

# Примеры вычислительных систем с архитектурой cc-NUMA

# Системы с неоднородным доступом к памяти (NUMA)

Архитектура	<p>Система состоит из однородных базовых модулей (плат), состоящих из небольшого числа процессоров и блока(ов) памяти. Модули объединены с помощью высокоскоростного коммутатора. Поддерживается единое адресное пространство, аппаратно поддерживается доступ к удаленной памяти, т.е. к памяти других модулей. При этом доступ к локальной памяти в несколько раз быстрее, чем к удаленной.</p> <p>В случае, если аппаратно поддерживается когерентность кэшей во всей системе (обычно это так), говорят об архитектуре <b>cc-NUMA</b> (cache-coherent NUMA)</p>
Примеры	HP <a href="#">HP 9000 V-class</a> в SCA-конфигурациях, SGI <a href="#">Origin2000</a> , Sun <a href="#">HPC 10000</a> , IBM/Sequent <a href="#">NUMA-Q 2000</a> , SNI <a href="#">RM600</a> .
Масштабируемость	Масштабируемость NUMA-систем ограничивается объемом адресного пространства, возможностями аппаратуры поддержки когерентности кэшей и возможностями операционной системы по управлению большим числом процессоров.
Операционная система	Обычно вся система работает под управлением единой ОС, как в <a href="#">SMP</a> . Но возможны также варианты динамического "подразделения" системы, когда отдельные "разделы" системы работают под управлением разных ОС (например, Windows NT и UNIX в NUMA-Q 2000).
Модель программирования	Аналогично <a href="#">SMP</a> .

# HP Superdome

Основу архитектуры компьютера HP Superdome составляют вычислительные ячейки (cells), связанные иерархической системой переключателей. Каждая ячейка является симметричным мультипроцессором, реализованным на одной плате, в котором есть все необходимые компоненты:

- \* процессоры (до 4-х)
- \* оперативная память (до 16 Гбайт)
- \* контроллер ячеек
- \* связь с подсистемой ввода/вывода.



# HP Superdome

Архитектура компьютера спроектирована таким образом, что в ней могут использоваться несколько типов микропроцессоров. Это традиционные для вычислительных систем Hewlett-Packard процессоры семейства PA: PA-8600 и PA-8700.

Вместе с тем, система полностью подготовлена и к использованию процессоров следующего поколения с архитектурой IA-64, разработанной совместно компаниями HP и Intel.

Ячейки являются главным строительным блоком Superdome. Аналогичные архитектурные элементы в ccNUMA-системах SGI Origin 3x00 и Compaq AlphaServer GS320 называются соответственно узлами (node) и QBB (Quad Building Block).

# HP Superdome

Основным элементом инфраструктуры этих блоков является коммутатор. В ячейке Superdome это – контроллер ячейки, в узле Origin 3x00 – концентратор (микросхема Bedrock), в QBB – локальный коммутатор (local switch).

Этот коммутатор связывает между собой процессоры, оперативную память, подсистему ввода-вывода (I/O), а также обеспечивает соединение блоков между собой (т.е. содержит порты к «межсоединению»).

# HP Superdome

Каждый процессор в ячейке имеет свой собственный выделенный порт в контроллере ячейки с пропускной способностью 1,6 Гбайт/с. Это предотвращает возможности конфликтов между процессорами за доступ к порту коммутатора, как это теоретически могло иметь место в серверах HP Класса V.

Такие конфликты могут иметь место в SGI Origin 3x00, где пары процессоров разделяют системную шину, а две эти системные шины имеют по одному порту на концентраторе.

В AlphaServer GS320, напротив, каждый процессор, как и в Superdome, имеет свой порт на локальном коммутаторе.

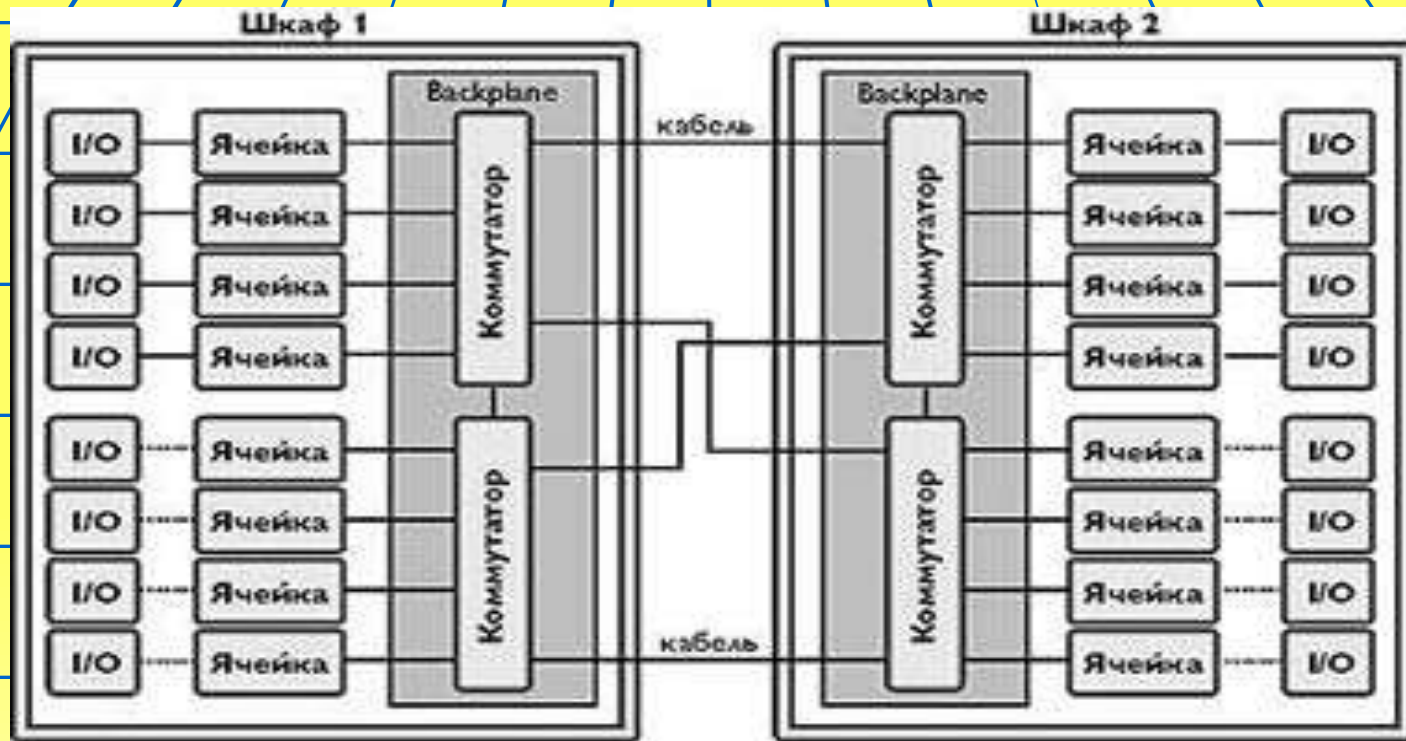
# HP Superdome

Ячейка Superdome представляет собой небольшую SMP-систему; правда, средства ввода-вывода в ней отсутствуют — они являются внешними по отношению к ячейке и подсоединяются через порт ввода-вывода контроллера ячейки с пропускной способностью 1,6 Гбайт/с. Ячейки Superdome конструктивно реализованы в виде платы.

Кроме процессора и контроллера ячейки, на этой плате располагается также оперативная память. Вся память разделяется между двумя банками, каждый из которых имеет по 1 порту в контроллере ячейки с пропускной способностью, равной 1,6 Гбайт/с.



# HP Superdome



Каждая стойка содержит по два восьмипортовых неблокирующих коммутатора. Все порты коммутаторов работают со скоростью 8 Гб/с. К каждому коммутатору подключаются четыре ячейки. Три порта коммутатора задействованы для связи с другими коммутаторами системы (один – в той же стойке и два коммутатора в другой). Оставшийся порт зарезервирован для связи с другим системами HP Superdome, что дает потенциальную возможность формирования произвольной конфигурации серверов с процессорами больше 64.



# HP Superdome

Одним из центральных вопросов любой вычислительной системы с архитектурой ccNUMA является разница во времени при обращении процессора к локальным и удаленным ячейкам памяти.

В идеале хотелось бы, чтобы этой разницы, как в SMP-компьютере, не было вовсе. Однако в таком случае система заведомо будет плохо масштабируемой.

# HP Superdome

В компьютере HP Superdome возможны три вида задержек при обращении процессора к памяти, являющихся своего рода платой за высокую масштабируемость системы в целом:

- \* процессор и память располагаются в одной ячейке, в этом случае задержка минимальна;
- \* процессор и память располагаются в разных ячейках, но обе эти ячейки подсоединены к одному и тому же коммутатору;
- \* процессор и память располагаются в разных ячейках, причем обе эти ячейки подсоединены к разным коммутаторам; в этом случае запрос должен пройти через два коммутатора и задержки будут максимальными.

# SGI Origin2000

Производитель	<u>Silicon Graphics</u>
Класс архитектуры	Модульная система с общей памятью ( <u>cc-NUMA</u> ).
Процессор	64-разрядные RISC-процессоры MIPS R10000, R12000/300MHz
Модуль	Основной компонент системы - модуль Origin, включающий от 2 до 8 процессоров MIPS R10000 и до 16GB оперативной памяти.
Масштабируемость	Поставляются системы Origin2000, содержащие до 256 процессоров (т.е. до 512 модулей). Вся память системы (до 256GB) глобально адресуема, аппаратно поддерживается когерентность кэшей.
Коммутатор	Модули системы соединены с помощью сети CrayLink, построенной на маршрутизаторах MetaRouter.
Системное ПО	Используется операционная система SGI IRIX.
Средства программирования	Поставляется распараллеливающий компилятор Cray Fortran 90. Поддерживается стандарт <u>CrayMPP</u> .

# SGI Origin2000



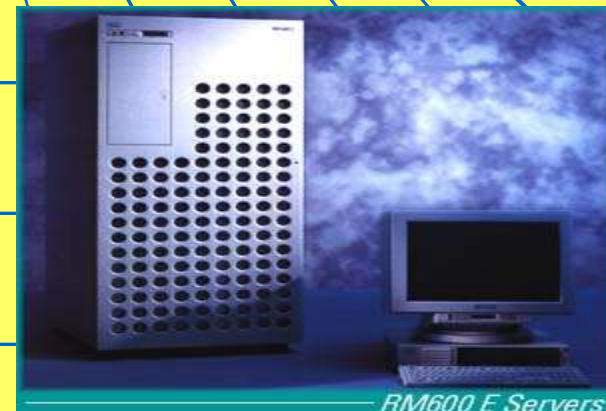
# SGI Altix3000

Производитель	<u>Silicon Graphics</u>
Класс архитектуры	Модульная система с общей памятью ( <u>cc-NUMA</u> ).
Процессор	Intel Itanium II 1.3GHz/1.5GHz
Модули	Вся система строится из модулей (вычислительных, коммутационных, проч.) Вычислительный компонент системы - модуль C-brick, состоящий из 2-х блоков, включающий 4 процессора (по 2 на блок), 4 слота памяти по 8DIMM (от 4 до 16Gb на C-brick).
Масштабируемость	Поставляются системы Origin2000, содержащие до 256 процессоров (т.е. до 512 модулей). Вся память системы (до 256GB) глобально адресуема, аппаратно поддерживается когерентность кэшей.
Коммутатор	Модули системы соединены с помощью сети NUMALink, построенной на собственных маршрутизаторах R-bricks.
Системное ПО	Используется доработанная ("открытые" доработки) операционная система Linux.

# SGL Altix3000



# RM600 E



Производитель	<u>Siemens Computer Systems (SNI)</u> , серия <u>RM Servers</u> .
Класс архитектуры	Многопроцессорная система с общей памятью ( <u>cc-NUMA</u> ).
Модификации	E60, E20
Процессор	Используются процессоры MIPS R10000 (200MHz).
Число процессоров	Процессорные платы с SMP-архитектурой объединяют до 4-х процессоров. Система включает в общей сложности до 24 процессоров в модели E60 и до 3 в модели E20.
Память	Общий объем оперативной памяти систем E60 - до 4GB, а систем E20 - до 2GB. Архитектура памяти системы - NUMA.
Масштабируемость	Несколько систем RM600 E могут объединяться в кластерную (MPP) систему - <u>Reliant Cluster Server</u> .
Системное ПО	Устанавливается операционная система <u>Reliant UNIX</u> .



# NUMA-Q 2000



Производитель	IBM (ранее - <u>Sequent</u> )
Класс архитектуры	Многопроцессорная система с общей памятью (cc-NUMA) Используется для организации сложных информационных систем.
Модификации	Model E410/E330/E320/E300/E200
Процессоры	Intel Pentium III Xeon (700 MHz в модели E410)
Узел	от 4 до 64 процессоров, до 64 GB оперативной памяти; узел состоит из базовых плат по 4 процессора (quads), соединенных между собой коммутатором IQ-Link.
Масштабируемость	Возможна организация кластеров, включающих до 4 узлов
Системное ПО	Используется операционная система DYNIX/px, версия UNIX от Sequent. Внутри одной системы могут одновременно исполняться UNIX и Windows NT

# Sun Fire 15K

Производитель	<u>Sun Microsystems.</u>
Класс архитектуры	Многопроцессорный <u>SMP-сервер.</u>
Процессор	UltraSPARC III/900MHz
Число процессоров	Система Sun Fire 15K объединяет до 106 процессоров.
Память	Система включает до 576GB памяти.
Системное ПО	ОС Solaris 8.

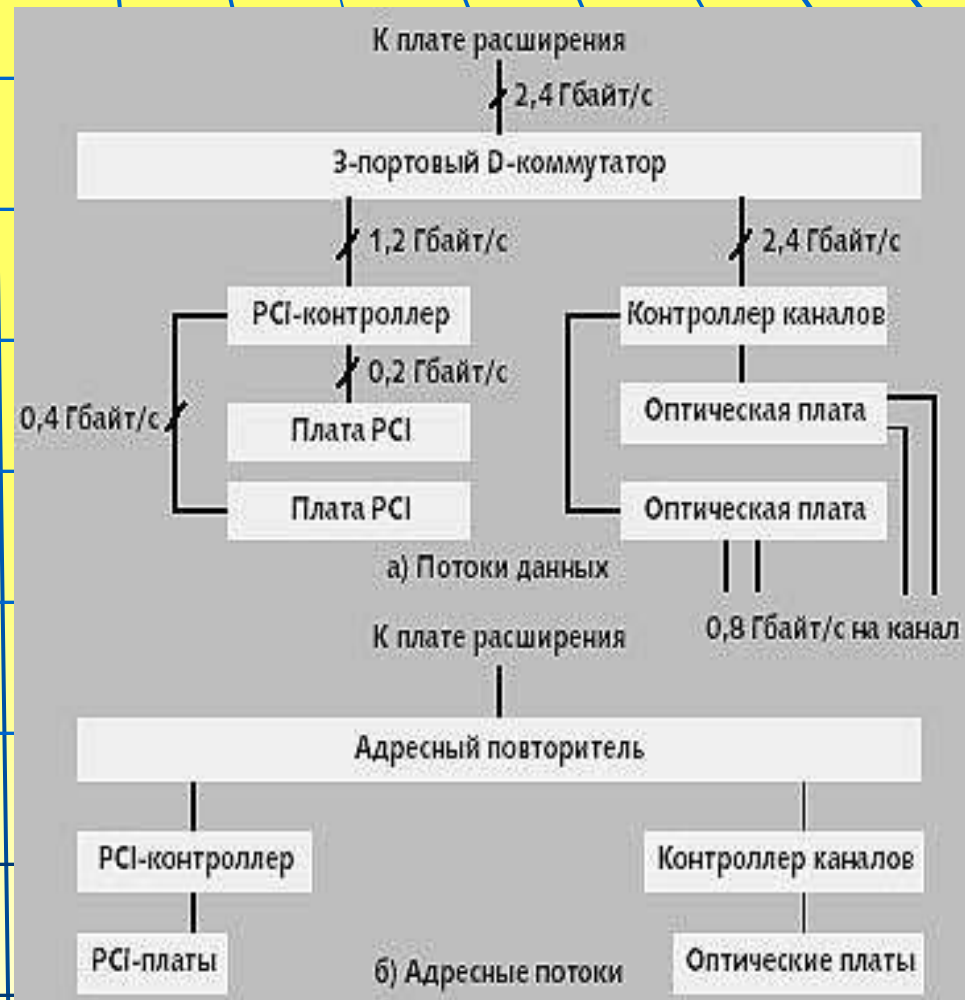
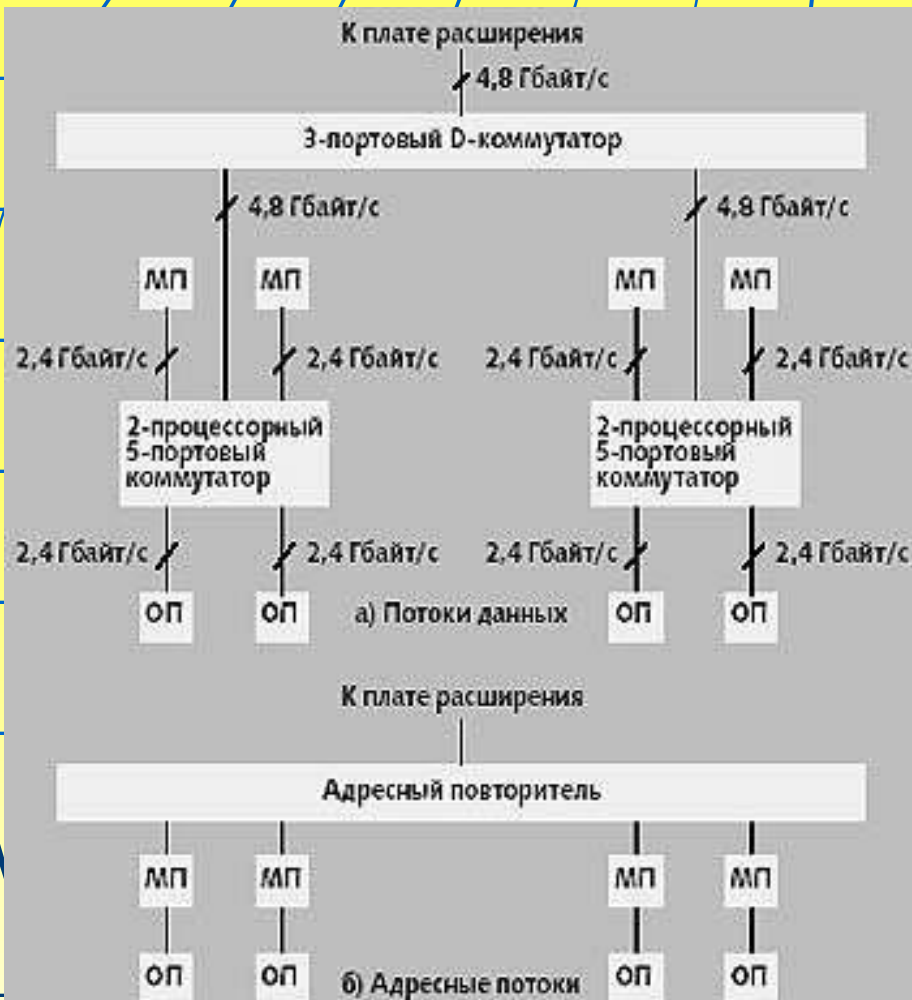


# Архитектура ccNUMA в исполнении Sun Microsystems (Sun Fire 15K)



Структура процессора UltraSPARC III

# Архитектура ccNUMA в исполнении Sun Microsystems (Sun Fire 15K)



Архитектура процессорной платы  
(плата Р/М)

Архитектура плат I/O