

Министерство цифрового развития, связи и
массовых коммуникаций Российской Федерации

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

Отчёт
по расчетно-графическому заданию
по дисциплине «**Архитектура Вычислительных Систем**»

Выполнил:

студент гр. ИС-142

«__» декабря 2023 г.

/Григорьев Ю.В./

Проверил:

старший преподаватель

кафедры ВС

«__» декабря 2023 г.

/Ревун А.Л./

Оценка « _____ »

Новосибирск 2023

Задание

Анализ иерархии коммуникационных сетей суперВС с оценкой структурных характеристик.

СуперВС для анализа: Aurora (HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11)

Операторы: Аргоннская Национальная Лаборатория, Министерство Энергетики США.

Локация: город Лемонт, штат Иллинойс, США.

Введение

Суперкомпьютеры, являясь вершиной технологического прогресса в области вычислительной техники, представляют собой мощные системы, способные обрабатывать и анализировать огромные массивы данных за доли секунды. Эти машины находят применение в самых разнообразных научных областях, от квантовой физики до биоинформатики, и их вклад в развитие науки и техники трудно переоценить. Суперкомпьютер Aurora, разработанный в рамках сотрудничества между Аргоннской национальной лабораторией и компаниями Intel и Cray, представляет собой одну из самых передовых вычислительных систем на сегодняшний день.

Aurora занимает особое место в истории суперкомпьютеров, будучи одним из первых в своем роде, кто приближается к достижению уровня производительности в 1 Eflops, то есть способности выполнять квинтиллион операций с плавающей запятой в секунду. Это не только техническое достижение, но и важный шаг вперед в возможностях моделирования и анализа данных для научного сообщества.

Основываясь на принципах модульности и масштабируемости, Aurora сочетает в себе инновационные технологические решения, такие как процессоры Intel Xeon (Sapphire Rapids), GPU архитектуры Xe (Ponte Vecchio) и высокопроизводительные коммуникационные системы, в частности сетевую технологию Slingshot. Эти технологии обеспечивают не только высокую производительность, но и эффективное управление данными и задачами на всех уровнях иерархии системы.



Aurora: A High-level View

- ☐ Intel-Cray machine arriving at Argonne in 2021
 - ☐ Sustained Performance > 1Exaflops
- ☐ Intel Xeon processors and Intel Xe GPUs
 - ☐ 2 Xeons (Sapphire Rapids)
 - ☐ 6 GPUs (Ponte Vecchio [PVC])
- ☐ Greater than 10 PB of total memory
- ☐ Cray Slingshot fabric and Shasta platform
- ☐ Filesystem
 - ☐ Distributed Asynchronous Object Store (DAOS)
 - ☐ ≥ 230 PB of storage capacity
 - ☐ Bandwidth of > 25 TB/s
 - ☐ Lustre
 - ☐ 150 PB of storage capacity
 - ☐ Bandwidth of ~ 1 TB/s




Рисунок 1 - Обзор аппаратного обеспечения суперВС Aurora

Сетевая архитектура Aurora, основанная на топологии Dragonfly, представляет собой сложную иерархическую структуру, обеспечивающую связь между вычислительными узлами с низкой задержкой и высокой пропускной способностью. Система Slingshot уникальна своей способностью к адаптивной маршрутизации и передовым методам управления перегрузками (congestions), что позволяет минимизировать влияние перегруженных приложений на производительность системы в целом.

Данная работа направлена на анализ коммуникационной иерархии и структурных характеристик суперкомпьютера Aurora, рассмотрением его

сетевых решений и возможностей, которые они предоставляют для решения сложных вычислительных задач.

Основная часть

Архитектура Aurora

Суперкомпьютер Aurora представляет собой высокопроизводительную вычислительную систему, которая сочетает в себе последние достижения в области процессорных технологий и инновационные подходы в области сетевых решений. Это позволяет Aurora стать одной из ведущих систем в мире по вычислительной мощности и эффективности.

Распределенное хранение данных

Суперкомпьютер Aurora включает в себя передовую систему хранения данных — Distributed Asynchronous Object Store (DAOS), которая также является продуктом с открытым исходным кодом (open source). DAOS предлагает высокую производительность по ширине полосы пропускания и операциям ввода/вывода, достигая объема хранения более 230 петабайт и скорости более 25 терабайт в секунду. Эта система критична для достижения высокой производительности ввода/вывода на Aurora, что подтверждает ее значимость для эффективной работы суперкомпьютера.

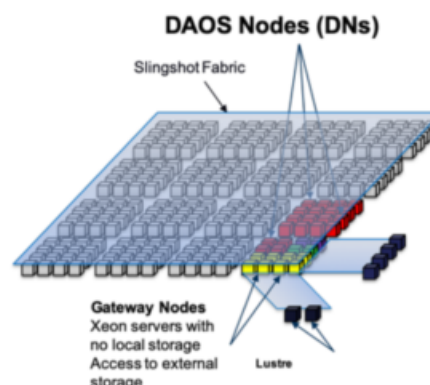
DAOS обеспечивает совместимость с существующими моделями ввода/вывода, такими как POSIX, MPI-IO и HDF5, что делает его универсальным решением для различных вычислительных задач. Кроме того, DAOS предоставляет гибкий API для хранения данных, который позволяет внедрять новые парадигмы ввода/вывода, облегчая интеграцию с современными приложениями и ускоряя процесс обработки данных.

Distributed Asynchronous Object Store (DAOS)

- ☐ Open source storage solution
- ☐ Offers high performance in bandwidth and IO operations
 - ☐ ≥ 230 PB capacity
 - ☐ ≥ 25 TB/s
 - ☐ Using DAOS is critical to achieving good I/O performance on Aurora
- ☐ Provides compatibility with existing I/O models such as POSIX, MPI-IO and HDF5
- ☐ Provides a flexible storage API that enables new I/O paradigms

DAOS

(Distributed Asynchronous Object Storage) for Applications:
Thursday 8:30-10:00AM



21

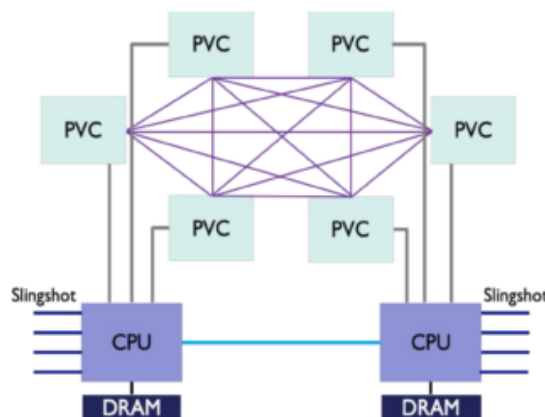
Рисунок 2 - Обзор распределённого хранения данных по технологии DAOS

Вычислительные Узлы

Каждый узел Aurora включает в себя два процессора Intel Xeon (Sapphire Rapids) и шесть графических процессоров (GPU) новейшей архитектуры Xe (Ponte Vecchio), что обеспечивает высокую производительность как для традиционных вычислительных задач (common scalable solutions) (например, подсчёт матриц, интегралов, генных алгоритмов), так и для задач искусственного интеллекта и машинного обучения. Следует отметить, что использование унифицированной архитектуры памяти между CPU и GPU упрощает разработку программного обеспечения и увеличивает общую производительность системы.

Aurora Compute Node

- ❑ 2 Intel Xeon (Sapphire Rapids) processors
- ❑ 6 X^e Architecture based GPUs (Ponte Vecchio)
 - ❑ All to all connection
 - ❑ Low latency and high bandwidth
- ❑ 8 Slingshot Fabric endpoints
- ❑ Unified Memory Architecture across CPUs and GPUs



Overview of the Argonne Aurora Exascale System
2:30pm - 3:30pm, Feb 5
Legends Ballroom

Рисунок 3 - Структура одного вычислительного узла в суперВС Aurora

Сетевая инфраструктура

Ключевым элементом архитектуры Aurora является её сетевая инфраструктура, базирующаяся на технологии Slingshot. Сеть состоит из 11 групп типа Dragonfly, включая 10 вычислительных групп и одну служебную. Каждая группа соединена друг с другом двумя связями, а внутри каждой группы – четырьмя связями, обеспечивая высокую плотность и надежность коммуникаций. Особенностью Slingshot является использование агрессивной адаптивной маршрутизации и продвинутого контроля за загруженностью сети, что позволяет минимизировать задержки и увеличивать пропускную способность. Как можно увидеть на рисунке, каждое соединение Slingshot предоставляет скорость в 25 ГБ/с в одном направлении, 50 ГБ/с в обоих направлениях. Таких финальных соединений (Slingshot Fabric) на каждый процессор предусмотрено по 4 штуки => со скоростью 200 ГБ/с. L0-кэш в данной системе составляет 1 модуль на 100 ГБ, L1-кэш - 4 модуля по 200 ГБ, L2-кэш - 2 модуля по 200 ГБ.

Slingshot Configuration

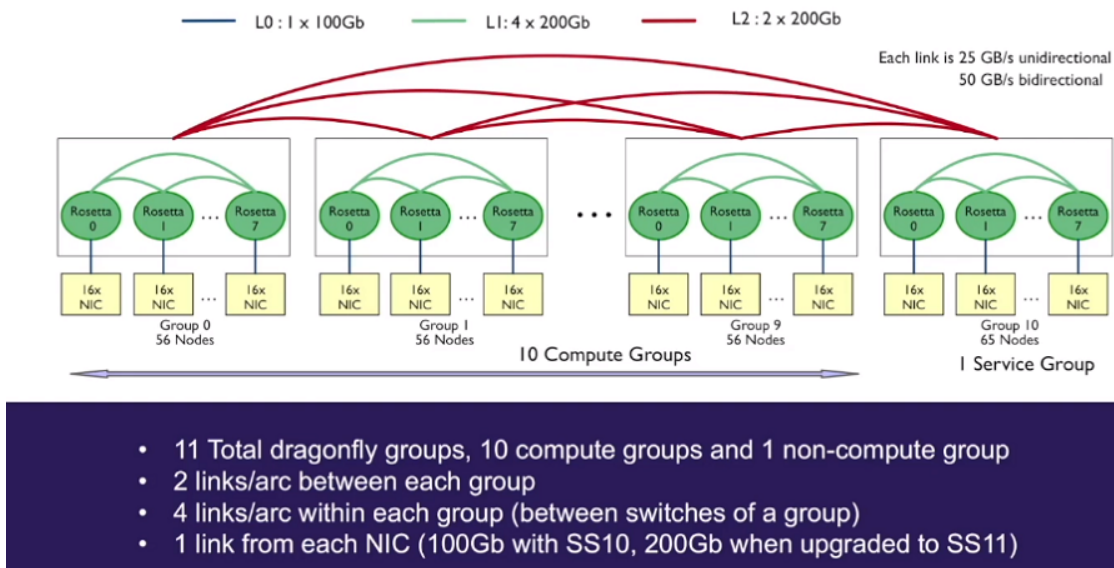


Рисунок 4 - Конфигурация Slingshot на суперВС Aurora

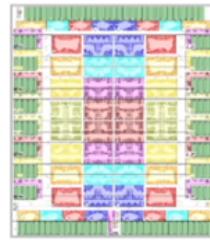
Связь между узлами

Коммуникационные сети Aurora обеспечивают связь между узлами с низкой латентностью и высокой пропускной способностью. Каждый узел соединен с сетью через 8 конечных точек Slingshot Fabric, что обеспечивает быстрый и надежный обмен данными. Slingshot поддерживает пропускную способность до 200 ГБ/с, что важно для задач, требующих интенсивного обмена данными между узлами.

Slingshot Interconnect

Rosetta Switch

- Multiple QoS levels
- Aggressive adaptive routing
- Advanced congestion control
- Very low average and tail latency
- High performance multicast and reduction



64 ports x 200 Gbps

SS-10 (100Gb)
Injection: ~14 TB/s
Bisection: ~24 TB/s

SS-11 (200Gb)
Injection: ~28 TB/s
Bisection: ~24 TB/s



Mellanox ConnectX NIC

Slingshot 10

- HPE Cray MPI stack
- Ethernet functionality
- RDMA offload



Cassini NIC

Slingshot 11

- MPI hardware tag matching
- MPI progress engine
- One-sided operations
- Collectives
- 2X injection bandwidth

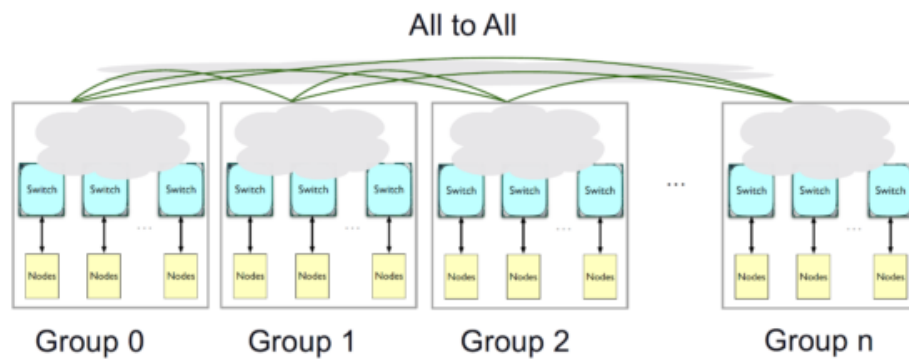
Рисунок 5 - Обобщенная информация о связи между узлами через Slingshot и коммутаторы Rosetta

Анализ иерархии сетей

Топология Dragonfly, лежащая в основе сетевой архитектуры Augora, позволяет оптимизировать связь внутри групп и между ними, использовать оптические кабели для дальних связей и низкочастотные электрические связи внутри группы. Это обеспечивает баланс между скоростью и стоимостью, а также увеличивает общую устойчивость системы к отказам.

Dragonfly Topology

- ❑ Several groups are connected together using all to all links.
- ❑ Optical cables are used for the long links between groups.
- ❑ Low-cost electrical links are used to connect the NICs in each node to their local router and the routers in a group



<https://www.cray.com/sites/default/files/resources/CrayXCNetwork.pdf>

12


Рисунок 6 - Структура топологии Dragonfly

Структурные характеристики

Сетевая инфраструктура Aurora обладает несколькими ключевыми структурными характеристиками:


- Пропускная способность: Сеть (прежде всего на базе Slingshot) спроектирована так, чтобы максимизировать пропускную способность и обеспечить эффективную передачу данных на скорости около 200 ГБ/с.
- Адаптивность: Агрессивная адаптивная маршрутизация и система распределенного хранения данных способствует более эффективной передаче данных, оптимизируя пути передачи в реальном времени.
- Управление перегрузками: Важной составляющей сетевой инфраструктуры Aurora является высокопроизводительный коммутатор Rosetta. Он отличается многоуровневым управлением перегрузками, целью которого является минимизация влияния загруженности

приложений на другие операции. С помощью Quality of Service (QoS) и агрессивной адаптивной маршрутизации, Rosetta обеспечивает низкую среднюю и пиковую задержку, что важно для задач, требующих быстрой обработки больших объёмов данных. Данный коммутатор способен обеспечивать высокую производительность мультикаста и редукации, что делает его ключевым компонентом в обеспечении эффективной работы всей системы Aurora. С пропускной способностью 25.6 Тб/с на коммутатор и портами от 64 до 200 Гб/с, Rosetta становится фундаментальным элементом для управления трафиком в сети Aurora.



High Bandwidth Switch: Rosetta

- ☐ **Multi-level congestion management**
 - ☐ To minimize the impact of congested applications on others
 - ☐ Very low average and tail latency
- ☐ **Quality of Service (QoS) – Traffic Classes**
 - ☐ Class: Collection of buffers, queues and bandwidth
 - ☐ Intended to provide isolation between applications via traffic shaping
- ☐ **Aggressive adaptive routing**
 - ☐ Expected to be more effective for 3-hop dragonfly due to closer congestion information
- ☐ **High performance multicast and reductions**



Rosetta Switch

25.6 Tb/s per switch, from 64 - 200 Gbs ports (25GB/s per direction)

13

Рисунок 7 - Обзор высокопроизводительного широкополосного коммутатора Rosetta

Примеры использования

Aurora предназначен для решения широкого спектра задач, от моделирования климата и исследования энергетических систем до расшифровки

генетического кода и анализа материалов / веществ посредством спектрального анализа.

Заключение

Выводы по анализу

Суперкомпьютер Aurora представляет собой пик современной вычислительной инженерии. Анализ архитектуры и коммуникационных сетей показывает, что Aurora разработан с учетом не только текущих вычислительных задач, но и будущих инноваций в области искусственного интеллекта, больших данных и машинного обучения. Топология Dragonfly и сетевая технология Slingshot обеспечивают этой системе высокую пропускную способность, низкую задержку и адаптивность, которые критически важны для современных научных и инженерных вычислений.

Перспективы развития

Способность системы к масштабированию и её модульная структура делают её идеальной платформой для развития и внедрения новых вычислительных технологий. В долгосрочной перспективе ожидается, что Аура и подобные ей системы будут играть центральную роль в решении глобальных проблем, таких как изменение климата, управление энергетическими ресурсами, здравоохранение (разработка новых вакцин / препаратов, предварительное моделирование их взаимодействия с клетками человека) и обеспечение национальной безопасности (прежде всего государства-разработчика и плейсхолдера данной системы).

Заключительные замечания

Данная работа подчеркивает значимость продолжающегося прогресса в области вычислительных систем и сетей, важность инвестиций в научные исследования и разработки, которые способствуют созданию мощных вычислительных инструментов. С появлением первых эксафлопсных вычислений, представленных системами вроде Auroga, научное сообщество стоит на пороге новых открытий, которые могут кардинально изменить мир.