Sub Clustering Modes with 3rd Generation Intel Xeon Scalable Processors

Explains sub clustering mode operations

Compares performance between different sub clustering modes

Provides recommendations for sub clustering mode setting

Includes instructions on setting sub clustering modes

Nathan Pham

Larry Cook

Kevin Huang

Abstract

Many sub clustering modes are available with the 3rd Generation Intel Xeon Scalable Processors. Understanding the operation of each sub clustering mode and its performance implication is important to maximizing the overall system performance.

This paper defines the under the hood operations of each sub clustering mode. It also provides performance comparison between these sub clustering modes on some of the most popular industry standard performance benchmarks. Sub clustering mode setting recommendation and instruction are also provided.

This paper is for customers and for business partners and sellers wishing to understand how to maximize the performance of Lenovo® ThinkSystem servers with 3rd Generation Intel Xeon Scalable Processors with sub clustering modes.

At Lenovo Press, we bring together experts to produce technical publications around topics of importance to you, providing information and best practices for using Lenovo products and solutions to solve IT challenges.

See a list of our most recent publications at the Lenovo Press web site:

http://lenovopress.com

> **Do you have the latest version?** We update our papers from time to time, so check whether you have the latest version of this document by clicking the **Check for Updates** button on the front page of the PDF. Pressing this button will take you to a web page that will tell you if you are reading the latest version of the document and give you a link to the latest if needed. While you're there, you can also sign up to get notified via email whenever we make an update.

Contents

Introduction

Intel introduced several memory subsystem enhancements with the 3rd Generation Xeon Scalable Processors, code named Ice Lake. Each Ice Lake processor has four integrated memory controllers (iMCs), and each iMC controls two double data rate (DDR) memory channels. Each DDR channel supports up to two DIMM slots and memory bus speeds operating at up to 3200 MHz.

Figure 1 shows that the processor package contains multiple Ice Lake processor cores (shown as "ICX cores" in the figure). Each core has a piece of the Last Level Cache (shown as "LLC slice"). Each processor socket has four iMCs, each of which supports two DDR memory channels.
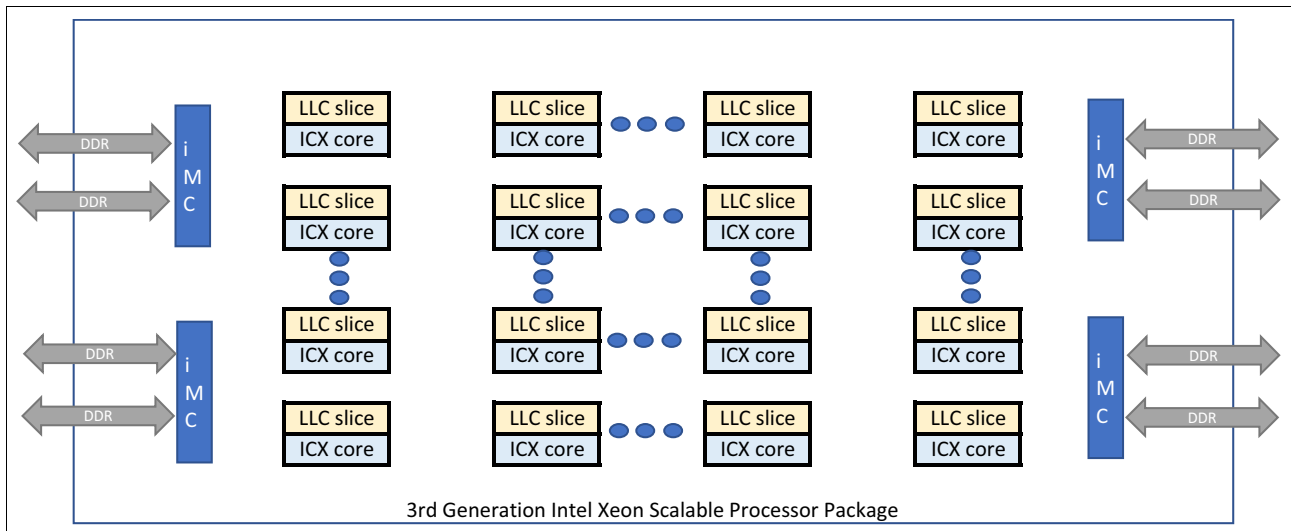

Figure 1   3rd Generation Intel Xeon Scalable Processor

Two sub-clustering modes are available with Ice Lake:

► Sub NUMA Cluster (SNC) mode
► Hemisphere (HEMI) mode, also known as UMA-Based Clustering (UBC) mode

In this paper, we will discuss these two modes, their memory configuration requirements, and their performance implications.

> **Tip:** Intel refers to Hemisphere mode as *HEMI mode* in architectural references, and as *UMA-Based Clustering (UBC) mode* in UEFI references.

Sub NUMA Cluster (SNC) mode

When SNC is enabled, the 3rd Gen Intel Xeon Scalable Processor socket is divided into two disjoint clusters, each having its own set of cores and LLCs, and memory controllers, as shown in Figure 2 on page 4.

This division results in two NUMA domains within a physical processor socket. Within each NUMA domain, cores, LLCs, and memory controllers are grouped based on their proximity to each other. Due to the improved proximity between the cores and the memory controllers

within a NUMA domain when SNC is enabled, memory latency and bandwidth performance are improved for local memory accesses.

Figure 2 shows an Ice Lake processor with SNC enabled. The dotted line denotes the division of the processor socket into two NUMA domain, each with its own set of cores, LLC slices, memory controllers, and DDR channels. This division is between the left and right sides of the processor socket.
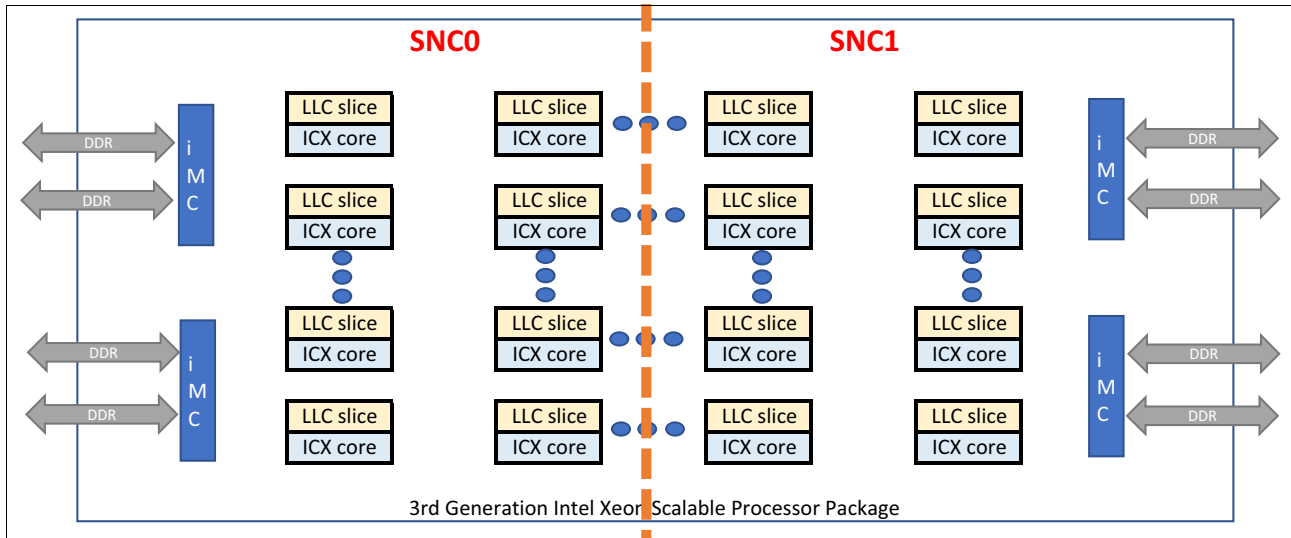

Figure 2   SNC sub clustering mode with 3rd Generation Intel Xeon Scalable Processor

SNC works well with highly NUMA optimized applications where software threads running on cores tend to access memory addresses in their local NUMA domain most of the time. There is also minimal or no sharing between caches in different NUMA domains.

SNC mode is only available under a very strict memory population condition. SNC can only be set when memory population in the processor socket is fully symmetric, i.e. the memory population is symmetric between top half and bottom half of the processor and between left half and right half of the processor. In other words, the memory populations of all memory controllers are identical to each other.

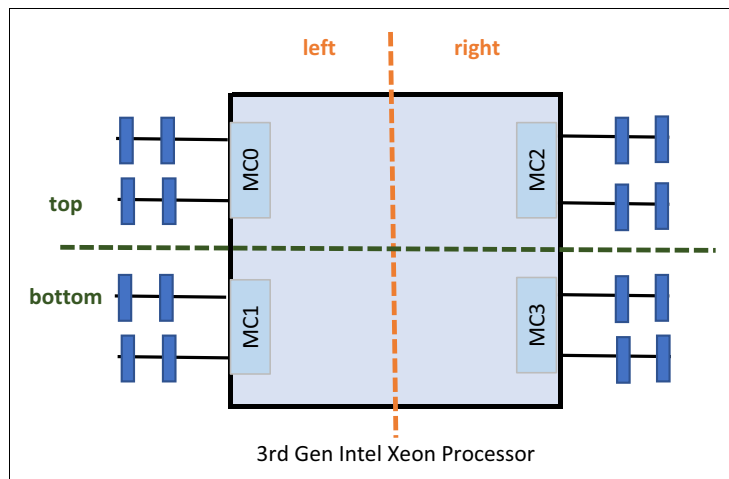Figure 3 shows an example of a fully symmetric memory population.

Figure 3   Example of a fully symmetric memory configuration

Hemisphere (HEMI) mode

When Hemisphere mode is enabled, there is an affinity between memory addresses and Last Level Cache (LLC) slices. Each memory address is mapped to one LLC slice and one of the two memory controllers on the same side of the core-cache die as the LLC slice. This means the data location of each memory address in the LLC when it is cached is predetermined. This affinity minimizes the distance between the memory location of a memory address and the LLC slice where it is cached.

Figure 4 shows an Ice Lake processor socket with its memory address space. Each memory address is assigned to an LLC slice with the associated iMC on the same side of the processor and is close to it. The result is a short distance between the LLC slice and iMC for each address line.
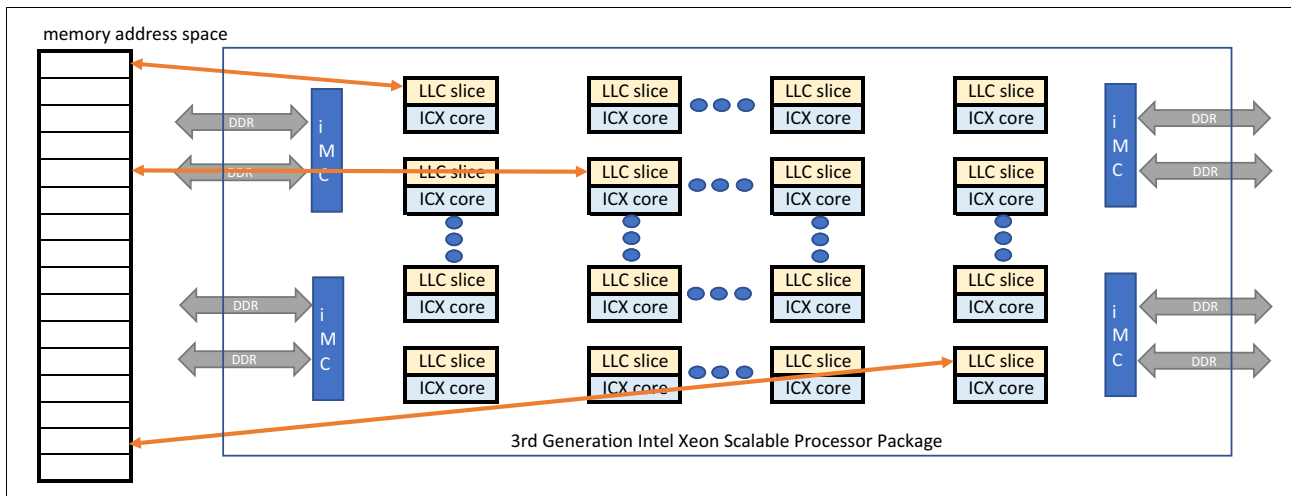

Figure 4   HEMI sub clustering mode with 3rd Generation Intel Xeon Scalable Processor

Hemisphere mode can only be enabled when the memory population is fully symmetric, or when there are only two DIMMs populated, the DIMMs must have left/right symmetry. Hemisphere mode is enabled by default if the memory configuration meets the above requirements and if SNC is disabled.

Hemisphere mode is also referred to as UMA-Based Clustering (UBC) in UEFI setting. As the name implies, UBC, or Hemisphere mode, is the suggested clustering mode when the processor is configured as Uniform Memory Access (UMA) node, i.e. SNC is disabled.

Figure 5 summarizes the SNC and HEMI support for each of the supported memory configurations with 3rd Generation Intel Xeon Scalable Processor.

DIMMs per socket	iMC0				iMC1				iMC2				iMC3				Sub Clustering Support	
	CH0 (A)		CH1 (B)		CH0 (C)		CH1 (D)		CH0 (E)		CH1 (F)		CH0 (G)		CH1 (H)			
	DIMM0	DIMM1	DIMM0	DIMM1	DIMM0	DIMM1	DIMM0	DIMM1	DIMM0	DIMM1	DIMM0	DIMM1	DIMM0	DIMM1	DIMM0	DIMM1	SNC	HEMI
1	DDR4																NO	NO
2	DDR4				DDR4												NO	NO
2	DDR4								DDR4								NO	YES
2	DDR4						DDR4										NO	NO
4	DDR4				DDR4				DDR4				DDR4				YES	YES
4	DDR4						DDR4		DDR4						DDR4		NO	NO
6	DDR4		DDR4		DDR4				DDR4		DDR4		DDR4				NO	NO
8	DDR4		DDR4		DDR4		DDR4		DDR4		DDR4		DDR4		DDR4		YES	YES
12	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4			DDR4	DDR4	DDR4	DDR4	DDR4	DDR4			NO	NO
12	DDR4	DDR4	DDR4		DDR4	DDR4	DDR4		DDR4	DDR4	DDR4		DDR4	DDR4	DDR4		NO	NO
16	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	DDR4	YES	YES

Figure 5   SNC and HEMI clustering mode support with 3rd Generation Intel Xeon Scalable Processor (The color of the DIMMs indicates the same capacity)

Performance implications

An experiment was done to compare performance between the following sub clustering modes:

1. SNC disabled + HEMI disabled

2. SNC disabled + HEMI enabled

3. SNC enabled + HEMI disabled

Note that SNC and HEMI are different sub clustering modes, and they are mutually exclusive. This means only either SNC or HEMI can be enabled at a time, provided that the memory population meets the requirements as described earlier.

In this experiment, each DDR channel was populated with one 32GB 2Rx8 RDIMM. This memory configuration meets the fully symmetry requirement, and the general memory subsystem performance was optimized. We compared performance between these sub clustering modes using well-known performance benchmarks:

► Linpack
► STREAM
► SPECrate2017_int_base (SIR17)
► SPECrate2017_fp_base (SFR17)

Table 1 and Figure 6 summarize the results.

Table 1   Performance comparison between sub clustering modes

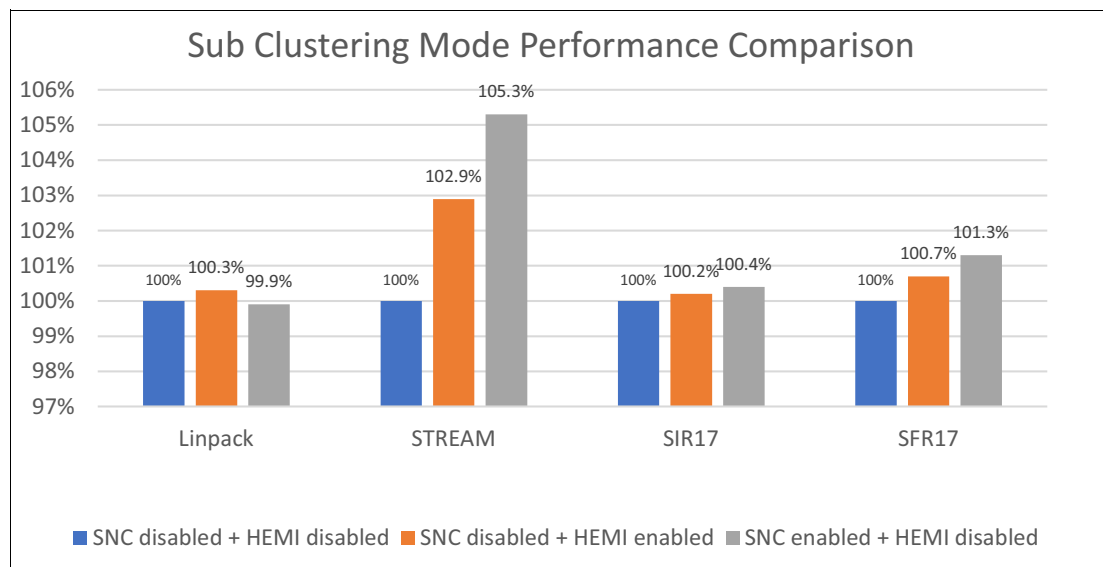|	SNC disabled + HEMI disabled	SNC disabled + HEMI enabled	SNC enabled + HEMI disabled
Linpack	100%	100.3%	99.9%
STREAM	100%	102.9%	105.3%
SPECrate2017_int_base	100%	100.2%	100.4%
SPECrate2017_fp_base	100%	100.7%	101.3%

Figure 6   Sub Clustering Mode performance comparison

These performance benchmarks are NUMA optimized and the results show performance improvement when a sub clustering mode is enabled. With these benchmarks, performance results are best with SNC enabled, followed by HEMI enabled, and lowest with no sub clustering. The one exception is Linpack where the result with SNC is on par with no sub clustering.

In general, NUMA applications should benefit from sub clustering modes. Between the two sub clustering modes, SNC tends to result in better performance improvement. However, due to the differences in the way the two sub clustering modes operate, users should evaluate both modes with their NUMA optimized applications to determine the optimal mode for their applications.

Setting up SNC and HEMI on Lenovo UEFI

The following sections provide instructions on how to set SNC and HEMI modes using either UEFI setup user interface or OneCLI command-line tool from within the OS.

Note: HEMI is also referred to as *UMA-Based Clustering* in UEFI context. Intel refers to this setting as HEMI in their architecture references, and UMA-Based Clustering in UEFI setting references.

How-to: Configuring SNC (Sub-NUMA Clustering) setting

There are two methods to configure SNC, using the UEFI menus accessible via F1 System Setup and using OneCLI locally after the server has booted.

Method 1: Change SNC setting under the UEFI setup UI (F1)
1. Press F1 when prompted during boot to enter the System Setup.

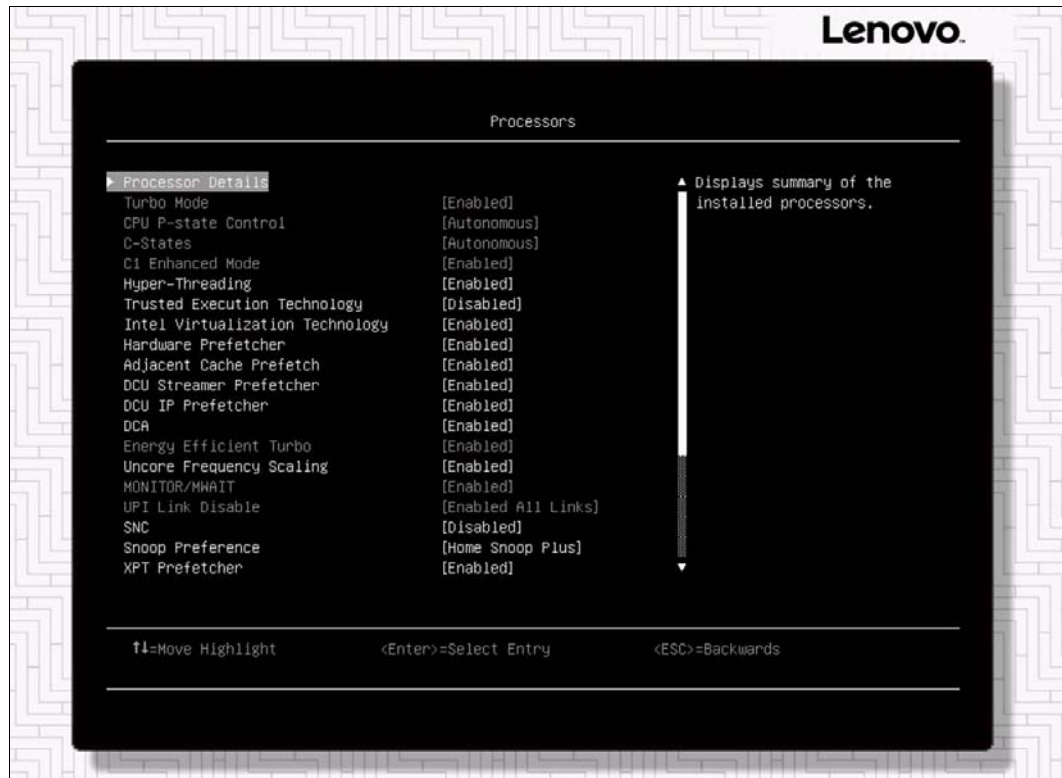2. Navigate to **System Settings** → **Processors** as shown in Figure 7.

Figure 7 Processor settings

3. Locate the SNC setting whereto enable or disable SNC as shown in Figure 8. The default is disabled.
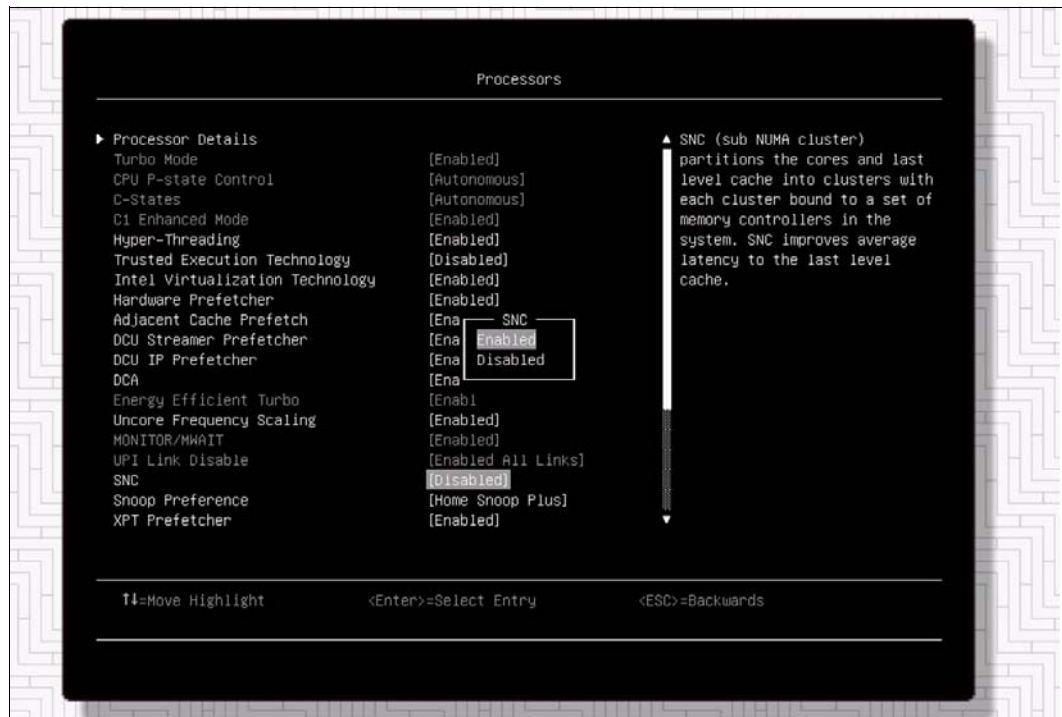

Figure 8 Enabling SNC

4. Return to the root menu and select **Save Settings** to save the changes.

5. Reboot the server to put the change into effect.

Method 2: Change SNC setting via OneCLI tool under the OS

In these steps, we are using RHEL8 as the example.

1. Boot to the OS and have an OneCLI tool package ready (make sure the OneCLI version supports the platform you're running on), and decompress the package to a newly created folder, as shown in Figure 9

```
[root@DEMO RHEL8]# ls
lnvgy_utl_lxce_onecli01o-3.1.0_rhel_x86-64.tgz
[root@DEMO RHEL8]# mkdir onecli_v3.1.0
[root@DEMO RHEL8]# ls
lnvgy_utl_lxce_onecli01o-3.1.0_rhel_x86-64.tgz  onecli_v3.1.0
[root@DEMO RHEL8]# tar -C onecli_v3.1.0/ -xvf lnvgy_utl_lxce_onecli01o-3.1.0_rhel_x86-64.tgz
```

Figure 9   Unpack the OneCLI package

2. Navigate to the folder directory, locate the `onecli` binary you'll be using, Figure 10

```
[root@DEMO RHEL8]# cd onecli_v3.1.0/
[root@DEMO onecli_v3.1.0]# ls
Lic     Sample  data   global.config  include  onecli   rdumount  utilities
OneCli  asu     dsa    help           libs     rdmount  ts_tools  uxspi
[root@DEMO onecli_v3.1.0]#
```

Figure 10   Directory listing

3. Run the binary with the following command to say enable SNC setting:

    # ./onecli config set Processors.SNC Enabled

The output is shown in Figure 11.

```
[root@DEMO onecli_v3.1.0]# ls
asu    dsa            help    libs  logs   OneCli   rdumount  ts_tools  uxspi
data  global.config  include  Lic   onecli  rdmount  Sample    utilities
[root@DEMO onecli_v3.1.0]# ./onecli config set Processors.SNC Enabled

Lenovo XClarity Essentials OneCLI lxce_onecli01o-3.1.0
(C) Lenovo 2013-2020 All Rights Reserved

OneCLI License Agreement and OneCLI Legal Information can be found at the following location:
    "/root/RHEL8/onecli_v3.1.0/Lic"

Invoking SET command ...
Connected to the node(1) by KCS
Processors.SNC=Enabled
Configure successfully, please reboot system.
Succeed.
```

Figure 11   OneCLI command

4. If something goes wrong, make sure the OneCLI variable for SNC setting matches the command, in this case `Processors.SNC`, and make sure to use the correct setting value. You may check the setting values using the following command:

    # ./onecli config showvalues Processors.SNC

The output is shown in Figure 12.


```
[root@DEMO onecli_v3.1.0]# ./onecli config showvalues Processors.SNC

Lenovo XClarity Essentials OneCLI lxce_onecli01o-3.1.0
(C) Lenovo 2013-2020 All Rights Reserved

OneCLI License Agreement and OneCLI Legal Information can be found at the following location:
    "/root/RHEL8/onecli_v3.1.0/Lic"

Invoking SHOWVALUES command ...
Connected to the node(1) by KCS
Processors.SNC=Enabled=<Disabled>

Succeed.
[root@DEMO onecli_v3.1.0]#
```

Figure 12   showvalues command

5. Reboot the system to put the change into effect.

How-to: Configuring UBC (UMA-Based Clustering) setting

There are two methods to configure UMA-Based Clustering (HEMI mode) using the UEFI menus accessible via F1 System Setup and using OneCLI locally after the server has booted.

Method 1: Change UBC setting under the UEFI setup UI (F1)

1. Press F1 during system boot to enter System Setup.

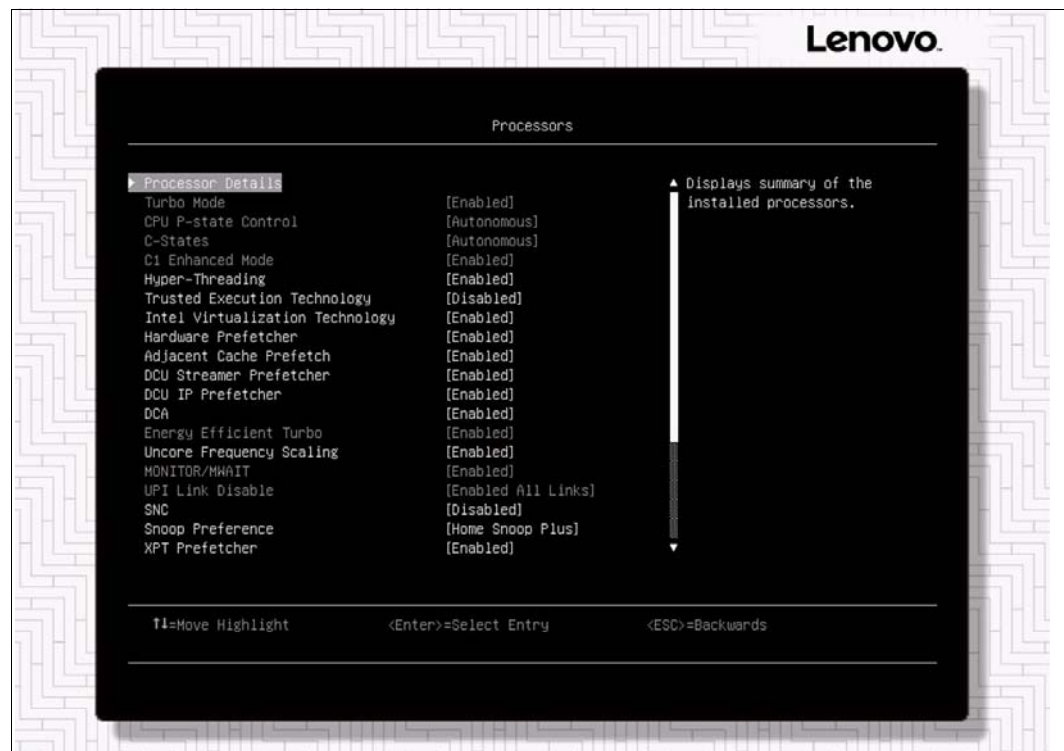2. Navigate to **System Settings** → **Processors** as shown in Figure 13.

Figure 13   Processor settings

3. Locate the UMA-Based Clustering setting and press Enter to set it to **Hemisphere** or **Disabled** (default Hemisphere), as shown in Figure 14.
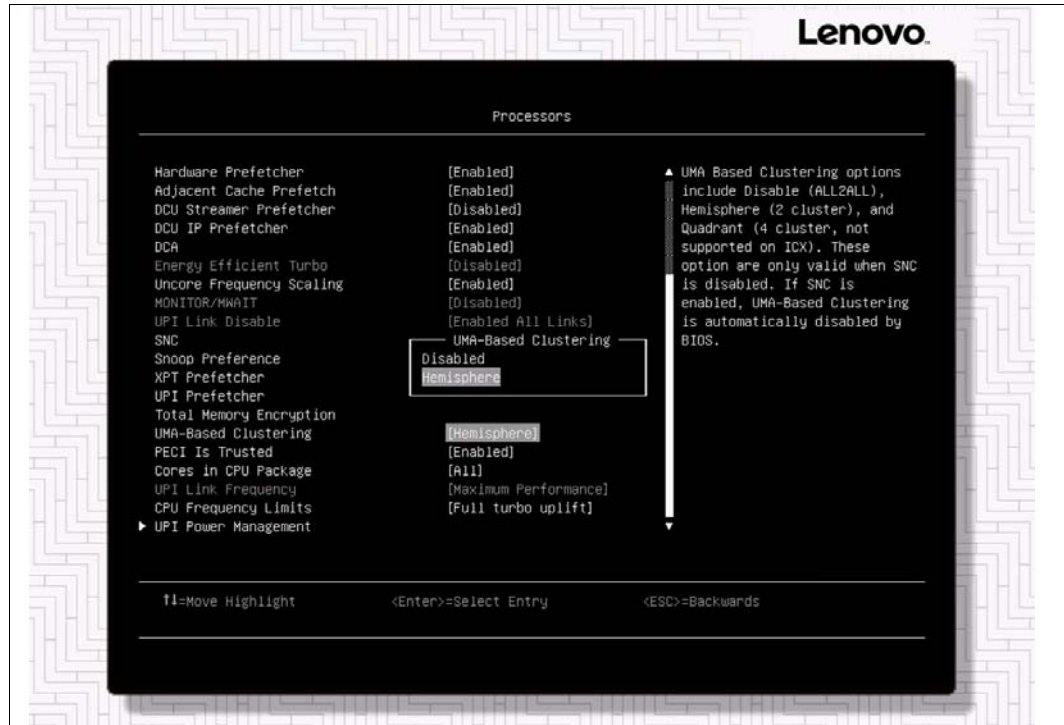

Figure 14   UMA-Based Clustering setting

Note: When SNC is enabled, UBC is greyed-out and fixed as **Disabled**.

4. Return to the root menu and select **Save Settings** to save the changes.

5. Reboot the server to put the change into effect.

Method 2: Change UBC setting via OneCLI tool under the OS

In these steps, we are using RHEL8 as the example.

1. Boot to the OS and have an OneCLI tool package ready (make sure the OneCLI version supports the platform you're running on), and decompress the package to a newly created folder, as shown in Figure 9

```
[root@DEMO RHEL8]# ls
lnvgy_utl_lxce_onecli01o-3.1.0_rhel_x86-64.tgz
[root@DEMO RHEL8]# mkdir onecli_v3.1.0
[root@DEMO RHEL8]# ls
lnvgy_utl_lxce_onecli01o-3.1.0_rhel_x86-64.tgz   onecli_v3.1.0
[root@DEMO RHEL8]# tar -C onecli_v3.1.0/ -xvf lnvgy_utl_lxce_onecli01o-3.1.0_rhel_x86-64.tgz
```

Figure 15   Unpack the OneCLI package

2. Navigate to the folder directory, locate the `onecli` binary you'll be using, Figure 10

Figure 16   Directory listing

3. Run the binary with the following command to change UBC setting to "Hemisphere":

   `# ./onecli config set Processors.UMA-BasedClustering Hemisphere`

   The output is shown in Figure 17.

Figure 17   OneCLI command

4. If something goes wrong and an error is returned, make sure the OneCLI variable for the SNC setting matches the command, in this case `Processors.UMA-BasedClustering`, and make sure to use the correct setting value. You may check the setting values using the following command:

   `# ./onecli config showvalues Processors.UMA-BasedClustering`

   The output is shown in Figure 18.

Figure 18   showvalues command

Just like UBC setting is greyed-out and fixed as "Disabled" when SNC is enabled in the System Setup interface, you will not be allowed to change UBC setting via OneCLI tool when SNC is enabled, as shown in Figure 19.

Figure 19   Setting is read only

5. Reboot the system to put the change into effect.

Authors

This paper was produced by the following team of specialists:

Nathan Pham is a Senior Engineer in the Lenovo Infrastructure Solutions Group Performance Laboratory in Morrisville, NC. He has spent over 19 years between IBM and Lenovo working on system performance. His focus has been in system performance architecture, modeling, analysis, and optimization. In his current role, he is responsible for all aspects of future Lenovo DCG system architecture and design for optimized performance. Nathan holds a Bachelor of Science degree in Computer Engineering and a Master of Science degree in Computer Engineering, both from North Carolina State University.

Larry Cook is an x86 Performance Engineer in the Lenovo Infrastructure Solutions Group Performance Laboratory located in Morrisville, NC. He has 13 years of experience between IBM and Lenovo with enterprise servers. For 8 of those years, his focus has been centered on system performance analysis, development, optimization, and benchmarking. In his current role within the System Performance Verification (SPV) group, he focuses on CPU and memory subsystem readiness on 2-socket servers in preparation for competitive benchmarking. Larry holds a Bachelor of Science degree in Electrical Engineering with a Concentration in Computers from Tennessee State University.

Kevin Huang is an x86 Enterprise Server Performance Engineer in the Lenovo Infrastructure Solutions Group Performance Lab in the Taipei Design Center. In his current role, he mainly focuses on the CPU and memory subsystem and is responsible for ensuring the subsystem performance to be reliable and competitive before product announcements, for every new generation of x86 server products developed by Lenovo. Prior to working on server performance, Kevin has also worked on automation, system back-end development, and software architecture.

Thanks to the following people for their contributions to this project:

► David Watts, Lenovo Press

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service.

Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
1009 Think Place - Building One
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

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

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document was created or updated on August 3, 2021.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/lp1499

Trademarks

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

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

Lenovo(logo)®   Lenovo®

The following terms are trademarks of other companies:

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

Other company, product, or service names may be trademarks or service marks of others.