Reference Architecture for Workloads using Lenovo ThinkAgile CP

Last update: 26 September 2019
Version 1.2

Provides a technical overview of Lenovo ThinkAgile CP

Illustrates how Splunk Enterprise software can be used in a cloud environment

Contains sizing recommendations

Explains reliability and performance features of cloud infrastructure

Ajay Dholakia

Bhavin Shah

Shyam Sareen

LENOVO PRESS

Table of Contents

1 Introduction

Lenovo® ThinkAgile™ CP is a ready-to-use private cloud designed to help IT leaders transform their organizations into innovative centers by freeing their staff from time consuming tasks. This all-in-one composable cloud solution includes an integrated application marketplace, and delivers a turn-key cloud experience in your data center. ThinkAgile CP uses modular compute, storage, and networking components along with Lenovo's cloud virtualization software to create pools of IT resources, independently scaling and allocating capacity, and automatically configuring resources to fulfil application requirements. ThinkAgile CP can scale as your business grows while helping you automate the management of your applications and security policies.

This document describes how application workload software can be deployed on ThinkAgile CP private cloud by selecting the appropriate compute, storage and networking components. In particular, this document focuses on Splunk Enterprise workload, which provides an application platform for real-time operational intelligence in an enterprise IT environment. As customers adopt private cloud for their enterprise IT needs, deploying analytics applications such as Splunk Enterprise within the private cloud is highly desirable. It enables exploiting the same cloud computing benefits that other business applications receive, thereby further simplifying the task of running the IT environment.

The intended audience for this document are Technical IT Architects, System Administrators, and IT Managers who are interested in deploying Splunk Enterprise on Lenovo's ThinkAgile CP to support the growing data requirements for their business. Running a collection of analytics algorithms requires a flexible infrastructure, with nodes capable of delivering compute-heavy, storage-heavy or balanced configurations. The Splunk Enterprise software design enables scaling along all these dimensions and the ThinkAgile CP configurations described in this document provide matching scalability to deliver the most cost effective solutions for state-of-the-art operational analytics.

2 Technical overview of appliances

This chapter provides a technical overview of ThinkAgile CP.

2.1 ThinkAgile CP

With its software-defined modular architecture, the ThinkAgile CP platform can be scaled quickly without requiring specialized skills by adding more compute and storage resources independently as your business grows.. ThinkAgile CP is designed to minimize downtime with all of the components engineered to work together, with tested, standardized, and automated code updates.

ThinkAgile CP offers the following key features:

- Modular and scalable on-premises cloud platform configurations designed to optimize workload performance and provide the IT agility for business demands.
- Modular and scalable physical and virtual compute, network, storage resources, and Lenovo cloud infrastructure software.
- Lenovo deployment services help get customers quickly install the solution with minimal impact to their business.
- Lenovo cloud infrastructure software helps simplify cloud deployments, application infrastructure management, and workload provisioning.
- Extensive security features, such as data at rest encryption, virtualized network and VM-level firewalls, and two-factor authentication help customers meet their security requirements.
- Centralized cloud-based management automates discovery, deployment, and configuration of cloud resources.
- Lenovo ThinkAgile Advantage provides a single point of contact for all support issues and integrates support chat and virtual technical assistance into your management interface.

ThinkAgile CP consists of the following components:

- CP Interconnect
  The Interconnect centralizes connectivity of your on-premises infrastructure to the Cloud Controller and acts as the entry point into your existing network, which connects the Cloud Controller to your environment. The CP Interconnect consists of one or two high-density, ultra-low-latency 10/40 GbE network devices and uses specialized software which provides the fabric for your cloud environment, automating discovery, onboarding, and device management. Integrated network virtualization secures virtual data center and applications.
- Compute Blocks
  The Compute Block is a modular 2U enclosure that contains up to four nodes, and it delivers processor and memory resources to the cloud. The compute nodes run a specialized hypervisor that combines open KVM-based virtualization software, hardware integration, and automation to orchestrate and deliver an end-to-end compute platform. ThinkAgile CP can support up to 10 Compute blocks.

- Storage Blocks
  The Storage Block is a 2U storage chassis with up to 24 PCIe NVMe SSDs and two controllers for high availability and redundancy. A storage block delivers data storage layer to the cloud. It simplifies storage management, provides enterprise-class storage functionality, and enables linear performance and capacity scaling. ThinkAgile CP can support up to 5 Storage Blocks.
- Management switch
  The 1 GbE management switch is required to provide out-of-band management for Compute Blocks and Storage Blocks. The switch can be supplied by Lenovo, or customers may provide their own switch.
- Cloud Controller
  The ThinkAgile CP Cloud Controller orchestrates and manages the on-premises infrastructure and workloads; however, it resides securely in the cloud. It automates and orchestrates infrastructure provisioning in real time to achieve complete and secure vertical hardware, software and workload integration. It federates and abstracts all physical hardware into a private cloud service.
  The Cloud Controller provides a single point of management across an unlimited number of ThinkAgile CP stacks with role-based access control, multi-factor authentication, and secure HTML and RESTful API interfaces.

For detailed information on ThinkAgile CP series, please see:

https://lenovopress.com/lp0872-lenovo-thinkagile-cp-series-cloud-platform.

2.2 Software components

This section gives an overview of the software components used in the solution.
ThinkAgile CP components ship with factory-installed software and a Cloud Controller license. ThinkAgile CP software is available in two editions: Standard and Guardian. The Standard edition provides data-at-rest encryption that is compliant with KMIP. The Guardian edition provides stringent security with encryption that is cimpliant with FIPS 140-2, STIG, Common Criteria, and USGv6.

The ThinkAgile CP software offers the following features:

- Cloud controller web-based interface
  - Provides a multi-site management portal for the cloud infrastructure
  - Manages compute, storage, and networking resources in the virtual data centers
  - Automates discovery, deployment, configuration, and composition of cloud resources
  - Provides real-time and historical analytics and monitoring
  - Delivers an application marketplace of pre-engineered application templates to speed provisioning
  - Provides a customizable service catalog that can be customized per virtual data center
- Security

    - Manages access to cloud resources with role-based access control
    - Provides two-factor user authentication to prevent unauthorized access
    - Encrypts communications between the cloud controller and the infrastructure
    - Protects cloud resources with Virtual Data Center (VDC) partitioning
    - Supports multi-tenancy with physical and logical resource isolation
    - Secures stored information with data-at-rest encryption
    - Provides network security with micro-segmentation and distributed firewalls
- Compute
    - Provides physical hardware abstraction layer for the processor and memory resources
    - Uses categories and tags to group and allocate compute resources to the applications
    - Dynamically distributes application instances across the Compute Nodes
    - Logically isolates compute resources with migration zones and virtual data centers
    - Supports live migration of application instances with zero downtime
- Storage
    - Provides distributed, all flash scale-out storage for cloud services resources
    - Scales easily by simply adding more Storage Blocks
    - Manages storage pools and allocates storage resources to the application instances
    - Protects from drive failures with RAID data redundancy and hot-spare drives
    - Provides enterprise-class storage management capabilities:
        - Thin provisioning
        - Cloning
        - Snapshots
        - Integrated local and remote backup and recovery to support business continuity
        - Asynchronous replication
- Networking
    - Provides network overlay virtualization and network function virtualization
    - Offers traditional bridged VLAN tagging-based networks for use by application instances
    - Enables seamless instance migration across the Compute Nodes with VNETs
    - Manages virtualized network functions: routers, firewalls, load balancing, NAT, and DHCP

3 Deployment models

This chapter provides recommended deployment models for a ThinkAgile CP (see Figure 1).

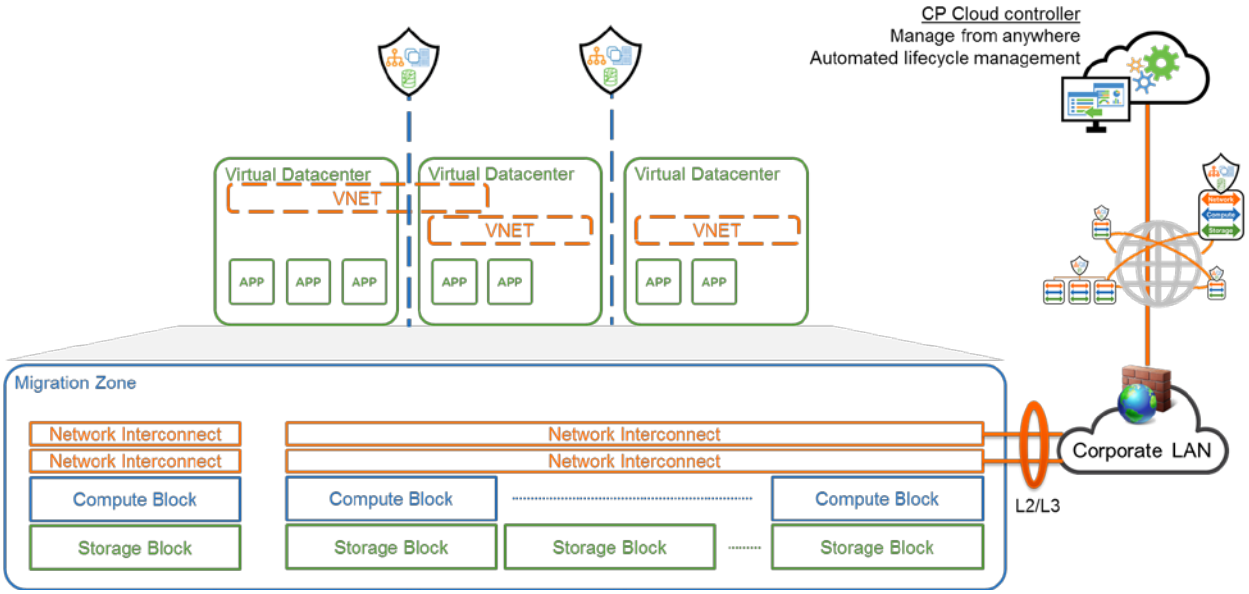

Figure 1: Overview of ThinkAgile CP series.

A ThinkAgile CP stack can be tailored to meet the requirements of a variety of different deployment scenarios. The software-defined modular architecture allows configurations that are optimized to deliver the compute, storage and connectivity requirements for different workloads. Desired deployments can be configurations that are compute-heavy, storage-heavy or achieve a balanced compute and storage capability. These configurations can also change during the life-cycle of a deployment to continue to meet growing and changing requirements.

Certain workloads have a variety of deployments driven by the capacity of data storage required to house the current and future data needs of an organization. The ability to add Storage Blocks allow for flexible scaling of a ThinkAgile CP deployment. Other workloads, like data warehouses and data analytics applications, need to support both a desired data capacity as well as a number of concurrent users, either of which can vary based on the business needs and size of an organization. To support these requirements, the deployment needs to accommodate a variety of combinations of Compute and Storage Blocks. Alongside the capacity, consideration must be given to the desired performance levels which, drive the CPU, memory, I/O and network connectivity of the various building blocks.

In summary, the ThinkAgile CP platform supports a broad range of capabilities while maintaining a common management framework and ease of scalability over the life-cycle of a deployment. For a specific workload deployed to meet a targeted number of users and data volume, see the sizing tables in the Operational Model section below.

The recommended model to use is ThinkAgile CP-6000. The configuration process for ThinkAgile CP includes the selectable components as explained below.

The 2U compute block is based on the Lenovo ThinkSystem™ SD530, which is a 4 bay chassis. Each bay has a node with 2x Xeon® Scalable processors and 2x 10Gb NICs. The exact processor model and memory configuration of each node is configurable. The ThinkAgile CP solution can up to 10 compute blocks.

The 2U storage block contains a minimum of 8 and a maximum of 24 SSDs. SSDs are added to a storage block in groups of 8. Dual-ported NVMe SSDs are used to ensure resiliency and reduce the risk of storage becoming the bottleneck for a workload. ThinkAgile CP Solution can support from 1 to 5 Storage Blocks. Different drive capacities supported are s 1.6TB, 3.2TB, 6.4TB and 12.8TB. In other words, this solution can support up to 1.5PB of storage at maximum configuration.

The 2U interconnect block has redundant SDN-enabled ONIE (Open Network Install Environment) Interconnects, each with 48x10 GbE and 6x40 GbE ports. 44 of the 10 GbE ports are used to connect to the Compute Nodes and Storage Blocks. The remaining 10 GbE ports are used for the management network that ties all the storage, compute and interconnect blocks together. Two of the 40 GbE ports are used to connect the 2 internal routers to each other for MLAG traffic. At least one of the 40 GbE ports is used to connect into the customer's data center switch. The remaining 3 may be used as needed for handling traffic between one router pair and another router pair, or for additional N-S traffic.

A 1GbE switch is used for management of the ThinkAgile CP.

Lenovo Deployment Services is mandatory and is used for deployment of ThinkAgile CP solution. Lenovo HW Installation Service is optional and customer can install themselves or buy installation service.

ThinkAgile CP hardware features a base 3 year warranty, as well as options additional warranty for protection.

ThinkAgile CP is designed for installation in a rack cabinet; the rack cabinet can be purchased from Lenovo (see Rack cabinets) or provided by the customer. Lenovo deployments services are included; they provide remote preparation and planning, configuring platform components, validating installation, transferring knowledge, and developing post-installation documentation.

Lenovo basic installation services are optional; they provide mounting the components in a rack cabinet, connecting components to network and electrical power, checking and updating firmware, verifying operations, and disposal of the packaging materials within the customer site.

4 Splunk Enterprise

Splunk Enterprise provides an application platform for real-time operational intelligence. It facilitates easy, fast and secure collection, analysis, and search of data from massive data streams generated by devices, applications, transactions, timed events, systems and technologies.

4.1 Business challenges and value

The type of business challenges that Splunk can address are as follows:

- Real-time identification and mitigation of advanced organizational security threats to the Enterprise by leveraging vigilant analysis and response capabilities
- Complexity of managing the abundance of systems prevalent in a data center, and ensuring high performance and availability of these systems, daily.

Splunk Enterprise provides an end-to-end, real-time solution for both of these business problems by delivering the following core capabilities:

- Universal collection and indexing of machine data and security data, from virtually any source
- Powerful search processing language (SPL) to search and analyze real-time and historical data
- Real-time monitoring for patterns and thresholds; real-time alerts when specific conditions arise
- Powerful reporting and analysis
- Custom dashboards and views for different roles
- Resilience and horizontal scalability
- Granular role-based security and access controls
- Support for multi-tenancy and flexible, distributed deployments on-premises or in the cloud
- Robust, flexible platform for big data apps

4.2 Solution overview

Splunk Enterprise provides an application platform for real-time operational intelligence. It facilitates easy, fast and secure collection, analysis, and search of data from massive data streams generated by devices, applications, transactions, timed events, systems and technologies.

Figure 1 below shows the architectural overview of Splunk Enterprise. Users can access one or more search head servers through a load balancer. The search head(s) provide access to information that is collected by forwarders from a variety of data sources possibly across multiple data centers.

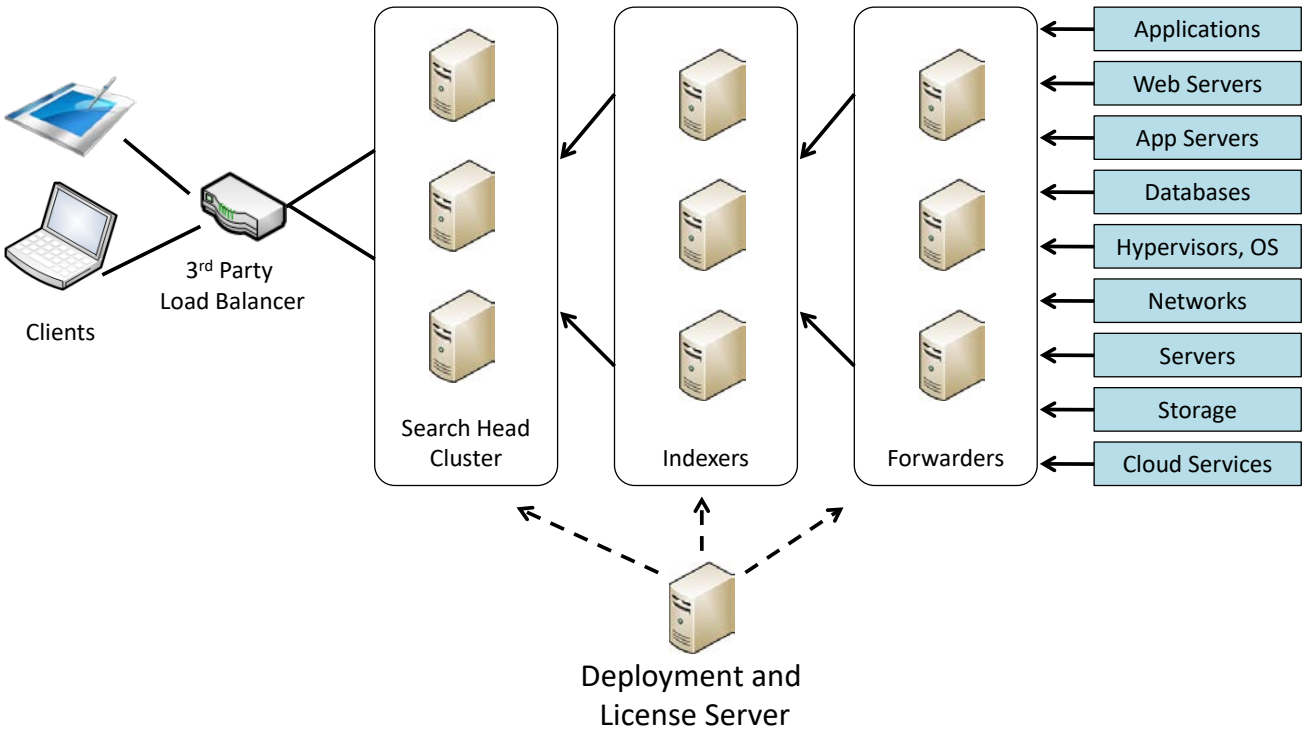

Figure 2: Architectural Overview of Splunk Enterprise

4.3 Component model

This section describes the component model for Splunk Enterprise. Figure 2 shows an overview of the major components.

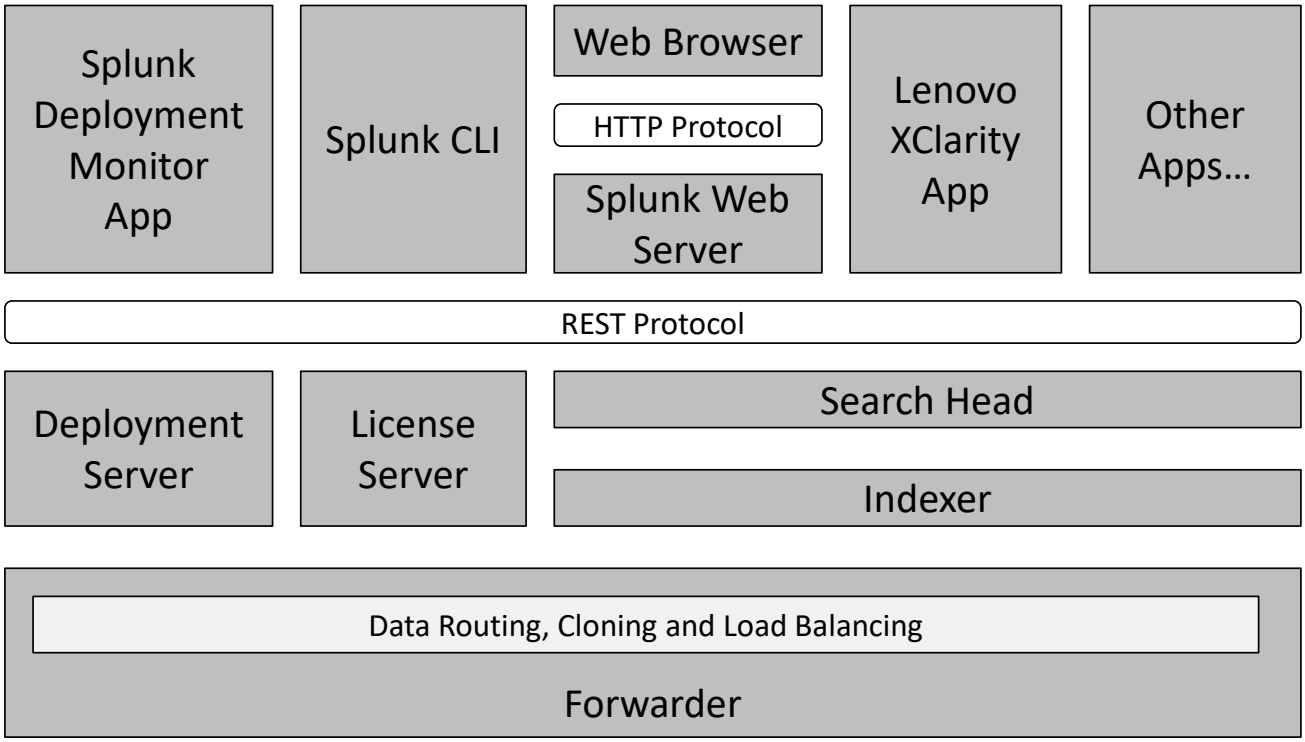

Figure 3: Component Model of Splunk Enterprise

4.3.1 Indexers

The indexer is the Splunk Enterprise component that creates and manages indexes. The primary functions of an indexer are:

- Indexing incoming data.
- Searching the indexed data.

4.3.2 Search heads

For large amounts of indexed data and numerous users concurrently searching on the data, it can make sense to distribute the indexing load across several indexers, while offloading the search query function to a separate machine. In this type of scenario, known as distributed search, one or more Splunk Enterprise components called search heads distribute search requests across multiple indexers.

4.3.3 Deployment server

Splunk Enterprise deployment server is used to update a distributed deployment. The deployment server pushes out configurations and content to sets of Splunk Enterprise instances (referred to, in this context, as deployment clients), grouped according to any useful criteria, such as OS, machine type, application area,

location, and so on. The deployment clients are usually forwarders or indexers. For example all of the Linux forwarders can be refreshed, after testing an updated configuration for a local Linux forwarder.

For small deployments, the deployment server can cohabit a Splunk Enterprise instance with another Splunk Enterprise component, either a search head or an indexer. For larger deployments it should run on its own Splunk Enterprise instance.

4.3.4 License server

The license server manages Splunk Enterprise licenses. It often runs in the same Splunk Enterprise instance as the Deployment server.

4.3.5 Splunk Webserver

Splunk provides a web user interface using a Python-based application server. It allows users to search and navigate data stored by Splunk servers and to manage the Splunk deployment.

4.3.6 Deployment monitor

Although it's actually an app, not a Splunk Enterprise component, the deployment monitor has an important role to play in distributed environments. Distributed deployments can scale to forwarders numbering into the thousands, sending data to many indexers, which feed multiple search heads. The deployment monitor can be used to view and troubleshoot these distributed deployments and it provides numerous views into the state of the forwarders and indexers.

4.3.7 Lenovo XClarity app

The Lenovo XClarity app for Splunk allows events to be forwarded from XClarity to the to the Splunk server listener. History and trends for different event can be viewed using built-in user interface.

4.3.8 Other apps

Because Splunk provides a rich RESTful interface into its data and functionality, there are a large number of Splunk and third party provided applications and add-ons.

4.4 Operational model scenarios

The following scenarios are considered in this chapter:

- Departmental server
- Small enterprise (1/4 rack)
- Medium enterprise (1/2 rack)
- Large enterprise (full rack)

Below is a list of items that can have a significant impact on Splunk Enterprise performance.

- Amount of incoming data – increases processes time
- Amount of indexed data – increases I/O bandwidth needed to store and search on data
- Number of concurrent users performing searches, creating reports, or viewing dashboards
- Number and types of searches
- Number and unique performance, deployment, and configuration considerations for each Splunk app

Table 1 below gives sizing information for Splunk Enterprise and shows how many search heads and indexers are needed for different combinations of incoming data size and number of concurrent users. This table is taken from the Splunk Capacity Planning website: docs.splunk.com/Documentation/Splunk/7.1.1/Capacity/Summaryofperformancerecommendations.

Table 1: Splunk Performance Recommendations

Users	< 2GB per day	2 to 300 GB per day	300 to 600 GB per day	600GB to 1TB per day	1 to 2TB per day	2 to 3TB per day
Less than 4	1 combined instance	1 combined instance	1 Search, 2 Indexers	1 Search, 3 Indexers	1 Search, 7 Indexers	1 Search, 10 Indexers
Max 8	1 combined instance	1 Search, 1 Indexers	1 Search, 2 Indexers	1 Search, 3 Indexers	1 Search, 8 Indexers	1 Search, 12 Indexers
Max 16	1 Search, 1 Indexers	1 Search, 1 Indexers	1 Search, 3 Indexers	2 Search, 4 Indexers	2 Search, 10 Indexers	2 Search, 15 Indexers
Max 24	1 Search, 1 Indexers	1 Search, 2 Indexers	2 Search, 3 Indexers	2 Search, 6 Indexers	2 Search, 12 Indexers	3 Search, 18 Indexers
Max 48	1 Search, 2 Indexers	1 Search, 2 Indexers	2 Search, 4 Indexers	2 Search, 7 Indexers	3 Search, 14 Indexers	3 Search, 21 Indexers

More data and more users can be supported by adding more search heads and indexers using the Splunk Enterprise scale-out architecture.

4.5 Servers

There are three kinds of servers for Splunk:

- Indexer
- Search head
- Deployment server

For very small deployments the search head can be combined into the indexer. For medium to large deployments a separate deployment server is needed which can also support license management for the Splunk system. Each section below explores the Lenovo recommended configuration for the three kinds of compute servers.

For indexers and search heads, the required VM sizing is 12 vCPU and 12 GB of RAM without any overcommit for either vCPU or memory. For deployment servers, the required VM sizing is 2 vCPU and 4 GB or RAM. Table 2 shows how each of the four deployment scenarios is mapped to a specific data and user combination.

Table 2: Mapping of Deployment Scenarios

Attribute	Departmental	Small Enterprise	Medium Enterprise	Large Enterprise
Incoming data per day	Less than 2 GB	300 to 600 GB	600GB to 1TB	1 to 2TB
Concurrent users	Less than 4	Maximum 16	Maximum 24	Maximum 48
Search Heads	Combined instance	1	2	3
Indexers		3	6	14
Deployment server	N/A	N/A	1	1
Total vCPU	12	48	96 + 2	204 + 2
Total memory (GB)	12	48	96 + 4	204 + 4

4.5.1 Indexer

An indexer needs to store a large amount of local data and each indexer can roughly handle 300GB of data per day. Table 3 lists the recommended Indexer storage configurations for each of the 4 deployment scenarios to store all the data including hot, warm and cold archival data.

Table 3: Indexer storage configurations

Attribute	Departmental	Small Enterprise	Medium Enterprise	Large Enterprise
Indexers	combined	3	6	14
Required storage (hot, warm and cold)	5.4TB	10.8TB	41.6TB	83.1 TB
Archival storage	14.5TB	29.1TB	145TB	290TB
Required storage +20%	23.88TB	47.88TB	223.92TB	447.72TB
Storage per indexer	23.88TB	15.96TB	37.32TB	31.98TB
Storage block configuration	Model CP-SB-D20 (with 1.6TB drives)	Model CP-SB-D20 (with 3.2TB drives)	Model CP-SB-D20 (with 6.4TB drives, 2 storage blocks)	Model CP-SB-D20 (with 12.8TB drives, 2 storage blocks)
Raw storage capacity	38.4TB	76.8TB	307.2TB	614.4TB
Usable storage capacity	28.8TB	57.6TB	230.4TB	460.8TB

4.5.2 Search head

The recommended configuration is included in Table 2 above.

.

4.5.3 Deployment server

The deployment and license server can use low performance processors. In order to provide redundancy for search heads, it is recommended to simply use the same configuration as a search head.

4.6 Racks and PDUs

For details of the Lenovo rack cabinets and power distribution units (PDUs) that can be used with ThinkAgile CP series, see:

https://lenovopress.com/lp0872-lenovo-thinkagile-cp-series-cloud-platform.

Resources

For more information, see these resources:

- Lenovo ThinkAgile
  http://www.lenovo.com/thinkagile
- Lenovo Data Center Solution Configurator (DCSC):
  http://dcsc.lenovo.com
- Lenovo Data Center Support
  http://datacentersupport.lenovo.com
- Reference Architecture: Splunk Enterprise with ThinkSystem Servers
  https://lenovopress.com/lp0908-reference-architecture-splunk-enterprise-with-thinksystem-servers

Document History

Version 1.0	18 March 2019	• Initial version
Version 1.1	29 March 2019	• Updated configuration information
Version 1.2	26 September 2019	• Updated deployment models

Trademarks and special notices

References in this document to Lenovo products or services do not imply that Lenovo intends to make them available in every country.

Lenovo, and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. These and other Lenovo trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by Lenovo at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of Lenovo trademarks is available from https://www.lenovo.com/us/en/legal/copytrade/.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

Lenovo®

Lenovo XClarity™

Lenovo(logo)®

ThinkAgile™

ThinkSystem™

The following terms are trademarks of other companies:

Xeon, and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.Information is provided "AS IS" without warranty of any kind.

All customer examples described are presented as illustrations of how those customers have used Lenovo products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information concerning non-Lenovo products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by Lenovo. Sources for non-Lenovo list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. Lenovo has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-Lenovo products. Questions on the capability of non-Lenovo products should be addressed to the supplier of those products.

All statements regarding Lenovo future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local Lenovo office or Lenovo authorized reseller for the full text of the specific Statement of Direction.

Some information addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in Lenovo product announcements. The information is presented here to communicate Lenovo’s current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard Lenovo benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the

storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Photographs shown are of engineering prototypes. Changes may be incorporated in production models.

Any references in this information to non-Lenovo websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this Lenovo product and use of those websites is at your own risk.