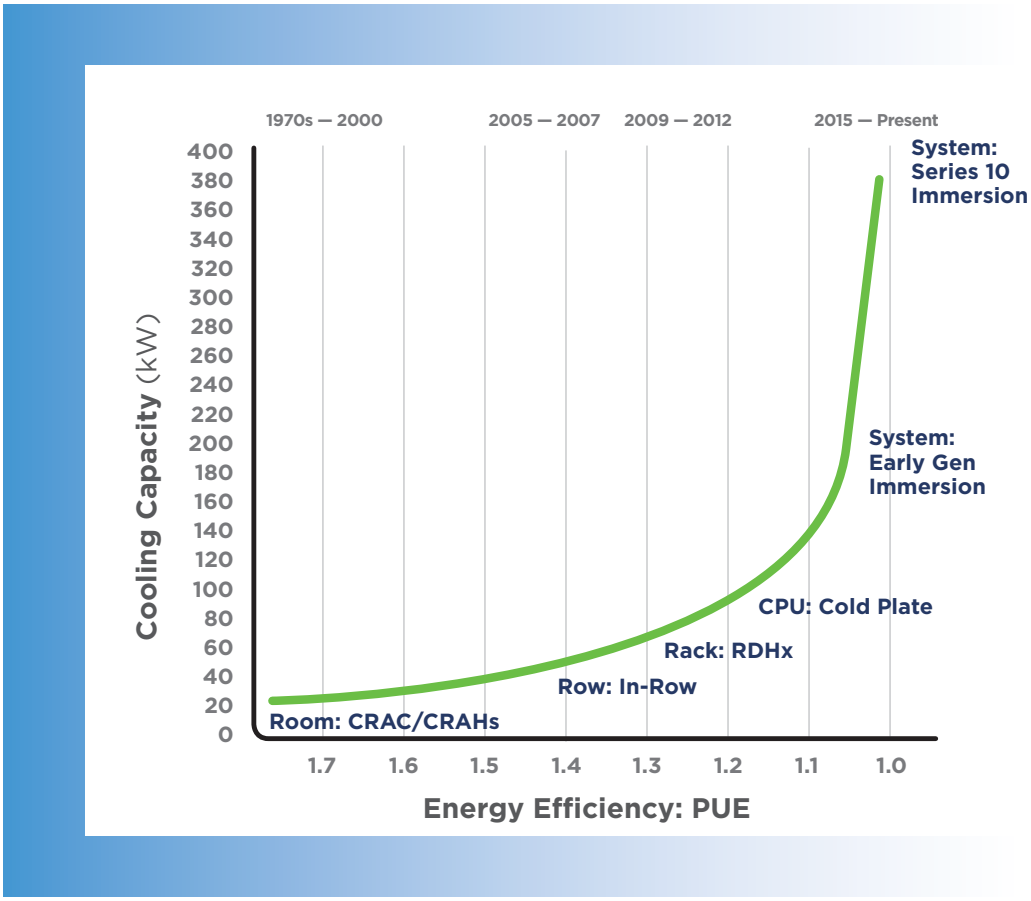


Figure 2.

Tackling the Density Dilemma: Four Scenarios

Given the inevitability of the compute density challenge, you may be compelled to handle the situation in a number of ways:

- Grow HPC capabilities on-premises, including:
  - Augmenting your legacy data center's capabilities with dedicated high-density zones
  - Adding capacity through new construction, expansion, or the addition of modular data centers for high-density apps
  - Completely rebuilding your existing facility
- Move operations to the cloud

As seen in Table 1 below, all these options have pluses and minuses with associated capital and operating costs that should be weighed carefully.

	Solutions	Pros	Cons
On-Premises	**Augmenting Your Legacy Data Center's Capabilities with Dedicated High-Density Zones**	Greater Control Most Cost-effective Resource Efficient (energy, space, and staffing) Short Lead Time Standalone/Non-disruptive Improved Utilization of Existing Facility and Assets Scalable	Disparate Infrastructure May Require Modifying O&M Practices Scalability Limited by Existing Space and Power
	Adding Capacity Through New Construction, Expansion, or the Addition of Modular Data Centers for High-Density Applications	Greater Control Non-disruptive Scalable	Space & Power Availability Cost-intensive Long Lead Time
	Completely Rebuilding Your Existing Facility	Greenfield Opportunity Total Design Flexibility	High CapEx Disrupts Existing Operations Requires Economies of Scale Longest Lead Time
Cloud	**Move Operations to the Cloud**	CapEx Savings Scalability Flexibility Risk Mitigation Least Disruptive	High OpEx Loss of Control: - Downtime - Security - Privacy Challenging to Migrate Applications Back On-premises

Table 1.

On-Premises Solutions

Overall, this approach offers flexibility plus the ability to maintain closer control of operations and security.

Whether you choose to augment your data center's capabilities with high-density zones through mixed racks/rows, add capacity through new construction, or completely rebuild, immersion cooling allows you to make your center HPC-capable without tossing the proverbial "baby out with the bathwater." Yet challenges do exist, most of which require A/C re-engineering.

Augmenting Capabilities with High-Density Zones

This trend toward HPC notwithstanding, AI and high-density applications are still a small fraction of most enterprise workloads, which explains the co-existence of high — and low-density hardware within some facilities, and why many operators are choosing to create "high-density zones" in their centers.

Yet, assuming high-density racks are allowed in your facility, you simply may not have enough physical space to achieve the proper compute density using conventional cooling methods. Or you may have maxed out your compute and not be able to draw computing power away from non-high-density applications. Furthermore, spreading hardware across racks or spreading racks out may not allow you to reach your target density (or anticipated density), will add networking complexities, will cost you valuable floor space. So will in-row cooling, which often requires 1:1 positioning, and can cut your floor space in half.

Setting up immersion-cooled, high-density zones gives you the ability to effectively leave your existing infrastructure unchanged while adding the advanced computing power you need to support these next-gen applications. Immersion can also help optimize your center's white space (compute) and lift constraints on grey space (supporting infrastructure).

Setting up immersion-cooled, high-density zones gives you the ability to effectively leave your existing infrastructure unchanged while adding the advanced computing power you need to support these next-gen applications.

Adding HPC Capacity Through New Construction, Expansion, or Modular Data Centers

Depending on current demands, breadth of mission and existing capabilities, you may feel compelled to create a dedicated HPC infrastructure outside your current data center (at the edge). Yet this approach could pose a significant increase in capital expenditures (CapEx) and operating expenses (OpEx).

Short of acquiring a new building or floorspace within the organization, modular immersion cooling solutions work very well for this scenario. These essentially provide a "data center in a box" and, according to Vertiv's Peter Panfil, Vice President, Global Power, "There is massive activity at the edge, and that will continue for the foreseeable future. It has been well documented that the fastest growth in our industry is happening at the edge and in hyperscale, and there are certain fundamentals that will continue to drive that trend." [5]

[5] https://www.datacenterfrontier.com/executive-roundtable/article/21438738/edge-computing-picks-up-steam

Because of latency concerns, high-density applications will drive more computing to the edge. Consequently, endpoint location and mobility issues will impact data center design decisions and add to the density dilemma. Modular immersion cooling solutions can address these concerns, too.

> "There is massive activity at the edge, and that will continue for the foreseeable future. It has been well documented that the fastest growth in our industry is happening at the edge and in hyperscale, and there are certain fundamentals that will continue to drive that trend."
>
> — Peter Panfil, Vice President, Global Power, Vertiv

Completely Rebuilding

With all the options previously discussed, it's evident that there are more viable alternatives to wholly rebuilding your data center. Immersion cooling allows you to grow your HPC capabilities flexibly and cost-effectively, and scale easily to keep pace with your operational growth plan.

However, if you do choose to start over, an immersion cooling infrastructure affords enormous advantages over traditional air-cooled schemes, slashing your CapEx up to 50%, and reducing cooling energy for ITE up to 90% — all while providing cooling capacity of up to 100 kW/rack.

Move to the Cloud

Because of its obvious cost-savings, flexibility, risk mitigation, and scalability, moving to the cloud might seem very attractive. Interestingly, some I & O leaders are currently using cloud AI portfolios for project experimentation and proof-of-concept trials before deciding whether to deploy solutions on-premises or the cloud.[6]

Yet, the cloud has drawbacks that may not make it a viable long-term solution. Ongoing costs of cloud services (OpEx) can snowball over time, negating the upfront cost savings in the long run. And apart from the challenges of adapting your applicationss to the cloud environment, migrating applications back from cloud to on-premises can be time- and cost-intensive, which effectively locks you in on the higher operating expense. Further, given the loss of control over infrastructure and the potential issues with security, uptime, and privacy, many businesses choose to make a strategic decision to keep mission critical apps on-premises.

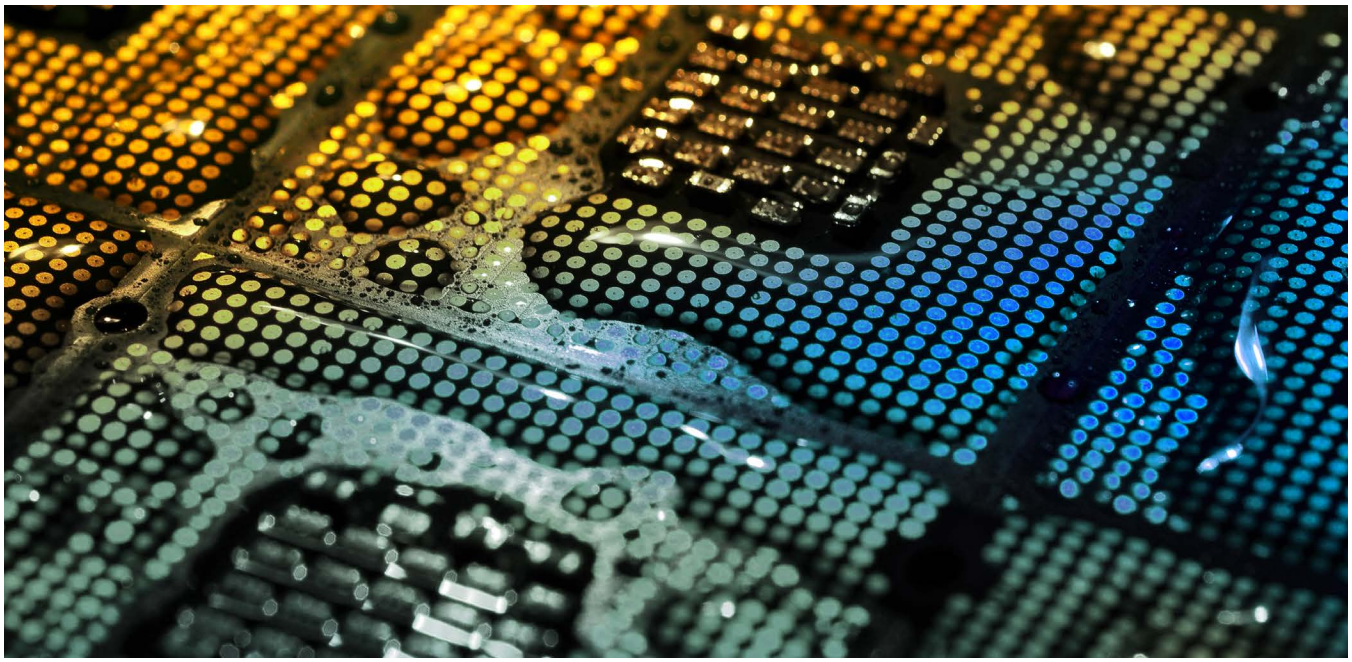
A Closer Look

Whether you're planning ahead or trying to catch up — and regardless of what expansion modes you may be contemplating, immersion cooling helps you solve the density dilemma.

Immersion Cooling Features

- Cool up to 368kW/Rack
- Lowers Upfront Costs by up to 50%
- Up to a 90% Reduction in Power for Cooling ITE
- Reduces Server Power Draw 11%
- Only 3% of Compute Load Required for Cooling

[6] Gartner (May 21, 2018) Market Guide for AI Portfolio Cloud Service Providers.



Immersion Cooling: The Solution to Your Density Dilemma — Delivered

The advance of high-density data center applications is inexorable. How I & O leaders accommodate it — whether through various on-premises solutions or in the cloud — depends on many factors pertaining to budget, vision and current infrastructure.

Regardless of the approach, immersion cooling offers many benefits that address not only this challenge but many others, enabling you to:

- Future-proof new or totally rebuilt data centers through a dramatic increase in compute density
- Grow compute density incrementally
- Easily create a high-density zone within a current site, with little to no impact on existing infrastructure
- Optimize data center white (compute) space and lift constraints on grey space (supporting infrastructure)
- Effectively cool different compute densities within the same data center

In addition, immersion cooling can help address the anticipated, rise of data center power costs, which are expected to climb 10% annually due to cost-per-kilowatt-hour (kWh) increases in underlying demand. For this reason, some data center systems are being pre-built with liquid cooling solutions, and experts expect the immersion cooling market to grow at a high rate through 2032.[7]

[7] https://www.futuremarketinsights.com/reports/immersion-cooling-market

About GRC

GRC is one of the pioneers of immersion cooling, and has been developing and enhancing the technology for nearly a decade and a half. Our first mainstream installation in early 2010 established our technology as the most powerful, efficient and cost-effective data center cooling system on the market.

Buoyed by endorsements from IT icons such as Intel® and the US National Security Agency (NSA), plus successful installations around the globe, we are now regarded as the world's Immersion Cooling Authority.

We offer both modular and micro-modular, rack-based immersion cooling solutions to help our clients overcome the density dilemma, as evidenced by one that is running an air-cooled data center at 20 kW/rack with two immersion-cooled 50 kW-density racks in the same room. This installation also repurposes exhaust water from an air-cooled system.

Look to GRC to help you master the new trend in high-density computing and truly future-proof your data center.

GRC
THE IMMERSION COOLING AUTHORITY®

Let GRC Help You Solve Your Density Dilemma
by Introducing Immersion Cooling into Your Data Center.

Call **+1.512.692.8003** • Email **ContactUs@grcooling** • Visit **grcooling.com**