
15th ANNUAL WORKSHOP 2019

# DISTRIBUTED ENDPOINT MANAGEMENT

## AN NVME-OF™ SCALE-OUT MANAGEMENT SOLUTION

Phil Cayton

**Intel Corporation**

**March, 2019**

# AGENDA

- NVMe over Fabrics (NVMe-oF) Overview
- Current State of NVMe-oF Management and Administration
- Distributed Endpoint Management (DEM) Project
- Brief Demonstration
- Development Opportunities and Wrap-up

# NVME OVER FABRICS

## NVMe-oF Overview

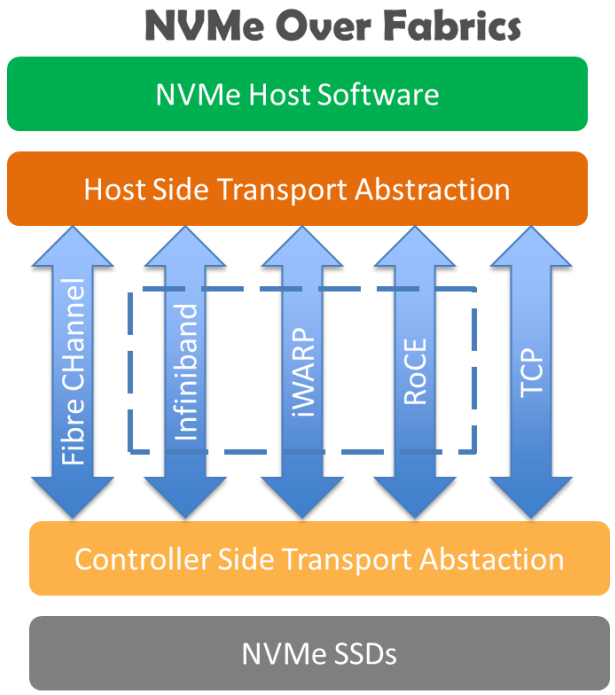
### NVMe: industry standard interface and storage protocol for PCIe SSDs

- High-performance, low-latency PCI SSD interface
- Eliminates unnecessary protocol translations (i.e., SCSI)
- Defines partitioning PCIe SSDs into one or more subsystems

### NVMe-oF: extends NVMe efficiency over Fabrics

- Builds on base NVMe architecture with thin encapsulation of base NVMe across a fabric
- Enables low-latency and high IOPS access to remote NVMe storage
- Defines end-to-end mechanisms to transfer NVMe commands and data structures

The NVMe-oF spec is not Fabric specific;
separate Transport Bindings are defined for each Transport



# NVME OVER FABRICS

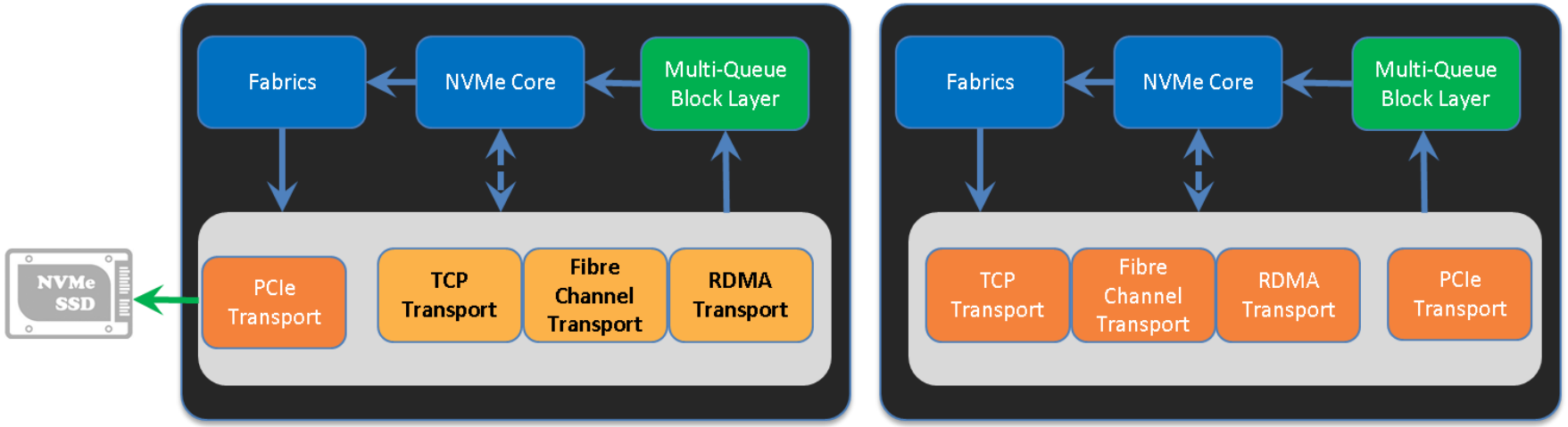
## NVMe-oF Overview

### Targets

- Create logical NVMe subsystems and Controllers that are presented to Hosts
- Logically map NVMe Namespaces to physical NVMe block devices
- Export NVMe subsystems virtualizing NVM Namespaces
- May be provisioned to allow individual Hosts access to specific resources

### Hosts

- Discover provisioned NVMe-oF resources from Targets
- Connect to provisioned resources



# DISTRIBUTED ENDPOINT MANAGEMENT

## Current State of NVMe-oF Management and Administration

**Linux In-kernel Implementation only supports local management**

**Each individual Target is either manually or statically configured**

- Configure each Fabric with Address, Port, …
- Define each NVMe-oF Subsystem
- Assign NVMe resources to each NVMe-oF Subsystem
- Set up Individual Host Access rights to each NVMe-oF Subsystem

**Each individual Host either accesses NVMe-oF resources through:**

- Static predefined configurations
- Manual resource discovery process (as defined by NVMe-oF specification)
  - Connects to each individual Target
  - Requests resources on that Target that they may access

**Limits usability flexibility, scale of dynamic installations of NVMe-oF**


# DISTRIBUTED ENDPOINT MANAGEMENT (DEM) PROJECT

# DISTRIBUTED ENDPOINT MANAGEMENT

## DEM Project

### Distributed Endpoint Management (DEM): An Open-Source Project

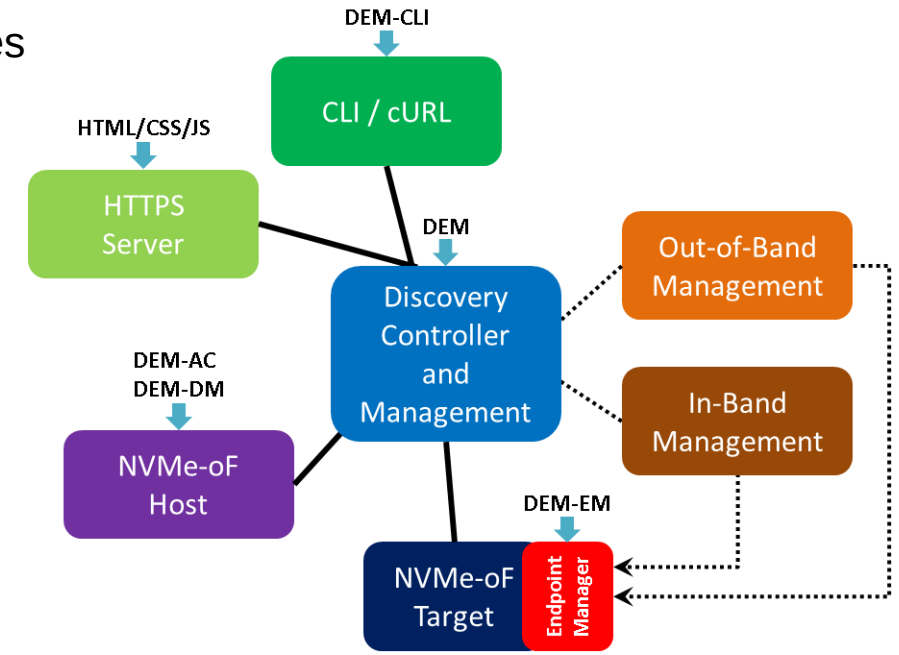
Enable efficient, dynamic configuration and provisioning of NVMe-oF Resources

Started out as a strawman driving specification changes into NVMe, NVMe-oF, and NVMe-MI

Management suite enabling:

- Remote configuration of NVMe-oF resources through RESTful interface
- Centralized enumeration of provisioned resources
- Single source for notification of changes to resources

Adopted by UNH-IOL with additional scripts written for Interoperability Test Suite



# DISTRIBUTED ENDPOINT MANAGEMENT

## DEM Project

**Distributed Endpoint Management (DEM): An Open-Source Project**

Enable efficient, dynamic configuration and provisioning of NVMe-oF Resources

Started out as a strawman driving specification changes into NVMe, NVMe-oF, and NVMe-MI

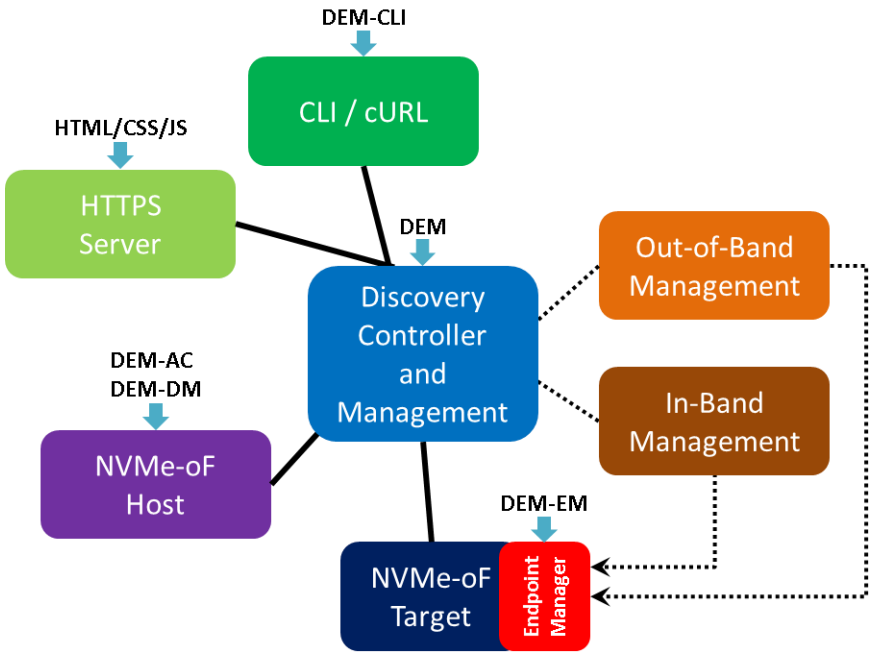
Management suite supporting:

NVMe-oF Transports

- RDMA - Validated on IB/iWARP/RoCE
- TCP

Configuration via

- In-band (i.e., using NVMe-oF protocol)
- Out-of-Band (i.e., using RESTful interface via JSON)



# DISTRIBUTED ENDPOINT MANAGEMENT

## DEM Project

## Project Components

### Discovery controller + Management (DEM)

### Endpoint Manager (DEM-EM) – customized for Target implementation

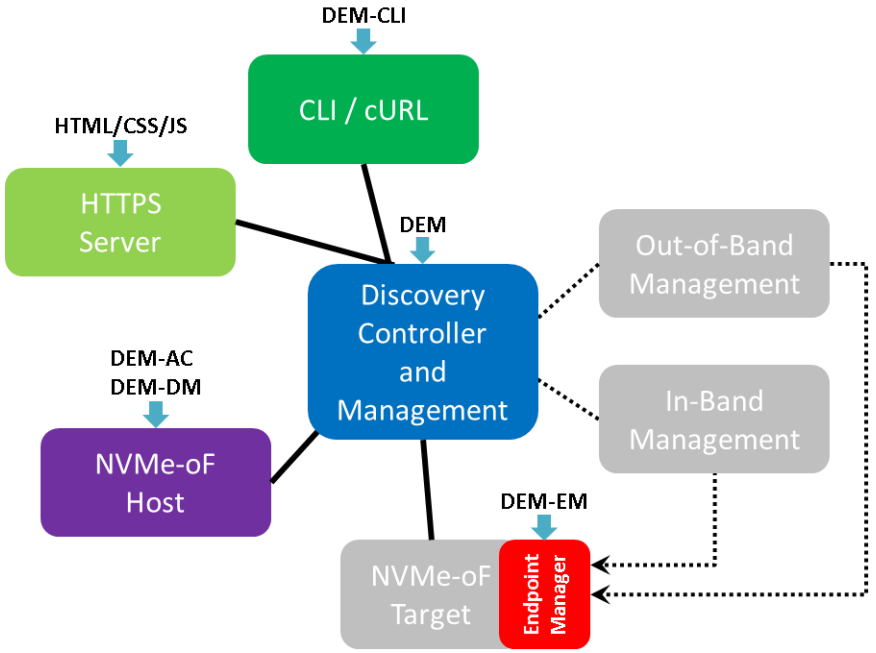
- In-Band Mode
- Out-of-Band Mode

### Optional Host Tools

- Auto Connect (DEM-AC)
- Discovery Log Page Monitor (DEM-DM)

### RESTful Management Interfaces

- Web Pages (HTML / CSS files)
- Command Line Interface (DEM-CLI)`



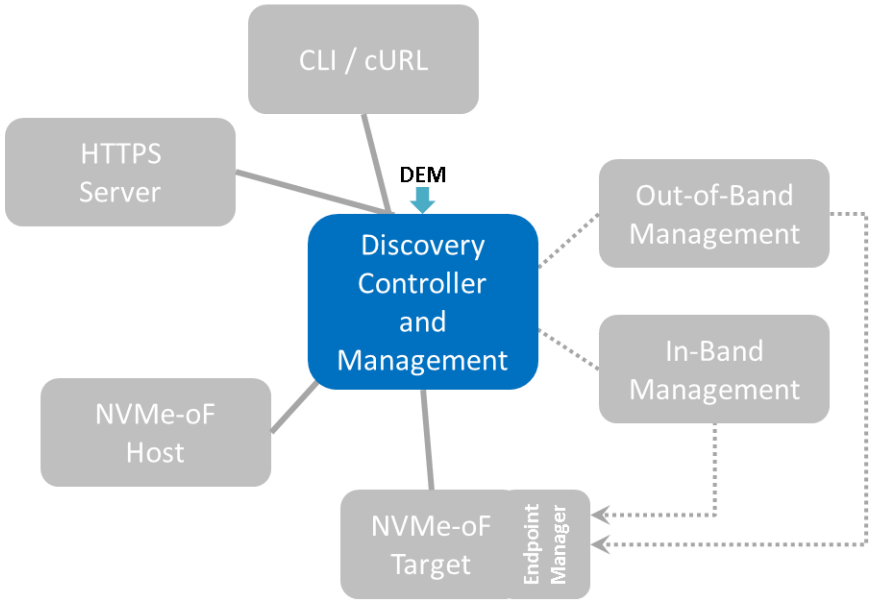
# DISTRIBUTED ENDPOINT MANAGEMENT

## Project Components

### Discovery controller + Management (DEM)

- Primary component for remote configuration and provisioning
- Plug-in module architecture for NVMe-oF supported Fabrics

- Configures remote NVMe resources via In-Band or Out-of-Band interfaces
- Collects & distributes tailored Discovery Log Pages to Hosts
- Receives notification of changes to NVMe-oF resources
- Reports changes to NVMe-oF resources to affected registered Hosts
- Enables additional access restrictions



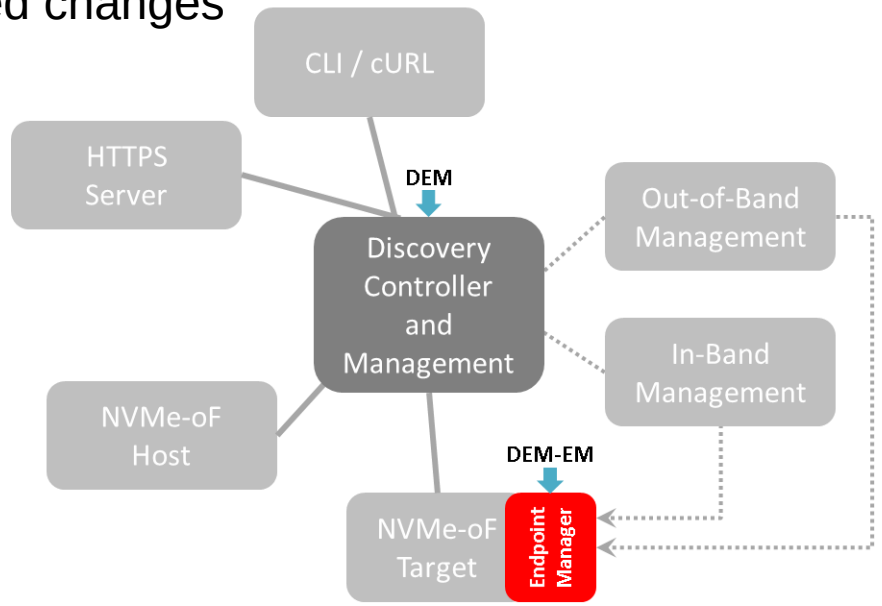
# DISTRIBUTED ENDPOINT MANAGEMENT

## Project Components

### Endpoint Manager (DEM-EM)

- Agent running on Target enabling remote configuration
- Reuses DEM plug-in module architecture for NVMe-oF supported Fabrics

- Plug-in configuration model enabling implementation-specific management of NVMe-oF resources
- In-Band configuration based on proposed changes to NVMe-MI specification
- Out-of-Band RESTful configuration based on proposed changes to RF/SF
- Used for Targets not managed by other means



# DISTRIBUTED ENDPOINT MANAGEMENT

## Project Components

### Optional Host Tools

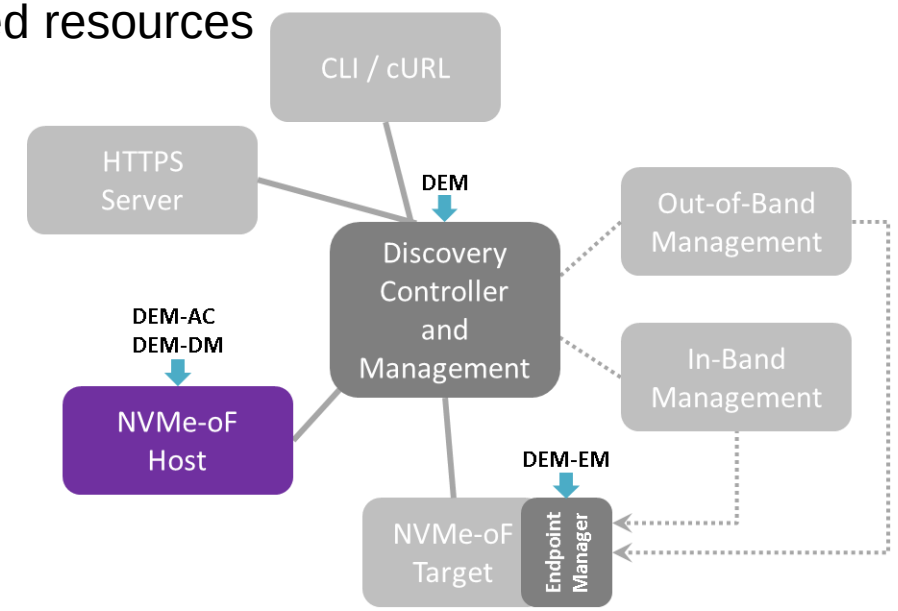
- Reuses DEM plug-in module architecture for NVMe-oF supported Fabrics

### Auto Connect (DEM-AC)

- Establishes persistent connection
- Collects Discovery Log Pages
- Automatically connects to its provisioned resources

### Discovery Log Page Monitor (DEM-DM)

- Establishes persistent connection
- Reports Log Change Events and displays updated Discovery Log Pages



# DISTRIBUTED ENDPOINT MANAGEMENT

## Project Components

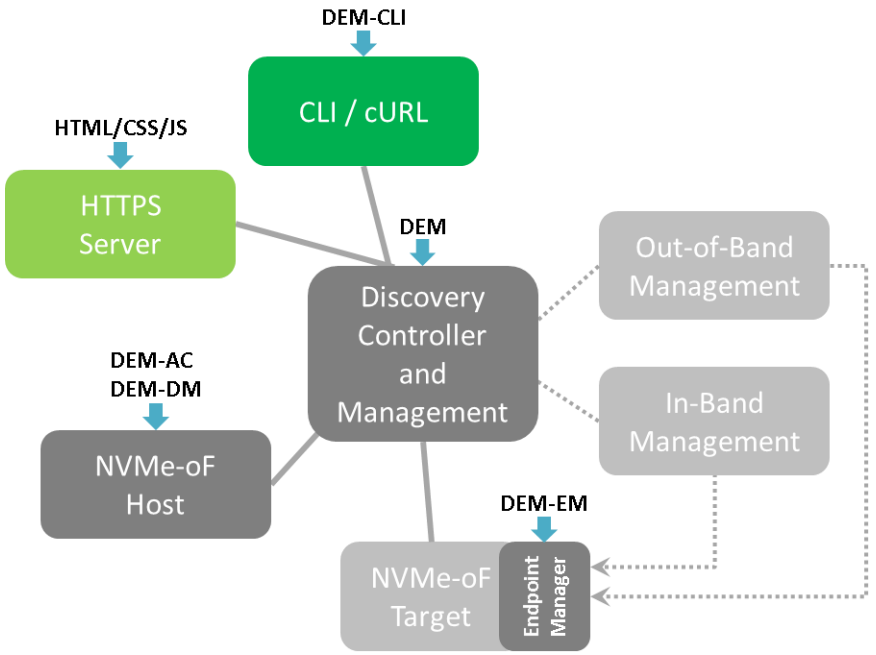
### RESTful Management Interfaces

#### DEM Command Line Interface (DEM-CLI)

- Local Interface to the DEM via console command line

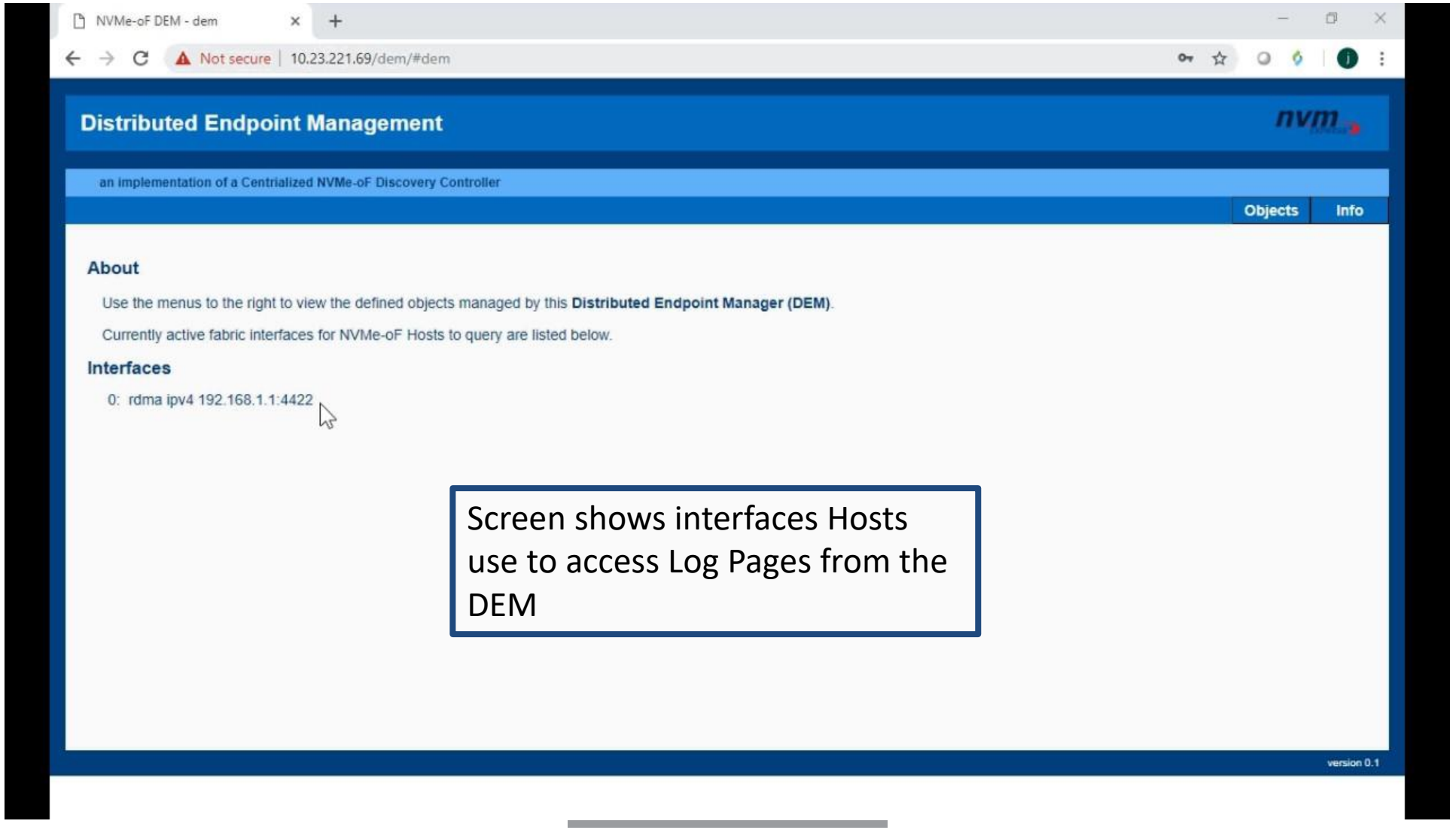
#### Web interface

- Interface to the DEM via web interface
- Project contains complete set of HTML, CSS, and JS files



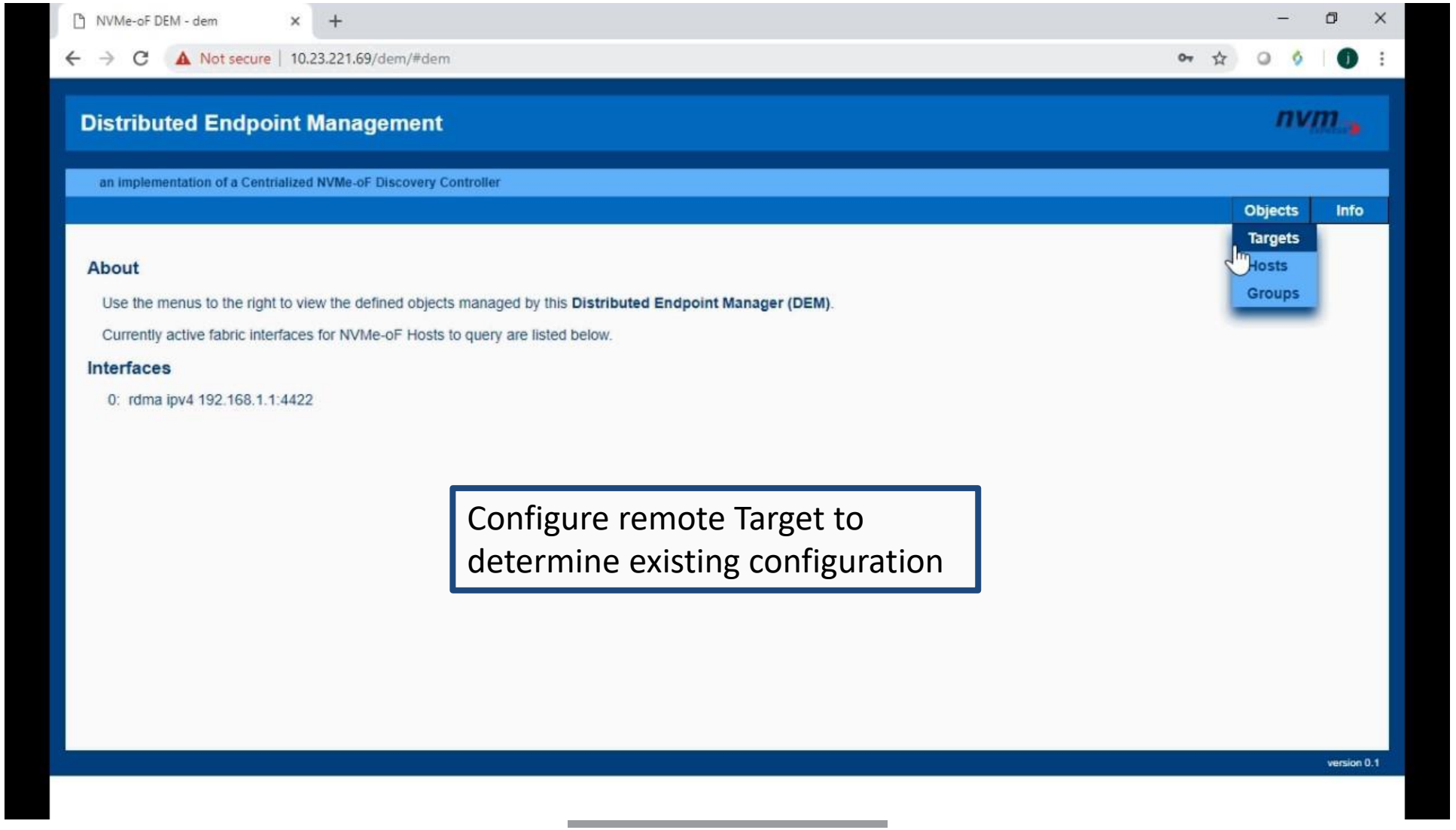

# BRIEF
# DEM-ONSTRATION

# DEMONSTRATION



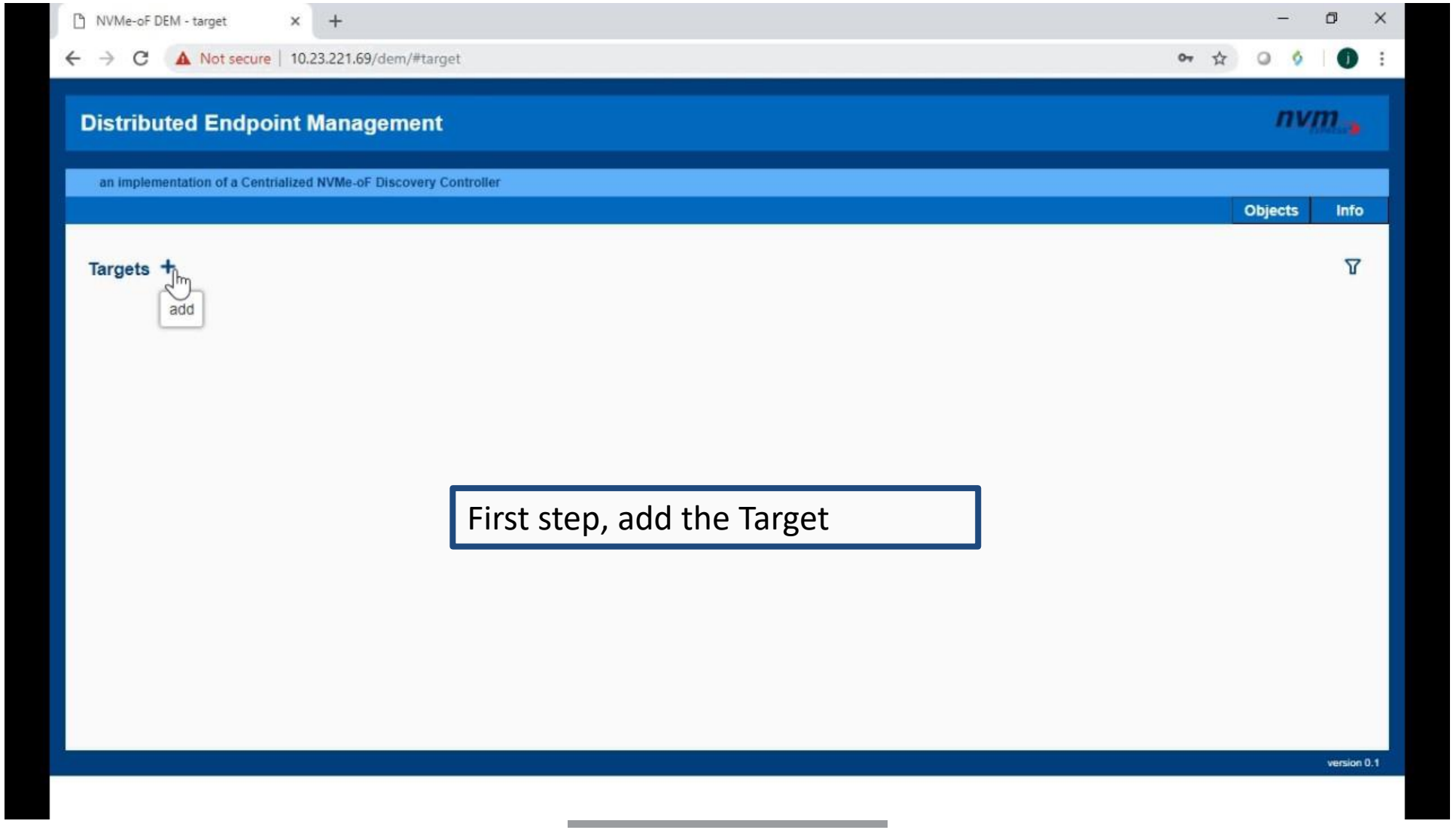

Screen shows interfaces Hosts use to access Log Pages from the DEM

# DEMONSTRATION

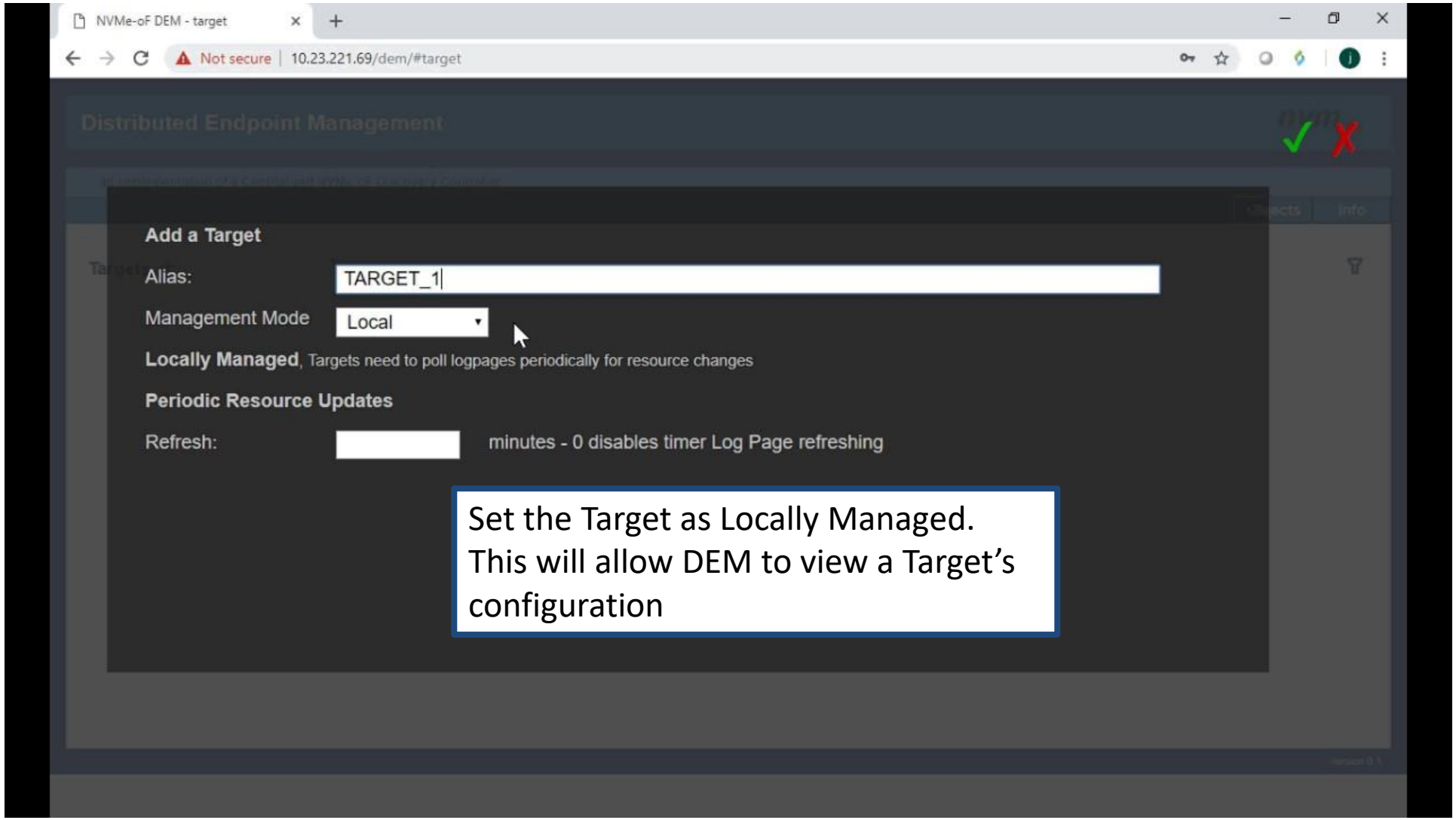



Configure remote Target to determine existing configuration

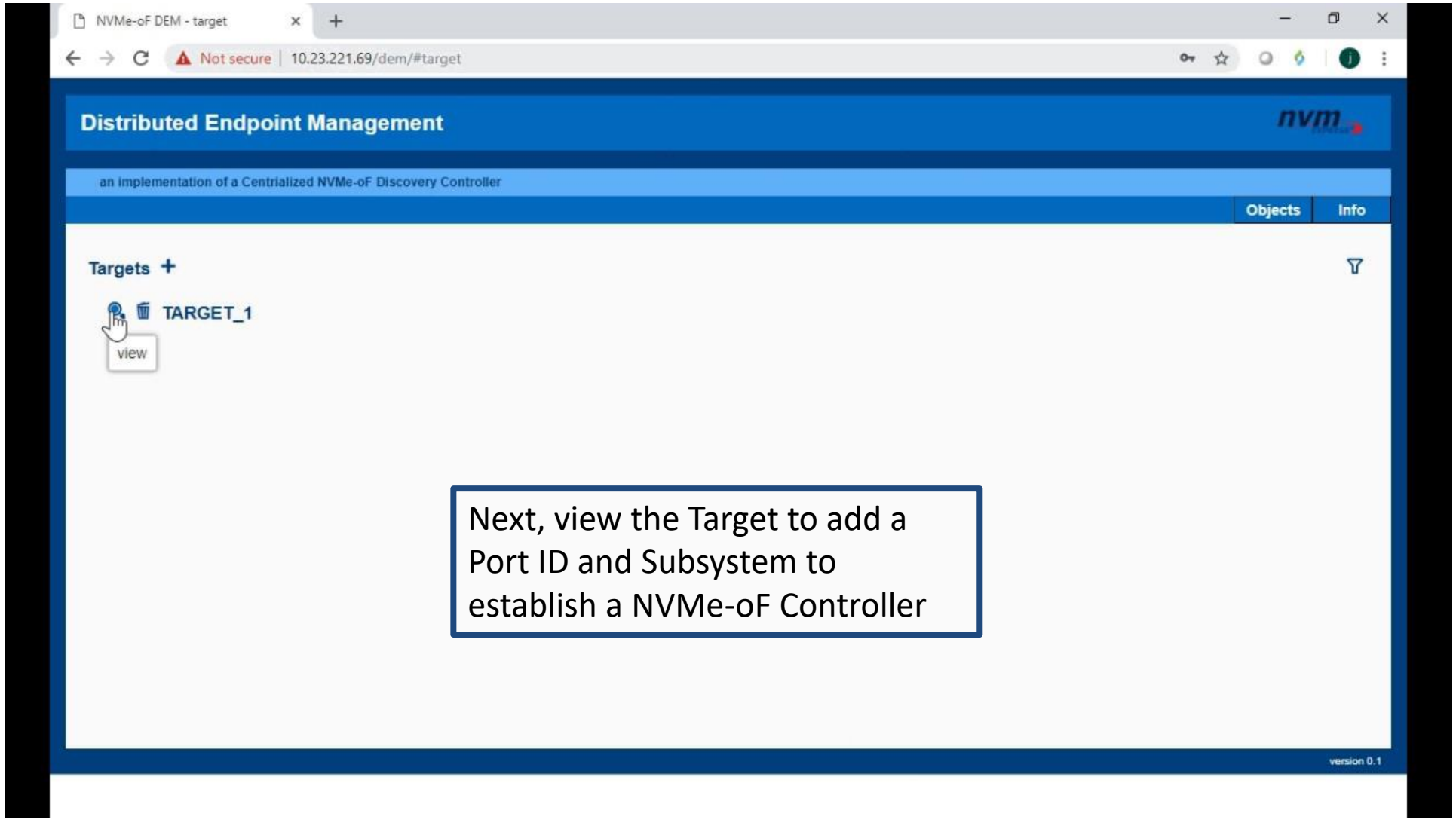
# DEMONSTRATION



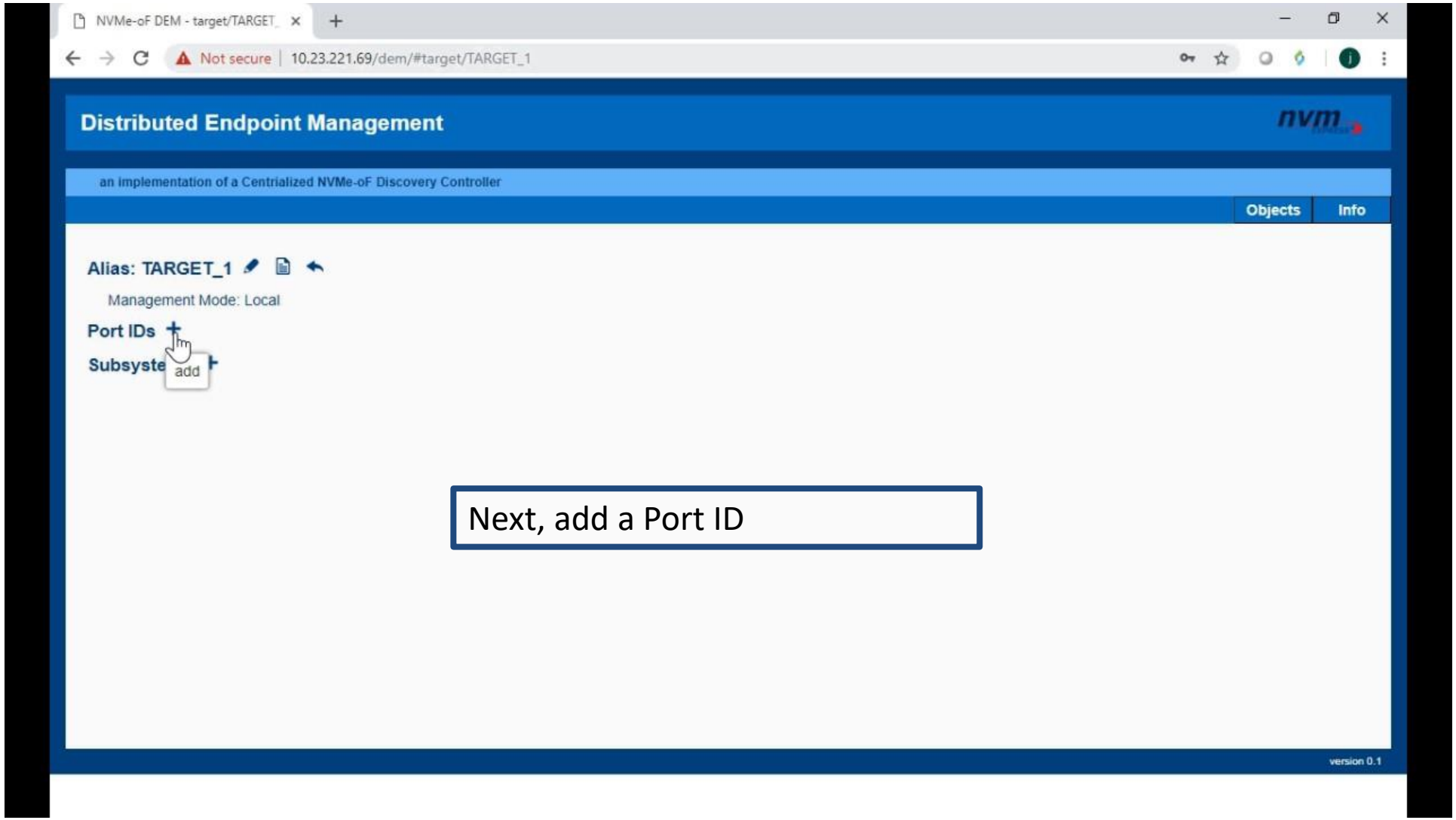
# DEMONSTRATION



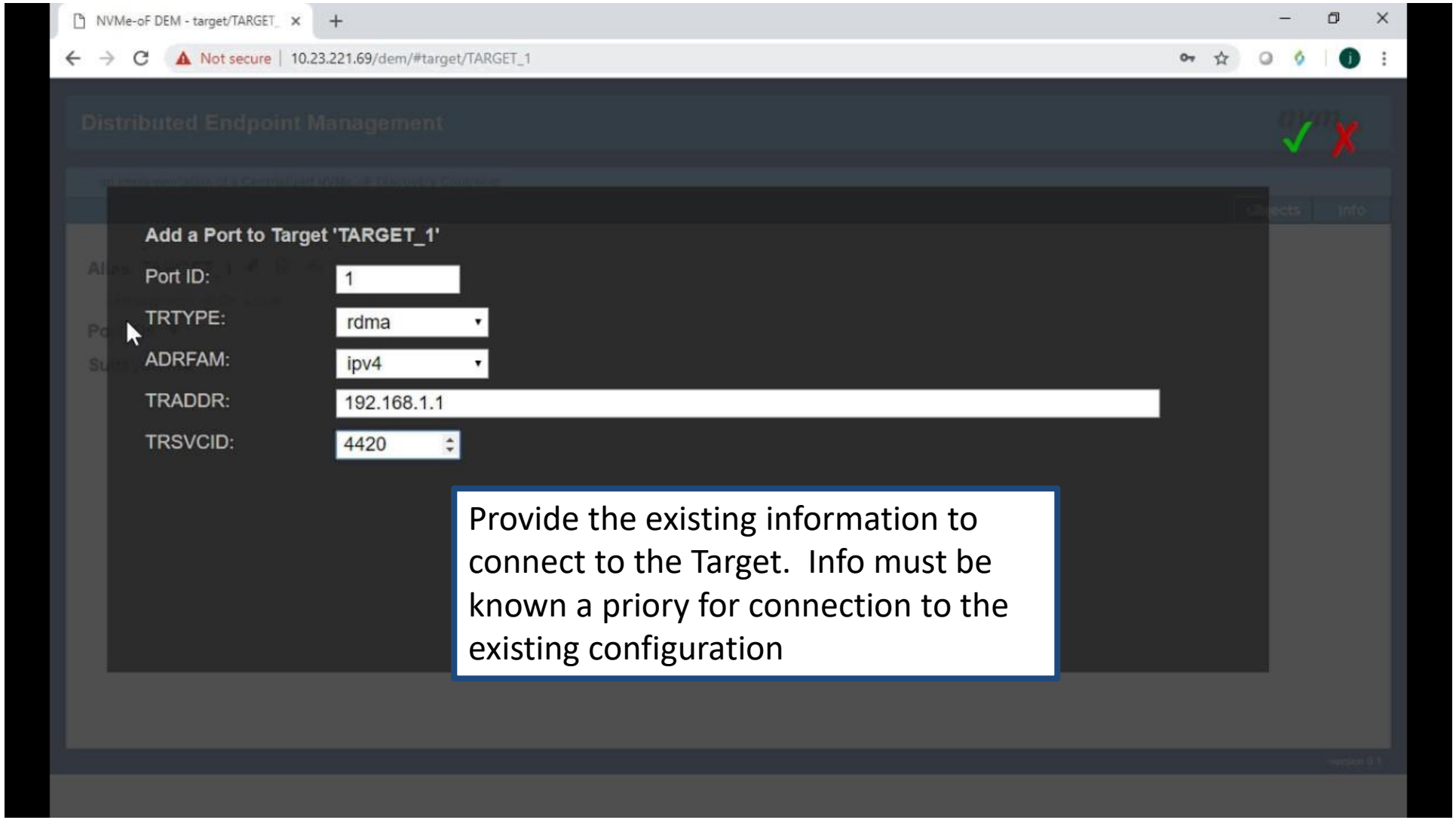
# DEMONSTRATION



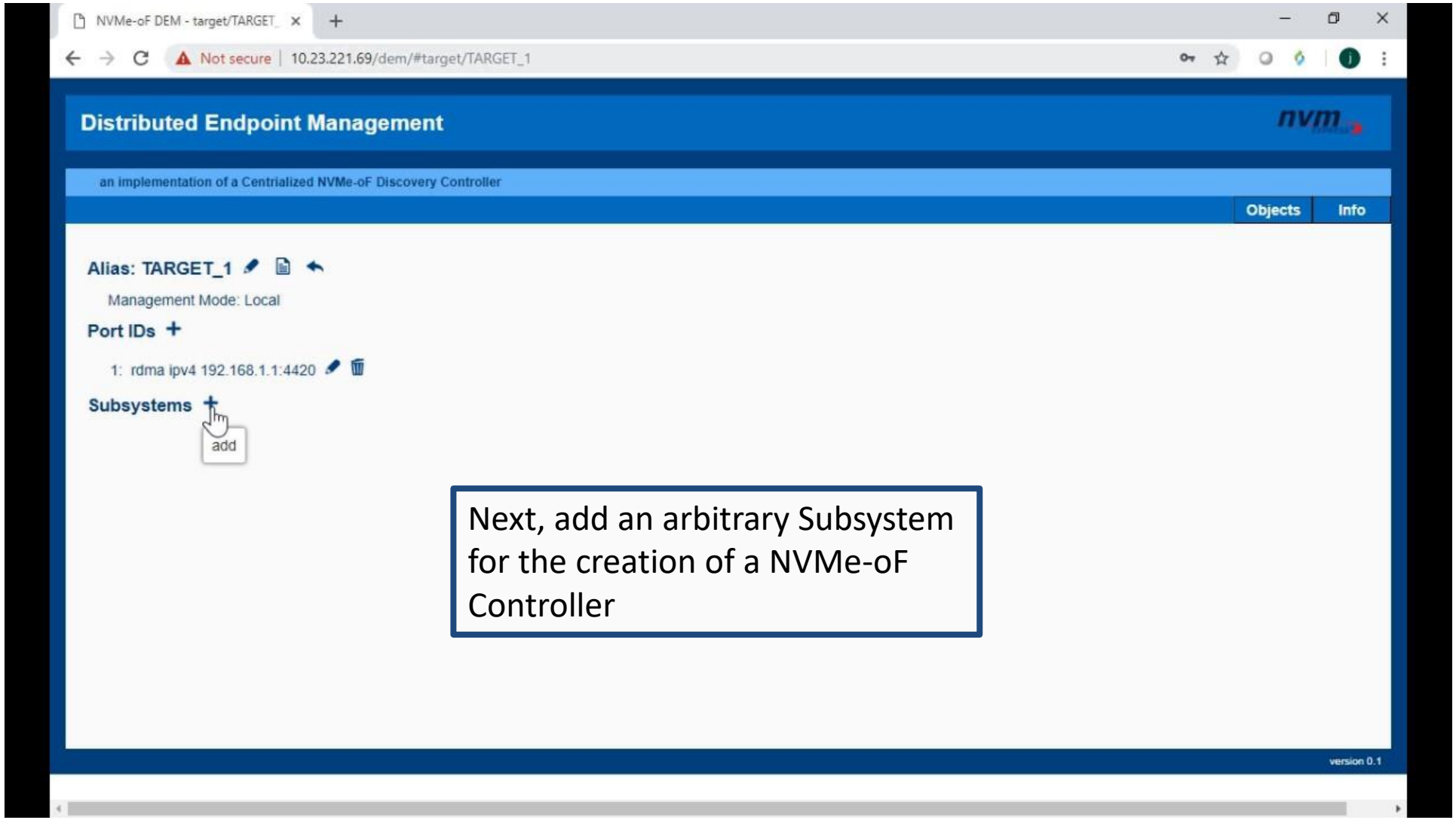
# DEMONSTRATION



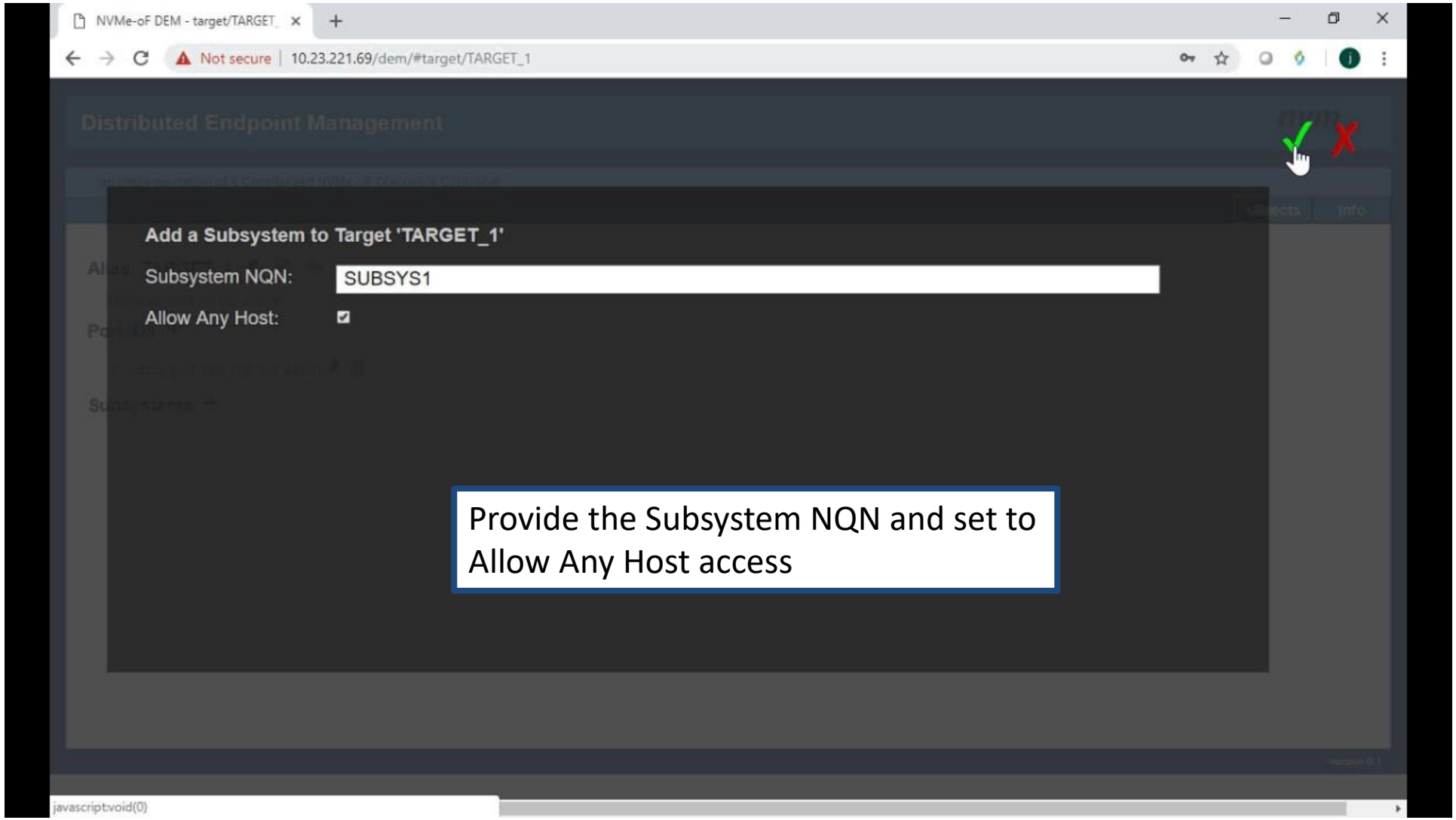
# DEMONSTRATION



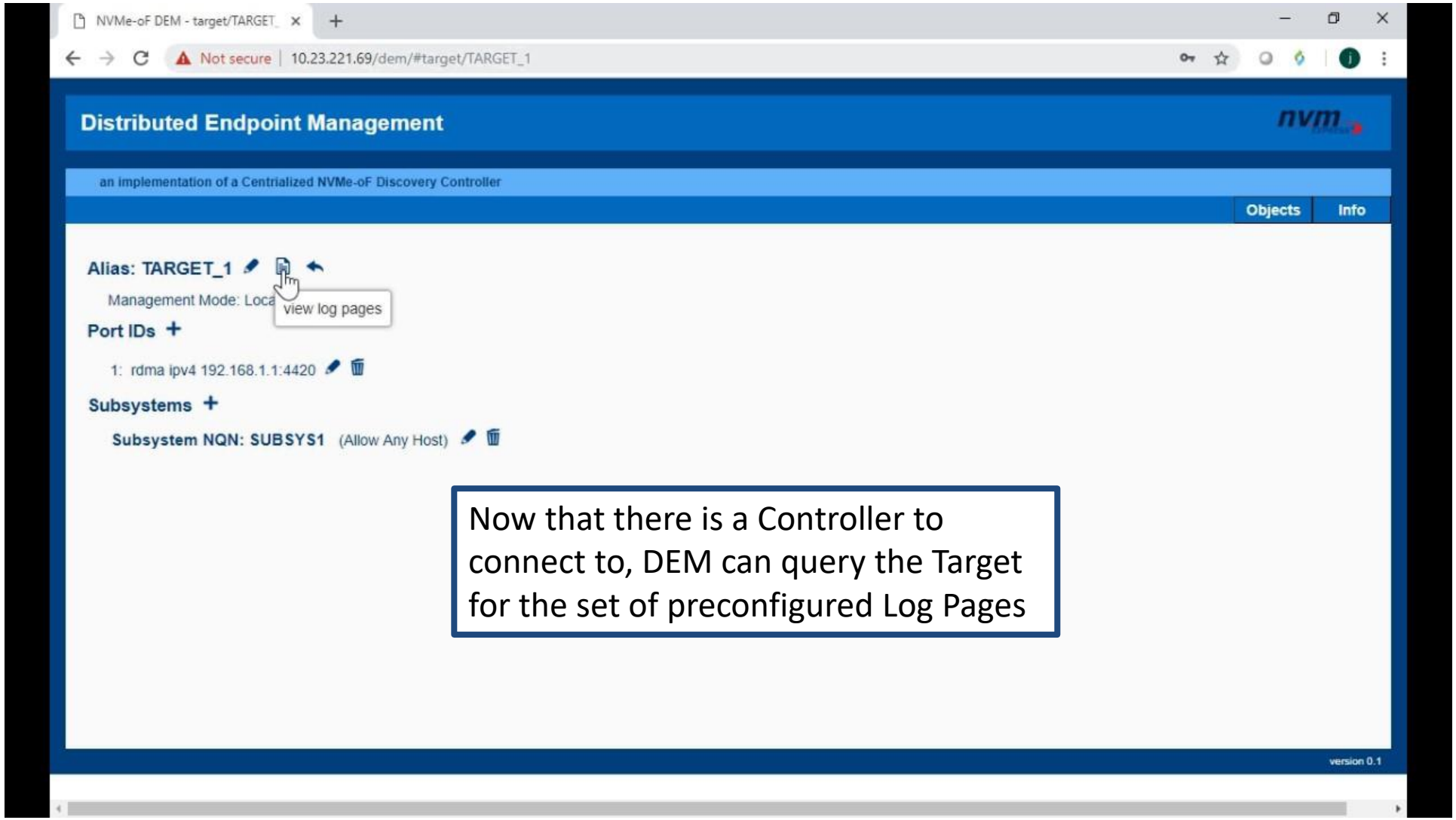
# DEMONSTRATION



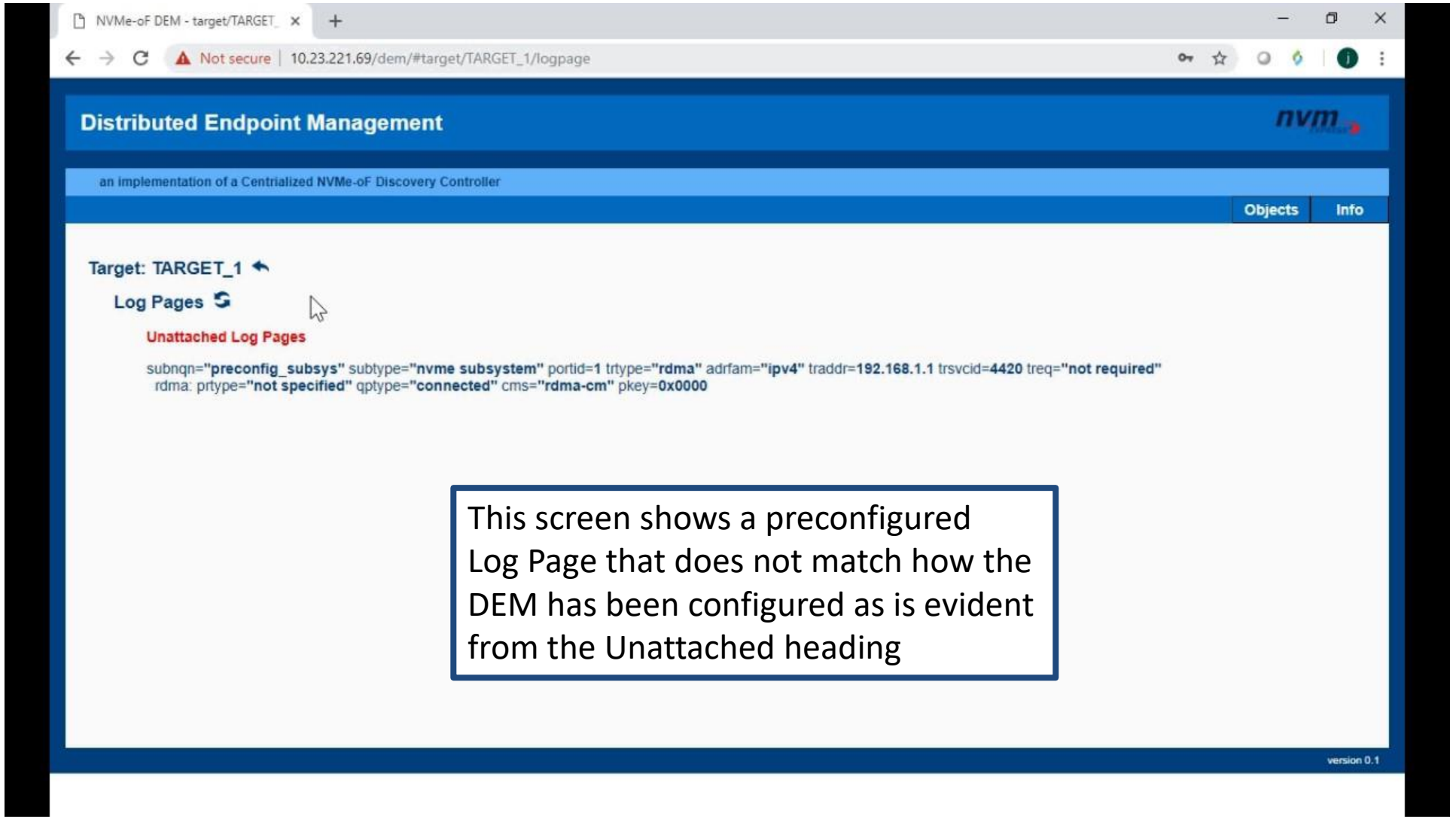
# DEMONSTRATION



# DEMONSTRATION

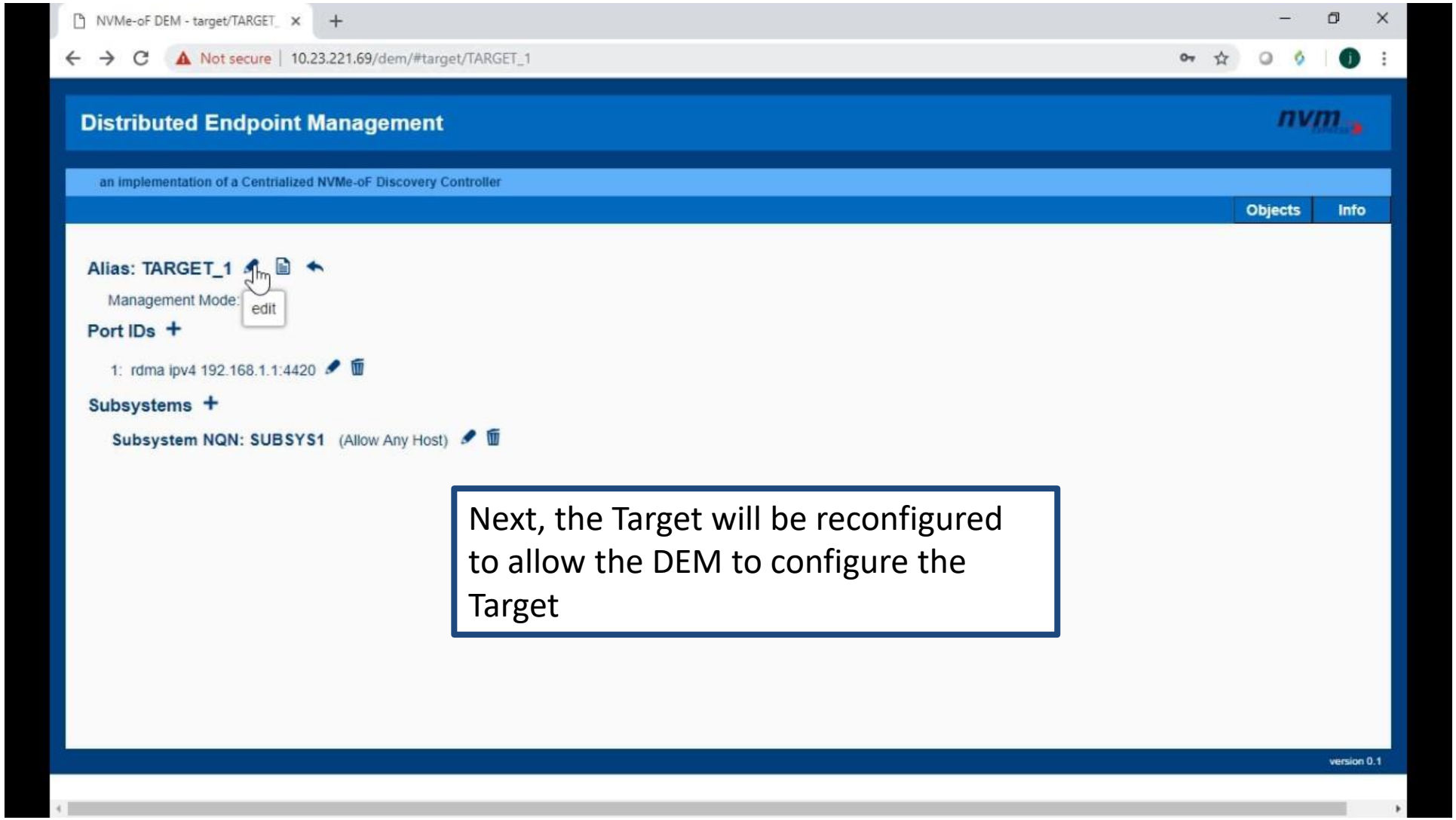


# DEMONSTRATION

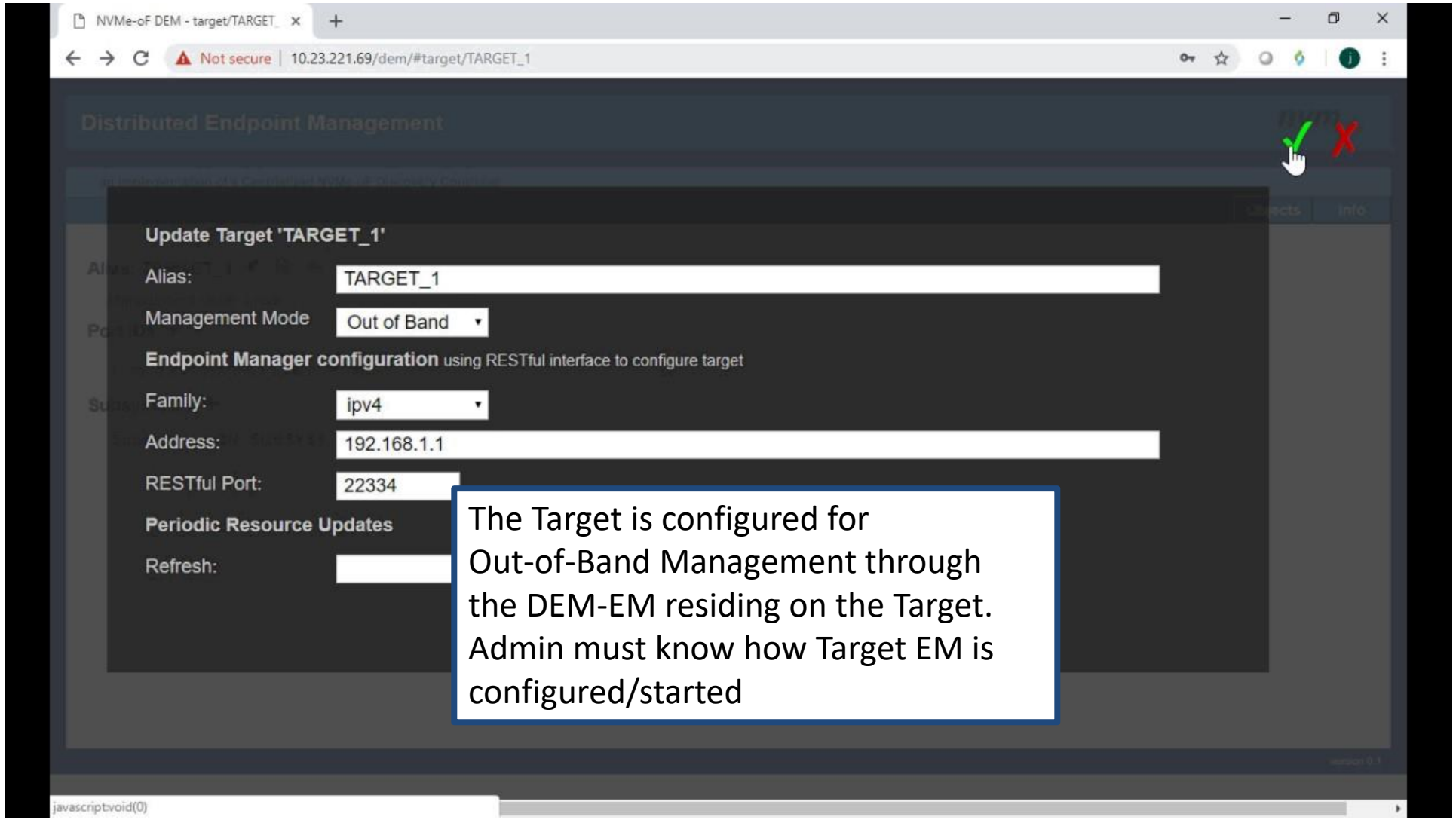



This screen shows a preconfigured Log Page that does not match how the DEM has been configured as is evident from the Unattached heading

# DEMONSTRATION

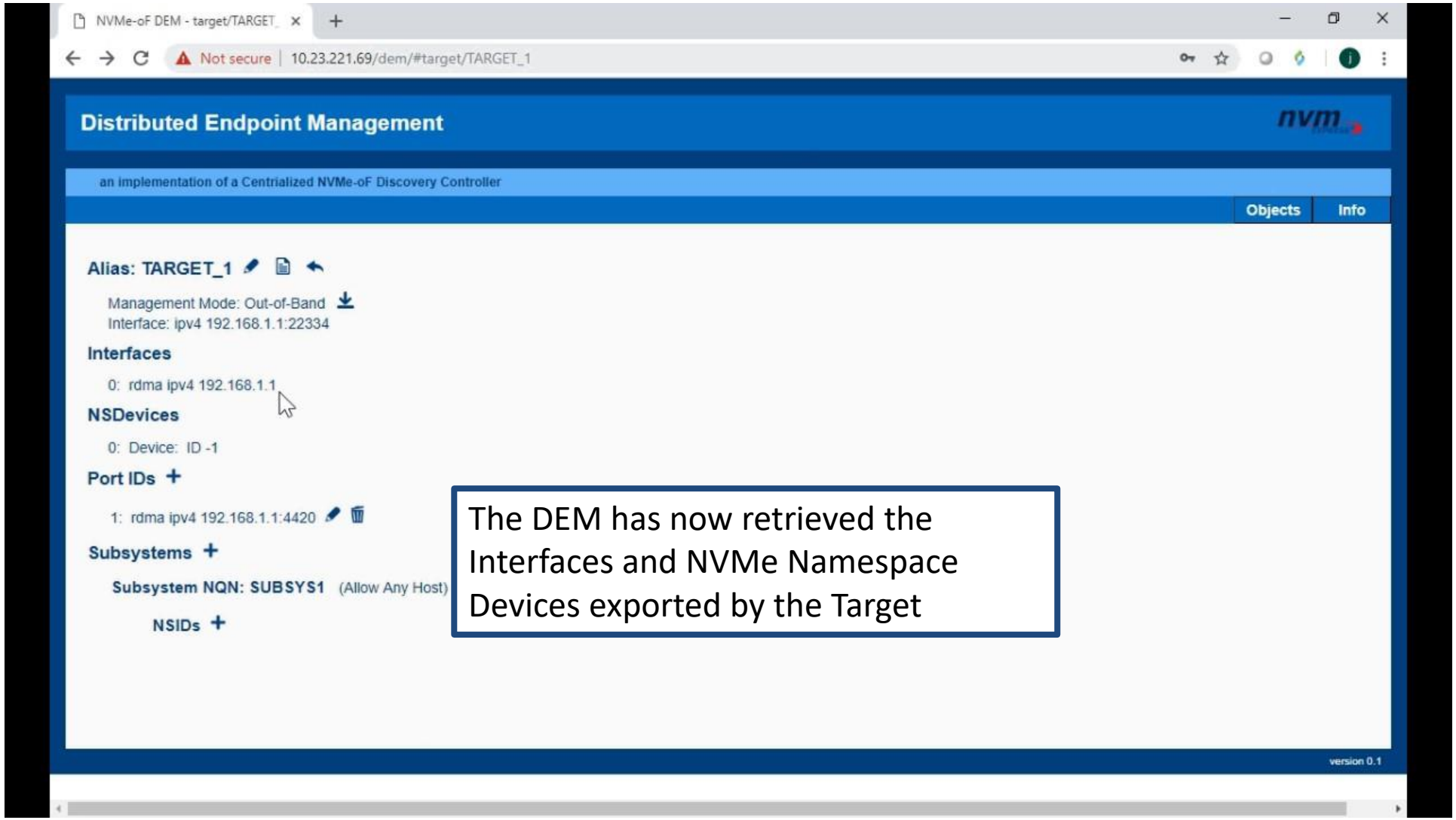


# DEMONSTRATION



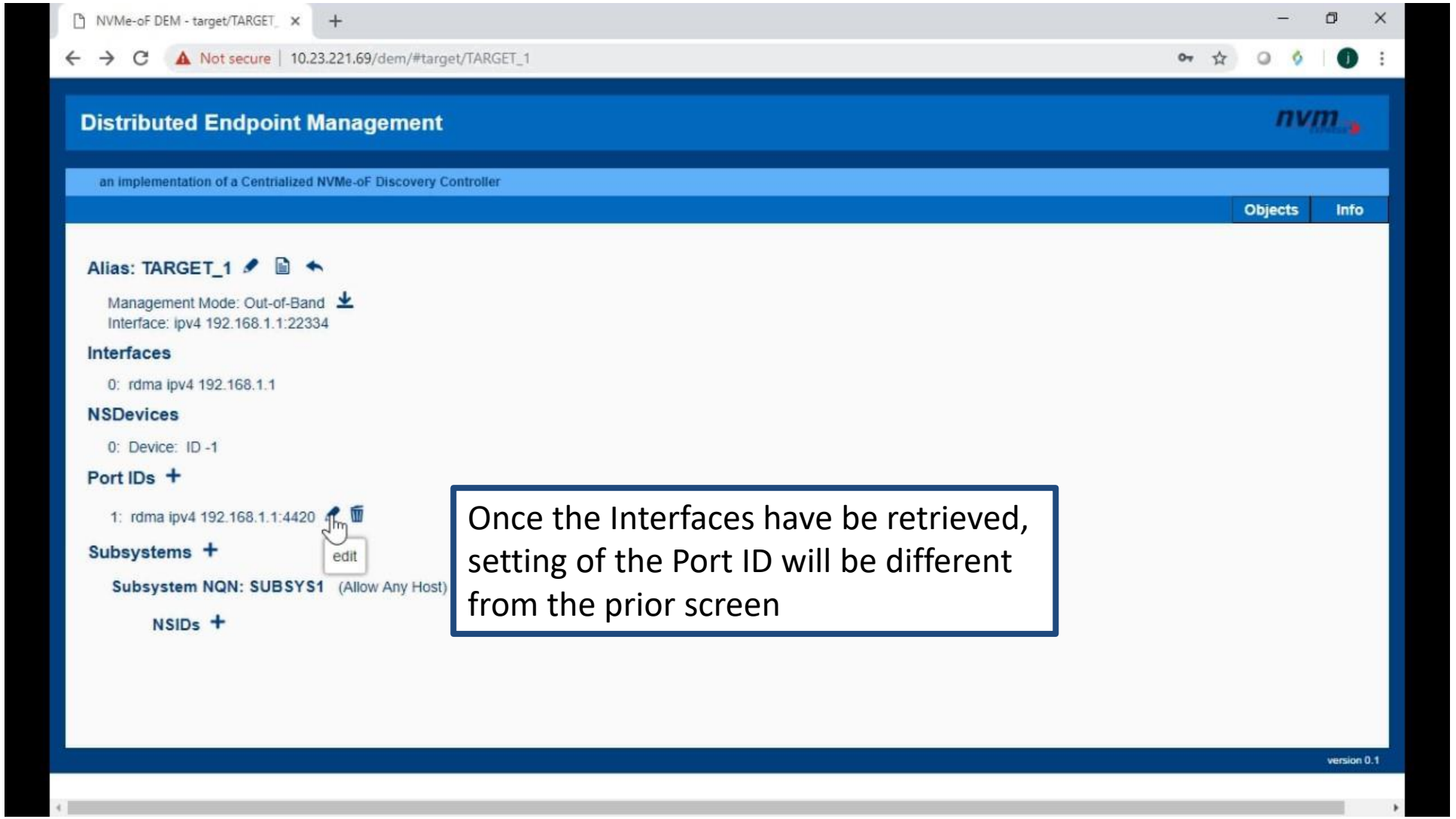

The Target is configured for Out-of-Band Management through the DEM-EM residing on the Target. Admin must know how Target EM is configured/started

# DEMONSTRATION



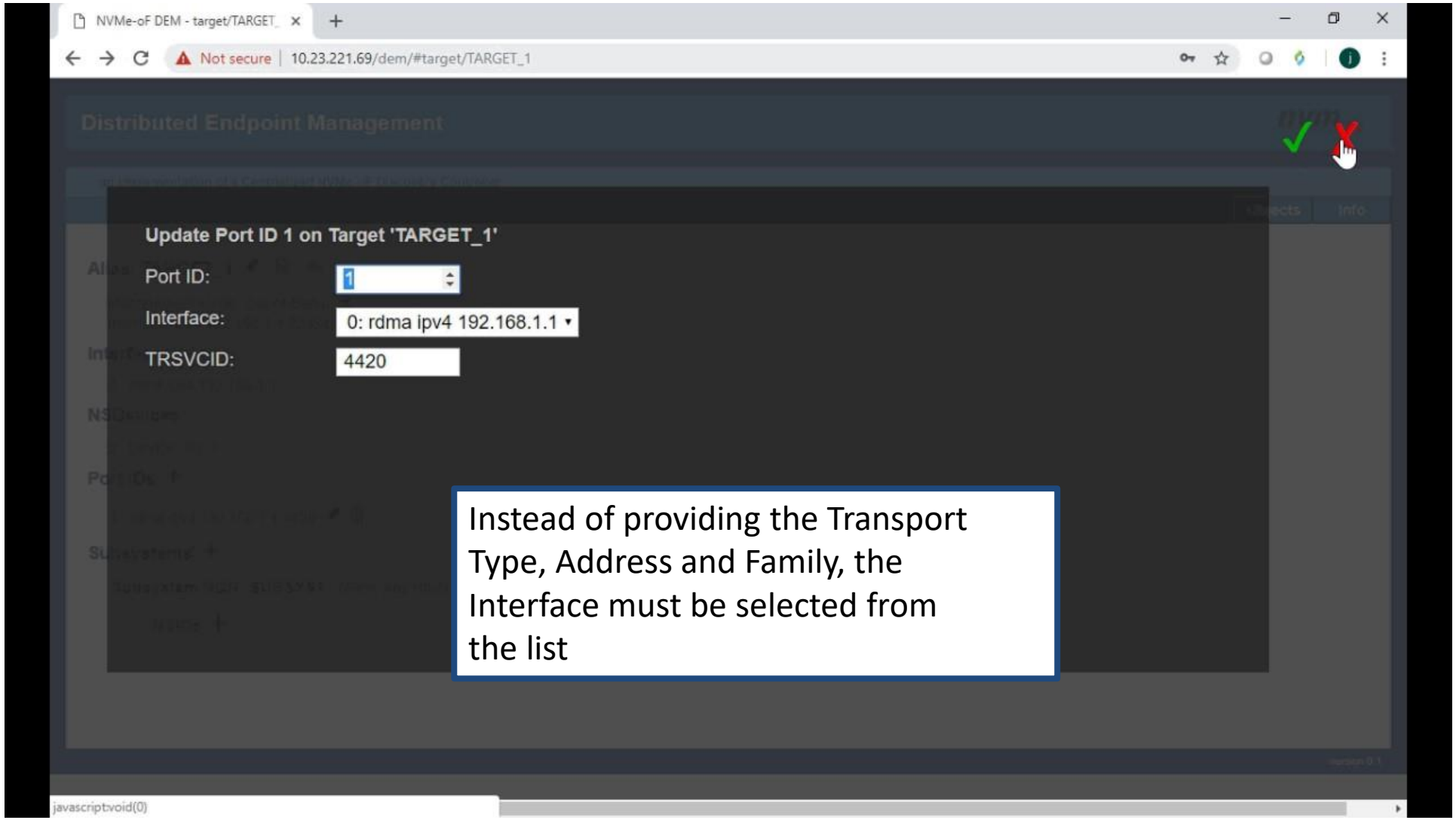

The DEM has now retrieved the Interfaces and NVMe Namespace Devices exported by the Target

# DEMONSTRATION

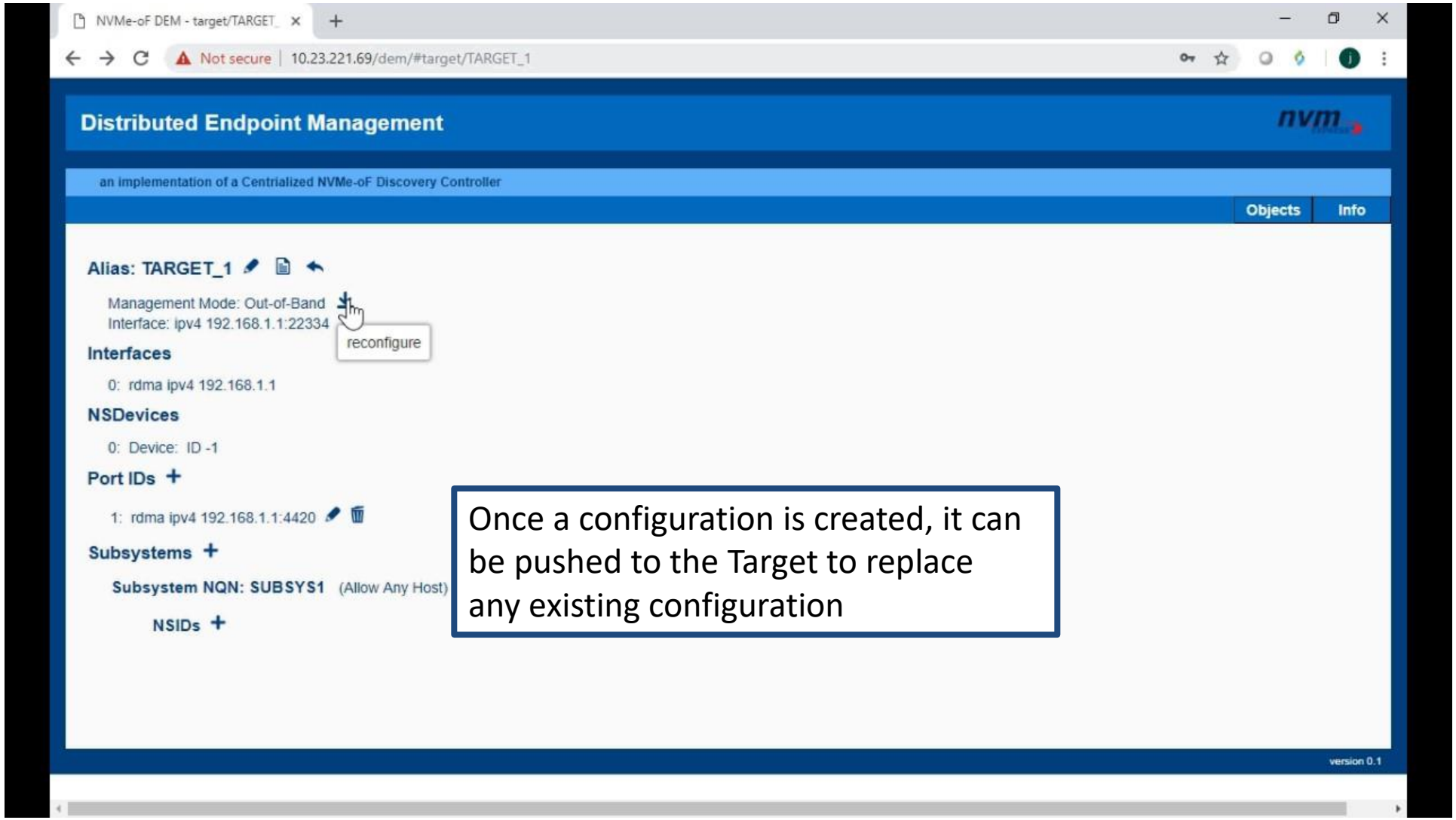



Once the Interfaces have be retrieved, setting of the Port ID will be different from the prior screen

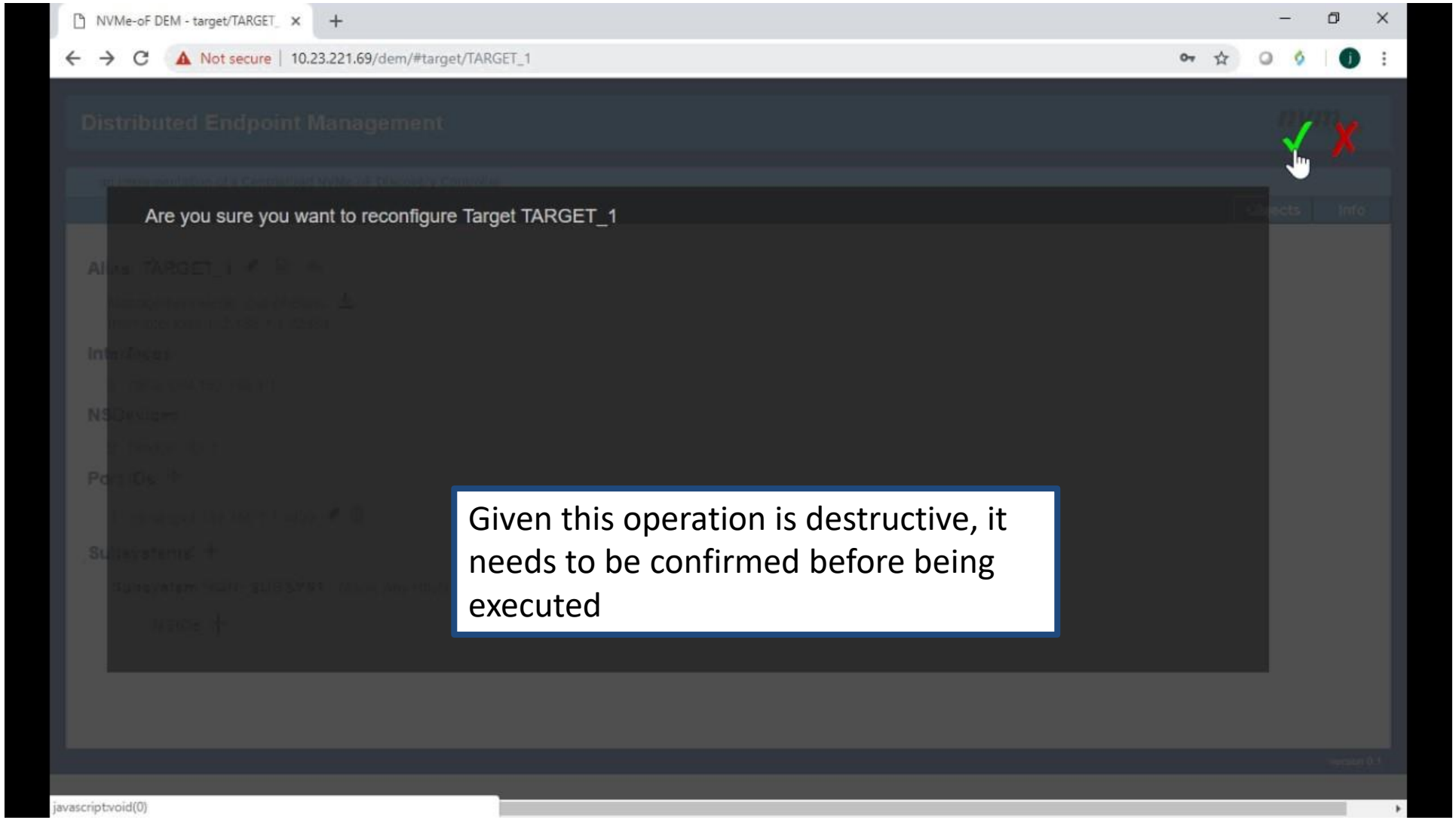
# DEMONSTRATION



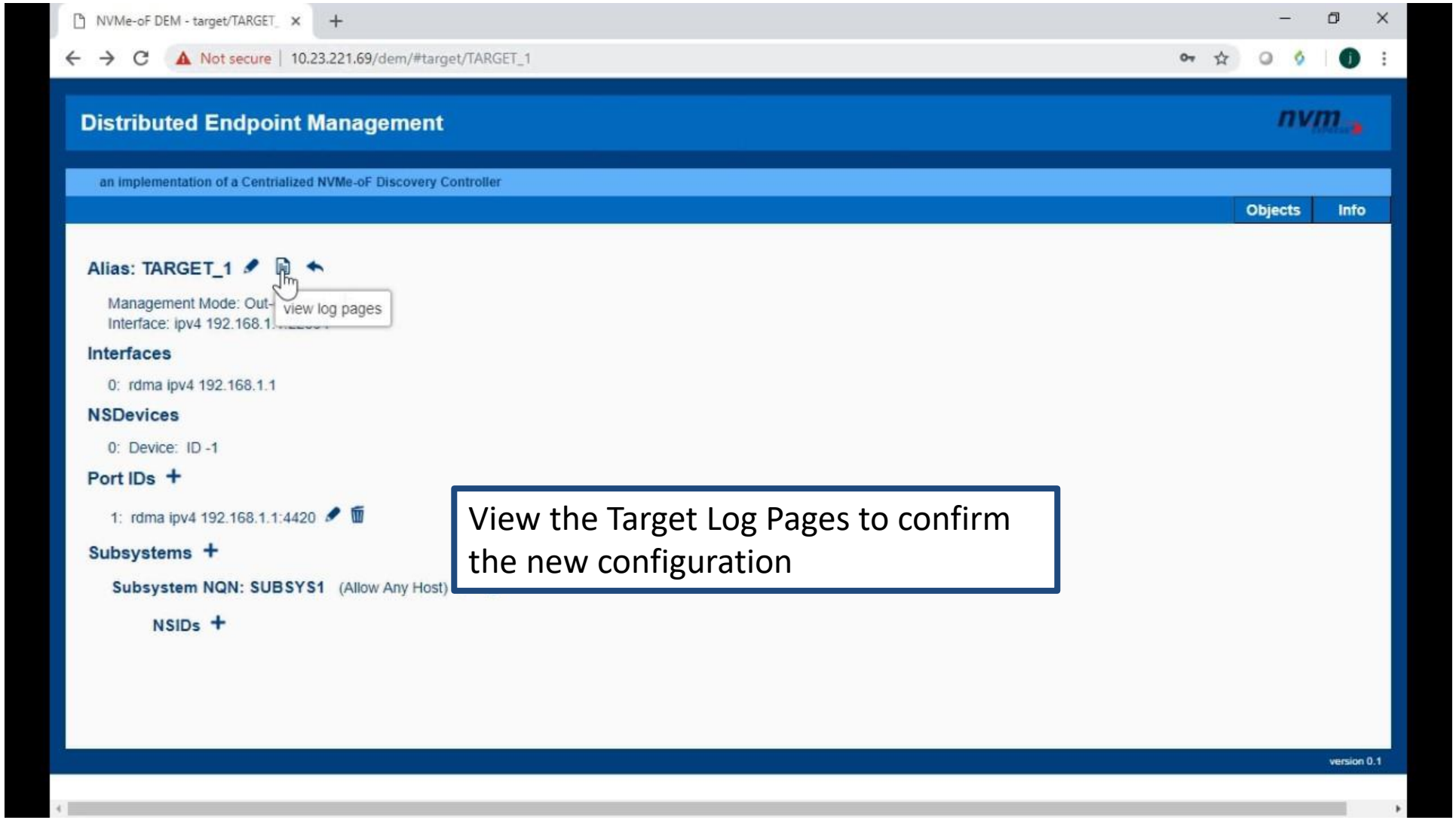
# DEMONSTRATION



# DEMONSTRATION

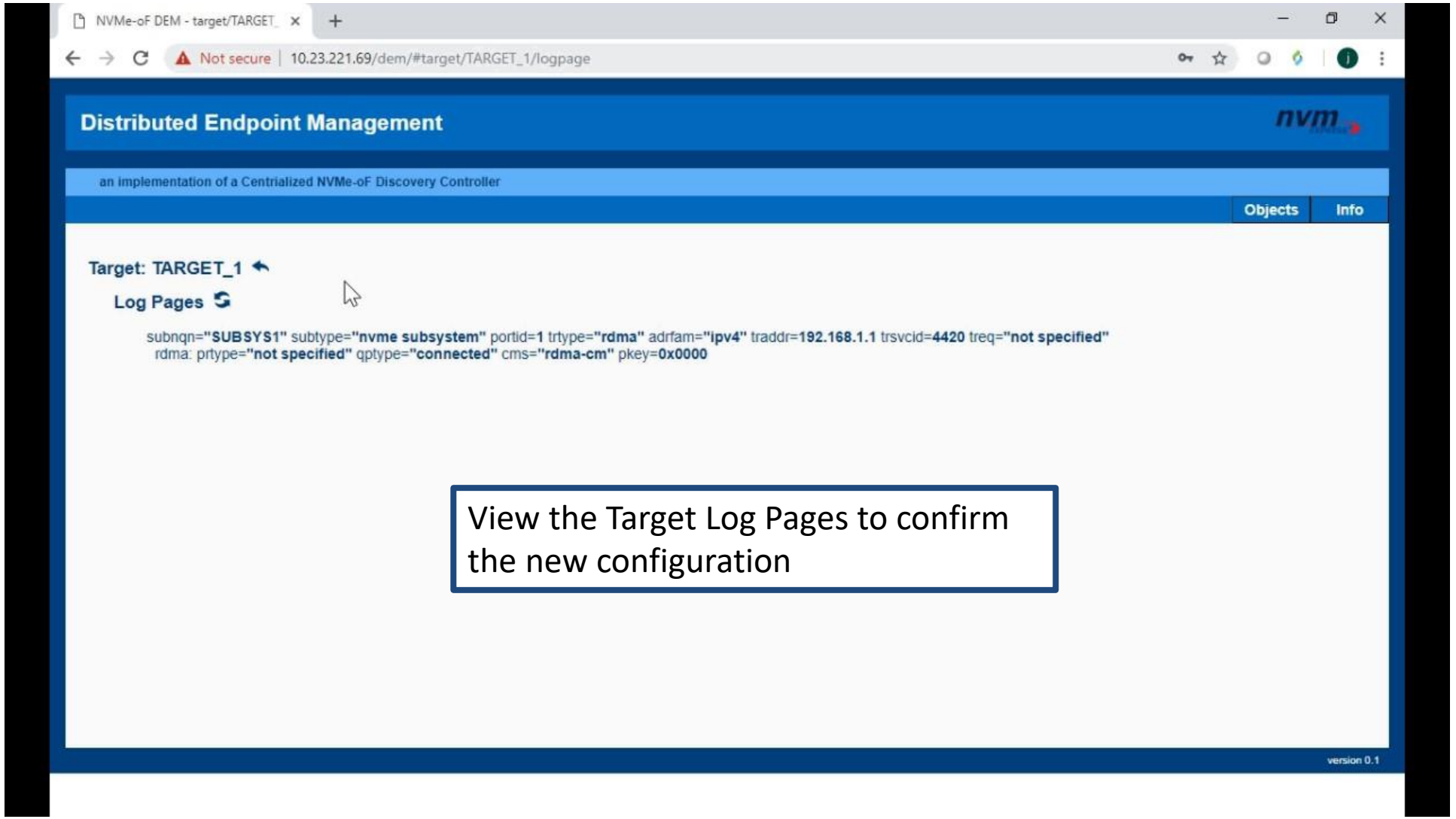


# DEMONSTRATION

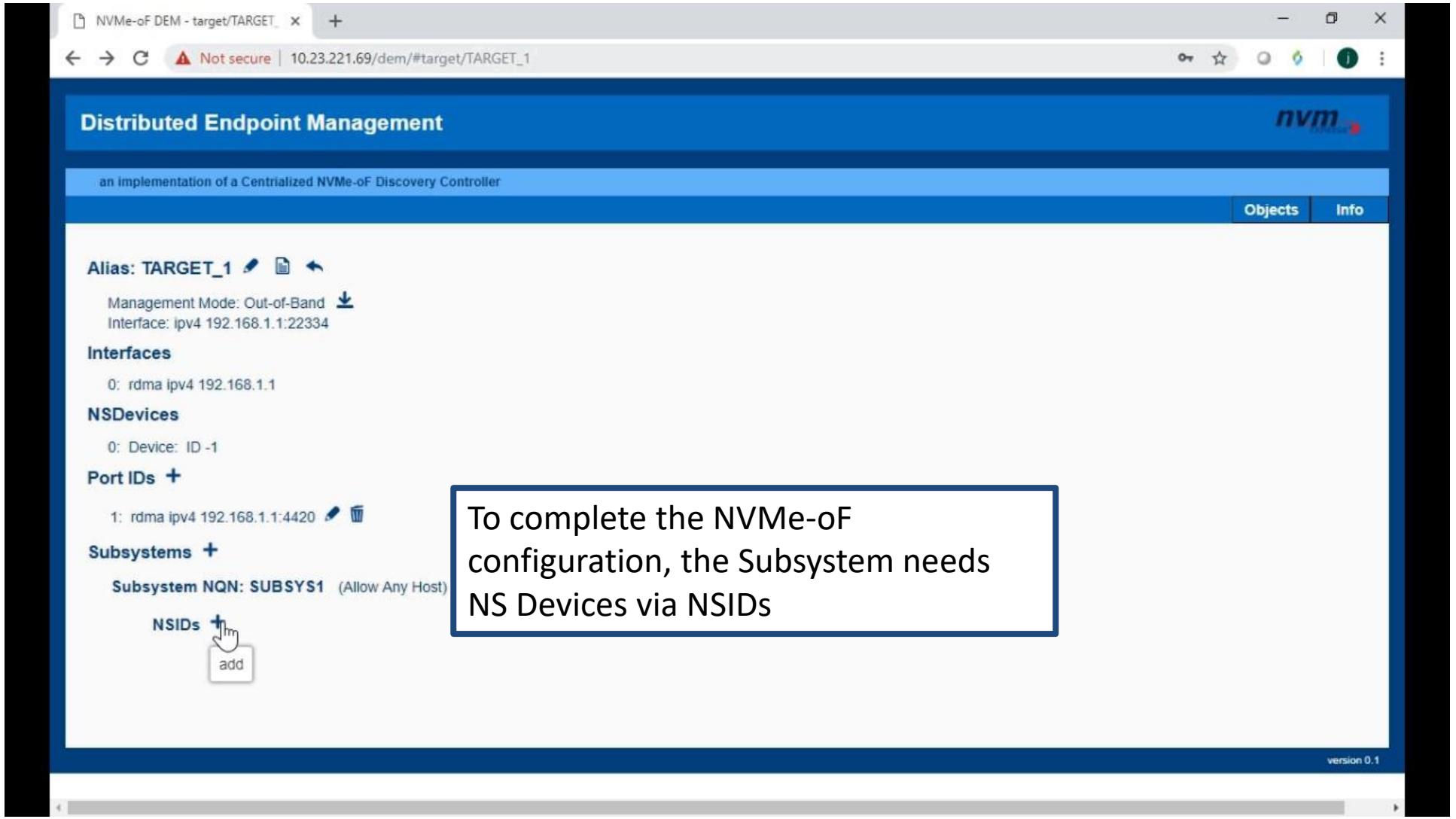



View the Target Log Pages to confirm the new configuration

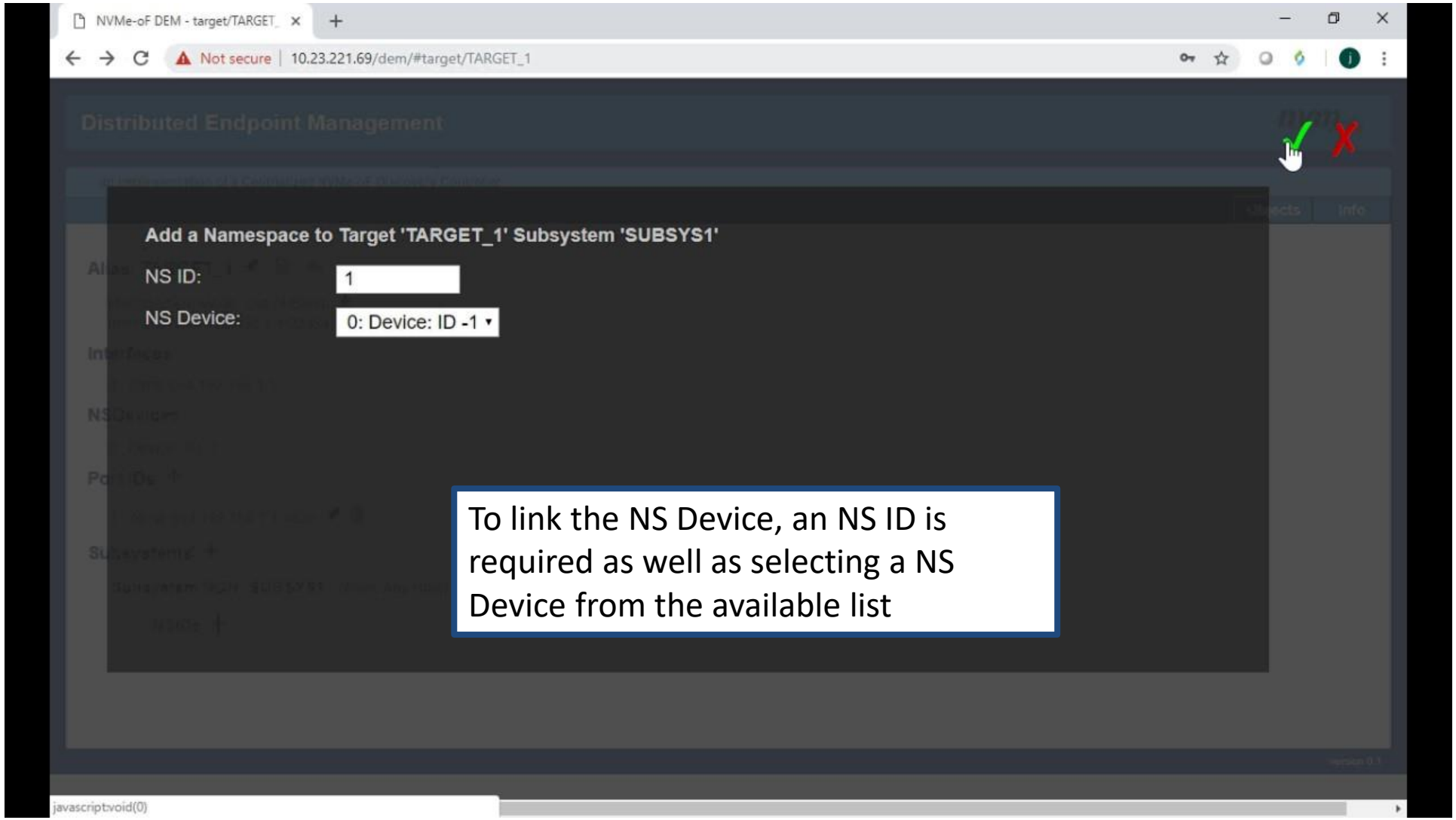
# DEMONSTRATION



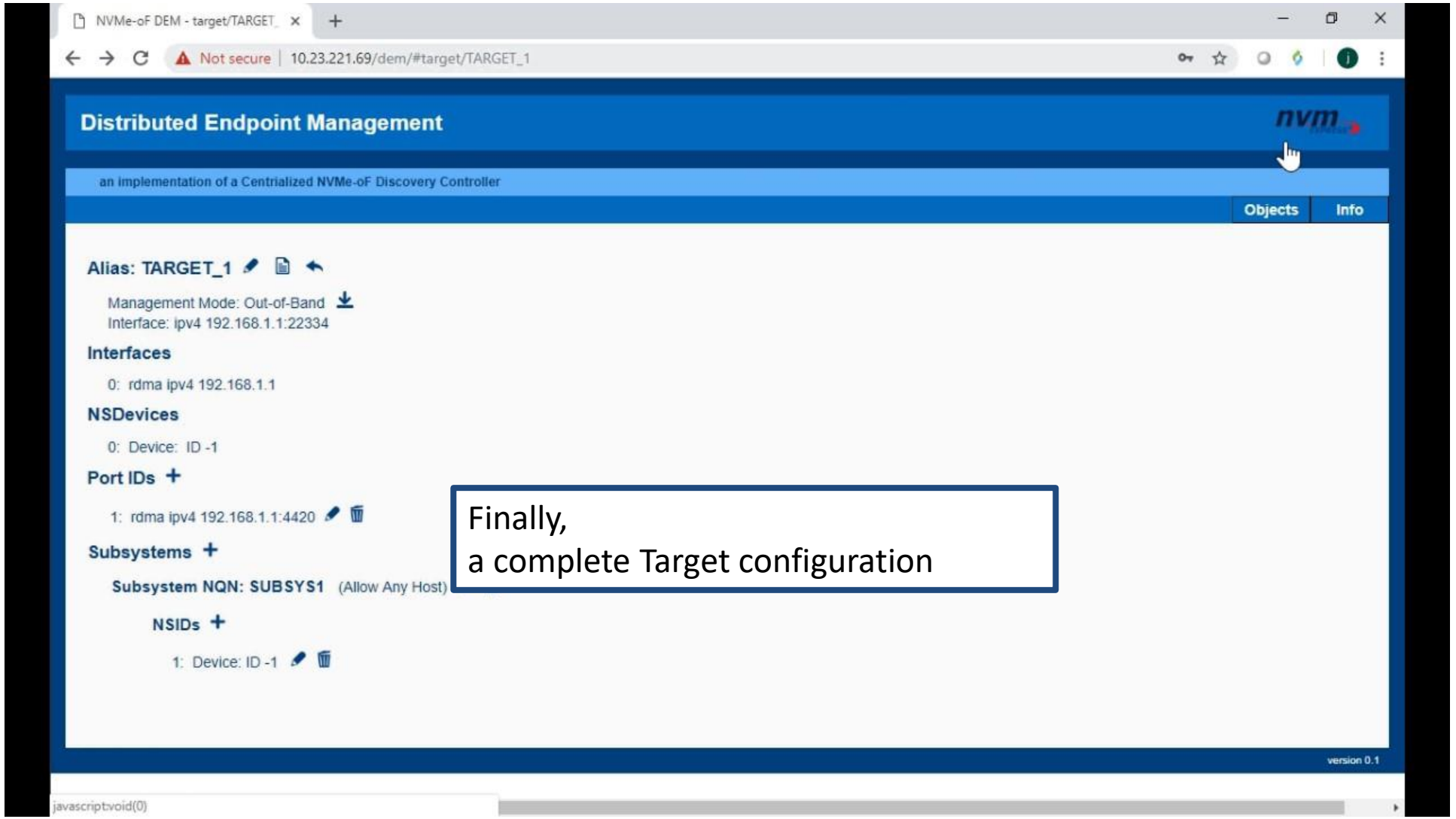
# DEMONSTRATION



# DEMONSTRATION



# DEMONSTRATION




Finally,
a complete Target configuration


# DEVELOPMENT OPPORTUNITIES AND WRAP-UP

# DISTRIBUTED ENDPOINT MANAGEMENT

## Development Opportunities

### Discovery controller + Management (dem)

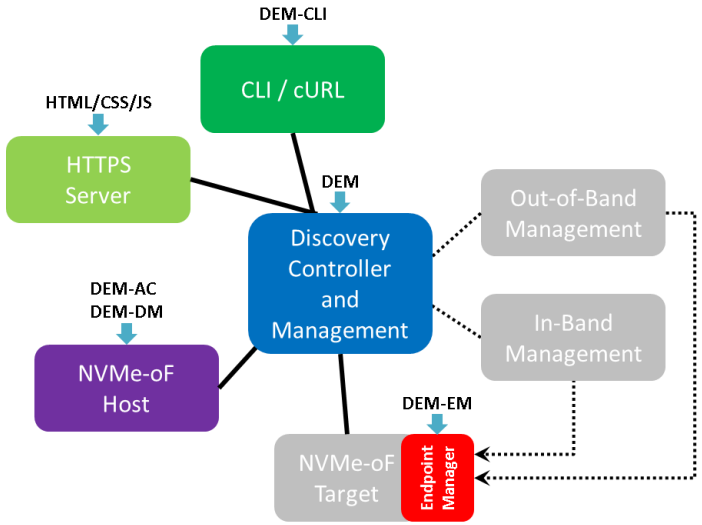
- Redundancy and Failover
- Generate Log Pages for Targets without a Discovery Controller
- Target usage monitoring
- Convert DEM to RedFish / SwordFish schema
- Convert DEM in-band configuration to current NVMe-MI Specification proposal

### Endpoint Manager (dem-em)

- Target usage monitoring
- Extend for other Targets

### Support:

- More NVMe-oF Transports (e.g., FC)
- Asymmetric Namespace Access Groups
- Namespace Subtypes/Partitions
- Transport Required (TREQ)
- Multipath



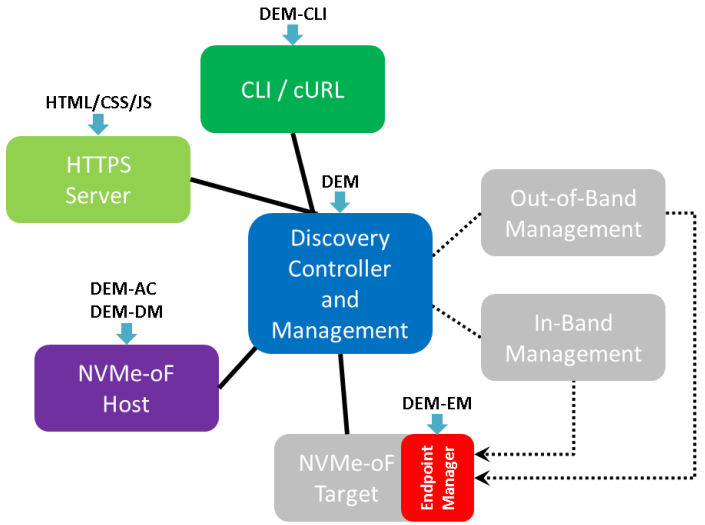
# DISTRIBUTED ENDPOINT MANAGEMENT

## Wrap-up

**Currently Available:**

- As a Dual GPL / BSD licensed project
- On GitHub (https://github.com/linux-nvme/nvme-dem)
- Wiki (https://github.com/linux-nvme/nvme-dem/wiki)

**Clone**
**Use**
**Contribute**




15th ANNUAL WORKSHOP 2019

# THANK YOU

Phil Cayton

Intel Corporation