

ТЕХНИЧЕСКОЕ ОПИСАНИЕ

Keysight AI Data Center Builder

Оценка и бенчмаркинг инфраструктуры AI / ML

Введение

Бенчмаркинг фабрики кластеров искусственного интеллекта (AI) и машинного обучения (ML) с реалистичными рабочими нагрузками обычно требует значительных инвестиций в вычислительные системы, такие как графические процессоры (GPU) и сетевые интерфейсные карты с удалённым прямым доступом к памяти (RDMA NIC). Эти системы не только дороги, но и требуют значительного времени на создание и эксплуатацию. Проблема заключается не только в финансовых затратах, но и в сложности обслуживания столь развитой инфраструктуры.

В этом техническом обзоре показано, как Keysight AI (KAI) Data Center Builder решает эти задачи, ускоряя циклы инноваций в сетях AI / ML и системах ввода-вывода. Программное решение использует высокоплотные генераторы нагрузки с портами 100–800GE для эмуляции AI-хостов и реалистичных шаблонов трафика AI / ML. KAI Data Center Builder помогает ускорить выпуск решений, снизить затраты, упростить настройку и обслуживание и повысить общую производительность.

Задачи

Для бенчмаркинга фабрики AI- и ML-кластеров с реалистичными рабочими нагрузками требуется комплексное решение, снижающее как стоимость, так и сложность. Ключевые компоненты такого решения:

- **Реалистичная эмуляция:** воспроизведение крупномасштабных AI-нагрузок с измеримой точностью и глубоким анализом производительности коллективных коммуникаций.
- **Упрощённый бенчмаркинг:** готовые приложения бенчмаркинга для более простой проверки AI-сетевой фабрики.
- **Гибкие тестовые движки:** возможность выбирать аппаратные генераторы нагрузки, программные конечные точки или реальные AI-ускорители и сравнивать результаты бенчмаркинга.
- **Автоматизированные методики тестирования:** квалификация эффективности AI-сетевой фабрики по времени завершения задания, изоляции производительности, балансировке нагрузки и механизмам управления перегрузкой.

Решение: KAI Data Center Builder

Keysight AI Data Center Builder (KAI DC Builder) — решение для оценки и бенчмаркинга, оптимизирующее проектирование систем AI / ML. Оно быстрее развёртывается и эксплуатируется, обеспечивает более глубокий анализ и более высокую точность измерений для каждого потока. На рисунке 1 показаны различные уровни центра обработки данных, на которых KAI DC Builder воспроизводит поведение AI-нагрузки.

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

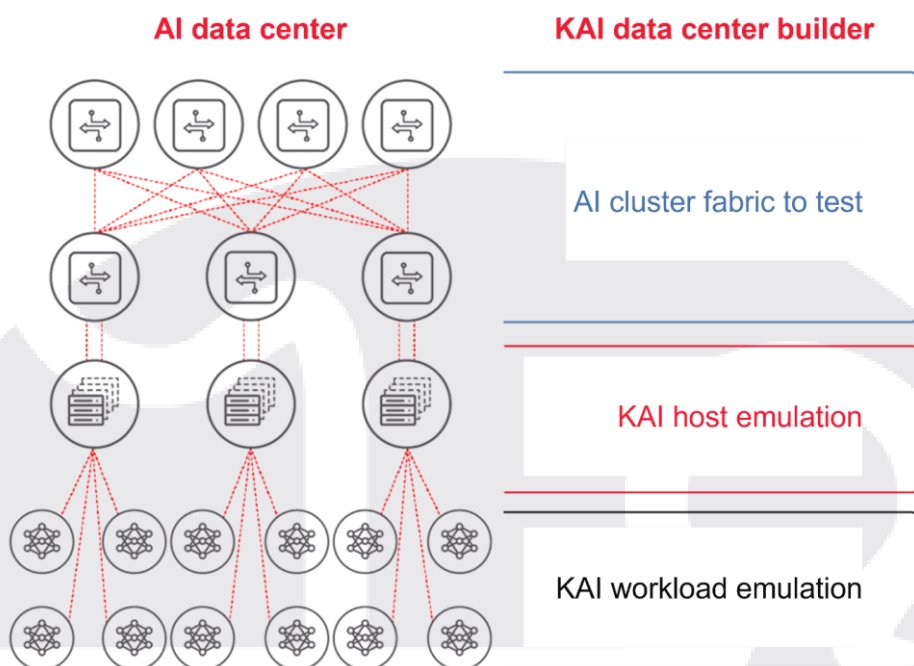


Рисунок 1. KAI DC Builder работает на нескольких уровнях, воспроизводя поведение AI-нагрузки

KAI DC Builder позволяет размещать и запускать приложения, специализирующиеся на измерении или анализе различных аспектов AI-инфраструктуры. Это даёт подробное представление о производительности, помогает выявлять узкие места и оптимизировать общую эффективность AI-систем.

Роль приложений заключается в сборе набора данных путём выполнения конкретного испытания. Например, одно приложение может измерять пропускную способность передачи данных, а другое — анализировать использование фабрики. Каждое приложение определяет цель испытания, инициализирует его собственной конфигурацией и формирует результаты на основе собранного набора данных. С приложениями KAI DC Builder можно работать через веб-интерфейс в визуальном интерактивном режиме или использовать сценарии автоматизации для пакетного выполнения из автоматизированного конвейера непрерывной интеграции.

KAI DC Builder включает два приложения, существенно повышающих производительность и эффективность AI-инфраструктуры:

KAI Collective Benchmarks выполняет микробенчмаркинг производительности алгоритмов распределённых коммуникаций, обычно используемых в масштабируемых AI-системах и называемых коллективными коммуникациями. Типичная цель приложения Collective Benchmarks — проверить, что сеть обеспечивает оптимальную и стабильную пропускную способность для диапазона обменов данными, характерных для распределённых AI-задач.

KAI Workload Emulation воспроизводит короткие последовательности шагов, через которые реальные AI-нагрузки — например, до или после обучения — проходят тысячи и миллионы итераций, называемых эпохами. Эти нагрузки представляют собой сочетание вычислительных и

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

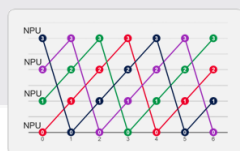
Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

коммуникационных шагов, главным образом определяемых схемой разделения AI-модели для распределения задачи между сотнями или тысячами GPU. Приложение Workload Emulation позволяет экспериментировать с производительностью выбранных схем разделения модели на различных сетевых топологиях и находить оптимальную комбинацию, уменьшающую время передачи данных и отдающую приоритет вычислениям на GPU, что в итоге сокращает Job Completion Time (JCT).

Комбинация этих двух приложений формирует более целостную и автоматизированную тестовую среду, снижая сложность и время настройки и анализа. Пользователи получают более точные данные и могут эффективнее оптимизировать AI-системы.

KAI Applications

Collective benchmarks

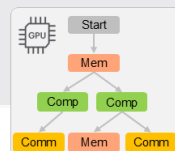


Network utilization

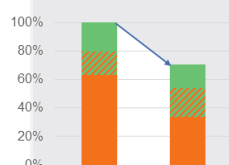


■ Measured ■ Idle

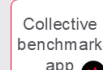
Workload emulation



Training time



Emulated AI hosts



Collective benchmark app



Workload emulation app

Рисунок 2. Приложения KAI Data Center Builder

Тестовые движки KAI

Приложения KAI могут выполнять испытания с использованием одного из тестовых движков KAI:

Аппаратное обеспечение Keysight AresONE управляет высокоплотной нагрузкой трафика с эмуляцией [RoCEv2](#) для точного воспроизведения шаблонов коммуникаций AI/ML в крупном масштабе через тестируемую фабрику. [RoCEv2](#) — транспортный протокол Remote Direct Memory Access (RDMA) over Ethernet, используемый узлами AI-кластера для коллективных коммуникаций.

Программные конечные точки Keysight работают на серверах общего назначения с RDMA NIC, но без GPU, что позволяет включить сетевые карты и их конфигурацию в состав тестируемой системы.

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

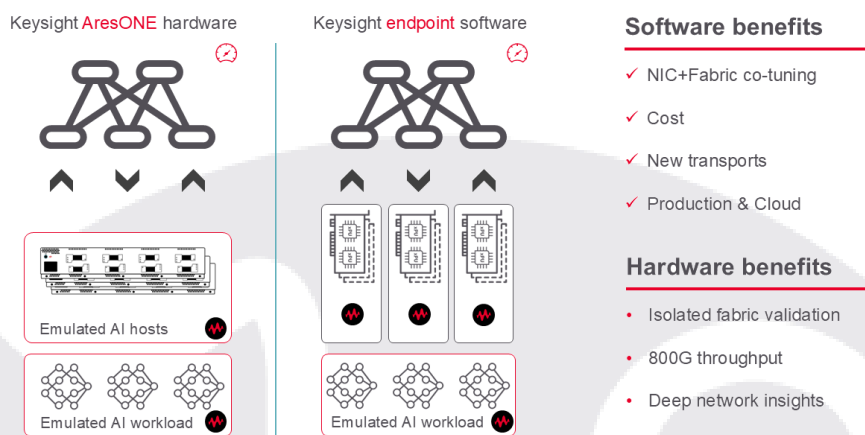


Рисунок 3. Варианты тестовых движков KAI — аппаратные платформы AresONE и программные конечные точки



AresONE-S 400GE



AresONE-M 800GE

KAI Collective Benchmarks

Приложение Keysight Collective Benchmark может выполнять микробенчмаркинг типичных алгоритмов AI-коммуникаций на предоставленной пользователем AI-сетевой фабрике, как показано на рисунке 4. Производительность фабрики можно измерять и улучшать её конструкцию в лабораторной или промежуточной среде без подключения AI-вычислительных узлов с GPU-ускорителями к тестируемой фабрике — с помощью эмуляции трафика [RoCEv2](#) на аппаратуре AresONE.

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

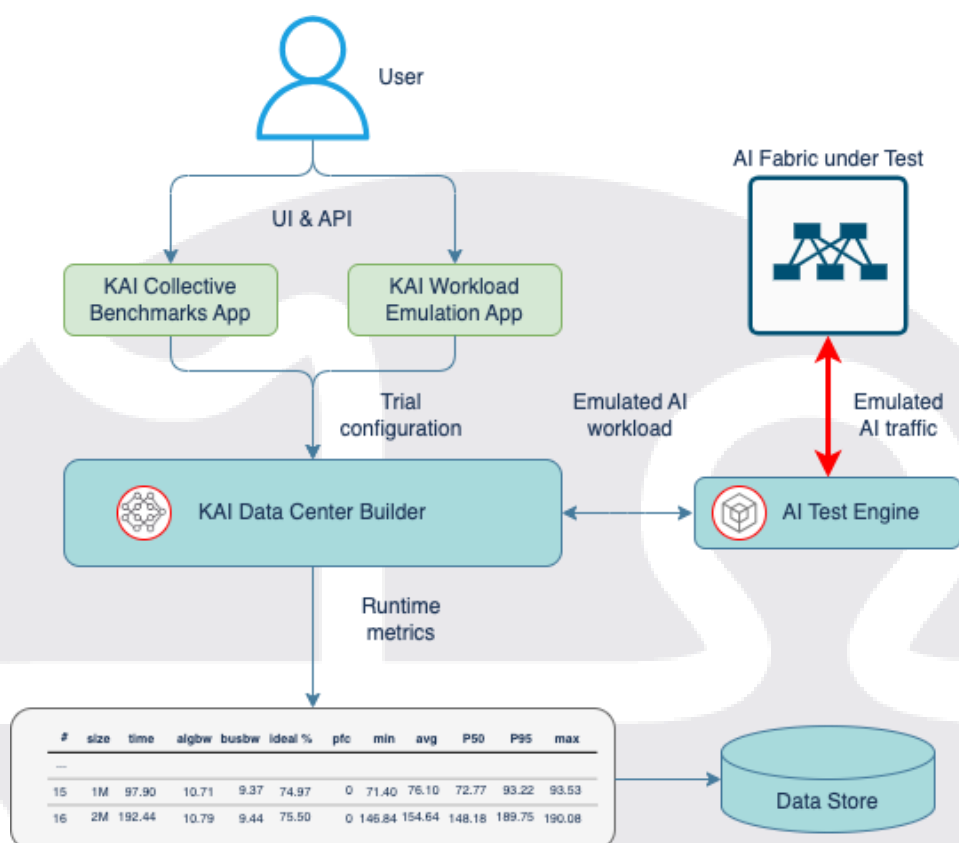


Рисунок 4. Рабочий процесс приложения KAI Collective Benchmarks с аппаратурой AresONE

Параметры конфигурации

Конфигурацию испытания KAI Collective Benchmarks можно параметризовать в соответствии с размером и свойствами AI-коллектива, на котором выполняется бенчмарк, а также профилем AI-нагрузки, эмулируемой генераторами трафика.

Параметр	Значения	Описание
Конфигурация логического кластера	- Количество GPU - GPU на вычислительный узел - Межсоединение NPU	Описывает конфигурацию AI-системы для эмуляции рабочей нагрузки
Коллективная операция: тип	- AlltoAll - AllReduce - AllGather - ReduceScatter - Gather - Broadcast	Тип коллективной операции

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

Коллективная операция: алгоритм	• Parallel • Unidirectional Ring • Bidirectional Ring • Halving-Double	Алгоритмы коллективной операции; применимость алгоритмов зависит от типа операции
Предустановка рабочей нагрузки: размер данных	• Начальное значение • Шаг итерации • Конечное значение	Диапазон размеров данных для выполнения бенчмаркинга
Предустановка рабочей нагрузки: транспортный профиль	• Транспортный протокол • Количество Q-Pairs на пару рангов • Размер RDMA-сообщения	Характеристики транспортного профиля; используются для тонкой настройки на уровне Q-Pair
Предустановка инфраструктуры: профиль NIC	• Скорость интерфейса NIC • MTU NIC • Настройки IP-интерфейса NIC • Traffic class / значение DSCP • Параметры PFC • Параметры DCQCN	Настройка управления перегрузкой и классификации трафика для эмулируемых NIC

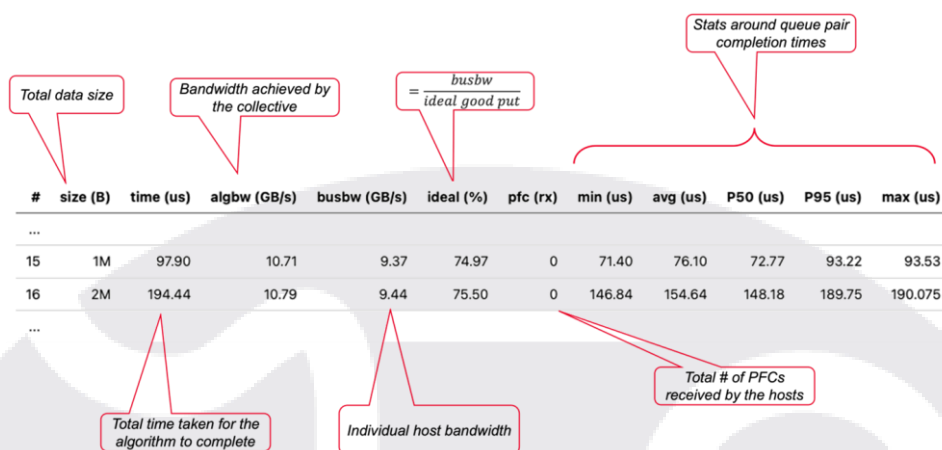
Метрики измерений

Приложение KAI Collective Benchmarks формирует результаты измерений в формате, широко используемом специалистами по AI. На основе параметров AI-коллектива и рабочей нагрузки оно собирает набор метрик и представляет их в табличном виде. Каждая строка содержит измерения для конкретного размера данных, которым должен обмениваться коллектив. Как показано на рисунке 5, при последовательном переборе настраиваемого диапазона размеров данных приложение оценивает производительность AI-фабрики по комбинации ключевых показателей, характерных для коллективных коммуникаций — времени завершения, пропускной способности алгоритма и шины — а также предоставляет дополнительную статистику распределения по отдельным очередям [RoCEv2](#) Q-Pair.

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

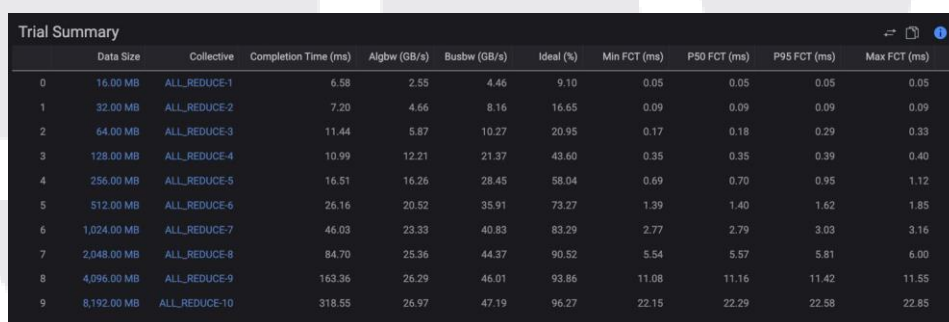


#	size (B)	time (us)	algbw (GB/s)	busbw (GB/s)	ideal (%)	pfc (rx)	min (us)	avg (us)	P50 (us)	P95 (us)	max (us)
...											
15	1M	97.90	10.71	9.37	74.97	0	71.40	76.10	72.77	93.22	93.53
16	2M	194.44	10.79	9.44	75.50	0	146.84	154.64	148.18	189.75	190.075
...											

Callouts:

- Total data size (points to size (B))
- Bandwidth achieved by the collective (points to algbw (GB/s))
- $\frac{busbw}{ideal\ good\ put}$ (points to ideal (%))
- Stats around queue pair completion times (points to min (us), avg (us), P50 (us), P95 (us), max (us))
- Total time taken for the algorithm to complete (points to time (us))
- Individual host bandwidth (points to busbw (GB/s))
- Total # of PFCs received by the hosts (points to pfc (rx))

Рисунок 5. Пример метрик измерения



	Data Size	Collective	Completion Time (ms)	Algbw (GB/s)	Busbw (GB/s)	Ideal (%)	Min FCT (ms)	P50 FCT (ms)	P95 FCT (ms)	Max FCT (ms)
0	16.00 MB	ALL_REDUCE-1	6.58	2.55	4.46	9.10	0.05	0.05	0.05	0.05
1	32.00 MB	ALL_REDUCE-2	7.20	4.66	8.16	16.65	0.09	0.09	0.09	0.09
2	64.00 MB	ALL_REDUCE-3	11.44	5.87	10.27	20.95	0.17	0.18	0.29	0.33
3	128.00 MB	ALL_REDUCE-4	10.99	12.21	21.37	43.60	0.35	0.35	0.39	0.40
4	256.00 MB	ALL_REDUCE-5	16.51	16.26	28.45	58.04	0.69	0.70	0.95	1.12
5	512.00 MB	ALL_REDUCE-6	26.16	20.52	35.91	73.27	1.39	1.40	1.62	1.85
6	1,024.00 MB	ALL_REDUCE-7	46.03	23.33	40.83	83.29	2.77	2.79	3.03	3.16
7	2,048.00 MB	ALL_REDUCE-8	84.70	25.36	44.37	90.52	5.54	5.57	5.81	6.00
8	4,096.00 MB	ALL_REDUCE-9	163.36	26.29	46.01	93.86	11.08	11.16	11.42	11.55
9	8,192.00 MB	ALL_REDUCE-10	318.55	26.97	47.19	96.27	22.15	22.29	22.58	22.85

Рисунок 6. Пример сводного отчёта испытания KAI Collective Benchmarks

KAI Workload Emulation

Операторы AI используют различные стратегии параллелизма, также называемые разделением модели, чтобы ускорить обучение AI-моделей. Практические результаты показывают, что согласование схемы разделения с топологией и конфигурацией AI-кластера обеспечивает выдающуюся производительность обучения. На этапе проектирования AI-кластера на критические вопросы лучше всего отвечать экспериментально. Многие из них связаны с эффективностью перемещения данных между графическими процессорами (GPU), например:

- Проектирование scale-up межсоединений GPU внутри AI-хоста или стойки.
- Проектирование scale-out сети, включая полосу пропускания на GPU и топологию.
- Настройка балансировки сетевой нагрузки и управления перегрузкой.
- Настройка параметров среды обучения для оптимизации производительности и эффективности.

Приложение KAI Workload Emulation воспроизводит шаблоны сетевых коммуникаций реальных задач обучения AI. Его цели — ускорить эксперименты, сократить время освоения и предоставить более глубокое понимание причин наблюдаемого ухудшения производительности. Такие сведения было бы значительно сложнее получить при использовании реального задания обучения AI.

Запуск KAI Workload Emulation в KAI DC Builder позволяет:

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

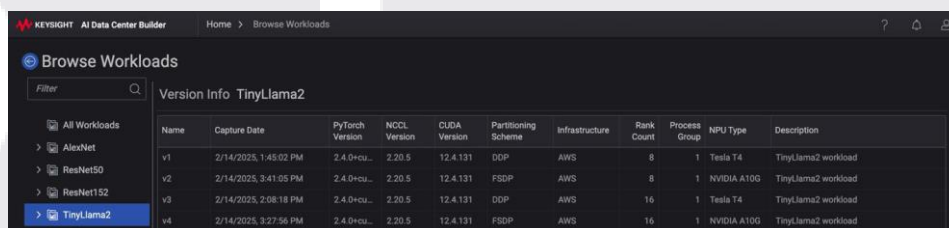
- Экспериментировать с параметрами разделения модели и планированием рангов в AI-инфраструктуре.
- Понимать вклад коммуникаций внутри разделов и между ними в общее время завершения задания (JCT).
- Выявлять типы коллективных операций с низкой производительностью и выполнять детальный анализ для поиска узких мест.
- Анализировать использование сети, tail latency, перегрузку и её влияние на JCT.

Библиотека рабочих нагрузок

KAI DC Builder включает библиотеку AI-нагрузок — от классификации изображений до больших языковых моделей (LLM), таких как Llama. Она содержит популярные схемы разделения модели, включая data parallel (DP), fully sharded data parallel (FSDP) и трёхмерный (3D) параллелизм.

Собственные трассы выполнения рабочих нагрузок можно импортировать в библиотеку в формате MLCommons Chakra.

На рисунке 7 показаны рабочие нагрузки в библиотеке, подготовленной Keysight, с подробными метаданными среды и ключевых гиперпараметров, использованных при записи трасс выполнения. Библиотека позволяет исследовать поведение нагрузок и понимать типы и временные характеристики коллективных операций, а также вычислений.



Name	Capture Date	PyTorch Version	NCCL Version	CUDA Version	Partitioning Scheme	Infrastructure	Rank Count	Process Group	NPU Type	Description
v1	2/14/2025, 1:45:02 PM	2.4.0rcu...	2.20.5	12.4.131	DDP	AWS	8	1	Tesla T4	TinyLlama2 workload
v2	2/14/2025, 3:41:05 PM	2.4.0rcu...	2.20.5	12.4.131	FSDP	AWS	8	1	NVIDIA A10G	TinyLlama2 workload
v3	2/14/2025, 2:08:18 PM	2.4.0rcu...	2.20.5	12.4.131	DDP	AWS	16	1	Tesla T4	TinyLlama2 workload
v4	2/14/2025, 3:27:56 PM	2.4.0rcu...	2.20.5	12.4.131	FSDP	AWS	16	1	NVIDIA A10G	TinyLlama2 workload

Рисунок 7. Библиотека рабочих нагрузок KAI Data Center Builder

Каждая рабочая нагрузка может содержать трассы, собранные с каждого отдельного ранга (NPU), фиксирующие шаги, которые задание выполняло на этом вычислительном узле, как показано на рисунке 8. За исключением простейших случаев, таких как DDP, поведение разных рангов может различаться в зависимости от их положения в разделе.

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

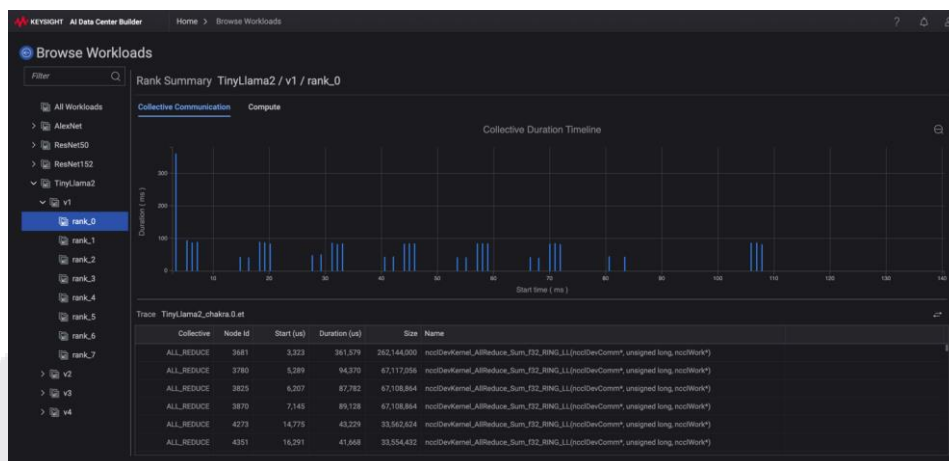


Рисунок 8. Подробности рабочей нагрузки для временной шкалы длительности коллективных операций

Параметры конфигурации

После выбора трасс рабочих нагрузок из библиотеки пользователь назначает их набору эмулируемых NPU (рангов) для воспроизведения. Как и в приложении KAI Collective Benchmarking, конфигурация содержит раздел, определяющий логическую структуру кластера: количество AI-хостов, количество NPU на хост и наличие межсоединений NPU. В тестовом стенде должно быть достаточно ресурсов эмуляции — аппаратных тестовых портов или RDMA NIC — для поддержки количества элементов, заданного в разделе логической инфраструктуры, как показано на рисунках 9, 10 и 11.

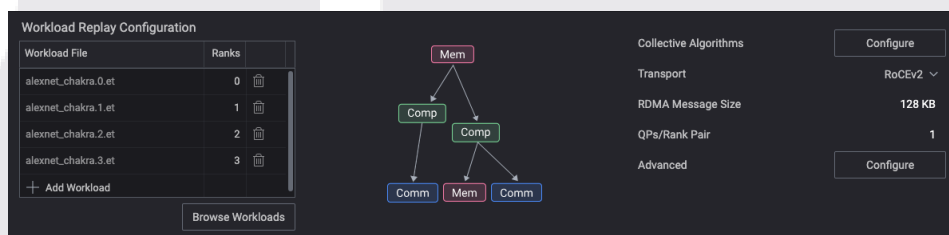


Рисунок 9. Планирование рабочих нагрузок

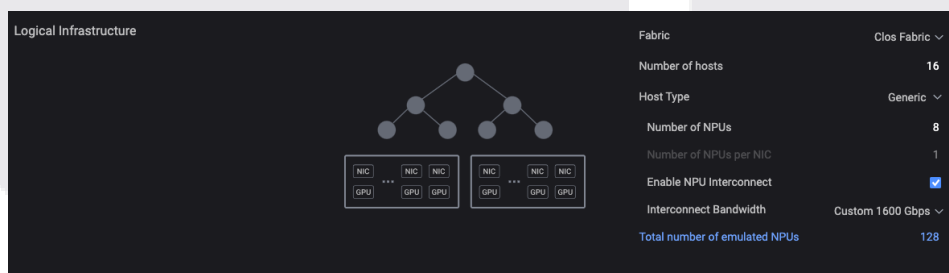


Рисунок 10. Моделирование AI-инфраструктуры: восемь стоек, 16 AI-хостов и восемь GPU

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

Collectives		
Broadcast	Broadcast Parallel	▼
All to All	All To All Parallel	▼
All Reduce	All Reduce Unidir Ring	▼
All Gather	All Gather Unidir Ring	▼
Reduce Scatter	Reduce Scatter Unidir Ring	▼
Gather	Gather Parallel	▼

Рисунок 11. Параметризация коллективных операций

Метрики измерений

После выполнения испытания Workload Emulation приложение сообщает общее время завершения для каждой трассы рабочей нагрузки, суммируя время по всем рангам, на которых она была запланирована. Кроме того, оно показывает производительность каждой коллективной операции для всех нагрузок в испытании, включая характеристики операции, время завершения, относительное время начала и сведения о разделе модели, как показано в таблице 1.

Таблица 1. Сводка измерений эмуляции рабочей нагрузки

#	Workload ID	Collective ID	Размер данных (GB)	CCT (с)	Ранг и	Старт (с)	PP Group	TP Group	DP Group
1	LLM_3DP_1	AllGather_1	0.125	0.32	8	3.56	-	1	-
2	LLM_3DP_1	AllGather_2	0.125	0.29	8	4.12	-	2	-
3	LLM_3DP_1	AllReduce_1	2.0	1.34	16	5.43	-	-	1

Итоги

Keysight AI Data Center Builder сочетает высокоточную эмуляцию AI-нагрузок, готовые приложения бенчмаркинга и инструменты анализа наборов данных, существенно повышая эффективность оценки производительности сетевой фабрики AI / ML-кластера.

Для ускорения проектирования сетей AI / ML Keysight AI Data Center Builder:

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

- **Эмулирует крупномасштабные AI-нагрузки с измеримой точностью** — предоставляет глубокое понимание производительности коллективных коммуникаций.
- **Упрощает процесс бенчмаркинга** — позволяет проверять AI-сетевую фабрику с помощью готовых приложений бенчмаркинга, созданных в сотрудничестве с крупнейшими AI-операторами и поставщиками AI-инфраструктуры.
- **Выполняет определённые модели поведения AI/ML** — позволяет обмениваться ими между пользователями и заказчиками для воспроизведения экспериментов.
- **Предлагает выбор тестовых движков** — между эмуляцией конечных точек [RoCEv2](#) на высокоплотных генераторах нагрузки AresONE и программными конечными точками на реальных AI-ускорителях для сравнения результатов бенчмаркинга.

KAI DC Builder обеспечивает крупномасштабную проверку и эксперименты с проектированием фабрики с использованием тестового оборудования [AresONE-S 400GE](#) и [AresONE-M 800GE](#) реалистично и экономически эффективно. Решение дополняет использование GPU для тестирования нагрузок AI / ML, позволяя AI-операторам сократить расходы, которые в противном случае полностью направлялись бы на GPU-системы бенчмаркинга, и использовать более масштабируемый, надёжный и интегрированный Keysight AI Data Center Builder.

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.