How to Use Edge Computing to Modernize Your Retail Store Infrastructure

Published 8 June 2021 - ID G00742986 - 7 min read

By Analysts Santhosh Rao, Miriam Burt, Jonathan Forest

Initiatives: Cloud and Edge Infrastructure; Industry Product Planning and Strategy; Retail Digital Transformation and Innovation

Video analytics, augmented/virtual reality, the Internet of Things and robotics in retail store modernization projects require retail I&O leaders to re-envision the operation of their infrastructures. Here, we outline mechanisms that simplify, consolidate and manage retail store infrastructures.

Overview

Key Findings

- Transformative technologies, such as video analytics, augmented reality/virtual reality, artificial intelligence and the Internet of Things, which enhance the customer experience and improve profitability, will require modernization of IT infrastructures at retail stores.

- Most retailers have limited or no local IT staff, which makes the monitoring, security and management of in-store applications and infrastructure a challenge.

- Simplified integration of networking and security is increasingly a main driver in infrastructure and operations leaders’ decisions regarding edge environments.

Recommendations

I&O leaders focused on cloud and edge infrastructure must:

- Avoid infrastructure silos by deploying edge-computing solutions that consolidate existing legacy applications and new-age applications on standardized infrastructure.

- Centralize site-level application and infrastructure life cycle management for better manageability, visibility and control.

- Consolidate functions by integrating networking and security in software-defined WAN branch solutions or implementing a secure access service architecture with lightweight SD-WAN at the edge, orchestrated with cloud security.
Strategic Planning Assumption

By 2025, 20% of all Tier 1 multichannel retailers will use edge computing to enable edge artificial intelligence (AI) use cases, which is an increase from fewer than 5% today.

Introduction

The COVID-19 pandemic has accelerated brick-and-mortar retail enterprises’ transition toward omnichannel retail. As brick-and-mortar retailers strive to establish their digital presence through e-commerce stores and marketplaces, they must focus equally on modernizing the physical retail store. Retailers should offer customers an integrated and consistent experience across both media by consolidating applications and associated data footprints in the data center or cloud. At the same time, to better manage operations, security, customer experience (CX) and customer analytics, certain applications must be deployed at the edge — i.e., the physical retail store.

Technologies such as video analytics, the Internet of Things (IoT), augmented reality (AR)/virtual reality (VR), live commerce and robotics that form the building blocks of applications at the edge will require infrastructure that supports low latency and high compute/memory. Therefore, a comprehensive approach to infrastructure design at the edge is required. Figure 1 represents the “retail store in the box,” in which traditional retail apps and emerging applications are consolidated onto a single platform.
This research outlines some of the challenges and risks that infrastructure and operations (I&O) leaders will face when modernizing retail store infrastructures, as well as ways to mitigate these risks. These challenges are not unique to the retail industry, but are often observed in such industries as manufacturing, energy and banking.

Analysis

Avoid Infrastructure Silos by Deploying Edge-Computing Solutions

I&O leaders must aim to consolidate applications deployed in-store onto a common infrastructure solution. The infrastructure should be modular, resilient, scalable and easy to manage. Hyperconverged infrastructure (HCI) platforms are increasingly becoming platforms of choice, because they’re capable of addressing most of these requirements. Depending on the performance and latency profile, the applications can be grouped and deployed on one or more hyperconverged clusters. Several such clusters can be centrally managed from the enterprise data center or cloud through a unified control plane.
As a first step, I&O leaders must group retail stores based on parameters such as size of the retail store, number of applications and its performance profiles. Doing so will simplify the process of standardizing on the hardware and software configurations for each of these groups, thus simplifying operations. I&O leaders can choose a three-node (FTT=2) or a two-node, “lightweight” HCI solution that supports open-source hypervisors. This can deliver a low total cost of ownership (TCO), particularly when scaling to hundreds of stores.

I&O leaders should pilot these infrastructure upgrade projects in a small number of stores to begin with, and then expand once the IT team demonstrates that it can meet store-level SLAs. Although a two-node solution will ensure low TCO, a three-node deployment ensures a higher level of resilience, particularly during software patching and hardware maintenance. Representative vendors that provide HCI-based solutions focused on edge environments include:

- DataCore
- Microsoft
- Nutanix
- Pivot3
- Scale Computing
- StarWind
- StorMagic
- VMware

Applications such as video analytics may require VPUs/GPUs to enable deep-learning algorithms to perform inference and, in some cases, training at the store. Ensure that video analytics or other such performance-intensive applications do not affect other applications that are colocated on the same server cluster. If required, host such applications on dedicated infrastructure at the edge.

Centralize Site-Level Application and Infrastructure Life Cycle Management

Retailers have a limited number of IT staff at stores; therefore, they must focus on centralized control for easier management and monitoring. HCI, hypervisors and hyperscalers provide platforms that can store template application profiles and images, which can be used to rapidly deploy applications to multiple retail stores remotely. These platforms also provide application and infrastructure performance and availability monitoring.

Although retailers’ use of containers in stores is still nascent, containers promise to significantly increase agility and modularity in application life cycle management. This is particularly beneficial when retailers implement digital transformation projects that improve and augment CX and business processes at the
store. Applications such as personalized advertising and stock calibration have a high daily change rate and require frequent application reconfiguration and complex integration with other in-store services.

Retailers can centrally manage these applications and apply uniform policies to groups of retail stores with identical profiles. For example, computer vision projects that analyze customer behavior are simpler to implement when containerized, given the “fail fast,” iterative nature of these projects.

A well-known retail chain in North America centrally manages application life cycle and system health across more than thousand stores through Kubernetes. Figure 2 shows a high-level architecture of the deployment. Each store hosts a Kubernetes cluster that runs check-out applications, product search and IoT applications. Because application requirements vary by store, those with near-identical requirements were grouped and assigned a unique namespace. In this way, application development teams were able to deploy changes to a specific group. These tasks were performed centrally through a platform with which software developers interfaced, to deploy and manage applications on Kubernetes clusters at the store.

**Figure 2: Tier 1 North American Retailer**

### Tier-1 North American Retailer

**In-Store Apps Running on K8 Clusters in Each Store**

- **Checkout**
- **In-Store Product Search**
- **Personnel Management**
- **Internet of Things**

**Centralized Control Pane**
- Centralize App Life Cycle Management
- Group Stores and Apply Common Policies Uniformly
- Monitor and Report on App Health

Source: Gartner 742986_C

Consolidate Functions by Integrating Networking and Security in SD-WAN Edge Solutions or Implementing SASE

As I&O leaders move more workloads to the cloud, they are implementing software-defined WAN (SD-WAN) with internet connectivity at the edge. This, in turn, moves the security perimeter out to the edge, whereas it was formerly at the data center. The result is the convergence of security and networking...
functions residing at the edge, driving consolidation of these functions for simplicity, agility and cost-effectiveness. We see two primary ways that enterprises accomplish this.

First, we see consolidated functions in a single device from a single vendor at edge locations. These devices support SD-WAN (routing, path selection, orchestration, etc.) and integrated advanced security, such as:

- intrusion prevention system (IPS)
- A/V
- Content filtering
- Layer 7 firewall
- Data loss prevention (DLP)

They are often cloud managed to simplify the on-premises footprint; however, the value is typically a less expensive price, a reduced footprint (consolidation of hardware), and simplicity, with fewer devices to manage. This is a “thicker” branch option, with most functionality residing at the edge location.

Typically, users that implement this solution prefer functionality from the same vendor, are cost-driven, want to retain control and are somewhat cloud-averse.

Second, we see a “thinner” SD-WAN networking capability at the branch with advanced security functionality offered as a cloud service. The cloud security offers the full complement of functionality, including all of the above capabilities, together with secure web gateways (SWGs) and cloud access security brokers (CASBs). This combined solution with SD-WAN networking at the edge and advanced security in the cloud is also known as secure access service edge (SASE). The cloud security functionality may be offered by the same vendor that offers the SD-WAN (although there are few vendors that can actually do this today) or a different vendor, where the functionality can be service chained.

The value here is that cloud security is more agile from an operational expenditure (opex) standpoint, centrally managed and also more aligned for remote/mobile users. If SD-WAN is offered from the same vendor that offers the cloud security, then typically there is tighter functional integration and orchestration and a simplified sourcing model. If there are two different vendors involved for the SD-WAN and the cloud security, then the orchestration may be more limited, the pricing is often higher, and sourcing and support models are more complicated.

Evidence

Client inquiries

Recommended by the Authors