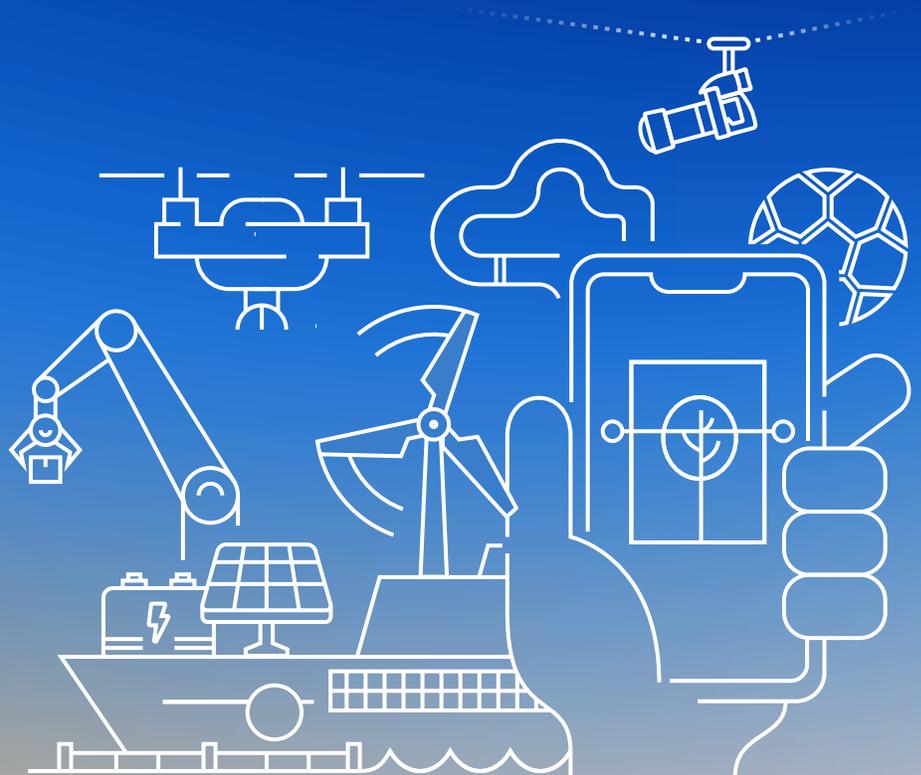




DECEMBER 2022

# Werner Vogels' Tech Predictions for 2023 and Beyond



Dr. Werner Vogels,  
VP & CTO at Amazon.com

# Contents

- Overview . . . . . 3
- Cloud technologies will redefine sports as we know it. . . . . 4
- Simulated worlds will reinvent the way we experiment . . . . . 6
- A surge of innovation in smart energy . . . . . 8
- The upcoming supply chain transformation. . . . .10
- Custom silicon goes mainstream . . . . .12

# Overview

With several global crises occupying our daily lives, it's important to see where we can leverage technology to solve these hard human problems. Today, we have more access to data from wearables, medical devices, environmental sensors, video capture, and other connected devices than we have had at any point in the past. When combined with cloud technologies, like computer vision, machine learning, and simulation, we are starting to get a glimpse of where that powerful blend of information and application can take us.

The next wave of innovators and inventors, a few of whom I was fortunate to have met while filming [Now Go Build](#), are already building solutions to reforest the planet, keep our youth active, and re-imagine the supply chain from the warehouse to delivery. And this is truly just the beginning. As access to advanced technology becomes even more ubiquitous—as every facet of life becomes data that we can analyze—we will start to see a torrent of innovation, and this will proliferate in 2023.

## PREDICTION ONE

# Cloud technologies will redefine sports as we know it

Like music and video, sports will become data streams that we can analyze. The insights that these will unlock in the coming years will transform the entire sports industry and redefine what it means to play—and experience—every game.

Sports are part of the human fabric. They transcend time, cultures, and physical borders. At this moment, one of the largest sporting events in the world is happening—the World Cup. An estimated 5 billion people are expected to watch. Until now, broadcast television has had the biggest impact on the evolution of professional sports, paving the way for what is today a \$500 billion industry. The next game-changing technology advancements are on the horizon. In the coming years, every facet of every sport will undergo a digital transformation, and this will happen at every level of play, from youth basketball to professional cricket.



Companies like [Veo](#) are leading this charge, making use of cloud technologies like machine learning, computer vision, and stream processing, to narrow the digital divide between amateur and professional athletes. Not only has Veo created a broadcast-like experience for viewers of amateur sports, but it has built a deep neural network that allows it to automatically create highlights from video streams. This allows players, coaches, and recruiters to easily find key plays, improve tactics, and share them in ways that simply weren't possible before. As technologies like Veo become more widely used across all levels of all sports, imagine what comes next.



Top leagues, like the Bundesliga and the NFL, have started using video streams, wearables, IoT sensors, and more, for real-time analytics and insights, but looking ahead, these capabilities will continue to advance, and the technologies will become an omnipresent force in nearly every sport and at every level. Imagine a scenario where a coach can use computer vision and biometric data that is analyzed in the cloud in real time to pull a player before they cramp or concede a goal, replacing them with the most well-rested teammate, something now quantifiable. This simultaneously improves player safety, and increases competitiveness of the game. At this point, the sports themselves will truly start to become a data stream that we can analyze and make decisions on in real time—player hydration, ball movement, field saturation—all of it, aggregated, and richer than anything we see today. And with more data comes further innovation. In the not-so-distant future, we will reach a point where teams are running constant what-if simulations in the background during every game, enabling them to better predict the impact of their decisions in the moment. Technology, itself, will become the competitive foundation for professional sports.

Whether in-person or on a screen, the fan experience will also change. Stadiums will rapidly adopt some of the innovations that we've seen in industries like retail, such as Amazon Go stores, where the use of computer vision, sensor fusion, and deep learning will enable ticketless entry and grab-and-go purchasing. We will also start to see the next generation of data overlays and real-time insights that go down to the player level, augmenting the game and bringing sports closer to what we expect in the most visually informative video games today. Co-viewing and personalized viewing experiences will continue to evolve, more closely connecting those 5 billion viewers than ever before.

The sports world is currently on the verge of the biggest revolution it has ever seen, and cloud technologies are at the center of this change.

## PREDICTION TWO

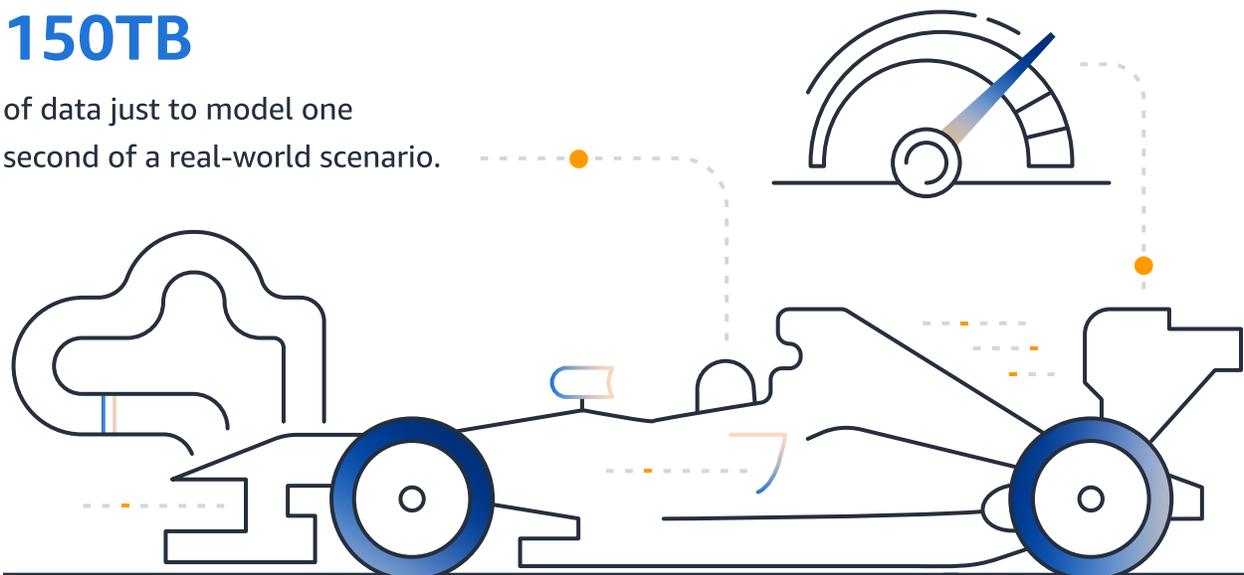
# Simulated worlds will reinvent the way we experiment

Spatial computing. Simulation. Digital twins. These technologies have been slowly maturing for years, but the everyday impact has been limited. This is quickly changing, and in 2023, the cloud will make these technologies more accessible, in turn enabling a new class of use cases that will be unbound by physical constraints.

Simulations are used to build better race cars, predict weather, and model the stock market. While the problems that simulations can solve are significant, the difficulty of building and running simulations is a barrier for everyday use cases. Companies are constrained by the need for high-powered hardware and a specialized workforce. Take a fluid dynamics simulation for a jet wing or race car as an example, where it may take up to 150 TB of data just to simulate one second of a real-world scenario.

## 150TB

of data just to model one second of a real-world scenario.





However, this is quickly changing with technologies like the recently launched AWS SimSpace Weaver, the first of many simulation technologies that will pave the way for a future where nearly anything in our world can, and eventually will, be simulated. Simulations will help us make better decisions about the roadways we construct, the ways we organize our warehouses, and the ways that we respond to disasters. With simulation, we can peer into the future to see the impacts of our efforts, running numerous what-if scenarios that answer our questions without having to wait and see what the impact might be many years down the line. With a technology like SimSpace Weaver, a company like [Terraformation](#) can model the growth of entire forests on their way to achieving the goal of planting 1 trillion trees. As a result, it can ensure a biodiverse, and healthy forest that has the most carbon offset possible.

Another area where I'm seeing a rapid uptick in innovation is spatial computing. Companies are already building specialized hardware and using cloud technologies to capture and create 3D models of nearly any environment. Doing this with just a mobile device will soon be a reality. This democratization will inspire a new wave of innovations in the architecture, construction, commercial real estate, and retail industries. Like video did for the internet, spatial computing will rapidly advance in the coming years to a point where 3D objects and environments are as easy to create and consume as your favorite short-form social media videos are today. Static 2D product images on the internet will become a thing of the past, replaced by 3D models that you can pick up, rotate, and place in your living room as seamlessly as you can see them in a web browser today. But expect more to emerge from these models, such that their intrinsic features can be simulated in your virtual home. A virtual lamp will not only be placed on the floor of your living room — you will be able to turn it on and off, watching how the ambient light interacts with your virtual furniture in real time, and understand the impact it has on your energy consumption. All of this, before ever pushing a “buy now” button.

In 2023, technologies like these will begin to converge. With the increasing integration of digital technologies in our physical world, simulation becomes more important to ensure that spatial computing technologies have the right impact. This will lead to a virtuous cycle of what were once disparate technologies, that begin to be used in parallel by business and consumers alike. The cloud, through its massive scale and accessibility, will drive this next era.

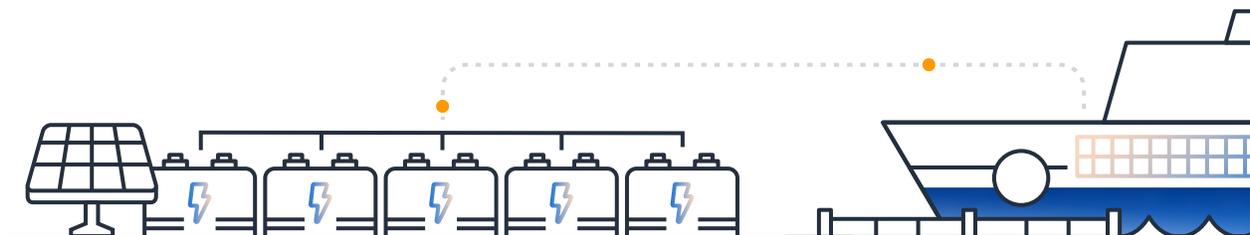
## PREDICTION THREE

# A surge of innovation in smart energy

Energy-storing surface materials. Decentralized grids. Smart consumption technologies. In 2023, we will see rapid development on a global scale that improves the way we produce, store, and consume energy.

We are in the midst of another energy crisis. Rising costs and reliable access to energy are global problems—they impact everyone. While this isn't the first time that we have faced an energy crisis, there are several maturing technologies that are beginning to converge, and together, they will enable us to address this like never before.

The environment around us produces more than enough renewable energy. The challenge is actually with storage and on-demand delivery to the systems that need to consume that energy. Amazon is doing work in this space, take for example the [150 MW battery storage system](#) in Arizona that's providing clean, reliable energy to our facilities in that area. But we're not the only ones. Companies across the globe are also quickly innovating in this space. The cloud is enabling materials research science for novel use cases, such as integrating energy storage into the structure of the objects they aim to power. Imagine a shipping vessel where the sides of the ship are actually the batteries that power it on its journey. This is just the tip of the iceberg—no pun intended. We are also starting to see breakthroughs in long-duration storage, like molten salt, stacked blocks, and fuel cells.





Another area is the decentralization of energy. With uncertainty around energy availability, some communities are turning to microgrids. I like to think of microgrids as community gardens (but for energy), where community members use these to sustain themselves, reducing their reliance on traditional energy companies and their aging infrastructure. In my neighborhood, we have a small microgrid, where solar is collected and shared among tenants. As we continue to see energy challenges amplified by geopolitical events and climate fluctuations, microgrids will become a viable solution for many communities around the world, and cloud technologies will play a role in enabling this. Data from solar panels, wind farms, geothermal, and hydroelectric power will be streamed, stored, monitored, enriched, and analyzed in the cloud. Machine learning will be used to analyze all energy data to predict usage spikes and prevent outages through redistribution of energy at a household-level of granularity.

We will also see IoT-based smart consumption devices really start to take off across the globe in the coming year. This will lead to the next wave of innovations that arise from the new observability capabilities that these devices provide for homes and businesses alike. Imagine the energy savings we can get by retrofitting historic buildings with energy saving technologies.

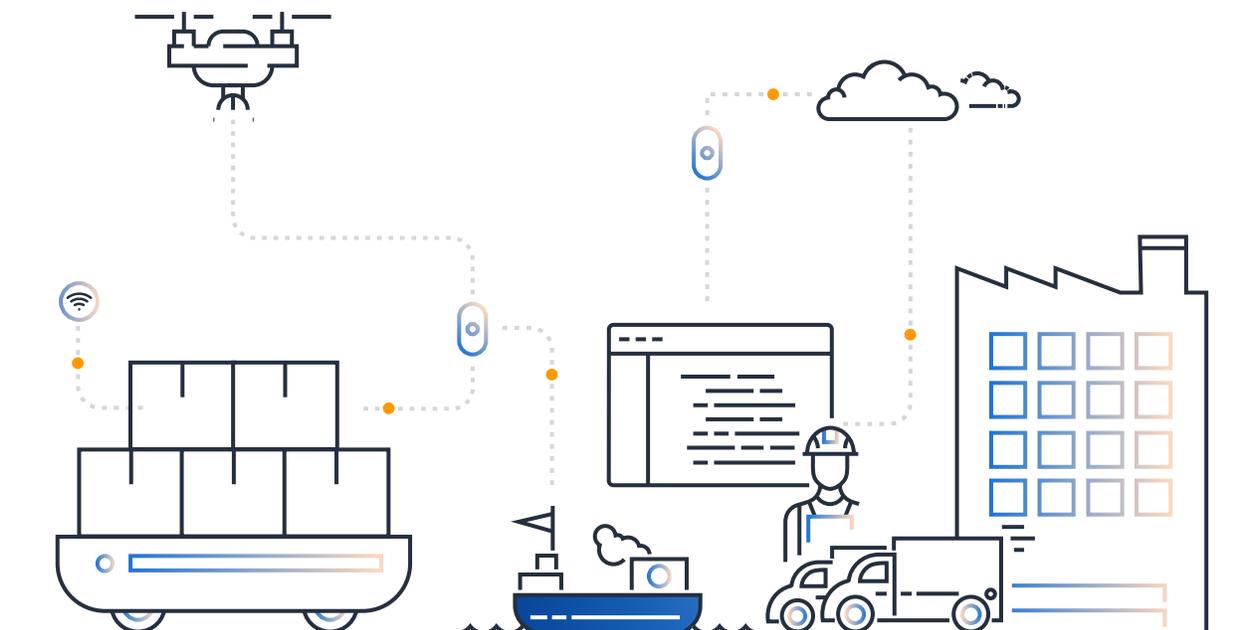
In the next few years, we will see a rapid convergence of all types of smart energy technologies, as we have finally met the threshold where our technology solutions can address our crisis. While this may not have the immediate impact that we all wish it would, together these technologies will, fundamentally and forever, change the way that we create, store, and consume energy in the future.

## PREDICTION FOUR

# The upcoming supply chain transformation

In 2023, adoption of technologies, such as computer vision and deep learning, will propel the supply chain forward. Driverless fleets, autonomous warehouse management, and simulation are just a few of the optimizations that will lead to a new era in smart logistics and global supply chain.

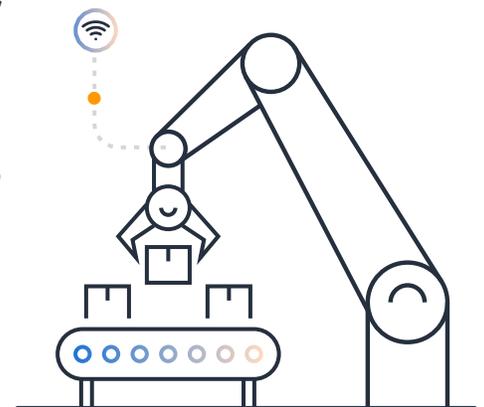
Something that I've reflected on regularly over the past few years is the fragility of the global supply chain. We are reminded of this daily—late deliveries, unavailable products, empty shelves. While Amazon has fine-tuned its supply chains with innovations, like digital freight matching and delivery stations, many companies have continued to struggle with logistical challenges. This is about to change.



This will start with the manufacturing of goods themselves. IoT sensors in factories will proliferate and machine learning will be used to not only predict machine failures, but prevent them. Less downtime means consistent production. Shipping those products across the globe is a whole other challenge. Digital freight networks powered by the cloud will traverse countries, even oceans, providing real-time data that allows carriers to optimize with the most efficient shipping routes and change course in response to inevitable events, such as equipment failures and weather disruptions. Think of it as having real-time insights about the current state and arrival time of goods, but at every level of the supply chain.

These freight networks will set the stage for the first cross-country autonomous truck shipments. The impacts will be felt immediately, with countries like the US currently experiencing a shortage of 80,000 drivers. Through the use of spatial computation, edge computing, and simulation, autonomous trucking is set to have a massive impact on our global supply chain. Why? A human driver can only spend so long behind the wheel before they become distracted, tired, and potentially dangerous. And this is before we consider each country's specific health and safety regulations. This means that fresh fruits being shipped from southern California can only hope to make it as far as Dallas before they begin to deteriorate. However, an autonomous truck can be on the road for 24 hours. There are no mandated breaks, and the technology never gets tired, impatient, or distracted. Products get where they need to faster, safer, and more efficiently.

Upon arriving at a local warehouse, robotic picking, order-sorting, and automated packing will become more commonplace. We will continue to see this evolve with new innovations in robotics that use AI, computer vision, and precision handling of individual products in a company's inventory. Autonomous robotics will also begin to play a bigger role in warehousing. Imagine being able to augment a forklift operator, who spends a good portion of time simply searching for products, with a real-time digital twin of the inventory, one that is constantly kept up-to-date using autonomous flying inventory drones.



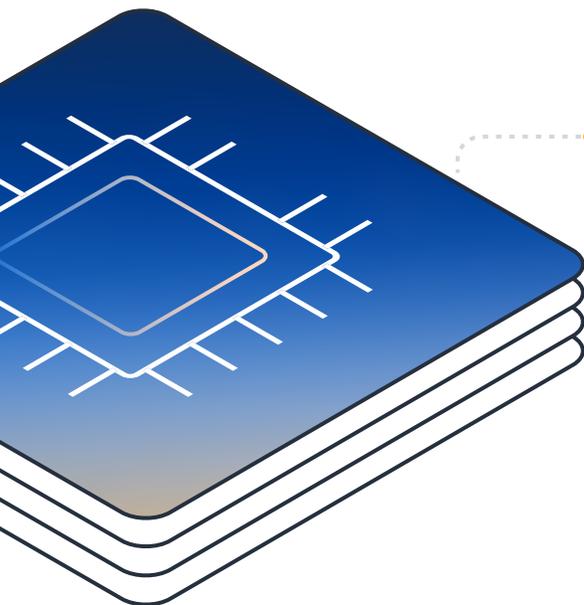
The key to transforming the supply chain is to use technology to optimize each step along a product's journey. Starting next year, we will see an acceleration in the development of smart factories, smart equipment, and smart shipping that does just that. Each will play a role in improving worker safety, optimizing inventory management, reducing maintenance costs, and streamlining production processes. The supply chain of the future is digital.

## PREDICTION FIVE

# Custom silicon goes mainstream

Usage of purpose-built chips will rapidly increase in 2023. As a result, the pace of innovation will accelerate as workloads take advantage of hardware optimizations that maximize performance, while lowering energy consumption and reducing cost.

Custom silicon and specialized hardware have been quickly gaining traction in the consumer technology industry. Everything from our laptops, to our cell phones, to our wearable devices are seeing significant leaps in performance with the fabrication and adoption of custom silicon. While adoption has been quick in the consumer space, the same hasn't been true for business applications and systems, where software and hardware traditionally have longer refresh cycles. However, this will quickly change in the coming years as the accessibility and adoption of custom silicon takes hold.



**100 million**

EC2 instances fired up every day.

**50%**

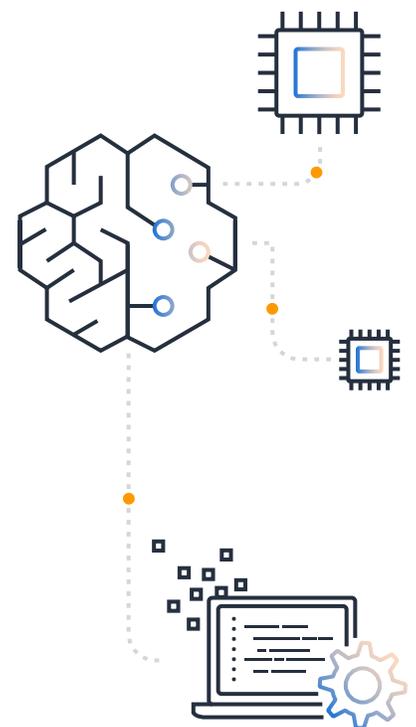
Cost-to-train savings with a Trainium-based instance.

**60%**

Graviton processors cost significantly less and use up to 60% less energy than comparable EC2 instances.

At AWS, there are an average of 100 million EC2 instances fired up every day (as of this writing). This is in large part due to how closely we've worked with customers over the years to understand the type of workloads they are running, and then, determine what we should build next. Like consumer devices, this has led AWS to invest heavily in chip design in recent years. That's because we know that the workloads companies are running in the cloud can be more performant and more cost-effective running on custom silicon, ones that are purpose-built for specific use cases.

Take machine learning workloads for example. Software engineers have traditionally relied on expensive, power-hungry GPUs to do everything from model building to inference. However, this one-size-fits all approach is not efficient—most GPUs aren't optimized for these tasks. In the coming years, more engineers will see the benefits of moving workloads to processors specifically designed for things like model training ([AWS Trainium](#)) and inference ([AWS Inferentia](#)). As this happens, a new wave of innovation will begin. By realizing a 50% cost-to-train savings with a Trainium-based instance, or 50% better performance-per-watt on an Inferentia2-based instance, engineers and businesses alike will take notice, and we will begin to see a massive migration of workloads. The same will be true even for generalized applications, where there are still benefits to moving to custom silicon, such as [Graviton3-based instances](#) that use up to 60% less energy for the same performance than comparable EC2 instances.



Cost savings and performance benefits will lead to more experimentation, more innovation, more adoption, and eventually, more custom silicon for other specific workloads. It's another virtuous cycle. Alan Kay once said, "*people who are really serious about software should make their own hardware.*" And in the coming year, people who are really serious about software will really begin to take advantage of all that custom silicon has to offer.

