

Гайд по оперативной памяти

для 1/2/4U серверных платформ с двумя процессорами AMD EPYC 7003

ITPOD-SR101-D04R
ITPOD-SR101-D10R

ITPOD-SR201-D08R
ITPOD-SR201-D12RE
ITPOD-SR201-D12R-NV
ITPOD-SR201-D25RE

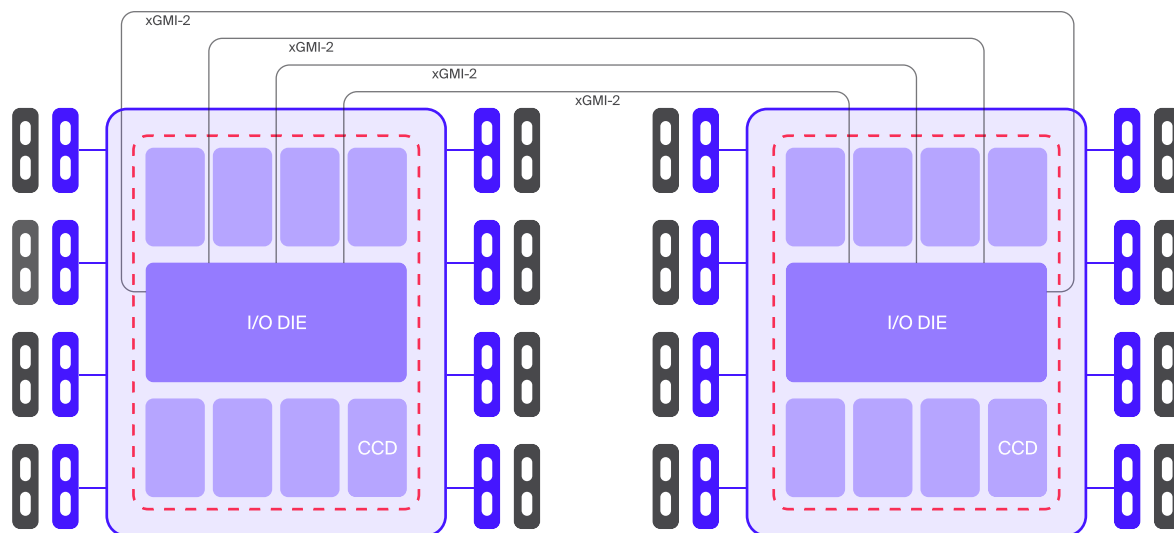
ITPOD-SR401-D24RE
ITPOD-SR401-D36RE

О процессорах AMD EPYC 7003

Процессоры AMD EPYC 7003 серии созданы на базе ядер Zen 3 и архитектуры AMD Infinity. Максимальное число ядер – 64 на процессор. Ядра могут выполнять два потока одновременно, что для 64 ядер дает 128 потоков, благодаря технологии AMD Simultaneous Multithreading (SMT).

Каждый процессор EPYC 3-го поколения состоит из:

- Core Complex Die (CCD) – от 2 до 8 шт, в зависимости от выбранного процессора.
- I/O Die (IOD) – 1 шт.



Каждый CCD содержит один Core Complex (CCX), в отличие от предыдущих поколений, поэтому каждый CCD может иметь до 8 ядер Zen 3, которые используют до 32 Мбайт общего L3 кэша и до 96Мбайт для CPU с поддержкой 3 DV-Cache.

Используя AMD Infinity Fabric, CCD подключаются к центральному чипу ввода-вывода (IOD) для соединения всех CCD и организации доступа к каналам оперативной памяти и PCIe линиям.

Поддерживается до 8 каналов памяти на процессор и 128 линий PCIe 4.0.

В двухсокетных серверах два процессора AMD EPYC 7003 серии соединяются между собой через соответствующие каналы xGMI. Также стоит учитывать, что в случае двух сокетов доступны 162 линии PCI Express.

Топология Non-Uniform Memory Access (NUMA)

В процессорах AMD EPYC 7003 серии используется архитектура памяти с неравномерным доступом к памяти (NUMA). I/O Die, который описан в предыдущем разделе, может быть разделен на 4 квадранта, что позволяют разделить процессор на различные NUMA домены.

Используя настройки BIOS, можно оптимизировать топологию NUMA для определенных ОС и рабочих нагрузок с помощью настройки параметра «NUMA Nodes per Socket (NPS)». Каждый сервер может быть настроен как NPS = 1 (NPS1), NPS = 2 (NPS2) или NPS = 4 (NPS4), с дополнительной опцией настройки кэша третьего уровня (L3CAN).

В зависимости от количества CCD, из которых состоит процессор, или определенных конфигураций RAM, могут быть недоступны варианты NPS2 и NPS4.

Параметр	Описание
NUMA Nodes per Socket	Auto (NPS1). Рекомендуется для большинства нагрузок.
	NPS1: один NUMA домен на сокет (если не включена опция «L3 Cache as NUMA»); Чередование обращений к памяти по всем восьми каналам в каждом сокете.
	NPS2: два NUMA домена на сокет (если не включена опция «L3 Cache as NUMA»); Чередование доступа к памяти по группам из четырех каналов (ABCD и EFGH) в каждом сокете. При использовании 200-гигабитных сетевых адаптеров может быть предпочтительнее использовать NPS2, чтобы обеспечить компромисс между задержкой памяти и пропускной способностью памяти.
	NPS4: Четыре NUMA домена на сокет (если не включена опция «L3 Cache as NUMA»); Чередование обращений к памяти по парам из двух каналов (AB, CD, EF и GH) в каждом сокете. Рекомендуется для высокопроизводительных вычислений (HPC) и других высокопараллельных рабочих нагрузок.

ACPI SRAT L3 Cache as NUMA domain

Если параметр «ACPI SRAT L3 Cache as NUMA Domain» в статусе «Enabled», то он определяет гранулярность чередования памяти (Memory interleaving). В таком случае каждый L3 кэш (по одному на каждый CCD) отображается как отдельный NUMA домен.

Например, один процессор с 8 CCD будет иметь 8 NUMA доменов: по одному на каждый CCD. Двухпроцессорная система будет иметь в общей сложности 16 NUMA доменов.

Этот параметр может повысить производительность рабочих нагрузок с высокой степенью оптимизации для работы с NUMA, если рабочие нагрузки или компоненты рабочих нагрузок можно закрепить за ядрами в CCX и, если они могут получить выгоду от совместного использования L3 кэш. Некоторые операционные системы и гипервизоры не поддерживают такой режим работы.

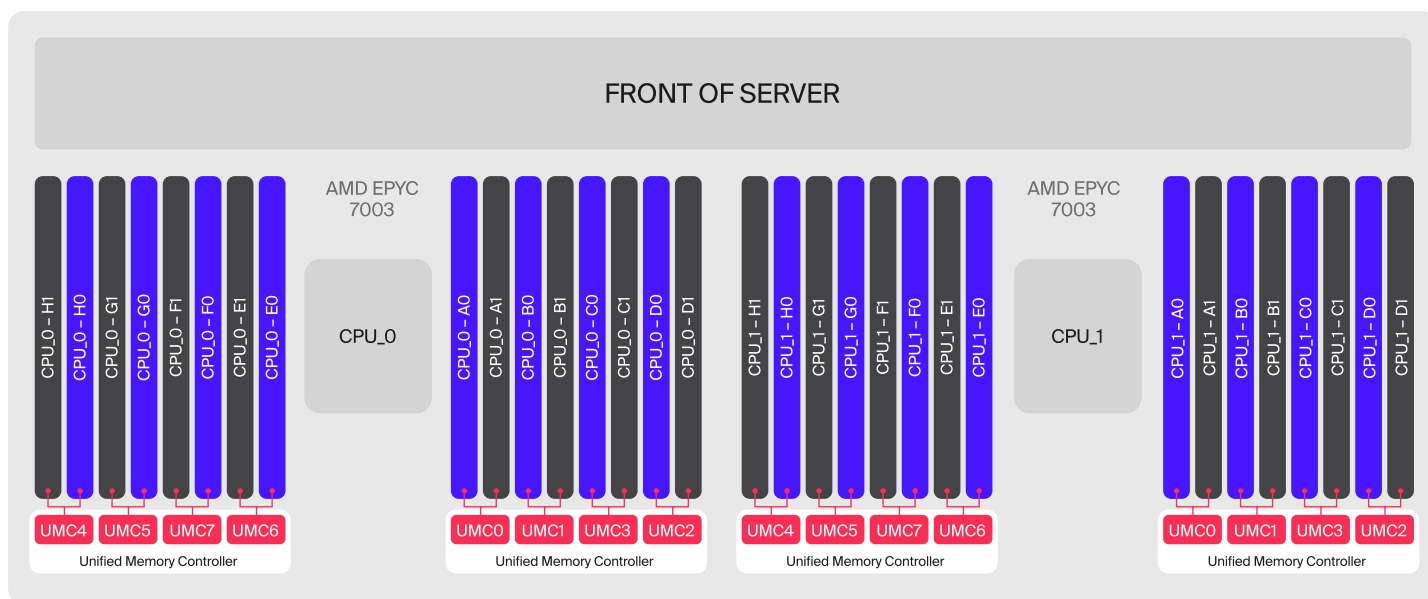
Если этот параметр отключен, домены NUMA идентифицируются в соответствии с настройкой параметра NPS.

Параметр	Описание
ACPI SRAT L3 Cache As NUMA Domain	Auto (Disabled)
	Disable: не сообщать ОС о каждом L3 кэше как о NUMA домене.
	Enable: сообщать ОС о каждом L3 кэше как о NUMA домене.

Организация памяти

Процессоры серии AMD EPYC 7003 характеризуются следующими параметрами, относящимися к RAM:

- Тактовая частота - 3200 МГц.
- Модули памяти Registered ECC DDR4 DIMM (RDIMM) и load-reduced DIMM (LRDIMM).
- 8 каналов памяти на CPU.
- 2 слота памяти на один канал памяти (максимум 16 слотов на один CPU или 32 слота в двухпроцессорной конфигурации сервера).
- Поддержка технологии Memory Interleaving для оптимизации работы только с шестью каналами памяти из доступных восьми.



Правила по использованию модулей памяти

- Сначала заполняются слоты 0 (синего цвета), потом слоты 1 (черного цвета).
- Не смешивайте модули RDIMM и LRDIMM.
- Модели 128GB non-3DS LRDIMM нельзя смешивать с модулями 256GB 3DS LRDIMM.
- Модули памяти в одном канале должны иметь одинаковый параметр DRAM width. Не допускается смешивание модулей памяти x4 и x8 в одном канале.
- Модули памяти в одном канале должны иметь одинаковый параметр DRAM Density. Запрещено смешивать модули памяти с DRAM 8Gb и DRAM 16Gb в одном канале.
- Модули памяти 256GB нельзя устанавливать вместе с GPU адаптерами, при этом окружающая температура не должна превышать максимум 28°C.
- Модули памяти для обоих процессоров всегда должны быть настроены зеркально.
- Модули памяти Non-buffered UDIMM, non-ECC DIMM и 4R RDIMM не поддерживаются.
- Память от серверов предыдущего поколения (DDR3) не совместима и не поддерживается серверами с процессорами AMD EPYC 7003.

Memory Interleaving

С процессорами EPYC 7003 AMD представила чередование памяти, что позволяет использовать только шесть каналов из доступных восьми. Функция чередования памяти решает проблемы с несбалансированными 6-канальными конфигурациями, позволяя получить пропускную способность до 75% от максимальной. Это позволяет более рационально расходовать средства на закупку памяти.

Конечно, и с предыдущими поколениями EPYC можно было использовать меньше восьми каналов памяти. Но максимальная пропускная способность достигалась только с восемью занятыми каналами. Четыре канала давали порядка 50% пропускной способности. Однако возникали проблемы при использовании шести каналов памяти из восьми. Половина контроллеров памяти работала с одним каналом, вторая половина - с двумя каналами. В итоге нагрузка получалась асимметричной. И вместо 75% от максимальной пропускной способности, на практике получалось значение между 40% и 60%, так как программному обеспечению, считывающим последовательные блоки памяти, необходимо дожидаться завершения операции передачи памяти, прежде чем начать следующий доступ к памяти. При включенном чередовании памяти последовательные блоки памяти находятся в разных «банках» (banks), поэтому все они могут вносить вклад в общую пропускную способность памяти.

В любом случае, AMD рекомендует, чтобы все восемь каналов памяти на процессор были заполнены, а каждый канал имел одинаковую емкость. Такой подход позволяет подсистеме памяти работать в режиме восьмиканального чередования, что в большинстве случаев обеспечивает наилучшую производительность.

Важное примечание относительно максимальной пропускной способности памяти

Максимально возможная пропускная способность памяти особенно актуальна в среде высокопроизводительных вычислений (HPC). В других областях применения влияние пропускной способности памяти на общую производительность меньше и зависит от соответствующего приложения. Например, тесты SPECint_rate_base2006 показывают, что даже при пропускной способности памяти 35% тест SPEC достигает производительности до 90%.

Конфигурации оперативной памяти с меньшим количеством модулей уменьшают максимально достижимую пропускную способность памяти, но также требуют меньше энергии и обеспечивают хорошую возможность расширения в будущем.

Порядок размещения модулей памяти

Кол-во каналов памяти	Канал памяти								Memory Interleaving для выбранного значения NPS							
	М = общая емкость модулей памяти на указанном канале; Может быть как один так два модуля памяти.															
	A	B	C	D	E	F	G	H	NPS=1		NPS=2		NPS=4			
1*	—	—	M1	—	—	—	—	—	C	C			C			
2*	—	—	M1	M1	—	—	—	—	CD	CD			CD			
3*	—	—	M1	M1	—	—	M1	—	CD, G	CD	G		CD		G	
4**	—	—	M1	M1	—	—	M1	M1	CDGH	CDGH			CD		GH	
4**	M1	M1	M1	M1	—	—	—	v	ABCD	ABCD		AB	CD			
4**	—	—	—	—	M1	M1	M1	M1	EFGH		EFGH			EF	GH	
5*	M1	—	M1	M1	—	—	M1	M1	A, CD, GH	A, CD	GH	A	CD		GH	
6***	M1	—	M1	M1	M1	—	M1	M1	ACDEGH	A, CD	E, GH	A	CD	E	GH	
7*	M1	M1	M1	M1	M1	—	M1	M1	AB, CD, E, GH	AB, CD	E, GH	AB	CD	E	GH	
8****	M1	M1	M1	M1	M1	M1	M1	M1	ABCDEFGH	ABCD	EFGH	AB	CD	EF	GH	

Первый столбец – указано количество каналов памяти, заполненных модулями памяти. Он не указывает на количество модулей памяти в канале памяти. В каждом канале может быть один (1DPC) или два модуля памяти (2DPC).

Например, конфигурация из 12 модулей DIMM, в которой заполнены все каналы памяти, шесть из которых имеют по одному модулю DIMM на канал, а два - по два модуля DIMM на канал, попадает в строку с восемью заполненными каналами.

Учитывайте данные по итоговой скорости работы памяти при выборе варианта заполнения сервера памятью.

Поддерживаемая CPU частота	RDIMM (16GB, 32GB, 64GB)		LRDIMM (128GB)		3DS LRDIMM (256GB)	
	1DPC	2DPC	1DPC	2DPC	1DPC	2DPC
DDR4-3200MHz	3200	2933	3200	2933	3200	2933

Обозначение **M1** указывает на общий объем памяти на конкретном канале и на то, что эти каналы могут чередоваться в соответствии с правилами, определенными в данном документе. Эта емкость может приходиться на один модуль памяти или два модуля памяти на канале.

Например, канал А с одним модулем памяти имеет емкость **M1**. Если в канале В установлено два модуля памяти, каждый из которых имеет емкость, равную половине емкости модуля памяти в канале А, канал В также будет обозначен как **M1**.

Комментарии

* Это пример заполнения для заданного количества заполненных каналов. Он предназначен для того, чтобы показать, как на чередование (Memory Interleaving) может повлиять заполнение каналов. Из-за неравномерного заполнения каналов памяти производительность ограничена двухсторонним чередованием для всех конфигураций.

Например, в строке с кол-вом заполненных каналов = 1 показан заполненным только канал С, но вместо него может быть заполнен любой другой канал с аналогичной производительностью чередования. Для достижения оптимальной производительности рекомендуется по возможности заполнять пары каналов одинаковыми модулями памяти.

** Если решено заполнить только четыре канала, AMD рекомендует ограничить выбор процессоров до вариантов с объемом L3 памяти 128 МБ или менее и заполнить каналы CDGH одинаково. Это позволит обеспечить четырехстороннее чередование. Все заполненные каналы памяти должны иметь одинаковую емкость.

*** Если решено заполнить только шесть каналов, AMD рекомендует заполнить каналы ACDEGH с одинаковой емкостью, где все каналы не превышают объем в 256 ГБ каждый. Это позволит обеспечить шестистороннее чередование. Обратите внимание, что для шестистороннего чередования параметр Memory Interleaving Size должен быть установлен в 2 или 4 KB при NPS=1. Все заполненные каналы памяти должны иметь одинаковую емкость.

**** Восьмистороннее чередование возможно в заполнении каналов ABCDEFGH, когда они имеют одинаковую емкость. Это рекомендуемый вариант, обеспечивающий наилучшую производительность во всех случаях.