

# Cloud Native Splunk Enterprise with SmartStore

## Predictive Maintenance for IT Operations Design Guide

### Abstract

This design guide provides architecture and design information for Predictive Maintenance for IT Operations with iDRAC telemetry using Cloud Native Splunk Enterprise with SmartStore.

Dell Technologies Solutions



## Copyright

© 2022 Dell Inc. or its subsidiaries. All rights reserved. Dell Technologies, Dell, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners.

# Contents

Revision history.....	5
<b>Chapter 1: Introduction.....</b>	<b>6</b>
Solution introduction.....	6
Overview.....	6
Document purpose.....	6
Audience.....	7
What's new?.....	7
Design Guide introduction.....	7
<b>Chapter 2: Solution concepts and requirements.....</b>	<b>8</b>
Overview.....	8
Solution concepts.....	8
Predictive Maintenance for IT Operations architecture.....	8
Solution requirements.....	10
<b>Chapter 3: Solution architecture.....</b>	<b>11</b>
Overview.....	11
Container network design.....	11
Operating components.....	11
Container communications.....	11
Robin CNP software defined network.....	12
Splunk networking.....	12
Physical network design.....	13
PowerSwitch networking.....	13
Design principles.....	13
Resilient networking.....	13
Network design.....	14
PowerSwitch configuration.....	15
High availability and load balancing.....	15
Server configurations.....	16
Overview.....	16
PowerEdge servers.....	16
Infrastructure nodes design.....	16
Worker nodes design.....	17
Storage design.....	18
ECS EX-Series storage.....	18
Storage design principles.....	18
Workload-specific storage architecture.....	19
Robin Cloud Native Storage.....	19
Container architecture.....	20
Overview.....	20
Robin CNP prerequisites.....	20
Container design.....	20

<b>Chapter 4: Predictive maintenance use case validation.....</b>	<b>24</b>
Overview.....	24
Software components.....	24
Platform validation.....	26
Deploy platform services.....	26
Deploy Deep Learning Toolkit.....	30
Use case validation.....	30
Live dataflow.....	30
SmartStore integration with ECS.....	34
Failure forecasting.....	36
Summary.....	38
<b>Chapter 5: Implementation, best practices, and sizing.....</b>	<b>39</b>
Overview.....	39
Splunk best practices.....	39
Baseline deployments.....	39
Production deployments.....	39
Extreme production deployments.....	40
Splunk integration best practices.....	40
SmartStore configurations.....	40
Configure indexers and search heads.....	43
Configure forecasting.....	46
Configure ECS.....	47
Configure the Robin CNP storage class.....	47
Protect PVCs with volume replication.....	47
Configure a remote client.....	48
Streaming with HAProxy.....	48
Predictive Maintenance for IT Operations specifics.....	49
Scaling up and out.....	51
<b>Chapter 6: Summary.....</b>	<b>52</b>
Overview.....	52
We value your feedback.....	52
<b>Chapter 7: References.....</b>	<b>53</b>
Dell Technologies documentation.....	53
Splunk Enterprise documentation.....	53
Intel documentation.....	54
Robin.io documentation.....	54
Dell Technologies Customer Solution Centers.....	54
Dell Technologies Info Hub.....	55
More information.....	55

# Revision history

**Table 1. Document revision history**

<b>Date</b>	<b>Document revision</b>	<b>Description of changes</b>
September 2022	1.0	Initial release

# Introduction

## Topics:

- [Solution introduction](#)
- [What's new?](#)
- [Design Guide introduction](#)

## Solution introduction

### Overview

Splunk is an analytics platform that enables you to gain visibility and insight into data, quickly and at scale, in order to take meaningful action for business and operational advantages.

Predictive maintenance is a methodology that uses data analytics and machine learning algorithms to make predictions about future outcomes, such as determining the likelihood of equipment and machinery breaking down. Using a combination of data, statistics, and machine learning and modeling, predictive maintenance can optimize when and how to run maintenance on IT or industrial machine assets.

Organizations are concerned with two types of maintenance: planned and unplanned. Planned or preventative maintenance happens before a failure occurs or prevents failures from happening. Unplanned or reactive maintenance happens when parts fail and need replacing. This type of maintenance is costly, as it affects operations due to asset availability and can impact customer service. The goal of forward-looking organizations is to shift from reactive to preventative to predictive maintenance; optimizing their operational availability and performance.

### Use cases

Splunk is used for a wide range of use cases, including:

- Advanced threat detection
- Application modernization
- Automation of the security operations center
- DevOps monitoring
- Incident investigation and forensics
- Insider threat detection
- IT modernization through end-to-end visibility

Splunk is used across nearly every industry, including aerospace and defense, energy and utilities, financial services, healthcare, higher education, and the public sector, to name a few. This guide presents a validated design for Splunk Enterprise on containers that is applicable to any of these industries and use cases. It focuses on the use case of predictive maintenance in IT operations, using Integrated Dell Remote Access Controller (iDRAC) telemetry data.

### Document purpose

The purpose of this document is to describe a validated design and reference architecture for Splunk Enterprise, delivered in containers with Kubernetes orchestration, running on Dell infrastructure. The document also describes the design and validation of a use case for predictive maintenance in IT operations.

For more information, see the companion document to this design guide: *Cloud Native Splunk Enterprise with SmartStore - Predictive Maintenance for IT Operations White Paper*.

**NOTE:** The contents of this document are valid for the described software and hardware versions. For information about updated configurations for newer software and hardware versions, contact your Dell Technologies sales representative.

## Audience

This document is intended for personnel who evaluate, acquire, manage, maintain, or operate Splunk Enterprise environments, including IT managers and administrators, IT architects, and system and storage administrators. In addition to general Splunk use cases, this guide is useful for IT operations personnel that are focused on predictive maintenance for IT infrastructure.

## What's new?

This design includes the following new content as compared to the previous reference architecture from Dell Technologies entitled *Ready Solutions for Data Analytics: Splunk Enterprise on Dell Infrastructure Reference Architecture Guide*:

- Splunk Enterprise 8.2.3.3
- Implementation for Splunk on containers, rather than bare metal, on a cloud-native Kubernetes platform
- Support for Splunk SmartStore data management capability. Splunk SmartStore enables indexed data to reside either locally or on an external or remote storage tier for scalability and performance.
- Validation of containerized applications: Splunk Machine Learning Toolkit (MLTK), Deep Learning Toolkit (DLTK), and Splunk Essentials for Predictive Maintenance
- The latest Dell infrastructure, including PowerEdge servers and PowerSwitch network switches
- Support for Dell Elastic Cloud Storage (ECS), an enterprise-class object storage platform used for SmartStore-based storage tiering and storage of cold-bucket indexed data.
- Focus on predictive maintenance as a primary use case, using data from the Integrated Dell Remote Access Controller (iDRAC) interface on PowerEdge servers as a sample dataset.

## Design Guide introduction

This design guide first presents some of the key concepts of predictive maintenance in the context of IT operations environments, along with the goals that the underlying design is intended to achieve. Then it describes the solution architecture for cloud native Splunk Enterprise with SmartStore, including the container network design, the physical network design, the server configurations, the storage design, and the container architecture.

In this design, Splunk Enterprise runs in containers with Kubernetes orchestration. The container platform that we have chosen is the Dell Validated Design for Analytics — Data Lakehouse. It includes the Robin Cloud Native Platform (Robin CNP) from Robin.io as the Kubernetes layer.

Then the predictive maintenance use cases are described, showing how using iDRAC telemetry data from Dell PowerEdge servers with Splunk can be used to gain predictive insights for preventative maintenance.

Lastly, Dell guidelines for implementation and deployment, best practices for Splunk and SmartStore implementation for this application, and infrastructure sizing recommendations are provided.

**NOTE:** Other Dell Validated Designs for container platforms, including Red Hat OpenShift Container Platform and VMware Tanzu, can be used with similar results. However, Dell Technologies has not validated these container platforms with the predictive maintenance use case as of the publication of this design guide.

# Solution concepts and requirements

## Topics:

- [Overview](#)
- [Solution concepts](#)
- [Solution requirements](#)

## Overview

This chapter presents some of the key concepts of predictive maintenance as an example use case, and the requirements that were established for this design.

## Solution concepts

### Predictive Maintenance for IT Operations architecture

Dell servers include Integrated Dell Remote Access Controller (iDRAC) modules, which transmit telemetry to external predictive analytics agents:

- The telemetry enables the agents to comprehend current and future failure possibilities.
- The agents act to mitigate failures before they surface.

This type of maintenance improves Mean Time to Failure (MTTF), Mean Time Between Failures (MTBF), and Mean Time to Recovery (MTTR).

The iDRAC telemetry specification is enriched as new failure models emerge:

- Relevant data becomes available for external agents to consume.
- The agents analyze the data for correlated anomalies and make forecasting decisions that predict individual or collective IT infrastructure failure.
- The agents use classified fault predictions to triage actions that keep the infrastructure safe, secure, reliable, highly available, and error-free.

In this control system, the substrate that consists of IT infrastructure components continuously sends data. The agents work on the data to forecast and mitigate substrate failures.

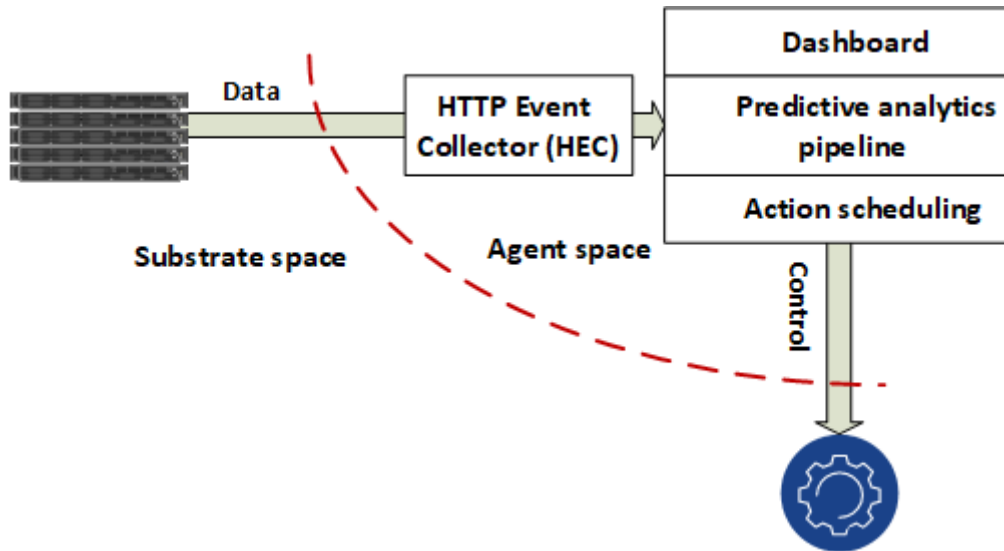
### Analytics agent workload

The agent workload consists of three operations:

- Anomaly detection
- Forecasting
- Action scheduling

Anomaly detection labels correlated anomalies for a fault group before forecasting decisions are made. The scheduler uses a time-to-failure estimate in the forecasting backend to triage actions. Substrate segmentation is based on the need for the agents, which can be distributed or accelerated, to respond in real time or near-real time. Deep-learning model training and retraining interleave with runtime execution. The figure below shows these interactions.





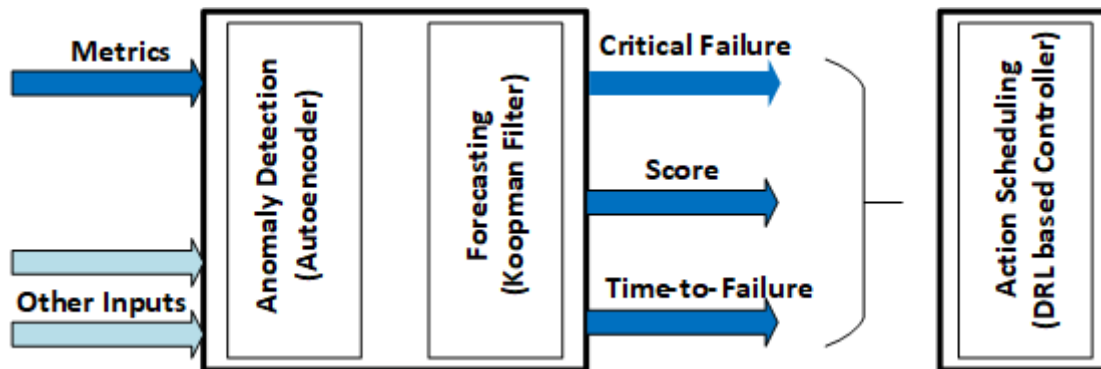
**Figure 1. Substrate-agent interaction**

The agents use native Splunk services such as:

- Splunk Data Stream Processor for data streaming
- Grafana for Splunk native data visualization
- HTTP Event (HEC) Collector for smart data ingestion
- Deep Learning Toolkit (DLTK) for data science

Robin.io includes these services in the Robin Cloud Native Platform on which Dell has based the Dell Validated Design for Analytics — Data Lakehouse. Remote iDRAC clients stream data that is encapsulated in HTTP events, which the Splunk HEC consumes.

Splunk SmartStore indexes and tiers the ingested data, which the agents consume using the Splunk DLTK. The agents schedule actions to be taken according to the data as alerts, which you can see in Grafana. Hot data lives in local storage, while warm data is sent to ECS object storage. For more information, see the Splunk document, [Managing Indexers and Clusters of Indexers](#).



**Figure 2. Predictive analytics model flow**

You can scale as follows:

- **Storage scaling**—Using a node or drive hot plug-in
- **Performance scaling**—Adding nodes or by integrating GPU slices
- **Fault scaling**—Adding new cells to the fault library as the iDRAC Telemetry specification grows
- **Warm storage scaling**—Adding higher-density drives to Splunk storage nodes

# Solution requirements

This design for Predictive Maintenance for IT Operations using Cloud Native Splunk Enterprise with SmartStore was developed with the following goals.

**Table 2. Solution requirements**

Requirement	Notes
The solution is deployed on the Dell Validated Design for Analytics — Data Lakehouse.	See the <i>Dell Validated Design for Analytics — Data Lakehouse Design Guide</i> for specifics.
Leverage Splunk indexing and searching.	Part of the Splunk core services
Leverage Splunk SmartStore tiering with Dell Elastic Cloud Storage (ECS).	Part of the Splunk core services
Warm data is stored in ECS.	See the <a href="#">ECS documentation</a> for specifics.
Predictive forecasting uses the Splunk Deep Learning Toolkit (DLTK).	The DLTK is in Splunkbase.
Streaming with HAProxy in the data path.	Required in production deployments
Integrate live telemetry.	Required for the Live telemetry test case
Streaming with the Splunk Event Generator (Eventgen) for fault injection.	Required for the Fault injection test case

# Solution architecture

## Topics:

- [Overview](#)
- [Container network design](#)
- [Physical network design](#)
- [Server configurations](#)
- [Storage design](#)
- [Container architecture](#)

## Overview

The components and operations that make up the container ecosystem require network connectivity. The network provides communication with all the other components within the cluster and the ability to respond to incoming network requests. The solution design that this guide describes uses Dell PowerSwitch networking infrastructure.

## Container network design

### Operating components

All hosts have an IP Address on the Data Cluster network. Applications run on hosts that are designated as compute nodes. Each compute node is equipped with resources such as CPU cores, memory, storage, NICs, and GPUs. Kubernetes provides a mechanism to enable orchestration of network resources through the Container Network Interface (CNI) API.

The Robin Cloud Native Platform (Robin CNP) uses both Calico and Single-root I/O Virtualization (SR-IOV) with Open vSwitch CNI drivers. The software-defined network (SDN) that Robin CNP uses is designed to enable network automation and support standard management interfaces and protocols. The Robin CNP CNI drivers allow for the retention of persistent IP Addresses. This design provides high availability and flexibility during management processes such as hardware maintenance or scaling and migration of workloads.

### Container communications

A pod, which is a basic unit of application deployment, consists of one or more containers that are deployed together on the same compute node. A pod shares the compute node network infrastructure with the other network resources that make up the cluster. As service demand expands, additional identical pods are often deployed to the same or other compute nodes.

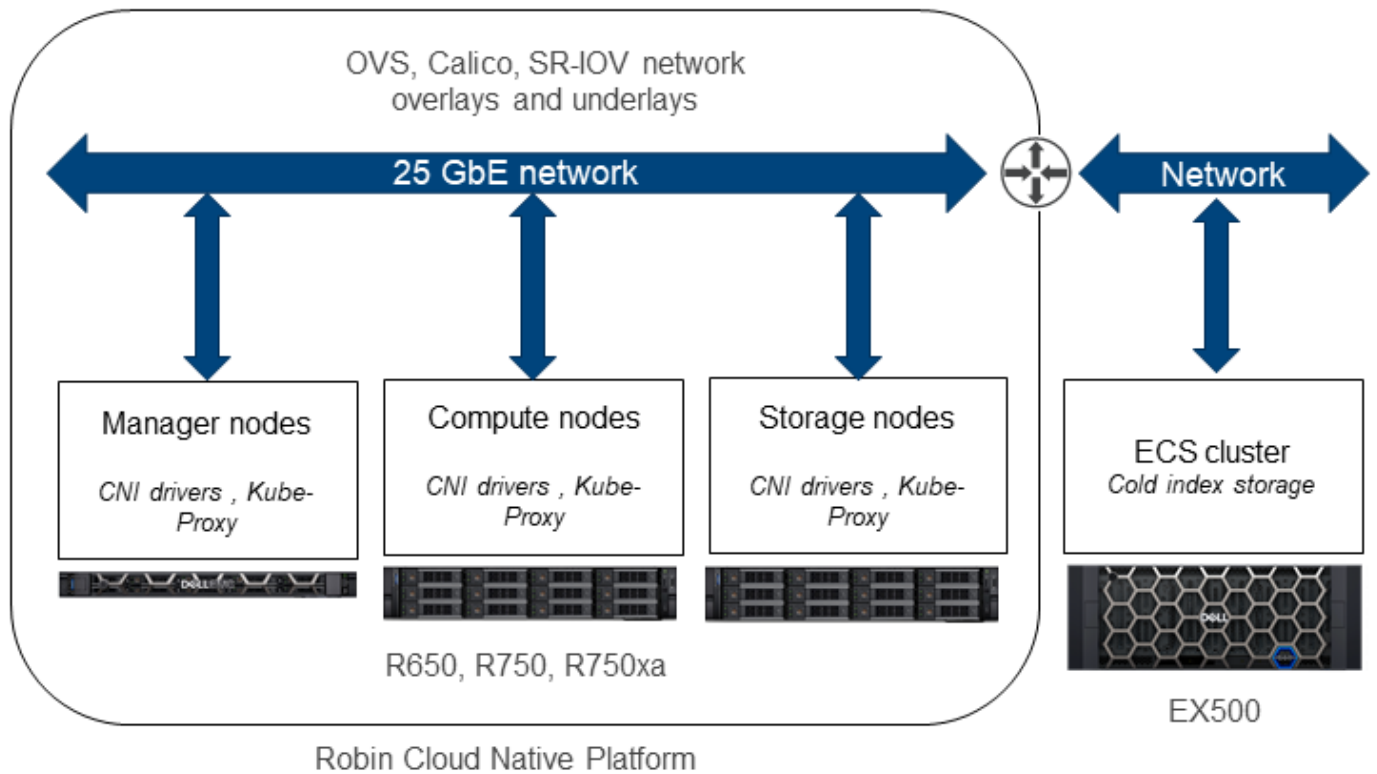
Networking is critical to the operation of a Robin Cloud Native Platform cluster. Three basic network communication flows occur within every cluster:

- Container-to-container connections
- Pod-to-pod connections
- Pod-to-service and ingress-to-service connections (handled by services)

Application containers use shared storage volumes (configured as part of the pod resource) that are mounted as part of the shared storage for each pod. Network traffic that might be associated with nonlocal storage must be able to route across node network infrastructure.

## Robin CNP software defined network

The Robin Cloud Native Platform (Robin CNP) software-defined network (SDN) is a flexible network overlay. It is built on the industry standard Open vSwitch and Calico software-switching technologies. Open vSwitch is a production quality, multilayer virtual switch licensed under the open-source Apache 2.0 license.



**Figure 3. Container network architecture**

For more detailed information see the following Robin.io and Kubernetes documents:

- [Managing Networking](#)
- [Network Policies](#)

## Splunk networking

The HTTP Event Collector (HEC) module in the Splunk core provides a listener service that accepts HTTP(S) connections on the server side, and a client-side API. It enables applications to post log data payloads directly to the indexing tier or a dedicated HEC. The HEC provides two endpoints that support data to be sent either in raw or JSON formats. Utilizing JSON you can add metadata to the event payload to facilitate greater flexibility when searching the data later. Authentication uses a shared secret token. You must specify:

- A service IP address
- Authentication bits
- Indexing metadata

For more details, see the Splunk document, [Splunk Validated Architectures](#).

# Physical network design

## PowerSwitch networking

PowerSwitch networking provides flexible, powerful top-of-rack (ToR) switches for data centers of all sizes. They are designed to deploy modern workloads and applications that are designed for the open networking era. They deliver low latency, superb performance, and high density with hardware and software redundancy.

This validated design uses the S5248F-ON and S3100 family, although other switch models can be used.

## Design principles

Dell PowerSwitch networking products are designed for ease of use and to enable resilient network creation. Dell Technologies recommends designs that apply the following principles:

- Meet network capacity and the separation requirements of the container pod.
- Configure dual-homing to two Virtual Link Trunking (VLT) switches.
- Create a scalable and resilient network fabric.
- Enable monitoring of container communications.

## Container network capacity and separation

Container networking takes advantage of the high-speed (25/100 GbE) network interfaces of the Dell PowerEdge server portfolio. To meet network capacity requirements, pods can attach to more networks by using available Container Network Interface (CNI) plugins.

Robin Cloud Native Platform supports the following network-based functions:

- Multiple interfaces
- Open vSwitch (OVS)
- Single-root I/O Virtualization (SR-IOV) device allocation

For more information, see [Advanced Compute and Networking Support](#) on the [Robin.io documentation website](#).

## Dual homing

Each node that is part of the Robin CNP cluster has at least two NICs, each connected to at least two switches. The switches must have Virtual Link Trunking (VLT) connections so that they operate together as a single unit of connectivity. This configuration provides redundant data paths for all network traffic. The NICs at each node, and the ports that they connect to on each switch, can be aggregated using bonding to assure High Availability (HA) operation.

## Network fabric

The microservices data traffic requires a nonblocking fabric. Dell Technologies recommends that you deploy a leaf-spine network.

## Resilient networking

Each server in the rack is connected to two S5248F-ON leaf switches with 25 GbE network interfaces and one S3148 management switch for iDRAC connectivity.

The Dell network design employs a VLT connection between the two leaf switches. All paths in a Virtual Link Trunking (VLT) environment are active. It is possible to achieve high throughput while still protecting against hardware failures.

VLT technology enables a server to uplink multiple physical trunks into multiple PowerSwitch switches by treating the uplinks as one logical trunk. A VLT-connected pair of switches acts as a single switch to a connecting server. Both links from the bridge network can forward and receive traffic. VLT replaces Spanning Tree Protocol (STP)-based networks by providing both redundancy and full bandwidth utilization using multiple active paths.

The major benefits of VLT technology are:

- Dual control plane for highly available, resilient network services
- Full utilization of the active Link Aggregation Group (LAG) interfaces
- Active/active design for seamless operations during maintenance events

The VLT configuration in this design uses two 100 GbE ports between each Top of Rack (ToR) switch. The remaining 100 GbE ports can be used for high-speed connectivity to spine switches, or directly to the data center core network infrastructure.

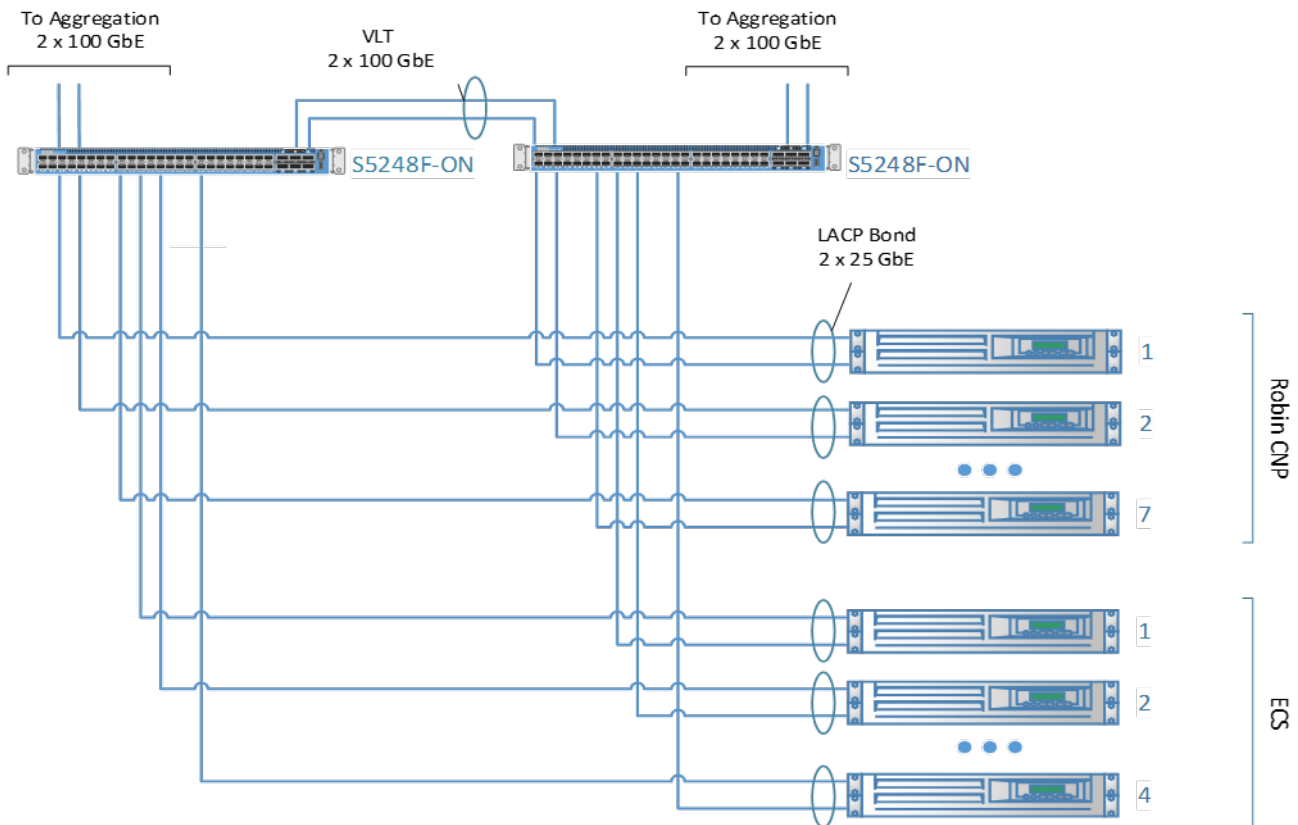


Figure 4. Resilient VLT networking

## Network design

This design includes three logical networks:

- |                             |  |
|-----------------------------|--|
| <b>Corporate network</b>    | The external network is used for the public API, the Robin Cloud Native Platform (Robin CNP) web interface, and applications exposed to the Corporate network.   |
| <b>Cluster Data network</b> | This internal network is the primary, nonroutable network for cluster management, internode communication, and server provisioning using Preboot Execution Environment (PXE) and HTTP. DNS and DHCP services also reside on this network to provide deployment functionality. Network Address Translation (NAT) configured on the bastion node provides communication with the Internet. |
| <b>iDRAC or Baseboard</b>   | The iDRAC or BMC network is a secured, isolated network for switch and server hardware management, including access to the iDRAC9 module and Serial-over-LAN. You can configure optional connections to  |

**Management Controller (BMC) network** Corporate network management, enabling more direct access to the hardware in this design guide. This network is also known as the Out-of-Band (OOB) network.

The figure below shows the Robin Cloud Native Platform logical network components.

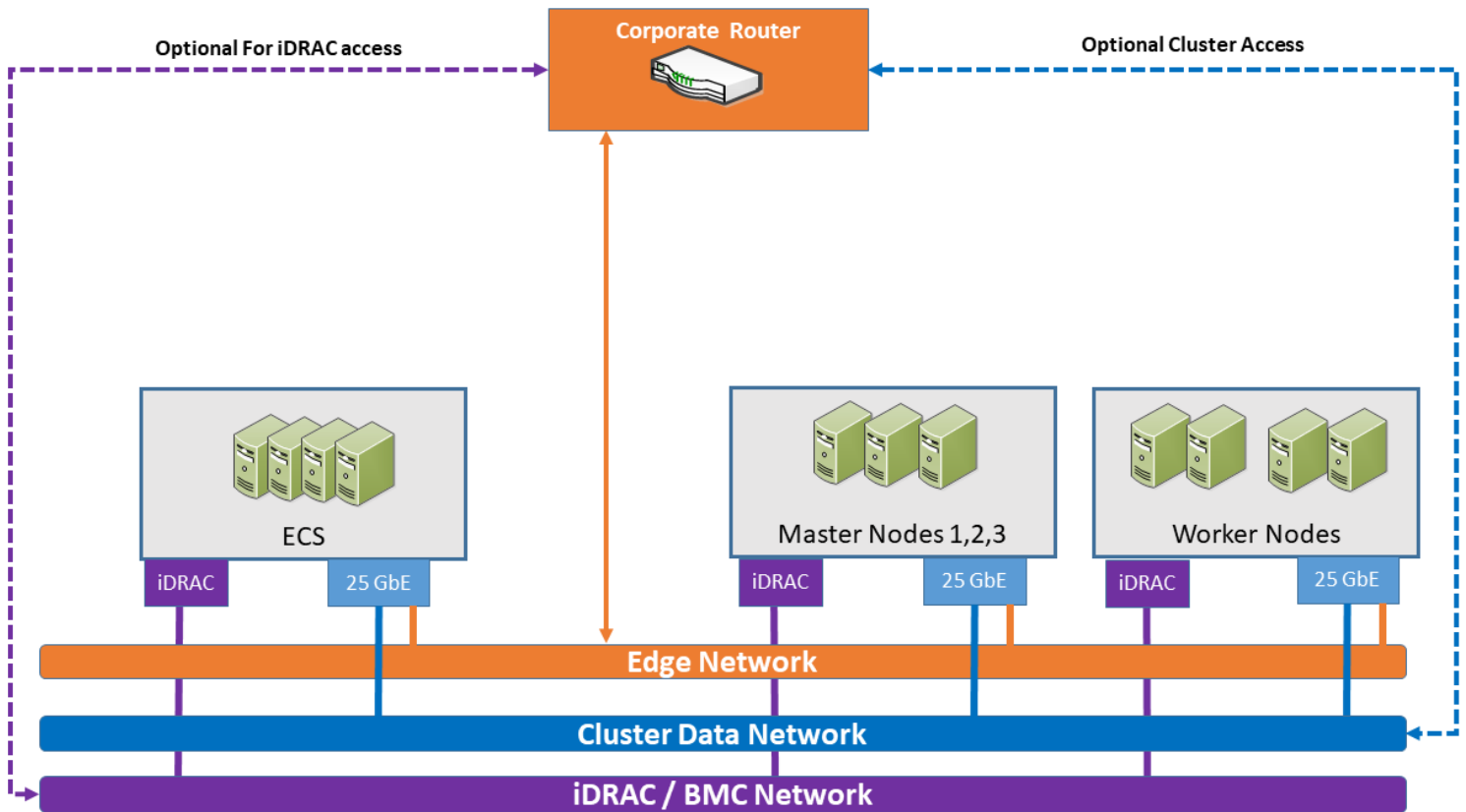


Figure 5. Logical network components

## PowerSwitch configuration

Dell PowerSwitch high-capacity network switches are cost-effective and easy to deploy. They provide a clear path to a software-defined data center, offering:

- High density for 25, 40, 50, or 100 GbE deployments in top-of-rack (ToR), middle-of-row, and end-of-row deployments
- A choice of 25 GbE and 100 GbE switches, including the S5048F-ON; S5148F-ON; S5212F-ON; S5224F-ON; S5248F-ON; S5296F-ON; and S5232F-ON
- A 10 GbE, 25 GbE, 40 GbE, 50 GbE, or 100 GbE modular switch - the S6100-ON
- S6100-ON modules that include:
  - 16-port 40 GbE Quad Small Form-factor Pluggable (QSFP)+
  - Eight-port 100 GbE QSFP28
  - Combination module with four 100 GbE C Form-factor Pluggable (CXP) ports and four 100 GbE QSFP28 ports

This solution design uses the PowerSwitch S5248F-ON configured with Dell SmartFabric OS10. SmartFabric OS10 enables multilayered disaggregation of network functions that are layered on an open-source Linux-based operating system. The following topic describes a high-level configuration of the PowerSwitch switches that are used for a Robin Cloud Native Platform deployment at various scales.

## High availability and load balancing

This design uses the following High Availability (HA) features:

- **Robin Cloud Native Platform**—Multiple control plane nodes and associated infrastructure
- **Resilient load balancing**—Three control plane nodes running HAProxy and Keepalived
- **Dell cloud native infrastructure**—PowerEdge servers with dual NICs
- **Dell PowerSwitch**—Spine-leaf fabric with Virtual Link Trunking (VLT)
- **HAProxy**—Manages API server requests and redirects them in a round-robin manner to the control plane nodes
- **Keepalived**—An open-source project that implements routing software using the Virtual Router Redundancy Protocol (VRRP)
  - If the primary server fails, VRRP enables a switchover to a backup server.
  - This switchover is achieved by using Virtual IP Address (VIP).

See the Robin.io document, [High Availability](#), for more information regarding the HA features of Robin CNP.

## Server configurations

### Overview

Dell Technologies recommends the following server configurations for use with Predictive Maintenance for IT Operations.

### PowerEdge servers

PowerEdge servers are built to support the work of IT organizations. They are engineered to handle the most demanding business applications and are designed with specific features to better run workloads, including:

- High Performance Computing (HPC)
- Collaboration
- Database
- Enterprise Resource Planning (ERP)
- Business intelligence
- Data warehousing and data analytics

As the foundation for a complete and adaptive solution, PowerEdge servers deliver exceptional performance and management advantages, powering the business applications that customers run most often. PowerEdge servers combine with the innovative OpenManage Essentials (OME) systems management portfolio and industry-leading workload solutions to provide intelligent yet simple technology. This combination gives you the power to do more in the most complex environments.

The latest generation of servers responds to customer needs in the following areas:

- **Memory capacity and scalability**—Larger memory footprints
- **Virtualization performance**—More processor cores and denser memory
- **System management**—Complete life cycle management using the integrated Remote Access Controller (iDRAC) with Lifecycle Controller, and monitoring and updating capabilities by using OME
- **Energy efficiency**—Comprehensive optimizations, including OpenManage Essentials Power Center
- **Infrastructure flexibility**—Innovations like select network interface controllers, offering more and better I/O options
- **Reliability**—More Remote Access Service (RAS) features, including a failsafe hypervisor option on most servers

This validated design uses the PowerEdge R650, PowerEdge R750, and PowerEdge R750xa servers.

### Infrastructure nodes design

Infrastructure nodes are used to host Robin Cloud Native Platform, the Kubernetes control plane, and HAProxy and `keepalived` for load balancing. Predictive Maintenance for IT Operations requires three infrastructure nodes.

Dell Technologies recommends the configuration below as a starting point. It is optimized for reliability and provides high performance.



**Table 3. Infrastructure nodes configuration**

Machine function	Component
Platform	PowerEdge R650
Chassis	2.5" chassis with up to eight hard drives and three PCIe slots
Processor	Two Intel Xeon Gold 6338 2 G, 32 C/64 T, 11.2 GT/s, 48 M cache, turbo, HT (205 W) DDR4-3200
RAM	24 32 GB RDIMM, 3200 MT/s, dual rank, eight 16 Gb BASE
Network daughter card	Nvidia ConnectX-6 Lx dual port 10/25 GbE SFP28, OCP NIC 3.0
Boot configuration	BOSS controller card + with two M.2 sticks 480 GB (RAID 1), full-height with S2 blank
Storage controller	PERC H755 Front front load
Disk - SSD	Six 3.84 TB SSD vSAS mixed use 12 Gbps 512e 2.5in, hot plug, AG drive SED, 3 DWPD
Drive configuration	Configuration 7, unconfigured RAID for HDDs or SSDs (mixed drive types allowed)

## Worker nodes design

Worker nodes provide compute resources for the workloads running on the cluster. They can be optimized for the intended workloads and services running on them.

Dell Technologies recommends the configuration below as a starting point. It is optimized for reliability and provides high performance.

**Table 4. Worker nodes configuration**

Machine function	Component
Platform	PowerEdge R750
Chassis	2.5" chassis with up to eight drives
Processor	Two Intel Xeon Gold 6338 2 G, 32 C/64 T, 11.2 GT/s, 48 M cache, turbo, HT (205 W) DDR4-3200
RAM	24 32 GB RDIMM, 3200 MT/s, dual rank, eight 16 Gb BASE
Network daughter card	Nvidia ConnectX-6 Lx dual port 10/25 GbE SFP28, OCP NIC 3.0
Boot configuration	BOSS controller card + with two M.2 sticks 480 GB (RAID 1), full-height, with S2 blank
Storage controller	PERC H755 Front rear load
Disk - SSD for persistent storage	Six 3.84 TB SSD vSAS mixed use 12 Gbps 512e 2.5in, hot plug, AG Drive SED, 3 DWPD,
Drive configuration	Configuration 7, unconfigured RAID for HDDs or SSDs (mixed drive types allowed)

# Storage design

## ECS EX-Series storage

Dell Elastic Cloud Storage (ECS) is an enterprise-class object storage platform that is designed for scalability, performance, resilience, and economics.

Deployable as a turnkey appliance or in a software-defined model, the Dell ECS EX-Series delivers rich S3-compatibility on a globally distributed architecture. It empowers organizations to support enterprise workloads such as cloud-native, archive, Internet of Things (IoT), artificial intelligence (AI), and big data analytics applications at scale.

The benefits of ECS include:

- **Cloud scalability**—A flexible, software-defined architecture promotes limitless scalability that supports both traditional and next-generation workloads.
- **Flexible deployment**—ECS can be deployed as an appliance or as a software-defined storage solution.
- **Reduced TCO**—ECS can dramatically lower total cost of ownership when compared to public cloud storage and traditional storage deployments. Low management overhead, smaller data center footprint, and high storage utilization all help to reduce storage costs.
- **Enterprise-grade technology**—ECS offers:
  - Data-at-rest encryption and encrypted intersite replication traffic
  - Reporting, policy-based and event-based record retention and platform hardening for SEC Rule 17a-4(f)
  - Authentication, authorization, and access controls with Active Directory and Lightweight Directory Access Protocol (LDAP)
  - Integration with monitoring and alerting infrastructure
  - Other enhanced enterprise capabilities such as multitenancy, capacity monitoring, alerting, and multipart uploads

ECS is optimized for a broad range of use cases, including:

- **Secondary storage**—ECS is an excellent option for secondary or tiered storage, enabling organizations to move infrequently accessed data away from more expensive primary storage.
- **Modern applications**—Designed for modern application development, management, and analytics, ECS supports next-generation web, mobile, and cloud applications.
- **Data Lake**—ECS establishes a data lake foundation for organizations of any size. It maximizes the value of user data with powerful HDFS services which enables in-place analytics capabilities that reduce risk, resources, and time-to-results.
- **Storage for IoT**—With no limits on the number of objects, the size of objects or metadata, ECS is the ideal platform to store IoT data.
- **Global content repository**—ECS enables any organization to consolidate multiple storage systems into a single content repository that is globally accessible.
- **Geo-protected archive**—ECS can serve as a secure and affordable on-premises cloud for archival and long-term retention purposes.
- **Video surveillance evidence repository**—ECS makes for a low-cost landing area or secondary storage site for video surveillance data, which has a high capacity footprint per file.

In this design, the ECS EX500 can be used for Splunk SmartStore-based storage tiering, for storage of cold-bucket indexed data.



Figure 6. Dell ECS EX500 chassis

## Storage design principles

Local drives in the PowerEdge servers provide primary storage for the environment. A RAID 1 mirror of two SSDs in each server is used for the operating system and installation. For each server assigned a storage role, the remaining drives are assigned to a

pool of storage managed by the Robin Cloud Native Storage (Robin CNS) layer. This pool is used to provision persistent volume claims made through the Kubernetes Container Storage Interface (CSI). You can organize the storage pool into storage classes with different characteristics, including media type, replication, fault domains, and workload characteristics.

An independent ECS cluster provides secondary storage for the environment. The ECS cluster is used as a remote S3-compatible object store.

## Workload-specific storage architecture

Storage allocation shows the recommended configuration for Splunk Enterprise. This configuration uses Splunk SmartStore for:

- **Index storage**—Providing automatic movement of index data between cold storage on the ECS cluster
- **Warm storage**—Provisioned from primary storage using the Kubernetes CSI

The Splunk indexes must be configured to enable SmartStore and provide the balance that you want between hot and warm storage for index data. The warm storage is provisioned from the object storage class. See [Splunk best practices](#) for more information.

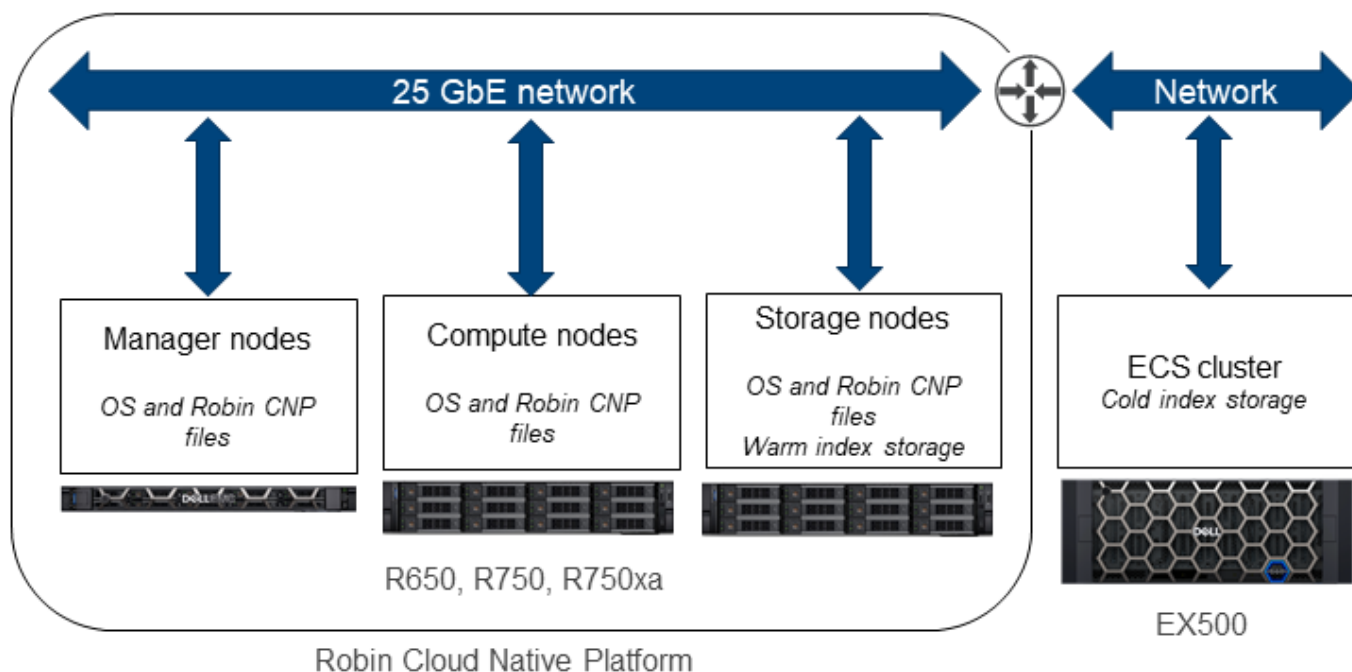


Figure 7. Storage allocation

## Robin Cloud Native Storage

Robin Cloud Native Platform (Robin CNP) includes a software-defined storage that supports a comprehensive set of application-aware services, including snapshots, clone, backup, encryption, and business continuity. All data services are application-aware. They track data storage, metadata, and the Kubernetes application configuration, protecting a wide range of datasets for “application-consistent” disaster recovery of complex network and storage intensive stateful applications.

With Robin Cloud Native Storage (Robin CNS) all life cycle and data protection operations are performed in the context of the entire application aware state (data, metadata, and Kubernetes configuration), not just storage volumes.

# Container architecture

## Overview

Splunk services like Streaming, Visualization, Smart Ingest, and Deep Learning Tools (DLTK) shape predictive maintenance outcomes. The first three services are packaged as Splunk Core Enterprise (SCE) software. DLTK is necessary for predictive maintenance. It includes deep-learning development tools and other machine learning capabilities. Containerized predictive maintenance entails deploying Splunk services on a cloud native environment.

## Robin CNP prerequisites

Requirements for deploying Robin Cloud Native Platform (Robin CNP) in a production environment include:

- At least one administration node for miscellaneous services such as HAProxy
- Three control plane nodes across all production deployment sizes
- One routable network
- One management network
- At least one cluster data network
- At least two Dell PowerEdge worker nodes with x86\_64 CPU architecture
- SSDs in the worker nodes sized per the use case requirement

## Container design

Containerized Splunk is deployed on the Dell Validated Design for Analytics — Data Lakehouse which is based on the Robin Cloud Native Platform (Robin CNP). The approach that this design takes is to:

1. Install containerized Splunk into the Dell Validated Design for Analytics — Data Lakehouse with bundled Splunk Core Enterprise (SCE) services.
2. Deploy peripheral services like Deep Learning Toolkit (DLTK) on which predictive maintenance intellectual property is developed.

Predictive analytics intellectual property is described elsewhere. It includes processing live telemetry from iDRACs attached to servers, for Anomaly Detection (AD) and Forecasting to capture the substrate fault map. You can leverage this fault map to write custom schedulers that shape your substrate needs. [Software stack delineation](#) shows the software stack delineation. Image sizes are approximately:

- **Advanced Deep Learning Algorithms (ADLA)**—1 GB
- **DLTK**—10 GB
- **SCE**—1 GB

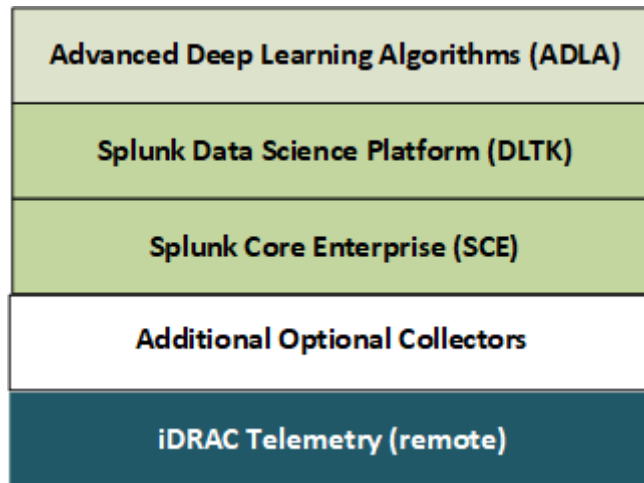


Figure 8. Software stack delineation

## Splunk Core Enterprise

Splunk Core Enterprise (SCE) is a collection of atomic indexers, search heads, and forwarders. Indexers index data, transform raw data into events, and attach them to the index. Search heads serve as the interface for Deep Learning Toolkit (DLTK) applications to use indexed data. Forwarders collect, forward, and load-balance to indexing tiers. External load balancers like HAProxy redirect the flows to multiple forwarders.

The number of indexers is calculated from the use case data rate and the predictive analytics window size. See [Splunk Storage Sizing](#). The indexing caps in the Splunk documentation guide the number of indexers. The number of required search heads is automatically calculated based on Intel benchmarks that show nearly linear growth with a five (indexers)-to-three search heads base. See the Intel document, [Intel Select Solutions: Containerized Splunk](#). The number of forwarders feeding an index cluster is based on the real-time needs of both DLTK and custom scheduling.

**NOTE:** Forwarder segmentation is outside the scope of this design.

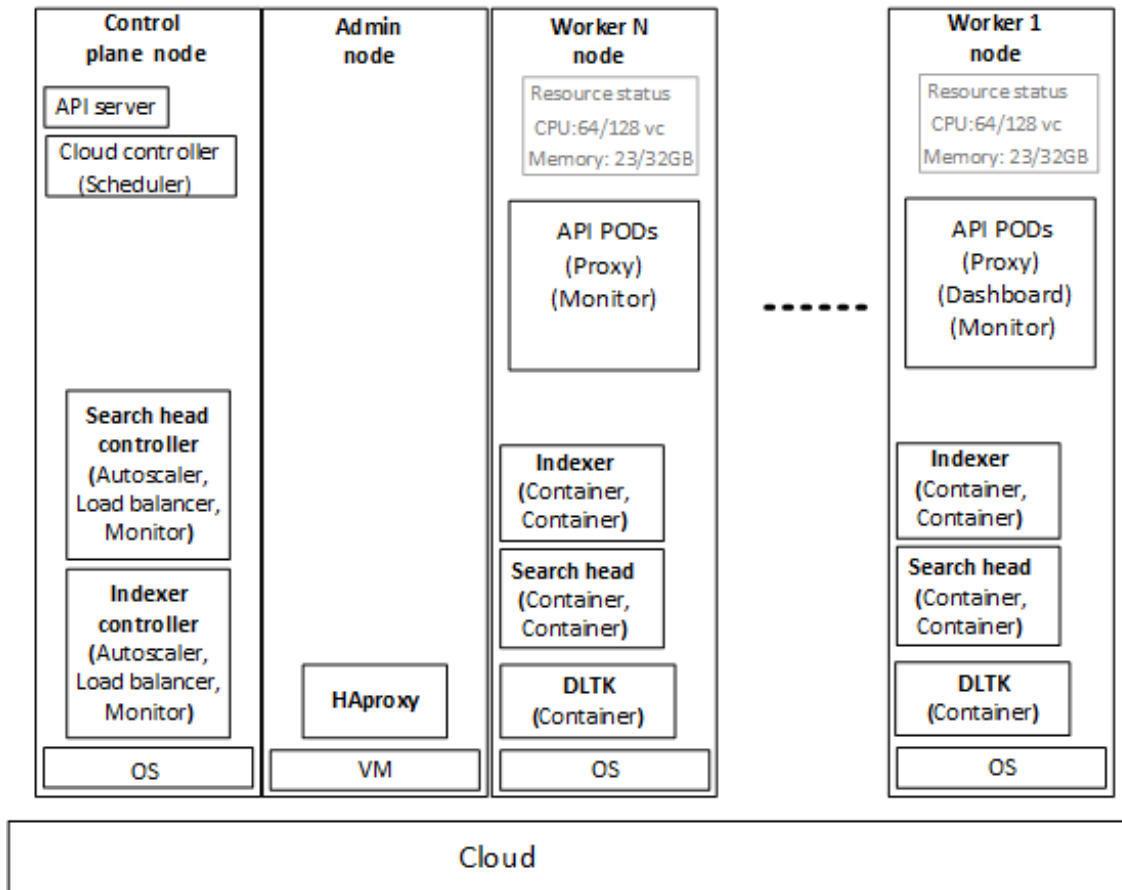
[Splunk container architecture](#) shows the Splunk Kubernetes components:

- Indexers
- Search heads
- Forwarders
- DLTK

The cloud controller managers reside on the control plane node, and the pods are distributed among the worker nodes.

**NOTE:** Calico-based pod networking is not shown.

Splunk DLTK is a collection of open-source tools such as Jupyter, MLflow, Spark, and so on, that are necessary to develop models. DLTK includes Machine Learning Toolkit (MLTK).



**Figure 9. Splunk container architecture**

The figure above shows an example of how worker nodes resource utilization can be viewed, and is not a snapshot:

- 64 out of 128 vCores used
- 23 out of 32 GB RAM used

## Container network framework

The Container Network Framework (CNF) holds the resilient and scalable Predictive Maintenance for IT Operations solution together. The Dell Validated Design for Analytics — Data Lakehouse is based on the Robin Cloud Native Platform (Robin CNP). Robin CNP is based on open-source Kubernetes. See the [Robin CNP documentation](#) for more details.

Robin CNP adds dynamic, software-defined storage provisioning using:

- Robin Cloud Native Storage (Robin CNS)
- Authentication capabilities
- Industry-standard operators and automation utilities

A standard, medium-size deployment includes:

- One administration node
- Three Kubernetes fault-tolerant control plane nodes
- Two to seven worker nodes to host the replicated application container components

Splunk services except Deep Learning Toolkit (DLTK) are bundled in the Dell Validated Design for Analytics — Data Lakehouse. DLTK is deployed separately because at the time of this publication, it is still a community-supported service.

## Container design summary

Dell Technologies made container design choices with the following concepts in mind:

- Three control plane nodes are necessary in order to support high availability (HA). This practice is standard in the cloud-native world.
- A baseline of two, to a maximum of seven, worker nodes saturate the proposed network architecture. For rare, larger deployments you can include more worker nodes if you add spine switches.
- For warm storage, Dell Technologies recommends a five-node ECS EX500 platform. You can scale this platform by increasing the number of drives, or by opting for higher density drives.
- Hot SSD storage requirements are calculated on a per-server basis. The exact number of servers in the substrate can vary, but the platform supports hot plug-ins for upscaling.
- The cluster includes one administration node to support housekeeping tasks such as HAproxy, DNS management, and so on. Production deployments can optionally include a standby administration node.

# Predictive maintenance use case validation

## Topics:

- [Overview](#)
- [Software components](#)
- [Platform validation](#)
- [Use case validation](#)
- [Summary](#)

## Overview

The Dell Validated Design for Analytics — Data Lakehouse provides the base platform for this example use case validation of Predictive Maintenance for IT Operations. It is built upon Robin Cloud Native Platform (Robin CNP), and includes PowerEdge servers, PowerSwitch networking, and ECS storage. PowerEdge servers include iDRAC modules, which transmit server telemetry.

In Predictive Maintenance for IT Operations, external predictive analytics agents receive and act upon the telemetry. These agents use native Splunk services such as:

- Splunk Data Stream Processor for data streaming
- Grafana for Splunk native data visualization
- HTTP Event (HEC) Collector for smart data ingestion
- Deep Learning Toolkit (DLTK) for data science

Dell Technologies ran validation tests that concentrated on:

<b>Software components</b>	Dell Technologies tested this specific use case with the software components that are listed in <a href="#">Validated software components</a> .
<b>Platform validation</b>	Robin CNP platform validation was divided into two parts: <ul style="list-style-type: none"> <li>• Platform services other than Deep Learning Toolkit (DLTK) are bundled at Robin CNP deployment time.</li> <li>• DLTK was deployed separately for this validation. At the time of this publication, it is still a community supported service.</li> </ul>
<b>Use case validation</b>	Splunk services validation was divided into three separate categories: <ul style="list-style-type: none"> <li>• Live dataflow</li> <li>• SmartStore integration with ECS</li> <li>• Failure forecasting</li> </ul>

## Software components

The software components and versions that are validated for this example use case validation of Predictive Maintenance for IT Operations are listed in the tables below. The validated components may not precisely match the Dell Technologies-recommended configurations. For more information, email [ai.assist@dell.com](mailto:ai.assist@dell.com), or contact your Dell Technologies sales representative.

**Table 5. Validated PowerEdge R650 firmware**

Component	Version
PowerEdge BIOS	1.4.4



**Table 5. Validated PowerEdge R650 firmware (continued)**

Component	Version
iDRAC	5.10.10.00
iDRAC service module embedded package	4.2.0.0, A00
Intel E810-XXV	20.5.13
PERC H755 Front	52.14.0-3708
PCIe SSD in slot 0, bay 1	1.2.3
BOSS-S2	2.5.13.4008
Backplane 1	3.57
Broadcom GbE BCM5720	22.00.6
Disk 0 on AHCI controller in slot 6	J004
Trusted Platform Module (TPM)	7.2.2.0
Dell operating system driver pack	21.10.02, A00
Dell 64-bit UEFI diagnostics	4301A66
System Complex Programmable Logic Device (CPLD)	1.0.5
Lifecycle controller	5.10.10.00

**Table 6. Validated PowerEdge R750 firmware**

Component	Version
PowerEdge BIOS	1.5.5
iDRAC	5.10.10.00
iDRAC service module embedded package	4.2.0.0, A00
PCIe SSD in slot 0, bay 1	VDV1DP25
BOSS-S2	2.5.13.4008
Backplane 1	3.57
Broadcom GbE BCM5720	22.00.6
Mellanox NIC 1	26.30.10.04
Mellanox NIC 2	26.30.10.04
Intel Ethernet converged network adapter X710	20.5.13
Intel Ethernet 10 G 4P X710 OCP	20.5.13
Disk 0 on AHCI controller in slot 6	J004
Dell operating system driver pack	21.10.02, A00
Dell 64-bit UEFI diagnostics	4301A68
System CPLD	0.4.2
Lifecycle controller	5.10.10.00

**Table 7. Validated software components**

Category	Component	Version
Server operating system	Red Hat Enterprise Linux Server	8.4
	Linux kernel	4.18.0-305.el8.x86_64
Server file system	XFS	N/A

**Table 7. Validated software components (continued)**

Category	Component	Version
Switch operating system	Dell SmartFabric OS10	N/A
Java Virtual Machine	Open JDK	8 (8u232-b09)
Automation platform	Robin Cloud Native Platform	5.3.11-217
Big data platform	Splunk Enterprise	8.2.3.3
Container engine	Docker	20.10.8
Container platform	Enterprise Kubernetes	1.21.5
Data lakehouse	Delta Lake	1.1.0
Programming language	Python	3.8

## Platform validation

This section describes how Dell Technologies validated the deployment of both Splunk bundled services and Deep Learning Toolkit (DLTK).

## Deploy platform services

### Prerequisites

Prerequisites include:

- Robin Cloud Native Platform (Robin CNP) must be installed on an infrastructure node.
- The Robin CNP web UI must be network-accessible.
- Proper Robin CNP user credentials must be created and assigned.

### About this task

To deploy the platform services:

### Steps

1. Log in to the Robin CNP UI with a web browser, to ensure that it is accessible.  
Your browser displays the Robin CNP Dashboard.

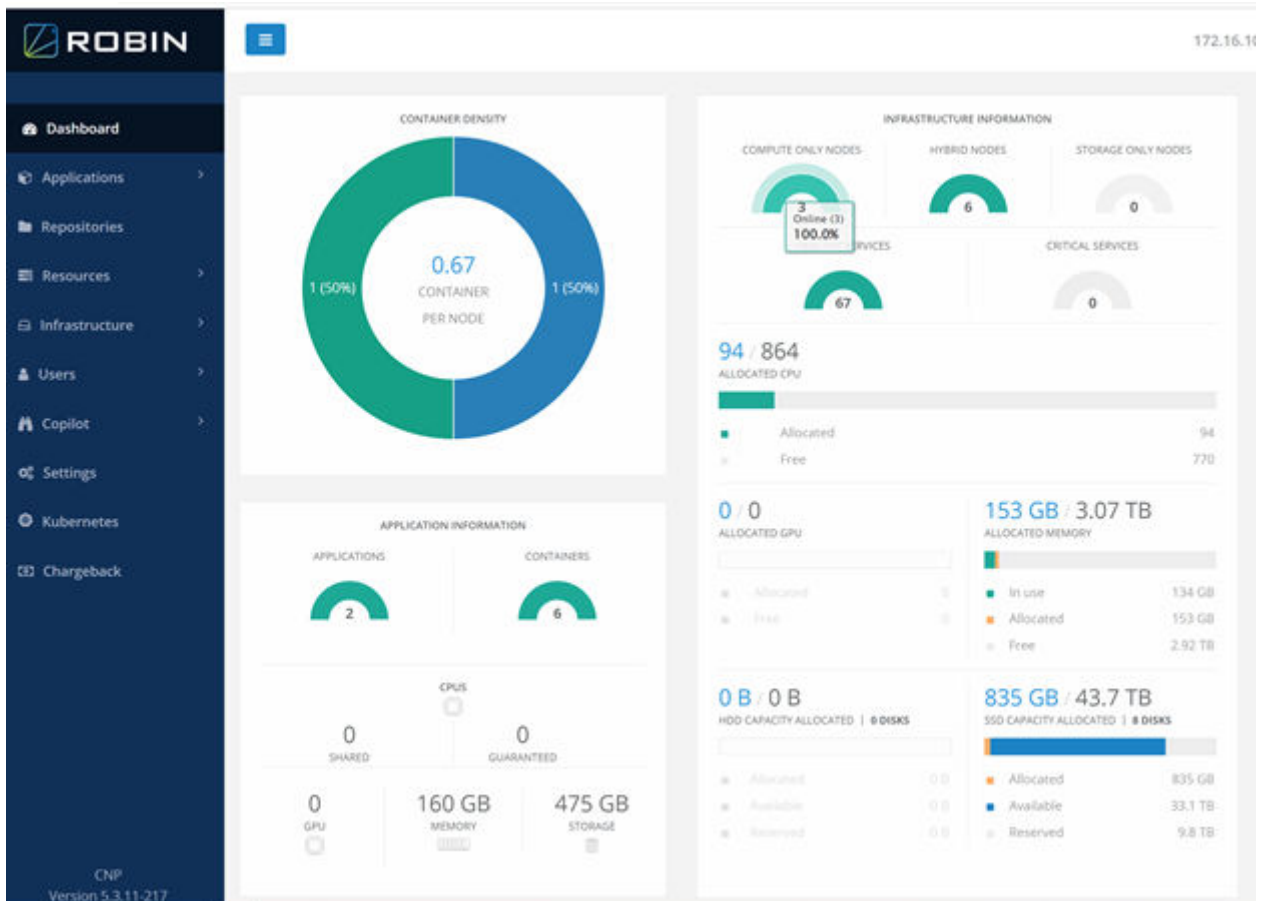


Figure 10. Robin CNP Dashboard

2. Access the Robin CNP infrastructure node node using SSH.
3. Log in by running the `robin login robin_admin` command.
4. Run the `robin cluster info` command to check the cluster status.

The output appears similar to the following example:

Server	#Port	Description
Robin HTTP	29443	HTTPS Port for Robin UI. Default: 29443
Client	29442	The REST port of the Robin server. Default: 29442
Agents REST	29450	The port where rest connections are made to the agents. Default: 29450
Database	29458	The port number of the database. Default: 29458.
Docker	2376	The port where docker TCP connections are made. Default: 2376
Consul HTTP	29462	Consul HTTP Port. Default: 29462

5. From the Robin CNP web UI, browse to **Infrastructure > Nodes** to check the status of physical nodes. Your browser displays the **Nodes** page.

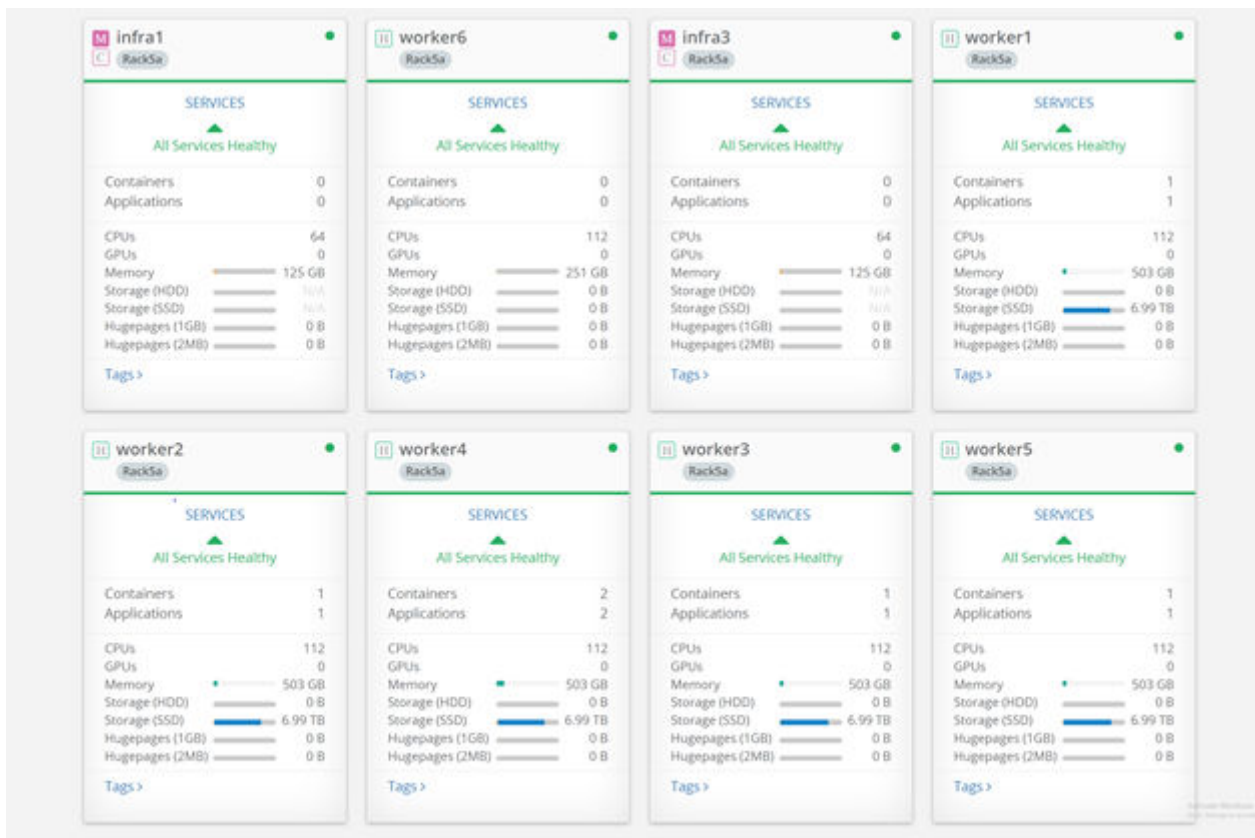


Figure 11. Robin CNP Nodes page

- Build a Robin CNP test bundle by following instructions in the Robin.io document, [Constructing an Application Bundle](#).
- From the Robin CNP web UI, browse to **Applications > Bundles**.  
Your browser displays the **Bundles** page.

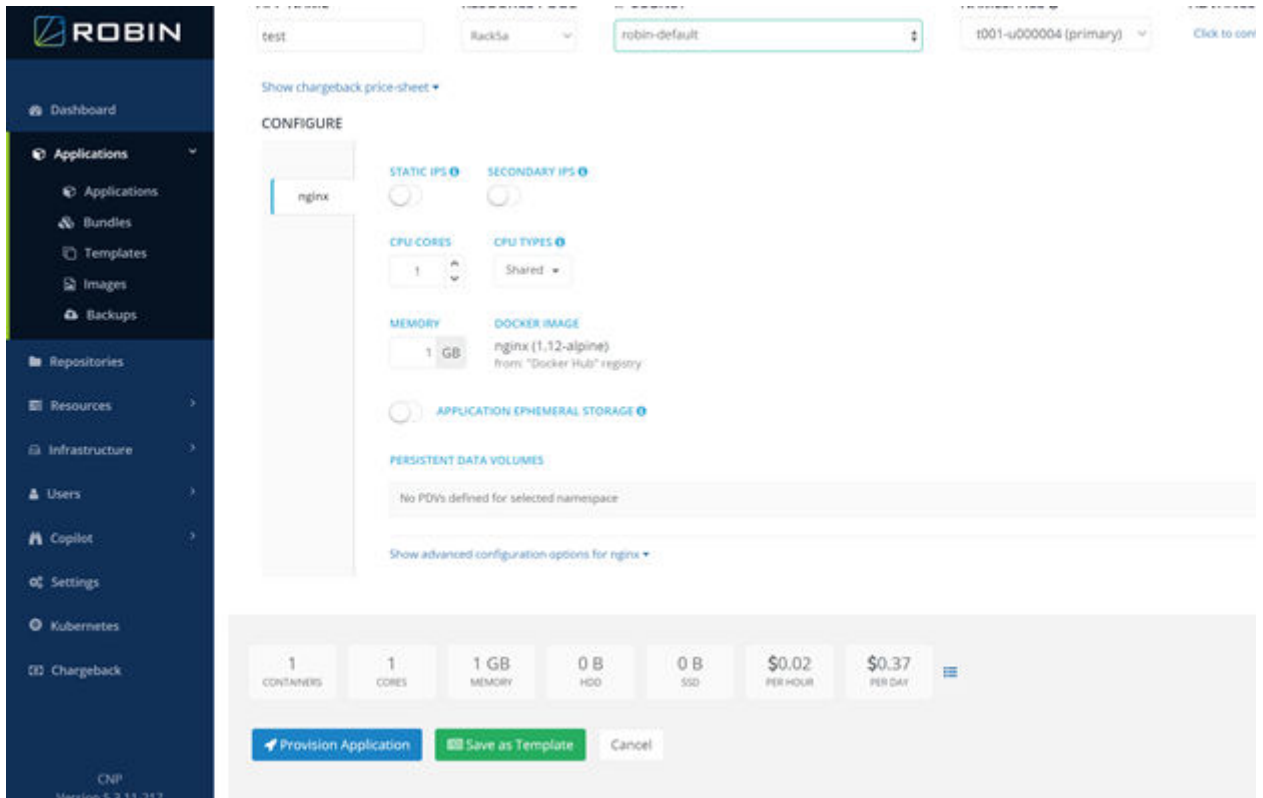


Figure 12. Robin CNP Bundles page

8. Select the bundle that you built in step 6.
9. Click **Provision Application** to deploy the bundle.
10. Browse to **Applications > Applications** to check the newly deployed bundle. Your browser displays the **Applications** page.

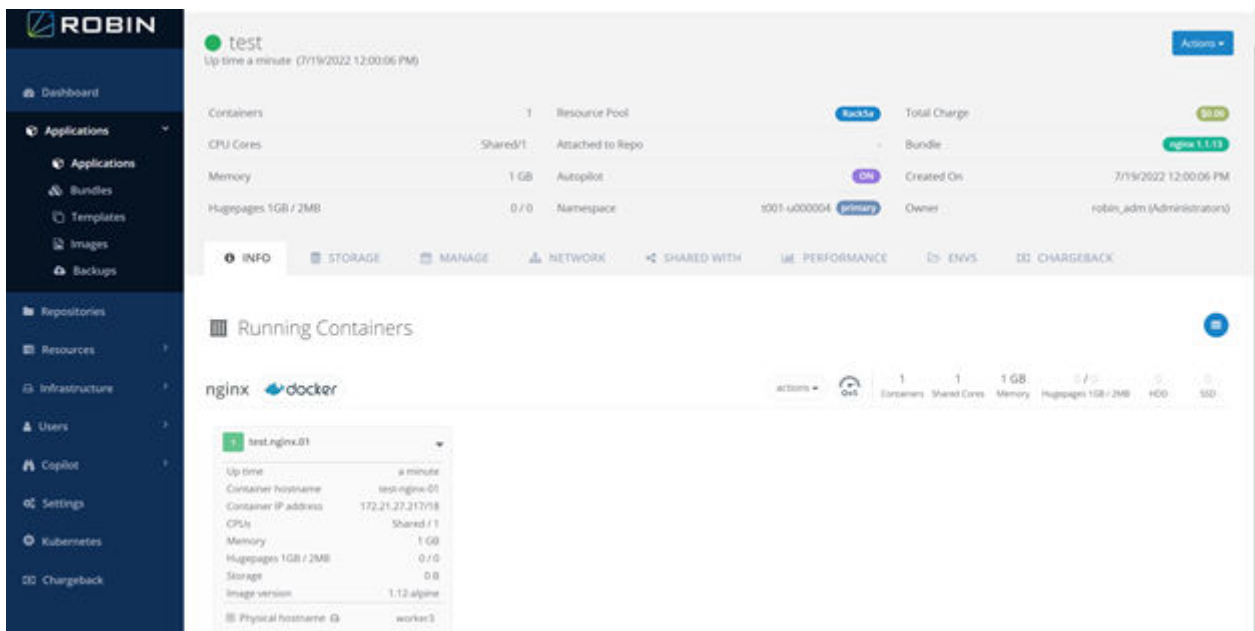


Figure 13. Robin CNP Applications page

11. Click the **down arrow** next to the container name, and then click **Console** to connect to the container.

12. In the console, ping the gateway or DNS server to test network connectivity. For example:

```
# ping 8.8.8.8
```

The output appears similar to the following example:

```
/ # ping 8.8.8.8
PING 8.8.8.8 (8.8.8.8): 56 data bytes
64 bytes from 8.8.8.8: seq=0 ttl=102 time=11.223 ms
64 bytes from 8.8.8.8: seq=1 ttl=102 time=10.521 ms
^C
--- 8.8.8.8 ping statistics ---
2 packets transmitted, 2 packets received, 0% packet loss
round-trip min/avg/max = 10.521/10.872/11.223 ms
/ # |
```

### Results

The bundled platform services are successfully deployed.

## Deploy Deep Learning Toolkit

Once the bundled Splunk services have been installed, deploy Deep Learning Toolkit (DLKT) by following the instructions in the Splunk document, [Installing Deep Learning Toolkit for Splunk](#).

## Use case validation

This section describes how Dell Technologies validated that the system performed the required workloads. Three simulated predictive maintenance use case tests were run: live dataflow, SmartStore integration, and failure forecasting.

### Live dataflow

#### Prerequisites

Prerequisites include:

- Robin Cloud Native Platform (Robin CNP) must be installed on an infrastructure node.
- Deep Learning Toolkit (DLTK) must be installed on the Robin CNP infrastructure node.
- The Robin CNP web UI must be network-accessible.
- Proper Robin CNP user credentials must be created and assigned.
- Server telemetry reports must be streamed from iDRAC hosts to the Splunk application.
- HAProxy must be installed. HAProxy is used as a reverse proxy and load balancer for traffic coming into the Splunk cluster.

#### About this task

To validate that server telemetry reports can be streamed from iDRAC hosts to the Splunk application:

#### Steps

1. Ensure that the cluster status is normal by running the following command in the main Splunk cluster:

```
sudo /opt/splunk/bin/splunk show cluster-status
```

The output appears similar to the following example:

```
[ansible@splunk8-master-01 ~]$ sudo /opt/splunk/bin/splunk show cluster-status
Replication factor met
Search factor met
All data is searchable
Indexing Ready YES

splunk8-indexer-02 A156DD31-817C-427A-BA53-0ADDE2360328 default
  Searchable YES
  Status Up
  Bucket Count=257

splunk8-indexer-03 A1DBF5E5-84F3-4C2E-ADCB-99F6B90E6132 default
  Searchable YES
  Status Up
  Bucket Count=257

splunk8-indexer-01 C1F52B1D-A350-4530-9868-1DF297D3A023 default
  Searchable YES
  Status Up
  Bucket Count=257
```

2. Configure the HAProxy backend to connect to the Splunk REST API by editing the `/etc/haproxy/haproxy.cfg` file as in the following example.

```
frontend manage
  bind *:80
  default_backend splunk_api

backend splunk_api
  server master 172.16.105.48:30313 check ssl verify none
  server indexer1 172.16.105.48:32498 check ssl verify none
```

This configuration validates that HAProxy is functional.

3. Confirm that the backend is reachable by running the following command:

```
curl -I http://<HAProxy FRONTEND IP>:<HAProxy FRONTEND PORT>/
```

The response should be `200 OK`.

4. To enable streaming of metric data from the iDRAC host to Splunk, create a custom metric index in the cluster. This index enables any indexer in the cluster to store metric data. Follow the instructions in the Splunk document, [Get started with metrics](#).

The following example shows a configuration entry for a metric index that is named `idrac_live` in `/opt/splunk/etc/master-apps/_cluster/local/indexes.conf`, located within the Splunk Control node container:

```
[idrac_live]
repFactor = auto
coldPath = $SPLUNK_DB/idrac_live/colddb
datatype = metric
homePath = $SPLUNK_DB/idrac_live/db
thawedPath = $SPLUNK_DB/idrac_live/thaweddb
```

5. Browse to the **Indexer Clustering** page of the Splunk interface on the Control node to verify that the newly created index is displayed.

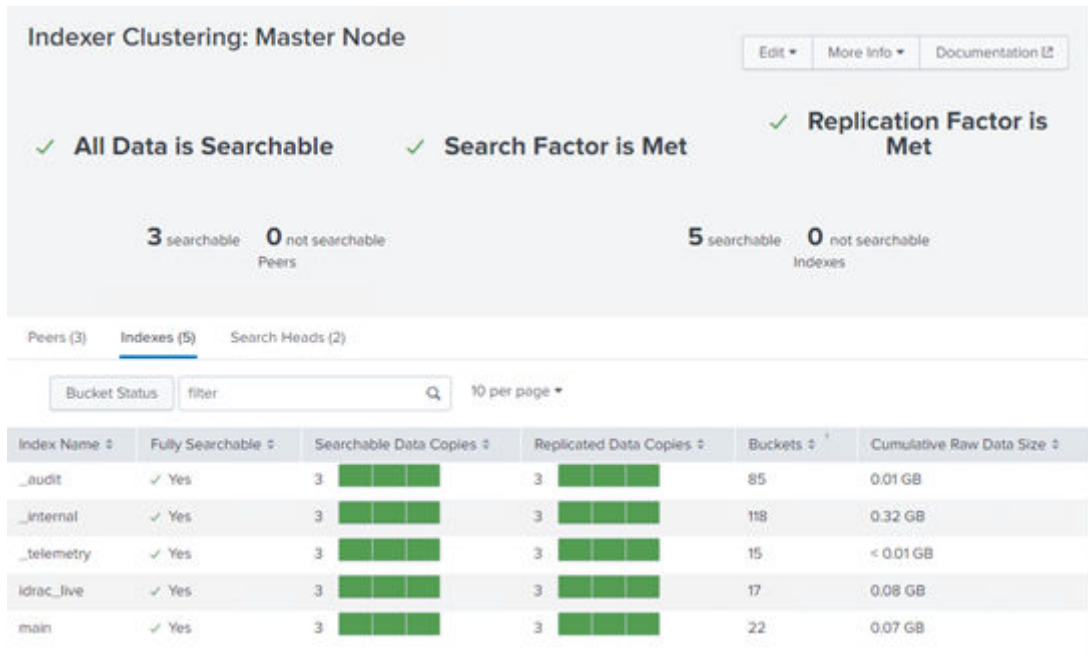


Figure 14. Splunk indexer clustering page

6. To allow data collection through HTTP, configure an HTTP Event Collector (HEC) endpoint on one of the indexer nodes.
7. To verify that the HEC endpoint is reachable, send an HTTP request to the endpoint with sample data, as in the following example:

```
[root@localhost ~]# curl -H "Authorization: Splunk 42f79af4-23cb-4467-b03f-db3fb8d41a8e" http://172.16.105.49:32269/services/collector/event -d '{"sourcetype": "my_sample_data", "event": "this is a sample data"}'
{"text": "Success", "code": 0}[root@localhost ~]#
```

8. Set HAProxy as a load balancer for one or more HEC endpoints.
9. To verify that the sample data was received, indexed, and searchable in the Splunk cluster, query it from the **Search & Report** application on the Splunk Control node.

The output appears similar to the following example:

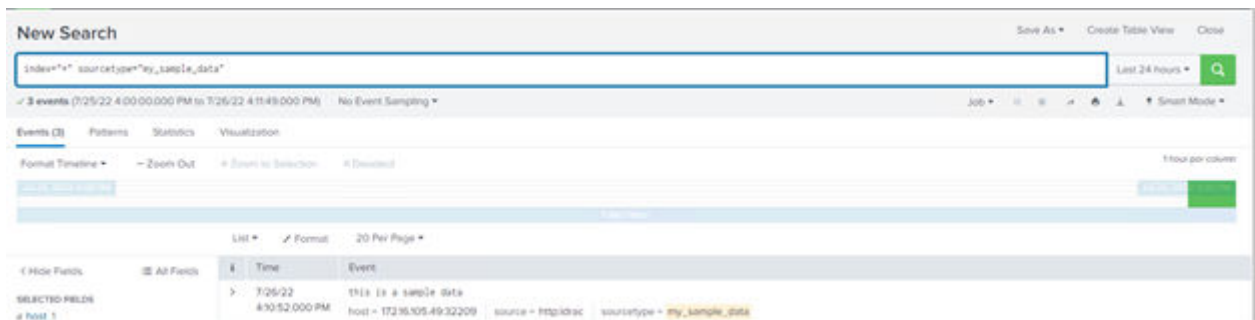


Figure 15. Splunk Search & Report application

10. To subscribe to telemetry reports from iDRAC hosts and forward them to the HEC through HAProxy:
  - a. Download iDRAC Telemetry Reference Tools from <https://github.com/dell/iDRAC-Telemetry-Reference-Tools>.
  - b. Install the tools on a separate server.
11. Verify that the telemetry data have been indexed and is searchable by querying with the `mpreview` and `mstats` commands in the Splunk **Search & Report** application.
  - a. Query the Splunk **Search & Report** application with the `mstats`.

The following example shows a query result for CPU usage telemetry:



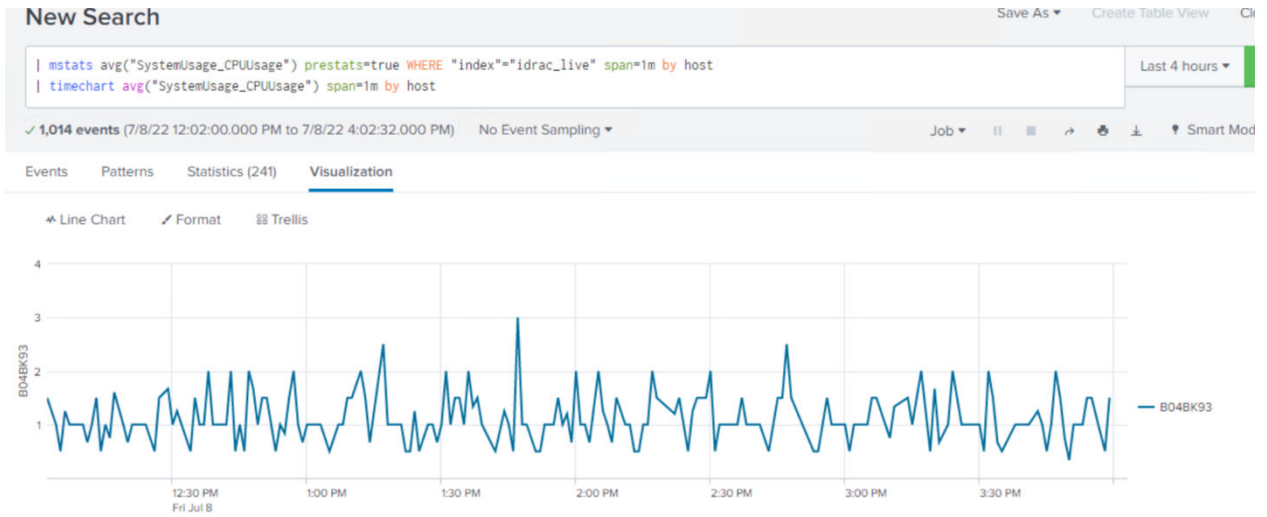


Figure 16. Splunk mstats query

b. Query the Splunk **Search & Report** application with the `mpreview` command.

The following example shows a query result for metric data points that are sent to Splunk:

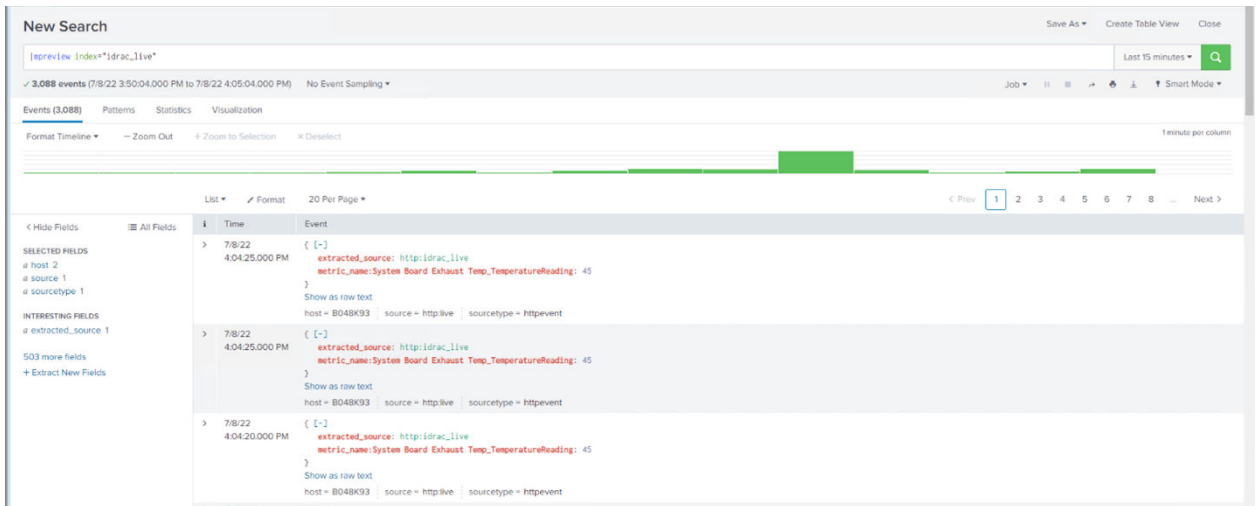


Figure 17. Splunk mpreview query

The same commands can be used to query the data to be sent to other Splunk applications, such as Deep Learning Toolkit (DLTK), for forecasting use cases.

12. Verify that graphic visualizations for selected metrics from multiple hosts are available from the **Analytics** tab in the Splunk **Search & Report** application.

The output appears similar to the following example:

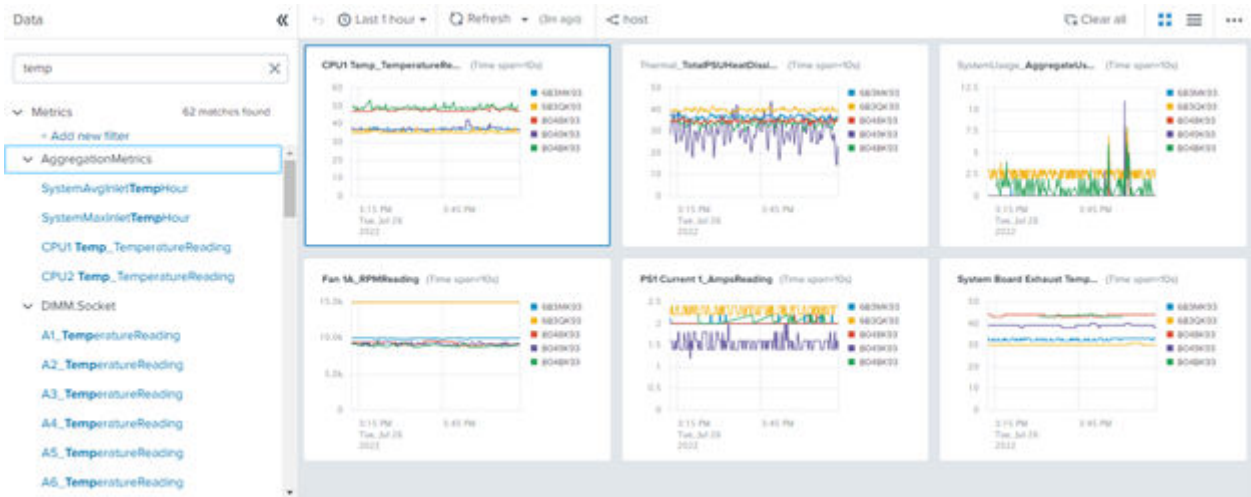


Figure 18. Splunk aggregation metrics

## Results

Dell Technologies successfully validated that server telemetry reports can be streamed from iDRAC hosts to the Splunk application.

## SmartStore integration with ECS

### Prerequisites

Prerequisites include:

- Robin Cloud Native Platform (Robin CNP) must be installed on an infrastructure node.
- Deep Learning Toolkit (DLTK) must be installed on the Robin CNP infrastructure node.
- The Robin CNP web UI must be network-accessible.
- Proper Robin CNP user credentials must be created and assigned.
- Server telemetry reports must be streamed from iDRAC hosts to the Splunk application.
- HAProxy must be installed. HAProxy is used as a reverse proxy and load balancer for traffic coming into the Splunk cluster.

### About this task

To validate that the Splunk SmartStore cache manager can be configured to use Dell ECS as warm and cold data tiering:

### Steps

1. Add ECS as an S3 volume in the cluster.
2. Create an index to use as the remote store for SmartStore.
3. In the Splunk UI, verify that SmartStore has a connection to ECS by browsing to **Settings > Monitoring console > Indexing > SmartStore Activity: instance**.
  - a. Set the SmartStore parameter, `hotlist_recency_secs`, to a duration in seconds.

This setting configures the cache retention period for hot buckets before evicting data to the connected ECS.

4. Stream arbitrary data to the index with ECS tiering using the HTTP Event Collector (HEC) with the **Splunk Event Generator** tool, at <https://splunk.github.io/eventgen/>.
5. Verify that the streamed data is indexed into Splunk.  
The output appears similar to the following example:

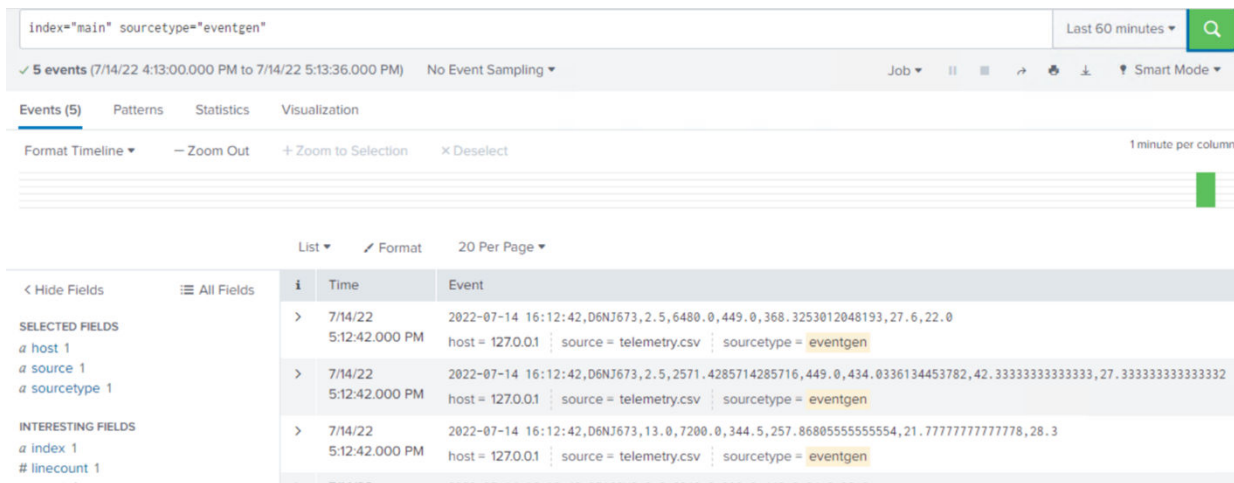
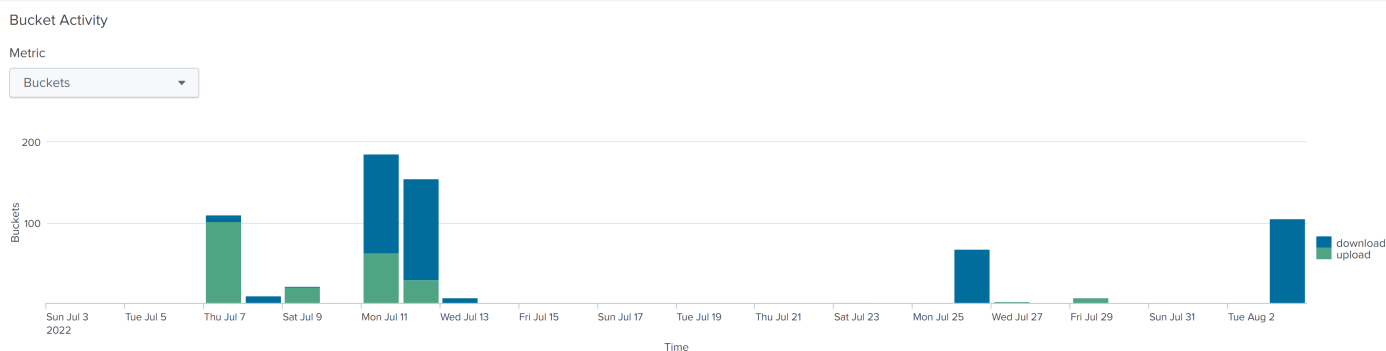


Figure 19. Streamed data indexed into Splunk

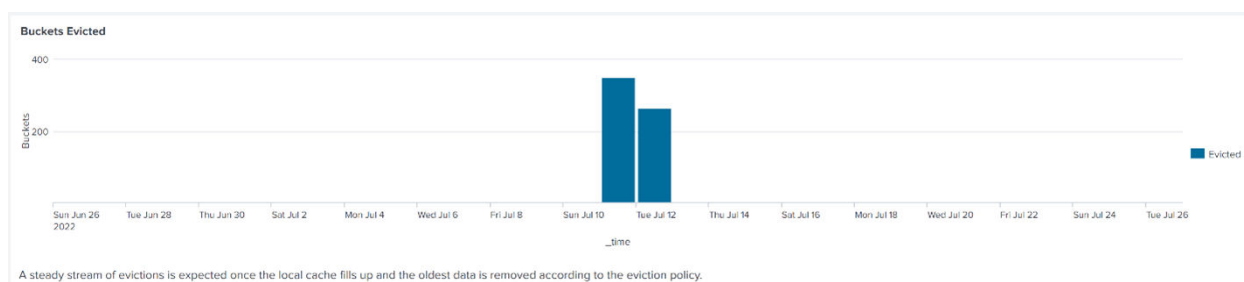
- In the Splunk monitoring console, verify that the data overflows to ECS after the configured retention period by browsing to the **Bucket Activity** application. The ECS download and upload activity appears similar to the following example:



Each new warm bucket gets uploaded to remote storage. Buckets are only downloaded if they're required for a search and are not already in the local cache. High download rates are a negative indicator of cache health

Figure 20. Data overflow to ECS

- Verify the eviction activity using the **Buckets Evicted** graph. The output appears similar to the following example:



A steady stream of evictions is expected once the local cache fills up and the oldest data is removed according to the eviction policy.

Figure 21. Buckets evicted to ECS

- Alternately, verify the eviction activity by querying `index=_internal sourcetype=splunkd bytes_evicted`. The output appears similar to the following example:

```
> 7/12/22 07-12-2022 10:48:40.670 -0400 INFO CacheManager [574758 IndexerService] - Eviction results: count=4, test_count=11, bytes_evicted=326561694, bytes_ne
10:48:40.670 AM eded=35794944, elapsed_ms=103
host = splunk8-indexer-02 | source = /opt/splunk/var/log/splunk/splunkd.log | sourcetype = splunkd
```

- Verify that the custom index is present, and new data has been written to the ECS S3 bucket for this index, with the **S3 Browser** tool, at <https://s3browser.com/>.

The screen appears similar to the following example:

Name	Size	Type	Last Modified	Storage Class
..				
._audit/				
._internal/				
._introspection/				
._metrics/				
._telemetry/				
.idrac_live/				
.main/				

**Figure 22. S3 Browser tool**

## Results

Dell Technologies successfully validated that the Splunk SmartStore cache manager can be configured to use Dell ECS as warm and cold data tiering.

## Failure forecasting

### Prerequisites

Prerequisites include:

- Robin Cloud Native Platform (Robin CNP) must be installed on an infrastructure node.
- Deep Learning Toolkit (DLTK) must be installed on the Robin CNP infrastructure node.
- The Robin CNP web UI must be network-accessible.
- Proper Robin CNP user credentials must be created and assigned.
- Server telemetry reports must be streamed from iDRAC hosts to the Splunk application.
- HAProxy must be installed. HAProxy is used as a reverse proxy and load balancer for traffic coming into the Splunk cluster.

### About this task

Customers use failure forecasting data to triage fixes and avoid server downtime. These tests validate that:

- Splunk can be used to deploy a deep learning failure forecasting model using Deep Learning Toolkit (DLTK).
- Splunk can generate an alert for forecasted failure events.

### Steps

- From the Splunk UI, verify that DLTK is configured with Kubernetes, and that the `__dev__` container can be deployed for DLTK.

The screen appears similar to the following example:

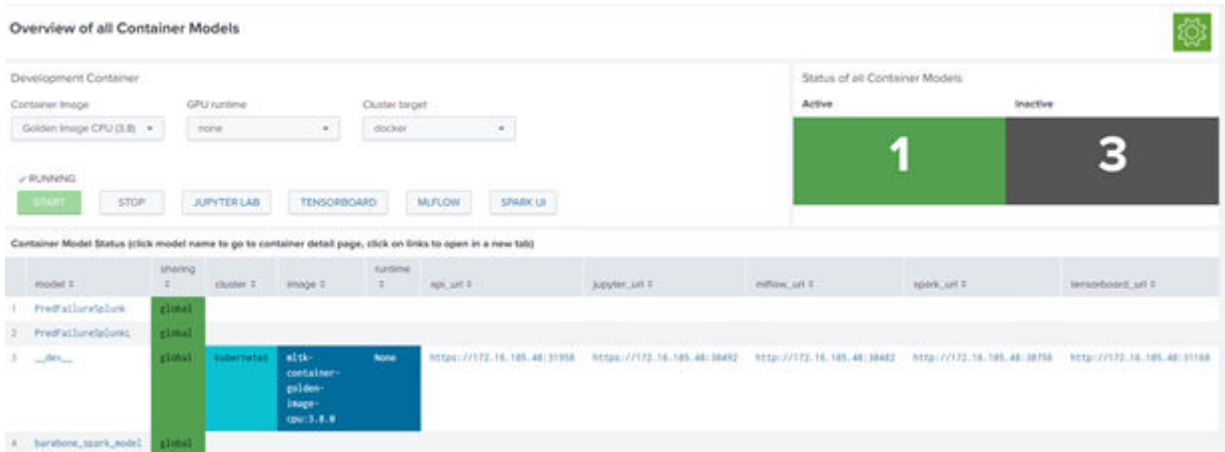


Figure 23. Container models overview

2. Add your failure forecasting model, ready for training, running as a container for data inference.
3. Collect real telemetry data from iDRAC hosts, and then inject artificial fault events.  
The added artificial events were for overloading, overheating, and hard disk failure throughout a timespan of multiple days. This generated telemetry data was streamed to HEC and ready to be queried.
4. Create an alert to periodically query and send this data to the DLTK container:
  - a. Browse to **Settings > Searches, reports, and alerts > New Alert**.
  - b. Configure it to trigger an alert according to the inference results.

The screen appears similar to the following example:

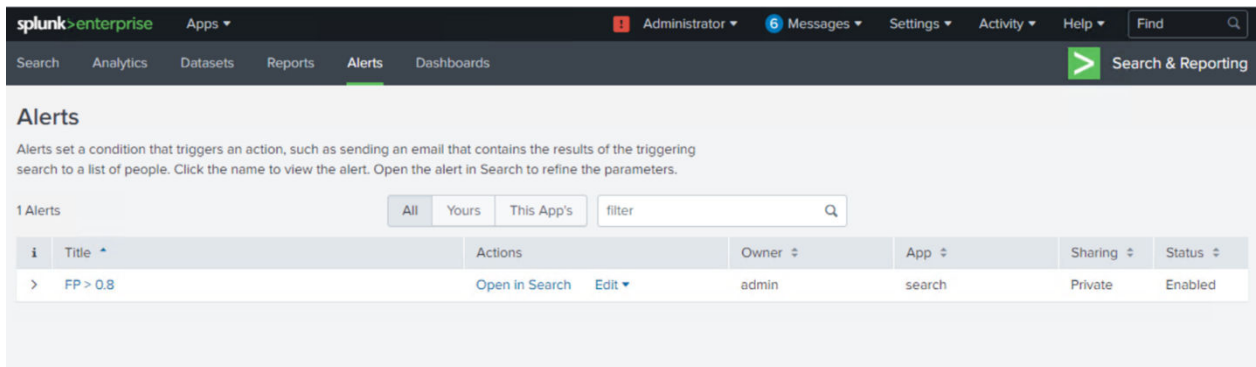


Figure 24. Alerts

5. Verify that the alerts triggered as expected:
  - a. Browse to **Activity > Triggered Alerts**.
  - b. Correlate with the injected faults in the generated data.

The screen appears similar to the following example:

App	Search & Reporting (search)	Owner	Administrator (admin)	Severity	All	Alert	All	Filter	
«Prev		Next»		Showing 1-6 of 6 results					
	Time	Fired alerts	App	Type	Severity	Mode	Actions		
<input type="checkbox"/>	2022-07-13 01:30:06 EDT	FP > 0.8	search	Scheduled	Medium	Digest	<a href="#">View results</a>	<a href="#">Edit search</a>	<a href="#">Delete</a>
<input type="checkbox"/>	2022-07-13 00:30:06 EDT	FP > 0.8	search	Scheduled	Medium	Digest	<a href="#">View results</a>	<a href="#">Edit search</a>	<a href="#">Delete</a>
<input type="checkbox"/>	2022-07-12 23:30:06 EDT	FP > 0.8	search	Scheduled	Medium	Digest	<a href="#">View results</a>	<a href="#">Edit search</a>	<a href="#">Delete</a>
<input type="checkbox"/>	2022-07-12 22:30:06 EDT	FP > 0.8	search	Scheduled	Medium	Digest	<a href="#">View results</a>	<a href="#">Edit search</a>	<a href="#">Delete</a>
<input type="checkbox"/>	2022-07-12 21:30:06 EDT	FP > 0.8	search	Scheduled	Medium	Digest	<a href="#">View results</a>	<a href="#">Edit search</a>	<a href="#">Delete</a>
<input type="checkbox"/>	2022-07-12 20:30:07 EDT	FP > 0.8	search	Scheduled	Medium	Digest	<a href="#">View results</a>	<a href="#">Edit search</a>	<a href="#">Delete</a>

**Figure 25. Triggered telemetry alerts**

### Results

Dell Technologies successfully validated that:

- Splunk can be used to deploy a deep learning failure forecasting model using Deep Learning Toolkit (DLTK).
- Splunk can generate an alert for forecasted failure events.

## Summary

All the tests ran successfully, with effective utilization of compute and memory resources. The objective of these tests was not to measure any performance metric data. They were meant to confirm that all the services deployed in the Robin Cloud Native Platform cluster were functioning correctly.

# Implementation, best practices, and sizing

## Topics:

- [Overview](#)
- [Splunk best practices](#)
- [Splunk integration best practices](#)
- [Predictive Maintenance for IT Operations specifics](#)
- [Scaling up and out](#)

## Overview

This chapter provides guidelines and best practices for sizing and scaling a Predictive Maintenance for IT Operations deployment as a use case example within the Dell Validated Design for Analytics — Data Lakehouse. It also provides relevant implementation and configuration information.

## Splunk best practices

### Baseline deployments

A baseline Predictive Maintenance for IT Operations deployment in the Dell Validated Design for Analytics — Data Lakehouse platform needs:

- A highly available Kubernetes control plane with three infrastructure nodes
- At least one worker node
- A switch to support networking

You can stand up this cluster for pilot deployments with standard configurations, to evaluate Splunk on the Dell Validated Design for Analytics — Data Lakehouse platform. Dell calls this configuration a "small" package.

### Production deployments

For production deployments you must consider the following in order to support your intended use cases:

- Failure tolerance of workers and storage elements
- Storage tiering
- Network and virtual infrastructure availability

Resource dimensioning in this case is a range. For example:

- Worker nodes can range in number from two to seven depending upon S5248F-ON port saturation.
- Storage can range to three times the calculated use case needs based on the required Kubernetes replication factor.
- External object storage (for example, Dell ECS) is added to facilitate economical access to storage services.
- Switches are added for high availability (HA).
- Upgrade to denser SSDs in Worker or ECS nodes.

**NOTE:** Scale-up possibilities, like adding a GPU (such as a NVIDIA Ampere A100) per worker node for training performance, will be available in a future release of Predictive Maintenance for IT Operations.

Dell recommends that you start with four worker nodes, and then build up to your use case needs by hot-plugging in as necessary. The Dell Validated Design for Analytics — Data Lakehouse platform does allow such hot plug-ins of storage and compute elements. Dell calls this configuration a "medium" package.

## Extreme production deployments

For extreme production deployments, you can add additional worker nodes. However, spine switches are required because this configuration exceeds the leaf switch saturation point. The requirements to support use cases that need higher throughput include:

- Higher density, 100 GbE NICs
  - Or, more NICs per worker node (four 25 GbE NICs)
- A switch upgrade from the S5248F-ON to the S5296F-ON

Dell calls this configuration a "large" package. This scenario is outside the scope of the current design and is addressed separately in a guidance addendum.

## Splunk integration best practices

### SmartStore configurations

Indexers maintain buckets for SmartStore indexes in hot and warm buckets:

- The hot buckets of SmartStore indexes reside on local storage, as with non-SmartStore indexes.
- Warm buckets reside on remote storage, although copies of those buckets might also reside temporarily in local storage.

### SmartStore cache controller

The cache manager maximizes search efficiency through intelligent management of the local cache. It favors retaining in the cache copies of buckets and files that have a high likelihood of participating in future searches. When the cache fills up, the cache manager removes, or "evicts", copies of buckets that are least likely to participate in future searches.

These settings in `server.conf` initiate eviction based on occupancy of the cache disk partition:

#### **max\_cache\_size**

Specifies the maximum occupied space, in megabytes, for the disk partition that contains the cache.

#### **minFreeSpace**

Specifies the minimum free space, in megabytes, for a partition.

#### **eviction\_padding**

Controls the amount of additional space, in megabytes, that the cache manager protects beyond the `minFreeSpace` value.

#### **max\_concurrent\_downloads**

Specifies the maximum number of buckets that can be downloaded simultaneously from remote storage. The default value is eight.

#### **max\_concurrent\_uploads**

Specifies the maximum number of buckets that can be uploaded simultaneously to remote storage. The default value is eight.

## Per-index settings

[SmartStore server.conf per-index settings](#) describes the per-index settings in `server.conf`.



**Table 8. SmartStore server.conf per-index settings**

Setting	Type	Description
maxGlobalDataSizeMB	Index or global	This setting specifies the maximum size, in MB, for all warm and cold buckets in a SmartStore index. When the size of the warm and cold buckets sets of an index exceeds this value, the system freezes the oldest buckets. When the size falls below this value, the buckets are unfrozen.
maxGlobalRawDataSizeMB	Index or global	This setting specifies the maximum size, in MB, of raw data residing in all warm buckets in a SmartStore index. When the raw data size in the set of warm buckets of an index exceeds this value, the system freezes the oldest buckets. When the size falls below this value, the buckets are unfrozen.
frozenTimePeriodInSecs	N/A	This same setting is used with non-SmartStore indexes. It specifies that buckets freeze when they reach the configured age. The default value is 188697600 seconds, or approximately 6 years.
hotlist_recency_secs	Index	This setting specifies the time period, based on the age of a bucket, that the cache manager attempts to protect a recent bucket from eviction. This setting operates on a per-index level, while the version of the setting in server.conf operates across all indexes.
hotlist_bloom_filter_recency_hours	Index	This setting specifies the time period, based on the age of a bucket, that the cache manager attempts to protect the non-journal and non-time series index (tsidx) files of a bucket, such as the bloomfilter file, from eviction. This setting operates on a per-index level, while the version of the setting in server.conf operates across all indexes.

## Predictive maintenance use case configurations

Configurations based on the Predictive Maintenance for IT Operations use case include:

- **Data Retention (Hot)**—Around five days
- **Data Retention (Warm) in ECS**—Around 14 TB (for N=55)

## Create and apply configurations

### About this task

Follow these steps to create and apply Predictive Maintenance for IT Operations configurations.

### Steps

1. Using the text editor of your choice, edit the `$SPLUNK_HOME/etc/master-apps/_cluster/local/indexes.conf` file as follows:

```
[default]
remotePath = volume:ecs_store/${_index_name}
repFactor = auto

[volume:ecs_store]
storageType = remote
path = s3://splunk-bucket/indexes
remote.s3.access_key = <s3 access key>
remote.s3.secret_key = <s3 secret key>
remote.s3.endpoint = <s3 end point>

[idrac_live]
repFactor = auto
```

```
coldPath = $SPLUNK_DB/idrac_live/coldddb
datatype = event
homePath = $SPLUNK_DB/idrac_live/db
thawedPath = $SPLUNK_DB/idrac_live/thaweddb

hotlist_recency_secs = 432000 #(5days)
maxGlobalRawDataSizeMB = 14680064 #(14TB)
```

2. Apply the configurations across the cluster by running the following command:

```
$ sudo /opt/splunk/bin/splunk apply cluster-bundle --answer-yes
```

## Results

The Predictive Maintenance for IT Operations configurations are ready for testing.

## Test configurations

### About this task

Perform these steps to test the configurations.

### Steps

1. Log in to the Splunk user interface.
2. Go to **Settings > Monitoring Console > Indexing > SmartStore Activity**.
3. Select the instance from the **Instance** drop-down.  
If SmartStore activity has occurred, you should see the status message, **ONLINE**.

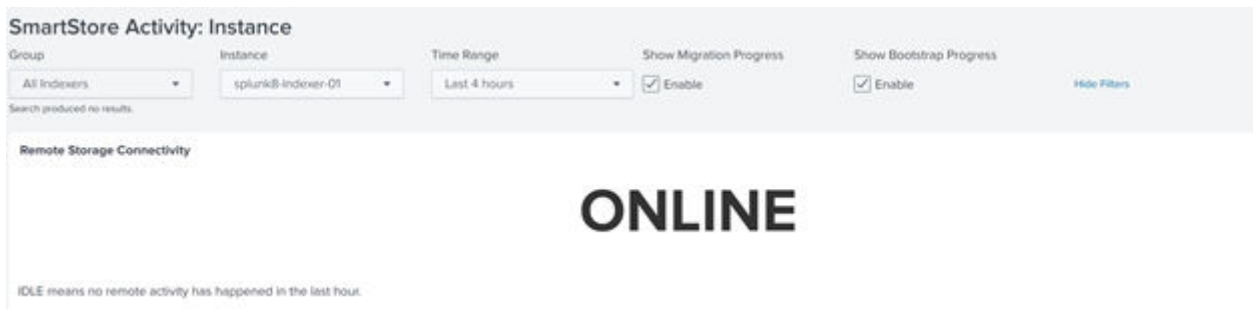


Figure 26. Instance activity

4. View the **Bucket Activity** screen to see downloads and uploads to SmartStore. Green bars show buckets that were uploaded to SmartStore. Blue bars show buckets that were downloaded from SmartStore. See [Bucket activity](#).



**Figure 27. Bucket activity**

**Results**

The configurations have passed the test and are operational.

## Configure indexers and search heads

You can configure indexers and search heads from the Robin Cloud Native Platform (Robin CNP) user interface. You can specify:

- Number of CPU cores
- Amount of memory
- Storage size
- Number of instances
- Several other attributes

APP NAME \*  RESOURCE POOL  IP SUBNET  NAMESPACE  ADVANCED OPTIONS [Click to configure](#)

[Show chargeback price-sheet](#)

**CONFIGURE COMPONENTS**

master

search\_head

indexer

STATIC IPS

SECONDARY IPS

**CPU CORES**

**CPU TYPES**

**MEMORY**

**DOCKER IMAGE** splunk/splunk (8.2.3.3) from: "Docker Hub" registry

**STORAGE**

configuration	1	volumes	50 GB	HDD	App Protected	No FaultDomain	/opt/splunk/etc
data	1	volumes	10 GB	SSD	App Protected	No FaultDomain	/opt/splunk/var

APPLICATION EPHEMERAL STORAGE

**Figure 28. Splunk control plane node configuration options**

## CONFIGURE COMPONENTS

search\_head service can be enabled/disabled by toggling the switch to the left (default setting is enabled)

master

search\_head

indexer

**INSTANCES**

1

**STATIC IPS**  **SECONDARY IPS**

**CPU CORES** 1 **CPU TYPES** Shared

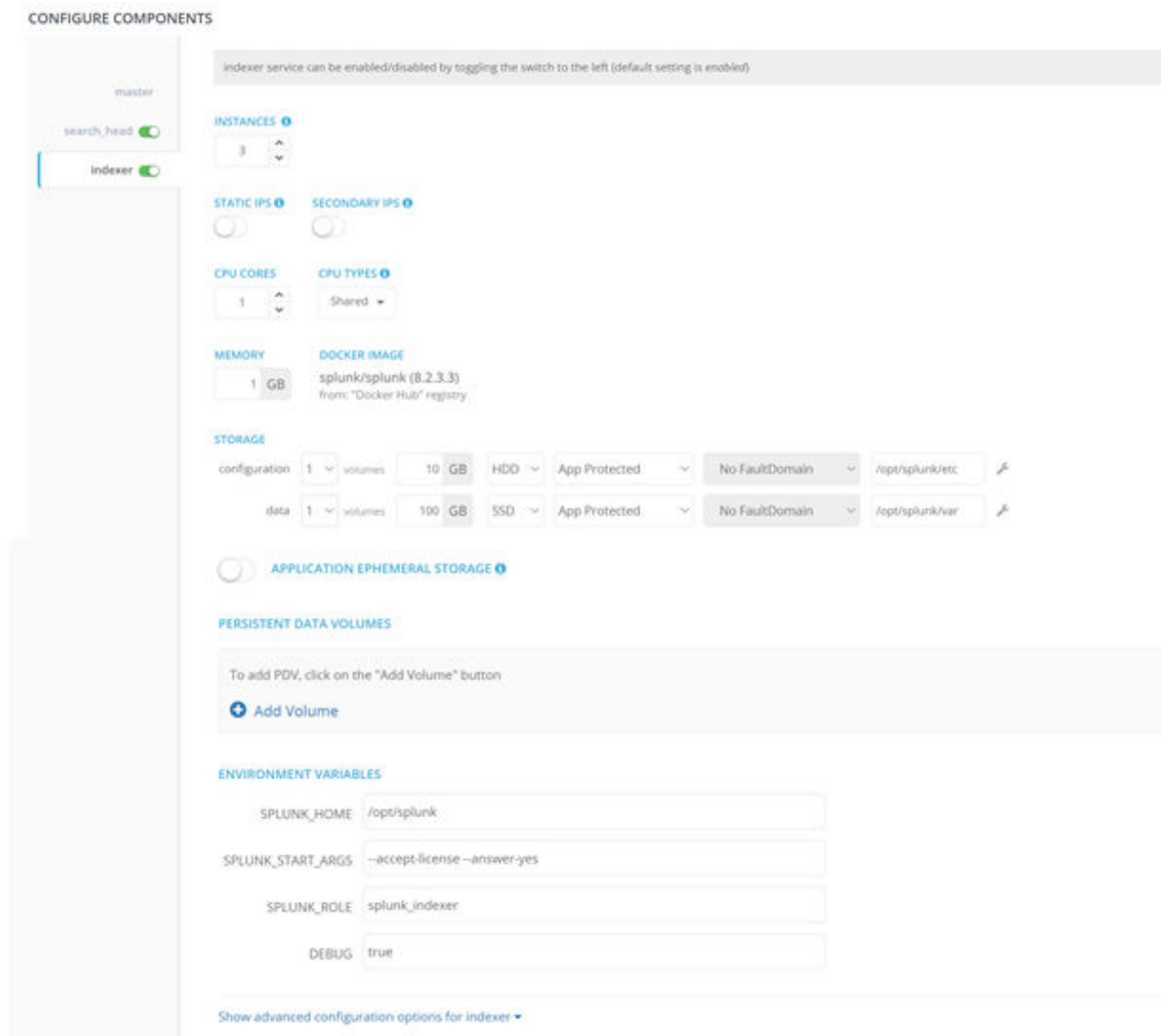
**MEMORY** 1 GB **DOCKER IMAGE** splunk/splunk (8.2.3.3) from: "Docker Hub" registry

**STORAGE**

configuration	1	volumes	50 GB	HDD	App Protected	No FaultDomain	/opt/splunk/etc
data	1	volumes	10 GB	SSD	App Protected	No FaultDomain	/opt/splunk/var

**APPLICATION EPHEMERAL STORAGE**

Figure 29. Splunk search head configuration options



**Figure 30. Splunk indexer configuration options**

Individual indexers configurations are defined in the `$(SPLUNK_HOME)/etc/system/local/indexes.conf` file. For more details, see the [Configure and deploy indexes](#) section of the *Splunk Installation and Upgrade Manual*.

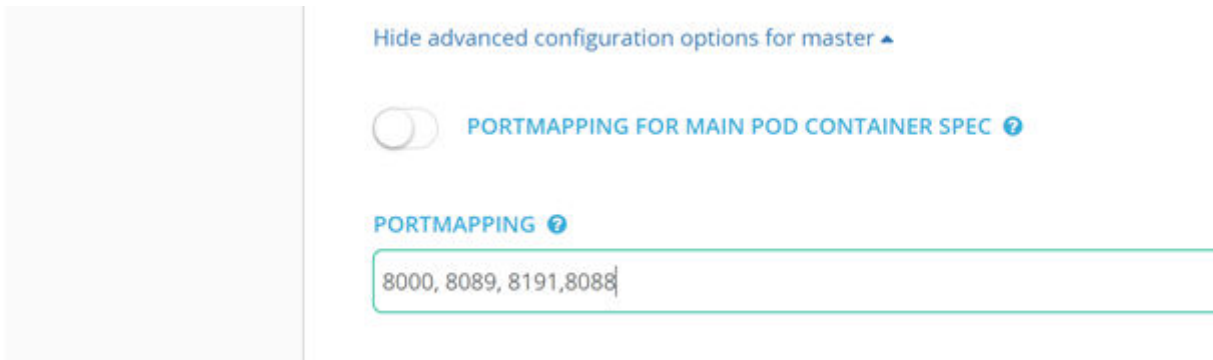
## Port forwarding

### About this task

Perform these steps to configure port forwarding:

#### Steps

1. Log in to the Robin Cloud Native Platform user interface.
2. Go to **Applications > Bundles > Splunk 8.2**.
3. Select **Advanced Options**.



**Figure 31. Advanced options - port mapping**

4. In the **PortMapping** text box, enter **8088**.  
The HTTP Event Collector uses port 8088.

**Results**

The indexers and search heads are configured and operational.

## Configure forecasting

Failure probability can be computed from the estimated survival probability over an observed time period. This probability is the residual after compounding survival probability over sections. This failure probability can be mapped to Time-to-Failure estimates using a sectional exponential decay model. This exponential decay curve is parameterized for accuracy, and varies slightly from one fault model to another. The more sections, the better the accuracy.

In the IT world, a difference of half a day in filing trouble tickets does not make a difference, so fewer sections are used. For other types of digital twins, finer sections are required for accuracy. A digital twin is designed as an accurate, virtual model of a physical device or object. Data scientists and IT personnel use digital twins to run simulations before committing to the time and cost of building physical devices or objects.

The following parameters can be tuned to adjust the Time-to-Failure:

```
A = 2000000
B = 14.72
B1 = 15.30
cutoff = 0.90
```

The following is an example python function:

```
def t2f_func(a, b, fp):
    t2f = a * np.exp(-b*fp)
    return t2f

failure_probability = 1 - np.prod(survival_probability_adjusted)
print ('Failure Probability: ', failure_probability)
if failure_probability >= cutoff:
    print ('Time-to-Failure: ', np.ceil (t2f_func (A, B , failure_probability)), 'Days')
else:
    print ('Time-to-Failure: ', np.ceil (t2f_func (A, B1, failure_probability)), 'Days')

Failure Probability: 0.9499373994767666
Time-to-Failure: 2.0 Days
```

# Configure ECS

## About this task

You can configure Dell Elastic Cloud Storage (ECS) from the ECS user interface.

## Steps

1. Log in to the ECS user interface.
2. Select **Manage > Storage Pool** to create a pool.  
A pool can contain single or multiple ECS nodes.
3. Select **Manage > Virtual Data Center** to create a virtual data center (VDC).  
A VDC contains management and replication end-point IP addresses.
4. Select **Manage > Namespaces** to create a namespace.
5. Select **User > New User Object** to create a user in a namespace.  
A user is required to authenticate Splunk SmartStore with ECS S3 Storage. Once created, the username and secret key that is used in SmartStore configurations is displayed, as in the following example.

```
remote.s3.access_key = <s3 access key>  
remote.s3.secret_key = <s3 secret key>  
remote.s3.endpoint = <s3 end point - ECS IP address>
```

6. Select **Manage > Buckets** to create a bucket.  
SmartStore uses this bucket to store data on ECS.

## Results

ECS is now configured. For more information, see the [ECS Administration Guide](#).

# Configure the Robin CNP storage class

The Container Storage Interface (CSI) is a standard for exposing storage to workloads on Kubernetes. A Kubernetes resource, **StorageClass**, must be created and registered within the Kubernetes cluster to enable automatic creation and deletion of volumes for CSI storage. Associated with the **StorageClass** is a CSI provisioner plug-in that provisions storage volumes based on the various attributes defined in the **StorageClass**.

By default a Robin Cloud Native Platform (Robin CNP) deployment is provisioned with two **StorageClass** resources:

- `robin`—The default **StorageClass**, it has no features that are enabled and can be used for standard RWO and RWX volumes.
- `robin-immediate`—The **StorageClass** that creates a volume when a Persistent Volume Claim (PVC) is created. It does not wait for the first consumer of that volume.

For more information about configuring the Robin CNP storage class, see the [Robin Hyperconverged Kubernetes Bundle Building Guide](#).

# Protect PVCs with volume replication

Robin Cloud Native Platform (Robin CNP) uses storage volume-level replication to ensure that data is always available, in case disk or nodes failures occur. When replication is configured to **2**, at least two copies of the volume are maintained on different disks. If replication is set to **3**, at least three copies are maintained. This configuration ensures that the volume data is available if one or two disk or node failures occur.

For more information about configuring replication, see the [Robin Cloud Native Storage \(Robin CNS\) documentation](#).

## Configure a remote client

A remote client can stream events and metrics to Splunk using the HTTP Event Collector (HEC). To connect and authenticate to HEC, a remote client requires:

- An HEC HTTP endpoint
- An HEC access token

The sample code below shows an example Python remote client streaming events to Splunk.

```
url='https://100.82.185.122:32401/services/collector/event'

authHeader = {'Authorization': 'Splunk
{}'.format('0617eea5-87a9-4d18-8ed4-6dc085ddb2c')}

requests.post(url, headers=authHeader, json=data, verify=False)
```

Splunk provides EventGen, a utility that helps users build real-time event generators. EventGen can be installed as a Splunk app or run separately on a Linux host or container. The `eventgen.conf` file is used to configure the generator. The sample code below shows an example that reads from the `cyclical.csv` file and streams using HEC.

```
[cyclical.csv]
mode=sample
interval=60
count=1
randomizeEvents=true
outputMode=httpevent
httpeventServers = {"servers": [{"protocol": "https", "port": "8088", "key": \
"D8B289CD-27BD-FF26-E508-0F5F1299668F", "address": "172.30.37.134"}]}
index=main
sourcetype=eventgen
sampletype=csv
source=eventgen_cyclical
```

## Streaming with HAProxy

HAProxy can be used for load balancing with a round-robin algorithm for Splunk. It is a separate software package that can be installed on a bare metal server, or on a VM on the administration node.

HAProxy configurations contain two main parts:

- **Front end**—Receives traffic before dispatching it to the back end.
- **Back end**—Receives the request, forms a response, and then sends the response back through HAProxy to the client.

The sample code below contains sample configurations for HAProxy.

```
global
  chroot      /var/lib/haproxy
  pidfile     /var/run/haproxy.pid
  maxconn     1000
  user        haproxy
  group        haproxy
  daemon
  stats socket /var/lib/haproxy/stats

defaults
  mode                http
  log                 global
  option              httplog
  option              dontlognull
  option http-server-close
  option              redispatch
  retries             3
  timeout http-request 10s
  timeout queue       1m
  timeout connect     10s
  timeout client      1m
```



```
timeout server 1m
timeout http-keep-alive 10s
timeout check 10s
maxconn 3000

frontend main 0.0.0.0:80
option forwardfor
reqadd X-Forwarded-Proto:\ http
default_backend splunk_backend

backend splunk_backend
balance roundrobin
server worker1 <IP address>:<port>

#We need to set SELinux to allow traffic on the port
#semanage port --add --type http_port_t --proto tcp <port>
```

For more information, see [HAProxy Configuration Basics](#).

## Predictive Maintenance for IT Operations specifics

Predictive Maintenance for IT Operations requires the "medium" package. Details are shown in the figure and tables below. Each worker node must have:

- 768 GB memory
- Thirty-two cores per socket
- More than 6 TB SSD capacity per node
- One dual-port 25 GbE NIC

This information is based on coarsely grained guidance from Dell Data Analytics baseline performance testing.

 **NOTE:** GPUs will be integrated into a future release of Predictive Maintenance for IT Operations.

In the figure below, a 48-port S3148 management switch is used to manage all the nodes.

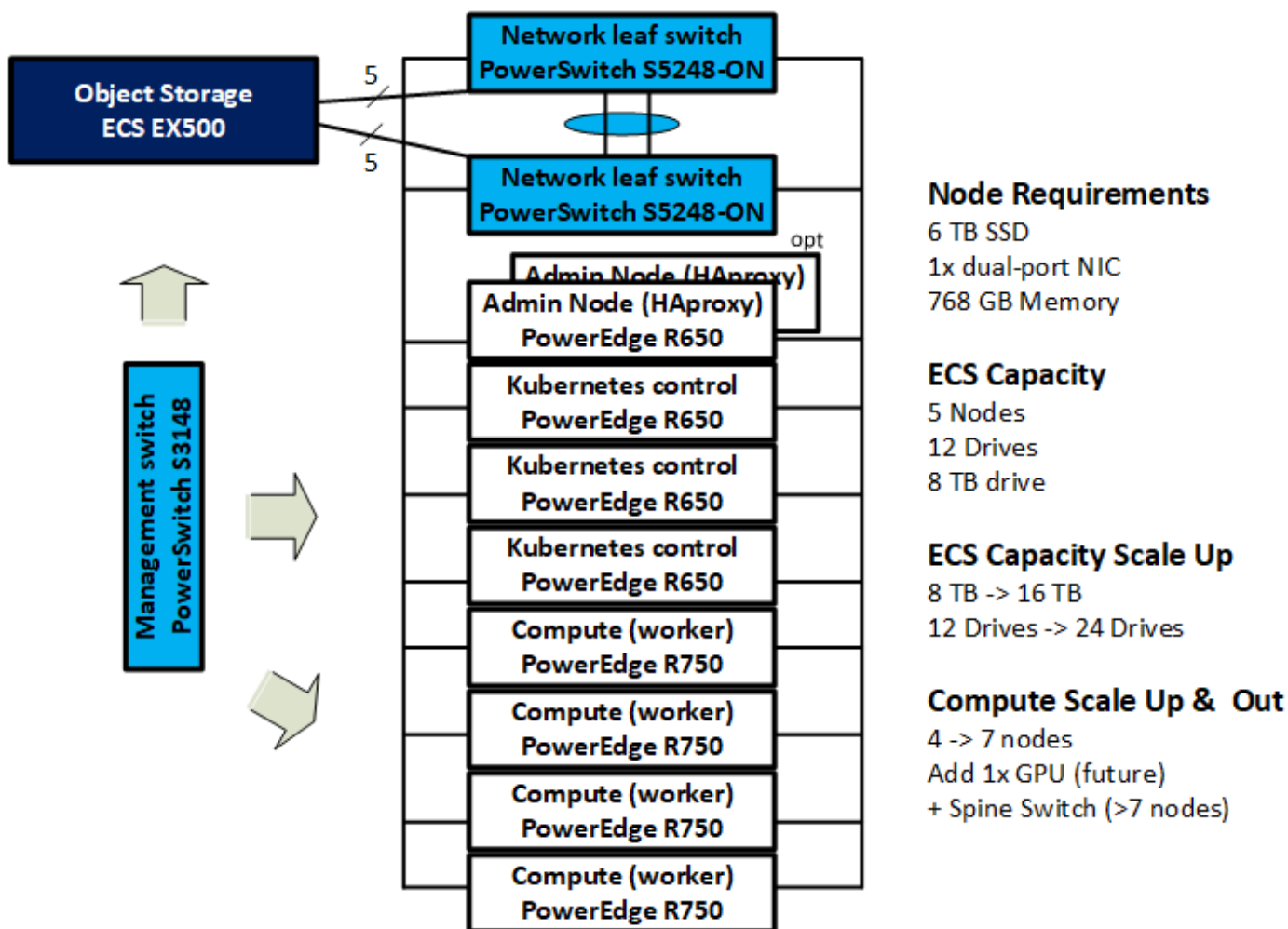


Figure 32. Predictive Maintenance for IT Operations with optional standby administration node

Table 9. Predictive Maintenance for IT Operations storage estimates

Component	Estimate
Daily data volume (DDV)	500 GB
Compression factor	0.15
Metadata size factor	0.35
Data retention hot (DRH)	5 days
Data retention warm	N days
Maximum volume per indexer	300 GB
RAID level	RAID 0
SSD (available) per node	6 TB

Table 10. Predictive Maintenance for IT Operations storage calculations

Calculation	Value
Volume calculation	$((0.15 + 0.35) * DRH * DDV) / (\text{indexers})$
Volume per indexer	160 GB [Assuming eight indexers]
Storage pool requirements	1.25 TB
Storage requirements after 3x redundancy	3.75 TB

**Table 10. Predictive Maintenance for IT Operations storage calculations (continued)**

Calculation	Value
Available SSD capacity	24 TB

For additional sizing tools and information, see:

- [Splunk Storage Sizing](#)
- [Intel Select Solutions: Containerized Splunk](#)

## Scaling up and out

You can scale up your compute resources by adding GPUs to worker nodes. Management nodes with virtual machines that host HAProxy can facilitate load-balancing of worker nodes, which can optimize performance.

You can scale out by adding more worker nodes, with the highest core densities.

Scaling storage up and out are relatively straight forward; you hot-plug higher density drives. As you add worker nodes to host those drives, you scale out. ECS is the Dell standard object storage solution for nonreal-time analytics of historical significance that leverage warm or cold tiers. ECS is fully integrated to the Dell Validated Design for Analytics — Data Lakehouse platform. The five ECS nodes can scale up to 16 TB per drive, and up to 24 drives.

The number of servers in the substrate may vary widely. The exact number does not matter in this context because Robin Cloud Native Platform enables you to hot plug-in additional SSD drives for upscaling. Substrate segmentation depends on the real-time needs of failure analysis. The DDV in [Predictive Maintenance for IT Operations storage estimates](#) is for 100 servers.

# Summary

## Topics:

- [Overview](#)
- [We value your feedback](#)

## Overview

This Design Guide has presented an architecture and design for predictive analytics in It operations. It used Splunk Enterprise for analysis of iDRAC telemetry data from Dell PowerEdge servers as an example of but one of many possible use cases.


In this design, Splunk Enterprise is deployed in containers with Kubernetes orchestration. The container platform we have chosen is the Dell Validated Design for Analytics — Data Lakehouse, which includes the Robin Cloud Native Platform (Robin CNP) from Robin.io as the Kubernetes layer.

Together, Cloud Native Splunk Enterprise with SmartStore, deployed on an optimized infrastructure of a Dell Validated Design for Analytics, enables data-driven organizations to rapidly gain visibility and real-time insight into their data at scale. Using this design, organizations can take meaningful action for business and operational advantages.

## We value your feedback

Dell Technologies and the authors of this document welcome your feedback on the solution and the solution documentation. Contact the Dell Technologies Solutions team by [email](#).

Authors: Dell Technologies Integrated Solutions Engineering, Technical Marketing, and Information Design & Development teams

 **NOTE:** For links to additional documentation for this solution, see the [Dell Technologies Info Hub for Data Analytics](#).

This document may contain language from third-party content that is not under Dell's control and is not consistent with Dell's current guidelines for Dell's own content. When such third-party content is updated by the relevant third parties, this document will be revised accordingly.

# References

## Topics:

- [Dell Technologies documentation](#)
- [Splunk Enterprise documentation](#)
- [Intel documentation](#)
- [Robin.io documentation](#)
- [Dell Technologies Customer Solution Centers](#)
- [Dell Technologies Info Hub](#)
- [More information](#)

## Dell Technologies documentation

The following Dell Technologies documentation provides additional and relevant information. Access to these documents depends on your login credentials. If you do not have access to a document, contact your Dell Technologies sales representative.

Additional information can be obtained at the [Dell Technologies Info Hub for Data Analytics](#). If you need additional services or implementation help, contact your Dell Technologies sales representative.

**Table 11. Dell Technologies documentation**

Document type	Location
Server specification sheets	<a href="#">PowerEdge R650 Spec Sheet</a>
	<a href="#">PowerEdge R750 Spec Sheet</a>
	<a href="#">PowerEdge R750xa Spec Sheet</a>
Storage specification sheets	<a href="#">ECS EX500 Spec Sheet</a>
Switch specification sheets	<a href="#">PowerSwitch S3100-ON Series Spec Sheet</a>
	<a href="#">PowerSwitch S5200-ON Series Spec Sheet</a>
Server manuals	<a href="#">PowerEdge R650 Manuals and Documents</a>
	<a href="#">PowerEdge R750 Manuals and Documents</a>
	<a href="#">PowerEdge R750xa Manuals and Documents</a>
Storage manuals	<a href="#">ECS EX500 Manuals and Documents</a>
Switch manuals	<a href="#">PowerSwitch S3100-ON Manuals and Documents</a>
	<a href="#">PowerSwitch S5200-ON Series Manuals and Documents</a>

## Splunk Enterprise documentation

The following documentation on the [Splunk website](#) provides additional and relevant information.

**Table 12. Splunk Enterprise documentation**

Document type	Location
Capacity planning	<a href="#">Splunk Enterprise Capacity Planning Manual</a>

**Table 12. Splunk Enterprise documentation (continued)**

Document type	Location
Installing Splunk	<a href="#">Splunk Enterprise Installation Manual</a>
System requirements	<a href="#">System requirements for use of Splunk Enterprise on-premises</a>
Splunk metrics	<a href="#">Get started with metrics</a>
Splunk storage sizing tool	<a href="#">Splunk Storage Sizing</a>
Splunk Operator repository	<a href="#">Splunk Operator for Kubernetes</a>
Splunk Operator storage guidelines	<a href="#">Splunk Operator Storage Guidelines</a>
Splunk Operator SmartStore resources	<a href="#">Splunk Operator SmartStore Resource Guide</a>
Splunk SmartStore Indexers	<a href="#">Managing Indexers and Clusters of Indexers</a>
Installing Deep Learning Toolkit (DLTK)	<a href="#">Installing Deep Learning Toolkit for Splunk</a>
Splunk validated architectures	<a href="#">Splunk Validated Architectures</a>

## Intel documentation

The following documentation on the [Intel website](#) provides additional and relevant information.

**Table 13. Intel documentation**

Document type	Location
Select Solutions for Containerized Splunk	<a href="#">Intel Select Solutions: Containerized Splunk</a>

## Robin.io documentation

The following documentation on the [Robin.io documentation website](#) provides additional and relevant information.

**Table 14. Robin.io documentation**

Document type	Location
Platform	<a href="#">Robin Cloud Native Platform</a>
Storage	<a href="#">Robin Cloud Native Storage</a>
Bundles	<a href="#">Robin Hyperconverged Kubernetes Bundle Building Guide</a>

## Dell Technologies Customer Solution Centers

Our global network of dedicated [Dell Technologies Customer Solution Centers](#) are trusted environments where world class IT experts collaborate with customers and prospects to share best practices; facilitate in-depth discussions of effective business strategies using briefings, workshops, or Proofs of Concept (PoCs); and help business become more successful and competitive.

Dell Technologies Customer Solution Centers reduce the risks that are associated with new technology investments, and can help improve speed of implementation.

All of the services of the Customer Solution Centers are available to all Dell Technologies customers at no charge. Contact your account team today to submit an engagement request.

# Dell Technologies Info Hub

The [Dell Technologies Info Hub](#) is your one-stop destination for the latest information about Dell Solutions and Networking products. New material is frequently added, so browse often to keep up to date on the expanding Dell portfolio of cutting-edge products and solutions.

## More information

For more information, including sizing guidance, technical questions, or sales assistance, email [ai.assist@dell.com](mailto:ai.assist@dell.com), or contact your Dell Technologies or authorized partner sales representative.