Федеральное государственное бюджетное образовательное учреждение высшего образования «Сибирский государственный университет телекоммуникаций и информатики»

Кафедра вычислительных систем

Расчетно-графическое задание по дисциплине

«Архитектура вычислительных систем»

Вариант 12

Выполнил:

Студент гр. ИП-813

Стояк Ю.К

Проверил:

к.т.н доцент кафедры ВС

Ефимов А.В.

Новосибирск 2020

ОГЛАВЛЕНИЕ

[Часть 1 3](#_Toc59456391)

[1.1 Задание 3](#_Toc59456392)

[1.2 Общие сведения 3](#_Toc59456393)

[1.3 Характеристики 4](#_Toc59456394)

[1.4 Иерархия структур коммуникационных сетей 5](#_Toc59456395)

[1.5 Технологии программирования 13](#_Toc59456396)

[1.6 Вывод 15](#_Toc59456397)

[Часть 2 16](#_Toc59456398)

[2.1 Задание 16](#_Toc59456399)

[2.2 Расчёт диаметра 16](#_Toc59456400)

[2.3 Расчёт среднего диаметра 17](#_Toc59456401)

[2.4 Расчёт бисекционной пропускной способности 17](#_Toc59456402)

[Список использованной литературы 18](#_Toc59456403)

# Часть 1

## Задание

Выполнить анализ архитектуры суперВС Sierra (№ 3 в списке Top500).

## Общие сведения

IBM Sierra — это суперкомпьютер, установленный в Ливерморской национальной лаборатории для нужд Национального управления ядерной безопасности. Он появился в списке top500 в июне 2018 года на 3ей позиции. С тех пор попал на 2ое и вернулся на 3ье место к текущему моменту времени.

Sierra похож по архитектуре на суперкомпьютер [Summit](https://ru.wikipedia.org/wiki/Summit_(%D1%81%D1%83%D0%BF%D0%B5%D1%80%D0%BA%D0%BE%D0%BC%D0%BF%D1%8C%D1%8E%D1%82%D0%B5%D1%80)" \o "Summit (суперкомпьютер)), но производительность меньше: теоретическая производительность 125 petaFLOPS против теоретической производительности [Summit](https://ru.wikipedia.org/wiki/Summit_(%D1%81%D1%83%D0%BF%D0%B5%D1%80%D0%BA%D0%BE%D0%BC%D0%BF%D1%8C%D1%8E%D1%82%D0%B5%D1%80)) 200 petaFLOPS.



Рисунок 1. Суперкомпьютерный комплекс Sierra

Sierra используется главным образом для построения прогнозов в приложении к задачам управления ядерным арсеналом, помогая обеспечить безопасность, надёжность и эффективность ядерного оружия США.

## Характеристики

Sierra имеет следующие технические характеристики и программное обеспечение:

* Процессоры: IBM POWER9 22C 3.1ГГц
* Графические процессоры: NVIDIA Volta GV100
* Узлов входа: 5
* Вычислительных узлов: 4 320
* Всего узлов: 4 474
* Ядер на узел: 44
* Всего ядер: 1 572 480 (из них CPU ядер: 190080)
* Количество CPU: 8 640
* Общий объём RAM: 1 382 400 GB
* Память процессора: 256 ГБ на узел
* Количество GPU: 17 280
* Память GPU: 64 ГБ на узел
* Пиковая пропускная способность памяти одного процессора: 170 ГБ/сек.
* Операционная система: Red Hat Enterprise Linux
* Параллельная файловая система: IBM Spectrum Scale
* Сеть внутренних коммутаций: IB Enhanced Data Rate
* Пиковая производительность CPU: 4 666 TFLOPS
* Пиковая производительность GPU: 120 960 TFLOPS
* Процессоры и графические процессоры: 125 626 TFLOPS

## Иерархия структур коммуникационных сетей

Начнем рассматривать Sierra c процессора.

IBM предлагает POWER 9 в двух различных вариантах исполнения: Scale-Out и Scale-Up.

* Scale-Out:
  + Предназначен для традиционных кластеров центров обработки данных, использующих серверы с одним сокетом и двумя сокетами.
  + Оптимизирован для серверов Linux
  + 24-ядерные и 12-ядерные модели
* Scale-Up:
  + Предназначен для серверов NUMA с четырьмя и более сокетами, поддерживающими большой объем памяти и пропускную способность.
  + Оптимизирован для серверов PowerVM
  + 24-ядерная и 12-ядерная модели

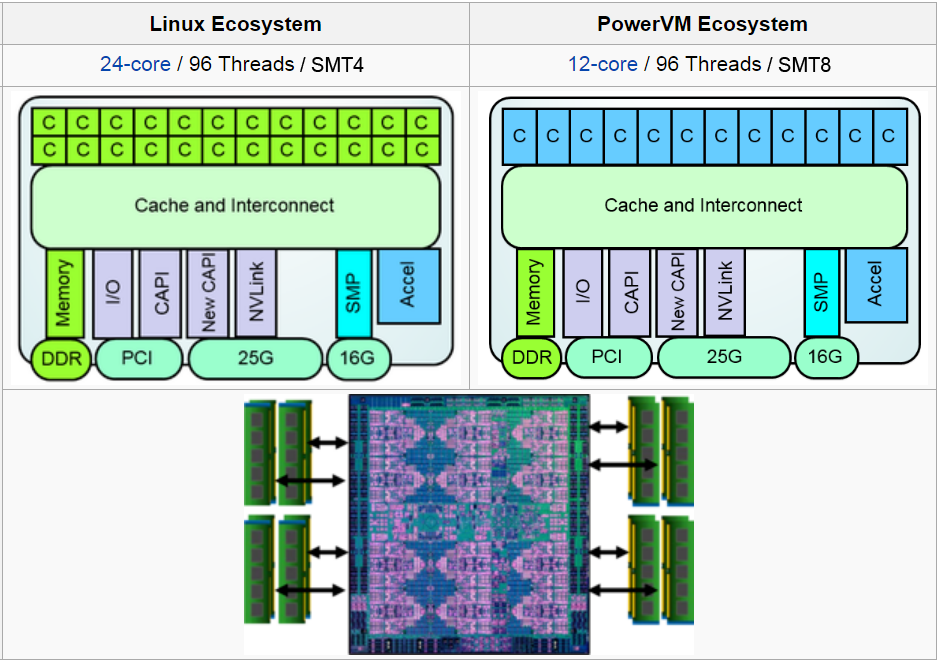


Рис.2. Scale-Out модель

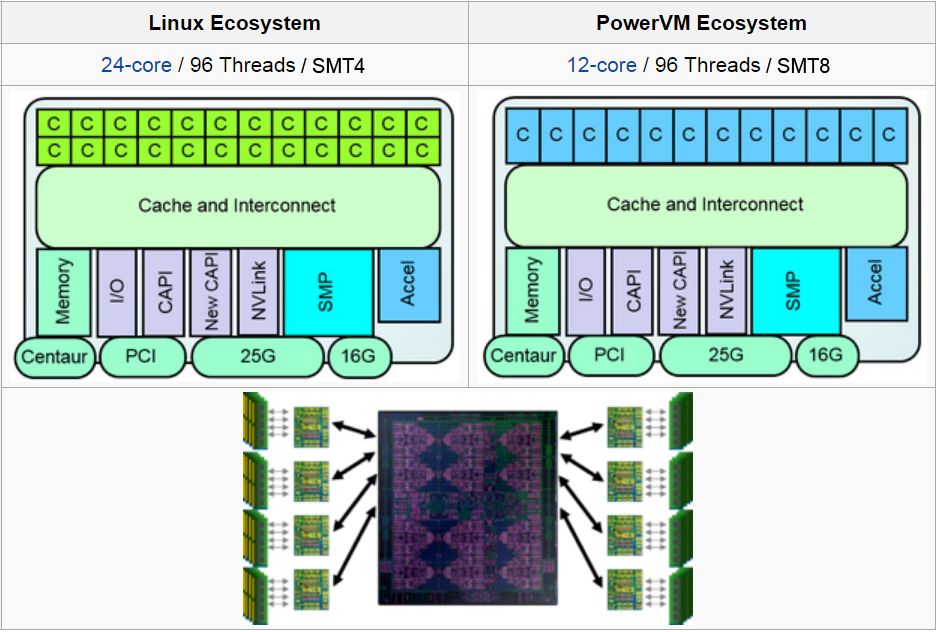


Рис.3. Scale-Up модель

POWER9 может иметь переменное число активных ядер: 16, 18, 20, 22. Каждый процессор POWER9 суперкомпьютера Sierra использует 22 ядра. Технологический процесс составляет 14 нм.

●      кэш:

○      данных L1: 32 KB на ядро, 8-way, private

○      инструкций L1: 32 KB на ядро, 8-way, private

○      L2: 512 KB на ядро, 8-way, private

○      L3: 120 MB, 20-way, shared as twelve 10 MB banks

Частота: благодаря адаптивным опциям управления питанием тактовая частота может варьироваться в зависимости от нагрузки системы. Скорости могут варьироваться примерно от 2,3 до 3,8 ГГц.

Высокая пропускная способность на кристалле: совокупная пропускная способность более 7 ТБ/с через встроенный коммутатор, соединяющий ядра с памятью, PCIe, графическими процессорами и т. д.

Архитектура Узла:

Каждый узел IBM POWER9 AC922 имеет два процессорных сокета IBM POWER9 и поддерживает до 6 графических видеокарт NVIDIA Volta V100. В качестве интерконнекта для CPU-GPU и GPU-GPU используется технология NVLink2.

Узел поддерживает суммарно до 16 ТБ основной (оперативной) памяти, для её размещения имеется 16 сокетов DDR4, по 8 DIMM DDR4 на процессорный сокет. Вычислительные узлы IBM Sierra имеют по 256 ГБ основной памяти. Пропускная способность канала CPU-RAM составляет 170 ГБ/с на процессор, или 340 ГБ/с на узел.

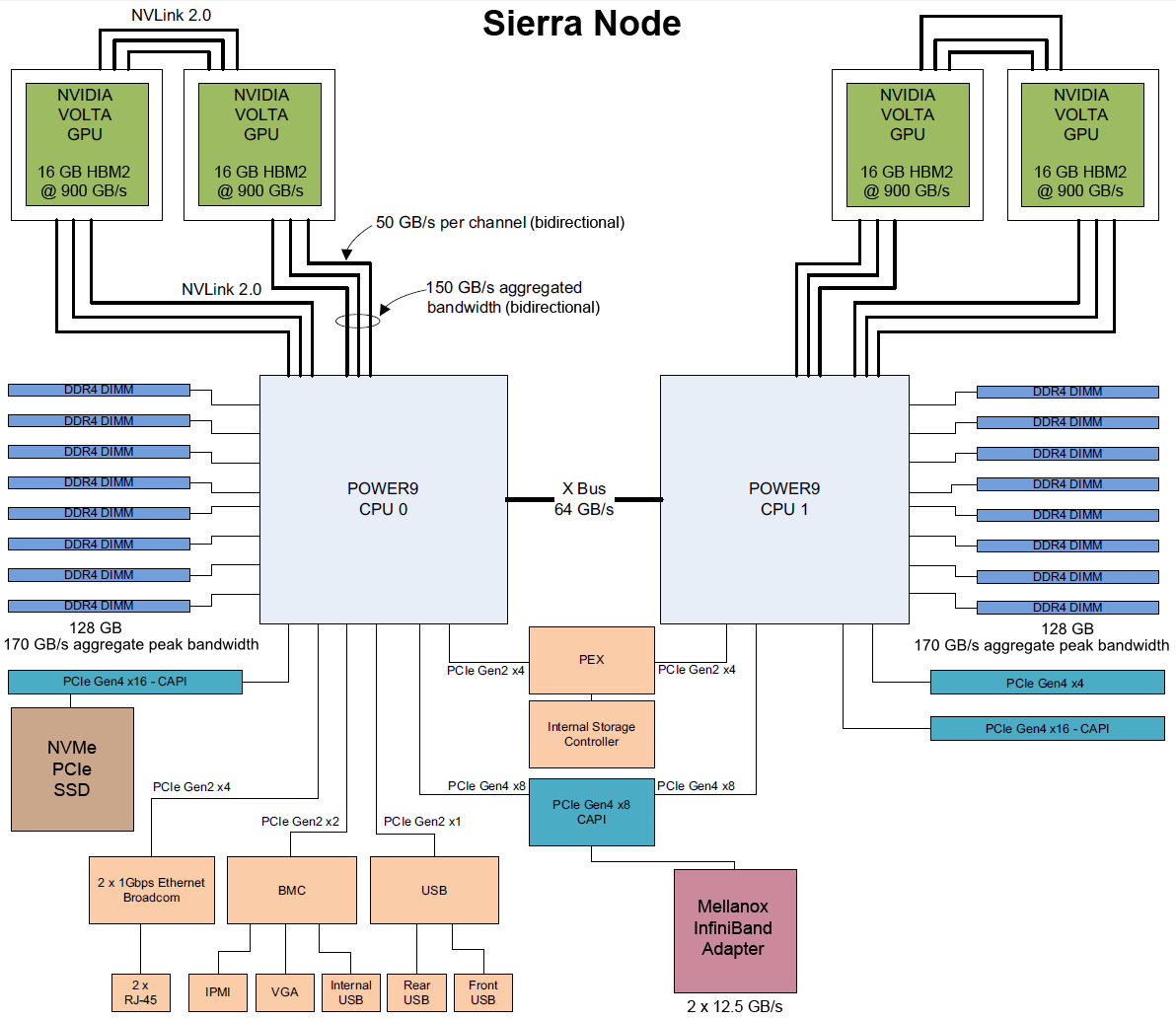


Рис.4. Диаграмма узла POWER9

Технология NVLink 2.0 поддерживает до 6 NVLink ссылок на один графический процессор. Каждая ссылка обеспечивает двунаправленное соединение со скоростью 50 ГБ/с к другому графическому процессору или центральному процессору, что обеспечивает суммарную пропускную способность 300 ГБ/с.

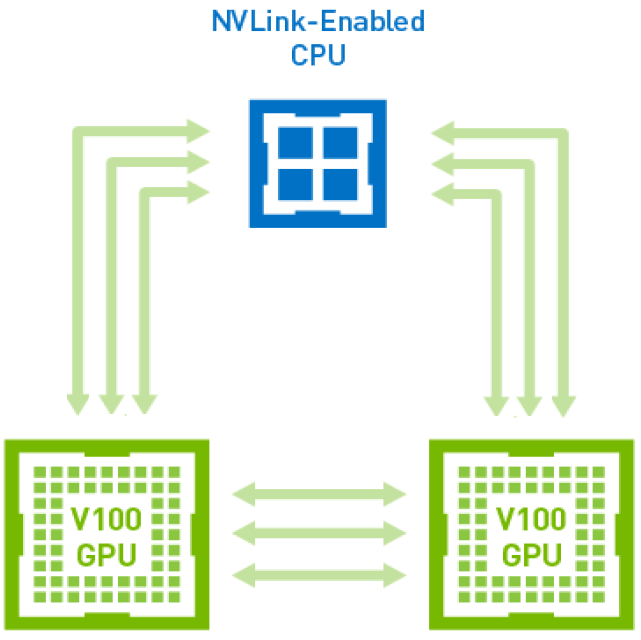


Рис. 5. V100 с подключенным по NVLink GPU-GPU и GPU-CPU

Центральные процессоры соединяются с контроллерами, устройствами ввода и вывода и пр. с помощью шины PCI Express 4.0 (PCIe) со скоростью обмена до 16 Гбит/с на линию. Использует высокопроизводительный физический протокол, основанный на последовательной передаче данных. Устройства PCI Express взаимодействуют между собой через среду, образованную коммутаторами, при этом каждое устройство напрямую связано соединением типа точка-точка с коммутатором.

Основным строительным блоком кластера Linux является узел. Узел — это, по сути, независимый компьютер.

* Основные характеристики:
  + Автономный, бездисковый, многоядерный компьютер.
  + Низкий форм-фактор — узлы кластеров очень тонкие для экономии места.
  + Монтаж в стойку — узлы монтируются компактно в виде выдвижного ящика для облегчения технического обслуживания, уменьшения занимаемой площади и т. д.
  + Удаленное управление — нет клавиатуры, мыши, монитора или другого устройства, обычно используемого для взаимодействия с компьютером. Все управление узлами происходит по сети с" управляющего " узла.

В общем случае производственный кластер имеет четыре типа узлов, основанных на функциях, которые могут отличаться деталями конфигурации: узлы входа в систему, отладочные, пакетные, узлы ввода-вывода и узлы обслуживания (недоступны пользователям).

Основной блок производственных Linux кластеров называется "масштабируемым блоком" (Scalable Unit) (SU). SU состоит из коммутаторов первой ступени, которые подключаются непосредственно к каждому узлу, разного оборудование для управления, рамки, достаточной для размещения всего необходимого оборудования.

Связь между узлами происходит с помощью коммутаторов первого уровня. А связь между несколькими SU происходит с помощью коммутаторов второго уровня.

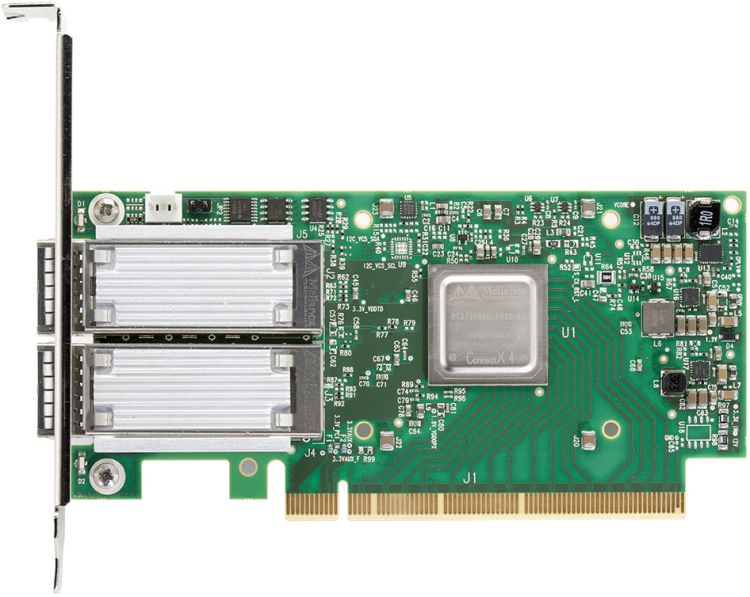


Рис. 6.Адаптер Mellanox EDR InfiniBand

Узлы имеют один двухпортовый адаптер Mellanox ConnectX EDR Infiniband с передачей данных 100Гбит/с. Порты адаптера подключаются к коммутаторам первого уровня Mellanox Switch-IB с 36 портами, которые через оптоволоконные кабели подключаются к коммутаторам второго уровня Mellanox CS7500 с 648 портами.



Рис. 7.Mellanox Switch-IB Top-of-Rack

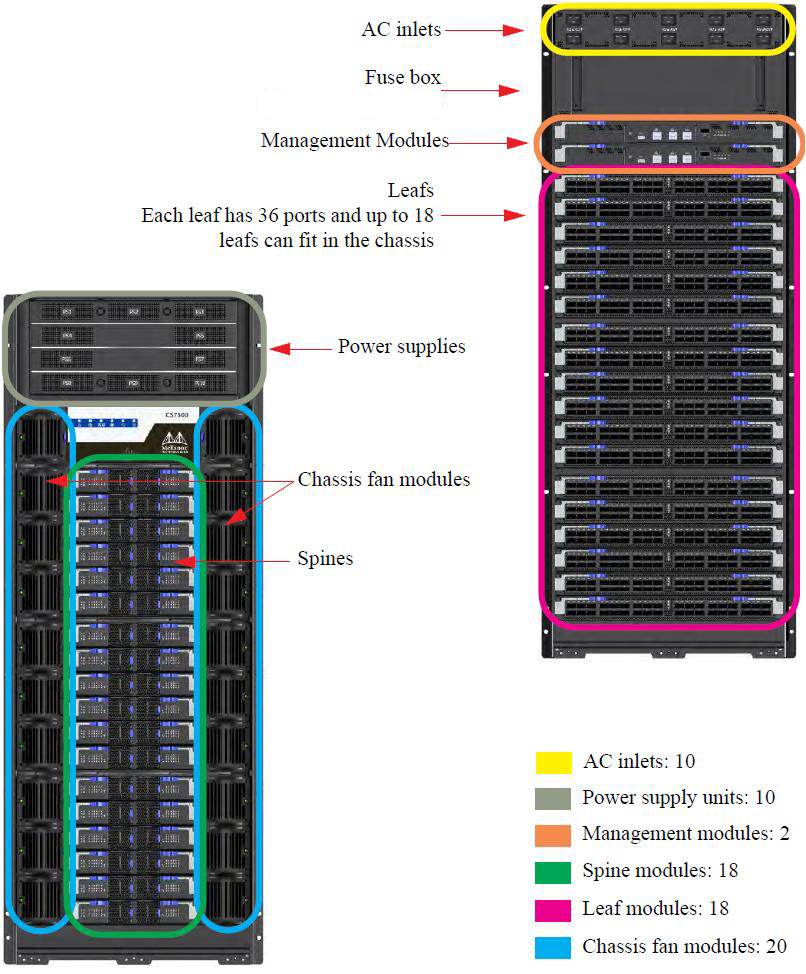


Рис. 8. Mellanox CS7500 с маркировкой

Коммутаторы образуют иерархию с коммутаторами более высокого уровня, имеющие большее количество соединений, чем коммутаторы более низкого уровня. Число отключений для переключателей нижнего уровня увеличивается в соотношении два к одному. Все узлы соединяются в единую сеть толстого дерева.

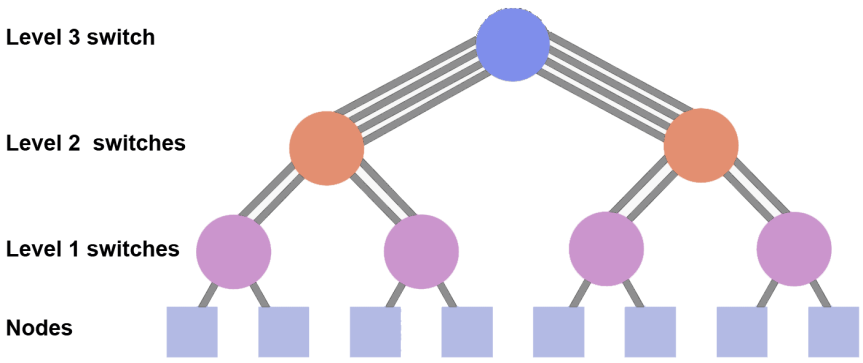


Рис. 9. Толстое дерево сети

Каждая стойка Sierra имеет 18 узлов и 2 переключателя TOR. Двухпортовый адаптер каждого узла подключается к обоим коммутаторам TOR с одним портом каждый. Каждый коммутатор TOR имеет 12 каналов связи с коммутаторами Director, по крайней мере, по одному на коммутатор Director. Существует девять переключателей Director. Поскольку каждый коммутатор TOR имеет 12 восходящих каналов и имеется только 9 коммутаторов Director, то на каждый коммутатор TOR существует ещё 3 дополнительных восходящих канала. Они используются для подключения к 3 из 9 коммутаторов Director.

Примечание: Sierra - "модифицированное" 2:1 каноническое толстое дерево. На самом деле это 1,5:1 (18 ссылок вниз, 12 ссылок вверх для каждого переключателя TOR).

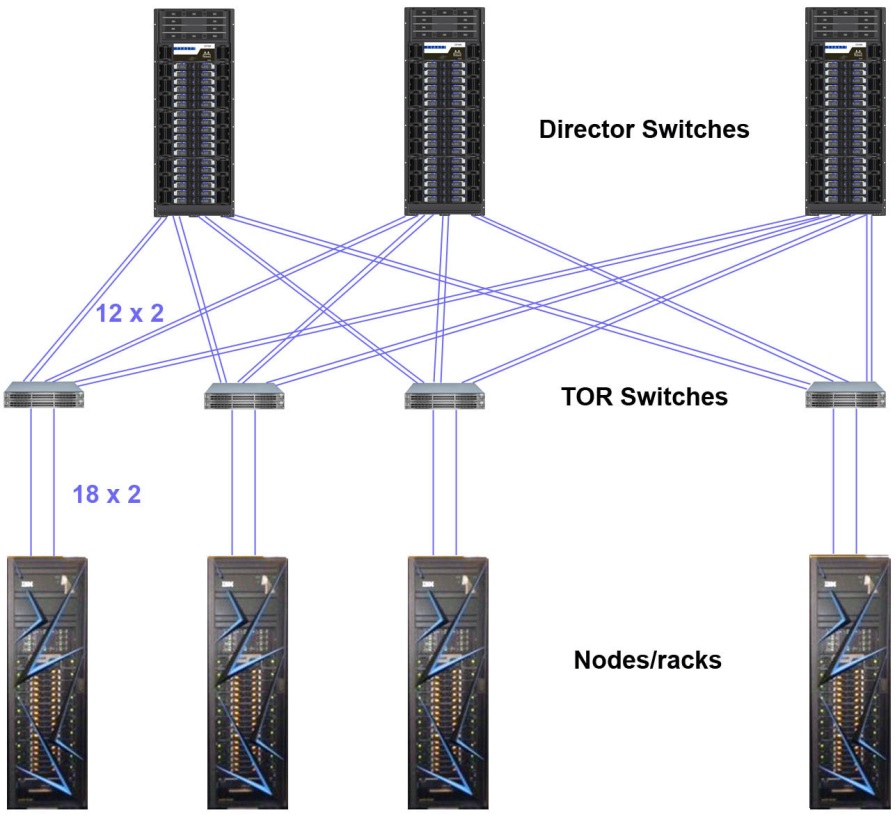


Рис. 10. Сеть Sierra

Проводились измерения пропускной способности между узлами проводились на работающих, сильно загруженных машинах с использованием простого неблокирующего тестового кода MPI. Одна задача на каждом из двух узлов. Размер сообщения = 1 МБ. С задержкой в 1 наносекуду пропускная способность оказалсь равна 21 ГБ/с.

## Технологии программирования

Программное окружение, предоставляемое пользователям суперкомпьютерной системы IBM Sierra, обладает следующими характеристиками:

Операционная система: Red Hat Enterprise Linux

Данный дистрибутив позиционируется для корпоративного использования. Новые версии выходят с периодичностью около 3 лет. Основная особенность дистрибутива — наличие коммерческой поддержки на протяжении 10 лет, с возможностью продления до 13 лет.

В системах TOSS 3 и CORAL EA LC предоставляет среду модулей Lmod, которая позволяет пользователям переключаться между различными версиями компилятора. Чтобы просмотреть список версий компилятора, доступных через модули, пользователь может запустить ключевое слово модуля "category: compiler".

Компиляторы:

clang: clang/3.9.0, clang/3.9.1, clang/4.0.0, clang/6.0.0, clang/7.0.1…

cuda: cuda/8.0, cuda/9.1.85, cuda/9.2.88, cuda/10.0.130…

gcc: gcc/4.8-redhat, gcc/4.9.3, gcc/6.1.0, gcc/7.1.0, gcc/7.3.0…

intel: intel/14.0.3, intel/15.0.6, intel/16.0.2, intel/16.0.3…

PGI: pgi/16.3, pgi/16.7, pgi/16.9 pgi/17.10, pgi/18.1, pgi/18.5…

Математические библиотеки: ESSL, CUBLAS 9.2, MASS, MASSV, BLAS, LAPACK, ScaLAPACK, FFTW, PETSc, GSL

MPI: IBM Spectrum MPI – единственная MPI библиотека, поддерживаемая на Sierra. Основана на Open MPI 3.0.0. Базовая архитектура и функциональность аналогичны.

Поддержка OpenMP: OpenMP поддерживается в системах Sierra для одноузлового параллельного программирования с общей памятью на C / C++ и Fortran.

В системах Sierra основной мотивацией для использования OpenMP является использование преимуществ графических процессоров на каждом узле:

* OpenMP используется в сочетании с MPI как обычно
* На узле: задачи MPI определяют вычислительно интенсивные разделы кода для разгрузки на графические процессоры узла
* На узле: параллельные области выполняются на графических процессорах узла
* Межузловые: задачи координируют работу по всей сети с помощью передачи сообщений MPI

Хотя системы Sierra имеют ряд общих черт с другими кластерами LC, выполнение заданий сильно отличается.

IBM Spectrum LSF используется для контроля доступа к системным ресурсам вместо Slurm:

* Совершенно новый набор команд для отправки, мониторинга и взаимодействия с заданиями.
* Совершенно новый набор команд для запроса конфигурации системы, очередей, статистики заданий и бухгалтерской информации.
* Новый синтаксис для создания сценариев заданий.

## Вывод

В результате исследования я выяснил, особенности устройства суперкомпьютера Sierra. Ядра процессора взаимодействуют с памятью, PCIe, графическими процессорами и т. д. с пропускной способностью 7 ТБ/с посредством встроенных коммутаторов. Центральные процессоры связаны с графическими с помощью технологии Nvidia NVlink 2.0, которая поддерживает до 6 соединений и суммарную пропускную способность 300 ГБ/с. Центральный процессор соединён с другими составляющими через шину PCIe. Все эти составляющие объединяются в независимые и автономные по совей сути узлы.

Узлы же, связываются в единую сеть посредством топологии Толстого дерева. Связь между узлами происходит с помощью коммутаторов первого уровня (Mellanox Switch-IB). А связь между несколькими SU происходит с помощью коммутаторов второго уровня (Mellanox CS7500).

Sierra работает на операционной системе Red Hat Enterprise Linux и поддерживает несколько разных компиляторов, математических библиотек и стандарты для распараллеливания программ.

Суперкомпьютер Sierra сейчас находится на 3ьем месте рейтинга top500, сразу же после ск Summit, созданного при помощи такой же архитектуры, но выполняющего другие задачи (Summit занимает 2 место top500). В теории, Sierra может быть увеличена для достижения такой же/большей производительности как у Summit.

# Часть 2

## Задание

Расчёт структурных характеристик включает в себя:

* Расчёт диаметра структуры
* Расчёт среднего диаметра
* Расчёт бисекционной пропускной способности

## Расчёт диаметра

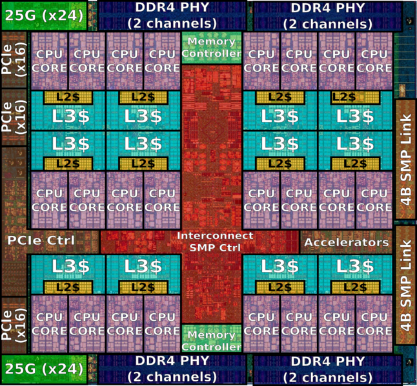


Рис. 11. SMP Архитектура процессора

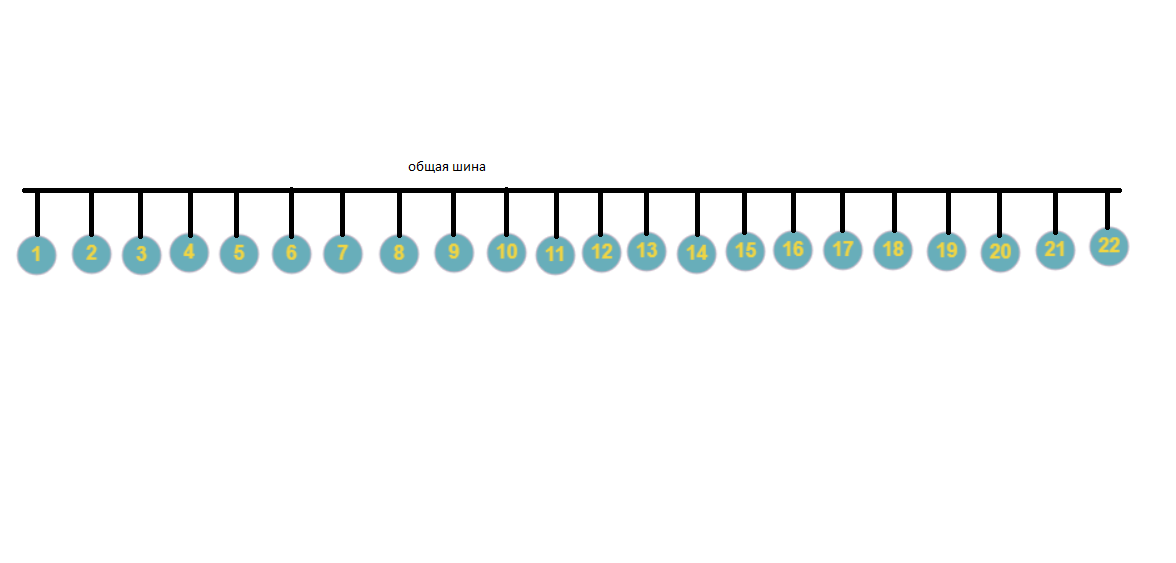


Рис. 12. Условный граф SMP архитектуры IBM POWER 9

Диаметр структуры - максимальное расстояние, определённое на множестве кратчайших путей между парами вершин структуры ВС

𝑑=𝑚𝑎𝑥𝑖𝑗{𝑑𝑖𝑗}

dij – расстояние, т.е. минимальное число рёбер, образующих путь из вершины i в вершину j ; i, j ∈ {0, 1, ..., N −1};

Расстояние между вершинами 2: 1⇒Общая шина⇒2

## Расчёт среднего диаметра

– число вершин, находящихся на расстоянии от любой выделенной вершины (однородного) графа G.

N – количество вершин

Выделим вершину 1 на рис 11. В таком случае:

Ответ: = 1,86

## 2.4 Расчёт бисекционной пропускной способности

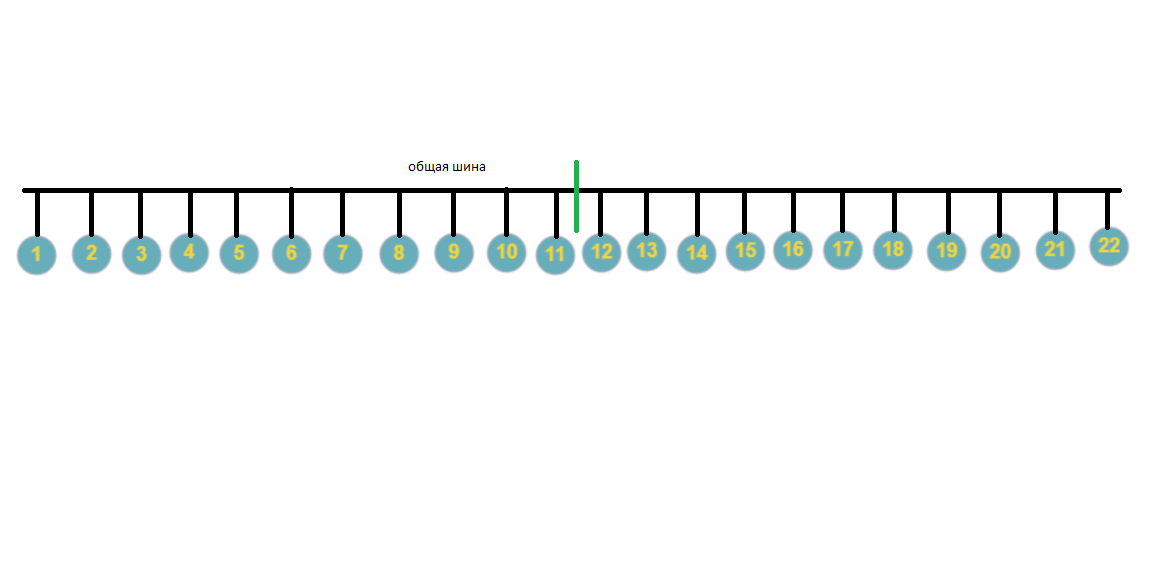


Рис. 13. Разделение на 2 непересекающихся подмножества

После разделения получились 2 одинаковых подмножества и 1 разрез  
Таким образом, бисекционная пропускная способность = 1

# Список использованной литературы

* + 1. Using LC’s Sierra Systems [Электронный ресурс]. – Электрон. текстовые данные. – URL: https://hpc.llnl.gov/training/tutorials/using-lcs-sierra-system#Hardware
    2. Суперкомпьютеры IBM Summit и IBM Sierra [Электронный ресурс]. – Электрон. текстовые данные. – URL: https://parallel.ru/computers/reviews/ibm\_summit\_sierra.html
    3. Sierra (Supercomputer) [Электронный ресурс]. – Электрон. текстовые данные. – URL: <https://en.wikipedia.org/wiki/Sierra_%28supercomputer%29>
    4. Sierra. Livermore Computing Center [Электронный ресурс]. – Электрон. текстовые данные. – URL: <https://hpc.llnl.gov/hardware/platforms/sierra>
    5. CS7500 InfiniBand Switch [Электронный ресурс]. – Электрон. текстовые данные. – URL: <https://hpc.llnl.gov/sites/default/files/Mellanox.CS7500switch.pdf>
    6. IBM Power System AC922 [Электронный ресурс]. – Электрон. текстовые данные. – URL: <https://www.ibm.com/ru-ru/products/power-systems-ac922>
    7. TOP500 the list [Электронный ресурс]. – Электрон. текстовые данные. – URL: <https://www.top500.org/system/179398/>
    8. Compilers at LC [Электронный ресурс]. – Электрон. текстовые данные. – URL: https://hpc.llnl.gov/software/development-environment-software/compilers