Sponsored by Lenovo

# Driving Automation and Cost Savings at the Edge

TCO Validation for Lenovo Open Cloud Automation

RESEARCH BRIEF

# Table of Contents





RESEARCH BRIEF

# Driving Automation and Cost Savings at the Edge

## TCO Validation for Lenovo Open Cloud Automation

### Executive Summary

As enterprises expand their distributed computing footprints to meet the demands of digital transformation, Industry 4.0, and AI initiatives, edge computing has emerged as a critical component of modern IT infrastructure. However, managing these distributed environments presents significant deployment, configuration, and ongoing maintenance challenges. Lenovo Open Cloud Automation (LOC-A) offers a compelling solution to these challenges, promising substantial improvements in efficiency and cost-effectiveness.

Our review of Lenovo's Total Cost of Ownership (TCO) model for a scaled-edge deployment scenario reveals robust potential savings. For a 900-site deployment (typical big box retailer) over three years, LOC-A demonstrates a 47% reduction in TCO, a 60% decrease in deployment time, and a 61% reduction in required personnel compared to traditional deployment methods. Furthermore, the model estimates up to an 84% reduction in CO2 emissions, aligning with corporate sustainability goals. These significant savings are driven by LOC-A's ability to streamline and automate deployment processes while leveraging zero-touch provisioning capabilities. The potential savings of LOC-A for enterprises facing large-scale edge computing deployments make it a worthwhile consideration.

### Introduction

With the ongoing build-out of networks globally, increasing fiber penetration, 5G adoption for mobile and fixed wireless access, and the rise of non-terrestrial networks (NTNs) led by low earth orbit (LEO) satellites, organizations worldwide face increasingly distributed footprints for computing.

In many situations, the new connectivity provides a path to cloud-based services where none previously existed. In some, performance, privacy, or compliance constraints may limit the use of cloud-hosted services beyond orchestration and communication.

Even as public cloud adoption by enterprises continues to grow, with analyst firm Gartner projecting a 20.4% growth in 2024 to $675.4B in total spending, there continues to be strong interest in enterprise private clouds and edge clouds. From our conversations with enterprises, hyperscalers, and carriers, we expect to see an ongoing expansion of edge computing clusters for enterprise workloads. Key verticals like retail, manufacturing, transportation, energy, utilities, and agriculture are the leading industries adopting edge computing. Edge computing could be called upon to host increasing AI workloads (predictive and generative AI).

While consumers of public cloud services benefit from their fully managed nature, deep cloud operator expertise, and centralized economies of scale, enterprises that need to coordinate computing, storage, and networking services at the edge might look to alternative approaches. This is where orchestration and automation products like Lenovo Open Cloud Automation (LOC-A) can provide value, reducing the complexity of deploying, configuring, and managing edge nodes.

**While consumers of public cloud services benefit from their fully managed nature, deep cloud operator expertise, and centralized economies of scale, enterprises that need to coordinate computing, storage, and networking services at the edge might look to alternative approaches.**

In this report commissioned by Lenovo, AvidThink examines the total cost of ownership (TCO) of LOC-A for edge workloads. We previously performed a cost analysis on the product for private cloud deployment in data centers (report download here). This time, we will validate Lenovo's TCO model to demonstrate the potential cost advantage of using LOC-A for a typical enterprise edge deployment versus a semi-automated approach. As part of the analysis, we will review the sustainability benefits of LOC-A as depicted by the CO2 emissions modeling in Lenovo's model.

## Why Edge?

While cloud platforms have achieved remarkable success over the past decade, centralized data centers only meet some application needs. Edge computing fills the remaining gap by offering the following unique benefits:

- **Latency:** Edge computing can provide latency measured in milliseconds or less. In contrast, transmitting data to the current network "edge"—typically localized content distribution networks or metropolitan points of presence—involves multiple hops and long distances, resulting in latencies of around 50 milliseconds. Latency to centralized data centers and public clouds can be even higher.
- **High Throughput**: Serving cached or locally generated content from the edge can offer throughput orders of magnitude greater than that from core data centers in public or private clouds.
- **Data Reduction:** Running data analytics at the edge allows enterprises to reduce the amount of data sent upstream significantly. This decreases costs and enables more efficient use of available bandwidth.
- **Security:** Enterprises with unique security requirements can secure on-premises edge computing resources similarly to their highly secure edge sites.
- **Resilience:** Edge sites can continue providing services even during degraded or lost connections to the cloud. Some edge platforms can function as isolated mini-data centers in locations permanently lacking significant connectivity (e.g., remote offshore sites).
- **Compliance and Privacy:** Edge applications can handle local data processing without transmitting data across state or national boundaries, helping organizations comply with privacy and data residency laws.

## Edge Rising — Industrial Workloads

We classify on-premises computing under edge computing in that there are cloud components that the on-premises equipment communicates with — hence the term "edge" of a centralized system. That contrasts with the network edge — either hyperscaler-managed edge and local zones or distributed infrastructure managed by other providers like telecom operators.

Retail companies with many warehouses and stores (and computing infrastructure at those locations) operate an edge infrastructure. Similarly, manufacturers and transportation companies with multiple locations across the country are classified similarly.

The difference for these companies in 2024 is that their on-premises computing infrastructure can be managed differently. While each was previously managed with hands-on staff in person or via remote login technologies by IT or managed services providers, newer orchestration technologies can provide cloud-like management constructs across distributed infrastructure.

However, while management platforms like VMware by Broadcom, Red Hat (OpenShift), or other Kubernetes-based solutions can manage VMs and containers, fresh computing equipment with the appropriate images that these other private cloud platforms can manage remains necessary.

This is where Lenovo's solution, Lenovo Open Cloud Automation (LOC-A), comes in.

## What is Lenovo Open Cloud Automation (LOC-A)?

Lenovo Open Cloud Automation (LOC-A) provides rapid deployment and lifecycle management of on-premises cloud infrastructure in the data center and at the edge site.

Based on the Infrastructure as Code (IaC) and GitOps philosophies in the data center and offering a no-code approach for the edge, LOC-A provides more consistent deployments with a lower rate of errors. This lets enterprises quickly deploy cloud platforms on their bare-metal servers and helps avoid configuration drift across large-scale distributed deployments.

LOC-A supports the deployment of bare-metal, CaaS (containers as a service), and IaaS (infrastructure as a service) platforms and works with solutions from Red Hat, VMware, and many OS distributions. It promises simpler edge deployments via its near-zero-touch provisioning (nZTP) capabilities, avoiding intermediate staging steps and establishing staging environments.

## Total Cost of Ownership — Lenovo's Model Explained

The Lenovo team has built a TCO calculator to show customers the potential savings from using LOC-A versus alternative approaches to deploying and managing infrastructure at the edges. At the highest level, the model compares a traditional workflow that entails shipping servers to a staging site, loading the appropriate operating systems or hypervisors onto servers at that site, and shipping servers to edge locations. It assumes a 3-node standard Kubernetes VMware cluster running at each edge location.

Along with a set of standard assumptions around cost estimates for labor, electricity prices, etc. (which are shown in Appendix A), the model considers possible "slow path" events where servers are delayed, arrive broken, or when initial loads of OS fail invisibly and arrive at edge site inoperable. A sample of these events are listed below:

- Lack of room or personnel in staging location (overpopulation) — queue of servers to be processed
- Shipping delays of units to either edge sites or staging sites
- Dead on arrival (DOA) servers

The model applies probability to events like these and combines the resulting time and costs in a weighted averaged blended time. While the model estimates a range of optimistic (no failures) to pessimistic (overly high rate of mishaps), we've picked a realistic scenario for our model validation. A more sophisticated model with multiple failure paths and branches could be employed, but we agree with Lenovo that this more straightforward probability-weighted approach suffices for a TCO comparison (since the same methodology applies to both with and without LOC-A paths).

On the following pages are pictorial depictions of the deployment process: the first diagram shows traditional deployment with a staging site, and the second shows direct drop-ship to edge sites and provision via LOC-A's nZTP capabilities.

While deployment scenarios will vary, we note that many today do not leverage bare metal nZTP solutions and require trained personnel to do the initial bring-up and verification of edge servers in a central location. Subsequently, the provisioned machines are shipped to edge locations where less technical staff can rack, stack, and cable them with simple instructions. LOC-A is then used to automatically provision the bare-metal servers and deploy OS and the cloud platforms. Since cloud platforms from vendors like VMware or Red Hat can be orchestrated from the cloud, engineers can then remotely manage workloads without a physical presence at the edge site.

**LOC-A provides rapid deployment and lifecycle management of on-premises cloud infrastructure in the data center and at the edge site. Based on the Infrastructure as Code (IaC) and GitOps philosophies in the data center and offering a no-code approach for the edge, LOC-A provides more consistent deployments with a lower rate of errors.**

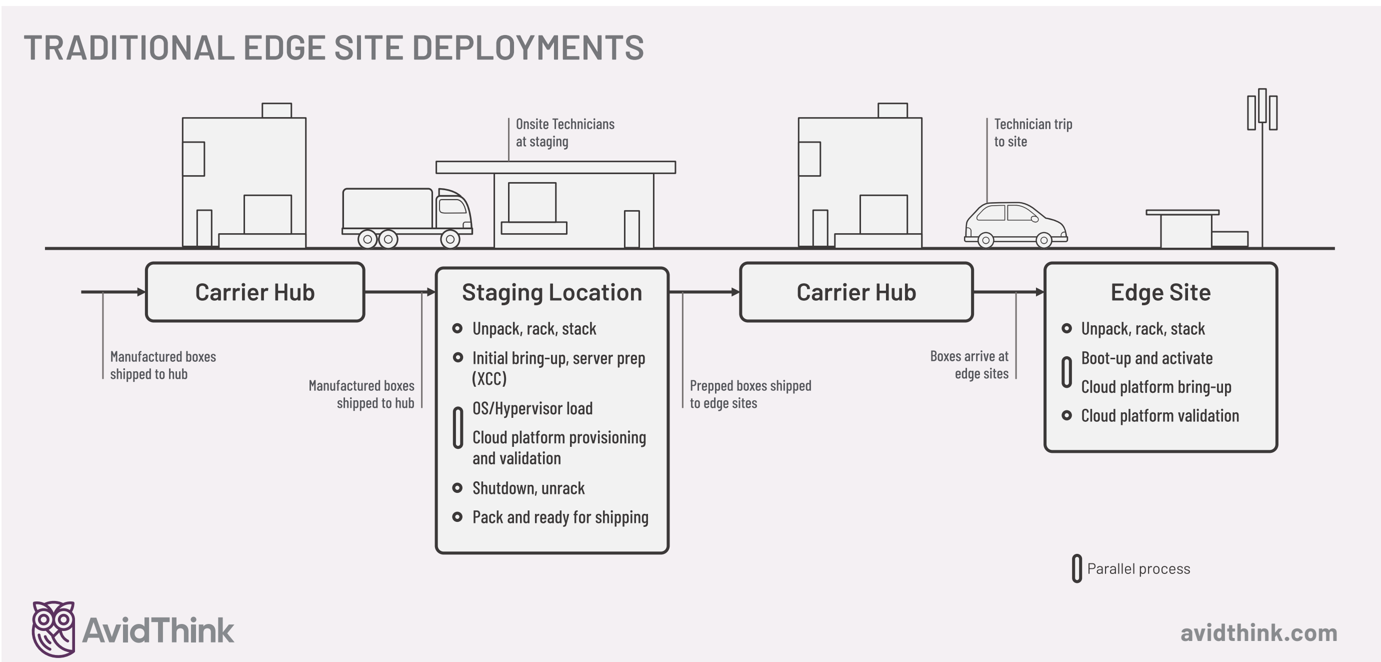
 

## Model Augmentation — Ongoing Provisioning for Hardware Failures

As part of AvidThink's engagement with Lenovo, we worked with Lenovo to extend the model to a common scenario — hardware failures in the edge sites requiring replacement hardware and re-provisioning. We assumed a 3% failure rate per year of hardware. Certain deployment scenarios could see higher failure rates; even premium, purpose-built servers are not infallible. Higher failure rates would improve the TCO in favor of LOC-A, which makes our conservative baseline estimate a good starting point. When hardware fails, the same workflow paths depicted above are repeated. Since the hardware RMA path is the same in both flows (with and without LOC-A), we can simplify our analysis by leaving out the RMA workflow.

## Scenario - 900 site deployment

The augmented Lenovo TCO model calculates the cost of the initial deployment of 2,700 servers (3 servers per location) across 900 sites. This could represent a retailer with large store locations around the country, a logistics company with a similar number of hubs, or any other business needing modest on-premises computing capabilities in distributed locations.



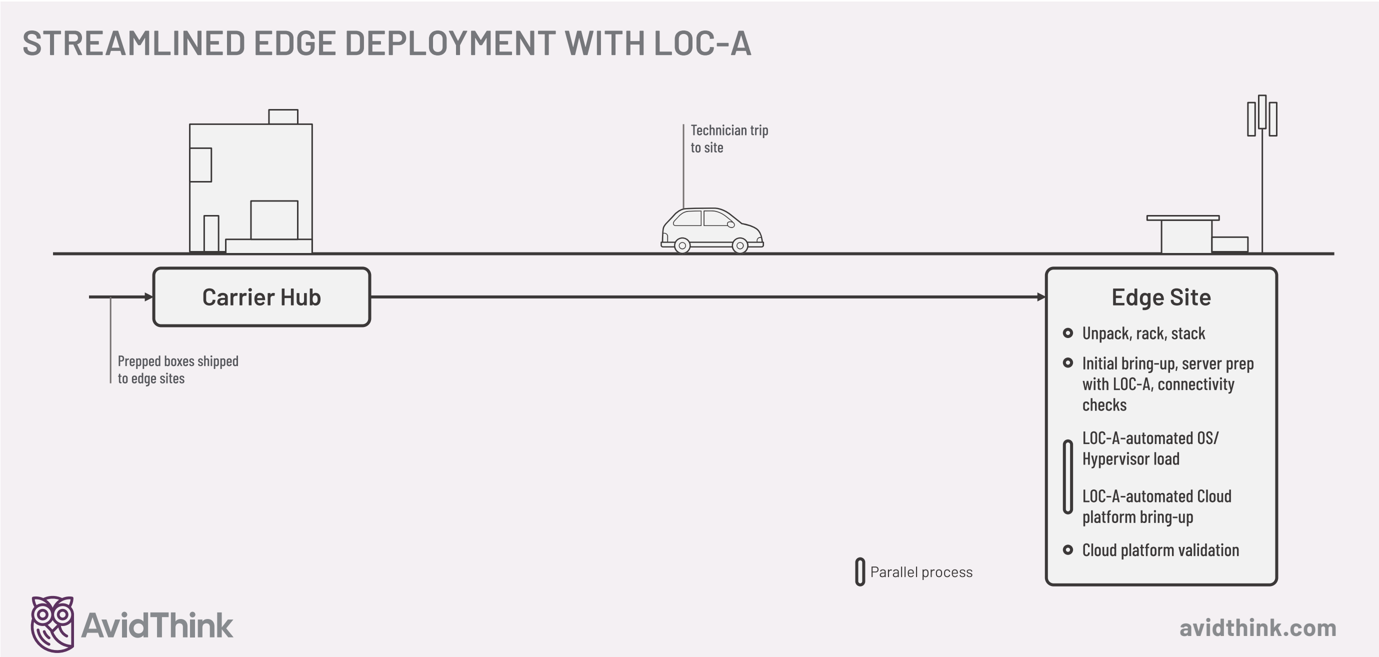

To help put this in perspective, the largest retailers in the US have between 1,000 super-sized warehouse stores and 4,000+ large warehouse stores, and the largest home improvement store chains have over 2,000 stores.

A three-server cluster could serve many vertical applications, such as providing inventory, retail POS, store operations support, video surveillance processing, and cached storage for retail. Similar sophisticated operations can be performed for warehouses and transportation or even healthcare, with the ability to operate disconnected if the wide-area network goes down.

In Lenovo's modeled scenario, the traditional path involves shipping all server units to a centralized location for staging and preparation before final shipment to each edge location. As described in the last section, the LOC-A optimized path involves sending bare metal servers directly to the edge site.

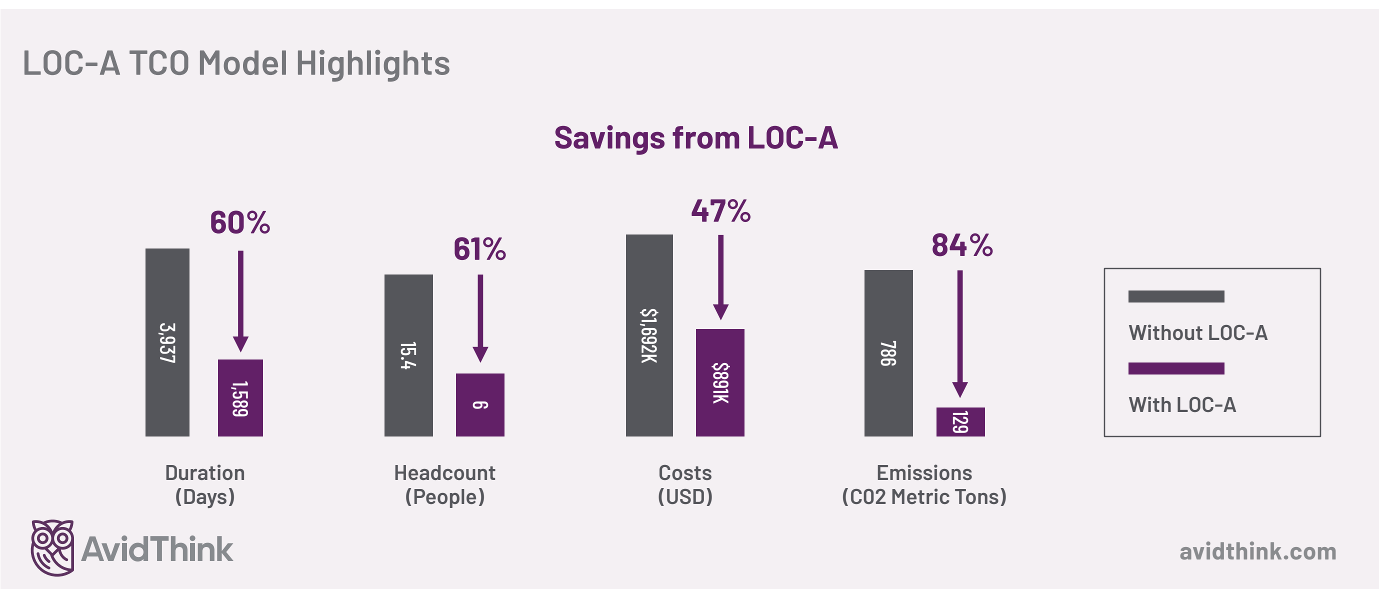
In our evaluation of Lenovo's TCO model, we have picked what Lenovo terms the "optimistic scenario," where the error-free path is encountered 70% of the time and the flawed paths 30% of the time. This is the most realistic scenario for validation.




## Examining the TCO Numbers

Here are key excerpts from the LOC-A TCO model:




When we compare the time required for executing the workflow across 2,700 servers for 900 sites with and without LOC-A, we see a 60% reduction in the total time spent on the initial deployment of the 900 sites and three years of redeployment of failed units. That translates to a 61% reduction in the number of people required.

From a cost perspective, factoring in cost savings from not having to maintain a staging site, cutting down the workflow steps involved in pre-staging, and faster bring-up with ZTP, but adding back license fees for LOC-A, we see a 47% reduction in TCO with LOC-A.

According to estimates of the impact of CO2 emissions, the LOC-A path could reduce CO2 emissions by up to 84%.

## AvidThink Analysis of the Cost Savings

Solutions that address taking an un-provisioned server to the state where the underlying cloud platform/hypervisor/operating system is installed are valuable in enabling increased automation across the lifecycle of computing resources. Our previous analysis of LOC-A for deployment in a private data center demonstrated the TCO in a scale deployment in a few locations. This time, as reviewed by AvidThink, Lenovo's TCO model demonstrates substantial potential cost savings (including personnel savings) for a distributed deployment in many locations. The new model also shows a likely reduction in CO2 emissions — which supports corporate sustainability initiatives.

Digging deeper into the model to better understand where the savings come from, we can identify the following key drivers of the cost differential:

- Eliminating a staging step removes the costs of maintaining and powering a staging site and obviates the need for another racking/stacking, power-on, pre-staging configuration cycle.
- Direct shipping to the remote sites instead of having an intermediate staging site cuts shipping costs, time, and CO2 emissions associated with the transportation of the server units.
- Having no staging site and removing a provisioning cycle reduces power consumption.
- Because there's one major step less, the number of things that can go wrong is reduced.
- Enabling nZTP at remote sites simplifies operations and streamlines the bring-up time at the edge site—even though more steps are performed at the edge, LOC-A automates them.
- The automatic deployment of OS and Cloud platforms brings more savings by eliminating the manual, error-prone process of keyboard-led installs, ensuring a consistent and secure rollout.

Unsurprisingly, the same cost reduction drivers identified above apply anytime a server unit fails at the edge. The end-to-end workflow is triggered when a replacement unit must be readied and shipped out. And so, during the maintenance (or Day 2) operations phase, the same savings accrue from using LOC-A.

## Other Considerations and Benefits of LOC-A

No model can perfectly capture every attribute of the different workflows, and that's not the intention of the Lenovo TCO model. We believe that it captures the main cost drivers and key events that happen in most deployments, as validated by field interviews conducted by the Lenovo team during the model's construction. The items identified are consistent with AvidThink's previous interviews with Lenovo partners/customers.

In evaluating the model, we would suggest that customers evaluating LOC-A consider the following additional factors in their utilization of the model in their decision-making:

- Emissions estimates can vary based on assumptions about transportation types: the type of vehicle used (gasoline, hybrid, electric), the average distance to staging and edge sites, and the source of power used at the staging and edge sites (renewables).
- Staff hourly rates and the cost of electricity will vary across regions and will have to be adjusted. While LOC-A use will likely generate cost savings compared to a situation without LOC-A, the magnitude of those savings depends on these rates.
- Hardware failure rates at edge sites may be higher than typical rates in data centers with better environmental controls and conditioned power. Edge locations may have intermittent power failures, be subject to poor line conditioning, and have inadequate or irregular environmental conditions. With higher failure rates, there will be more re-provisioning events during the 3-year TCO evaluation period, skewing cost savings towards the use of LOC-A.

**These savings and other potential benefits make LOC-A a worthwhile consideration for enterprises looking to deploy edge infrastructure at scale.**

Other factors have yet to be worked into the model (because they are more amorphous) but are likely to improve the business case for LOC-A. These include:

- Faster scale-up of new sites added in the future using LOC-A. The template-based methodology and low-code mindset automation provide a better foundation for improved scale than a less automated workflow with a staging step.
- Potentially faster troubleshooting when errors occur due to the organized planning that must be put in place to use LOC-A, which requires pre-planning and capturing all relevant information into a single document.
- In a disaster recovery situation, eliminating the staging site and streamlining the edge bring-up process results in a faster recovery time for key services.

## Conclusion

Enterprise-edge deployments continue to grow, driven by digital transformation, Industry 4.0, and AI initiatives. While cloud-orchestrated private cloud platforms like those from VMware and Red Hat provide the necessary automation and orchestration for the cloud platform, solutions like LOC-A from Lenovo provide the bootstrapping and zero-touch capabilities to take an out-of-the-box server to a manageable node in a VM or Kubernetes cluster.

AvidThink's review of Lenovo's TCO model for estimating the staffing, time, cost, and emissions savings from using LOC-A shows that LOC-A can provide significant savings across all those metrics. These savings and other potential benefits make LOC-A a worthwhile consideration for enterprises looking to deploy edge infrastructure at scale.

## Appendix A - Assumptions

| | | |
|---|---|---|
| **Site-Related** | | |
| Number of servers per site | 3 | Servers |
| Number of sites | 900 | Sites |
| | | |
| **Server Hardware** | | |
| Instant power consumed by 1 server in idle (https://lcp.lenovo.com/) (SE350) | 68 | W |
| Instant power consumed by 1 server at 50% utilization | 120 | W |
| Percentage of DOA servers | 0.5% | |
| Percentage of annual failures for servers at edge | 3% | |
| | | |
| **Server Configuration** | | |
| Time to unbox (re-pack) a server | 10 | Mins |
| Time to rack and cable (dismantle) a server | 10 | mins |
| Time for server boot (reboot) | 10 | mins |
| Time for server secure activation/deactivation | 10 | mins |
| Time to do initial XCC (BMC) setup | 20 | mins |
| Time to do initial XCC (BMC) setup, server registration and connectivity check | 10 | mins |
| Time to shutdown servers | 5 | mins |
| Number of clusters that can be deployed in parallel by one person using LOC-A | 10 | |
| | | |
| **Personnel** | | |
| Salary of HW engineer in Staging / year ($70K) | $ 23.97 | per hour |
| Salary of SW engineer in Staging / year ($100K) | $ 34.25 | per hour |
| Number of work days per year | 200 | days |
| Number of work hours per day | 8 | hours |
| Number of field teams | 10 | 2 members |
| Number of work hours for Transit | 16 | hours |
| Number of work hours for Staging | 16 | hours |
| Number of work hours for on-site | 8 | hours |
| Number of work hours for remote | 8 | hours |
| Number of work hours for LOC-A system | 24 | hours |
| | | |
| **Travel/Shipping** | | |
| Percentage of shipments that can be done in parallel (to multiple sites) | 70% | |
| CO2 Emissions per trip | 0.267 | Kg CO2 |
| Gas used per trip | 2.667 | L |
| Avg. L/100Km | 8 | |
| Trip length if duration is: 40 min | 33.3 | Km |
| Average speed during trip to site | 50 | Km/h |
| | | |
| **Emissions and Power** | | |
| CO2 Equivalent for 1 kWh (< 0.1 dPACA, < 1 % PACA) (*1) | 0.296 | Kg |
| 1 kWh cost (EU) (*2) | 0.2525 | EUR |
| SE350 Server BTU/Hr (idle) (https://lcp.lenovo.com/) | 232 | BTU/Hr |
| SE350 Server BTU/Hr (50%) (https://lcp.lenovo.com/) | 409 | BTU/Hr |
| 1 BTU/h to W conversion (*3) | 0.293 | W |

* Note: LOC-A software license fees available on request from Lenovo

*1: https://www.rensmart.com/Calculators/KWH-to-CO2; EU aggregate value
*2: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Electricity_price_statistics
*3: https://vmguru.com/2011/06/how-to-calculate-electrical-costs-for-cooling-and-power-consumption/

**AvidThink, LLC**
1900 Camden Ave
San Jose, California 95124 USA
avidthink.com