

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.

/Наумов А.А./

Проверил:

старший преподаватель

кафедры ВС

«__» декабря 2023 г.

/Ревун А.Л./

Оценка «_____»

Новосибирск 2023

Задание

Анализ возможностей средств отказоустойчивого выполнения параллельных программ на современных суперВС (№ 1-10 в списке top500.org, № 1-50 в списке top50.supercomputers.ru)

СуперВС для анализа: SUNWAY TAIHULIGHT - SUNWAY MPP, SUNWAY SW26010 260C 1.45GHZ, SUNWAY

Операторы: Национальный суперкомпьютерный центр.

Локация: Уси, провинция Цзянсу (Китай)

Введение

National Supercomputing Center in Wuxi представляет собой выдающийся центр современных вычислений, являющийся домом для некоторых из самых мощных суперкомпьютеров в мире. В свете стремительного развития технологий и постоянно увеличивающихся требований к вычислительным ресурсам, вопрос об отказоустойчивости и эффективности выполнения параллельных программ на этих масштабных системах становится более актуальным.

Провести анализ возможностей средств отказоустойчивого выполнения параллельных программ на современных суперкомпьютерах National Supercomputing Center in Wuxi.

Основные характеристики:

Sunway TaihuLight (кит. trad. 神威·太湖之光, пиньинь Shénwēi tàihú zhī guāng) — китайский суперкомпьютер, который с июня 2016 по июнь 2018 года являлся самым производительным суперкомпьютером в мире с производительностью 93 петафлопса, согласно тестам LINPACK. Такая скорость вычислений более чем в 2,5 раза выше по сравнению с предыдущим мировым рекордсменом Тяньхэ-2, у которого вычислительная мощность составляет почти 34 петафлопса.

Пиковая производительность Sunway TaihuLight — 125,43 петафлопса против 54,9 петафлопса у Тяньхэ-2.

Кроме того, по состоянию на июнь 2016 года Sunway TaihuLight занимал третье место в списке Green500 наиболее энергоэффективных суперкомпьютеров с результатом 6 гигафлопсов/ватт (на тесте Linpack).

Стоимость данной системы составляет примерно 1,8 миллиарда юаней (270 миллионов долларов США), финансирование предоставлено в равных долях центральным правительством, провинцией Цзянсу и городом Уси.

Таблица 1. Основные параметры суперкомпьютера Sunway TaihuLight

Вершина производительность	125 Пфлопс	
Линпакпроизводительность	93 Пфлопс	
Процессорчастота	1,45 ГГц	
Пиковая производительностьЦП	3.06 Тфлопс	
Общая Память	1310,72 ТБ	
Общая пропускная способность памяти	5591,5ТБ/с	
Сетевая ссылкапропускная способность	16 ГБ/с	
Сеть пропускная способность	70 ТБ/с	
Общее хранилище	20 ПБ	
Всего ввода-выводапропускная способность	288	ГБ/с
Энергопотребление при работе Linpackтест	15.371МВт	
Мощность соотношение	6,05Гфлопс/Вт	

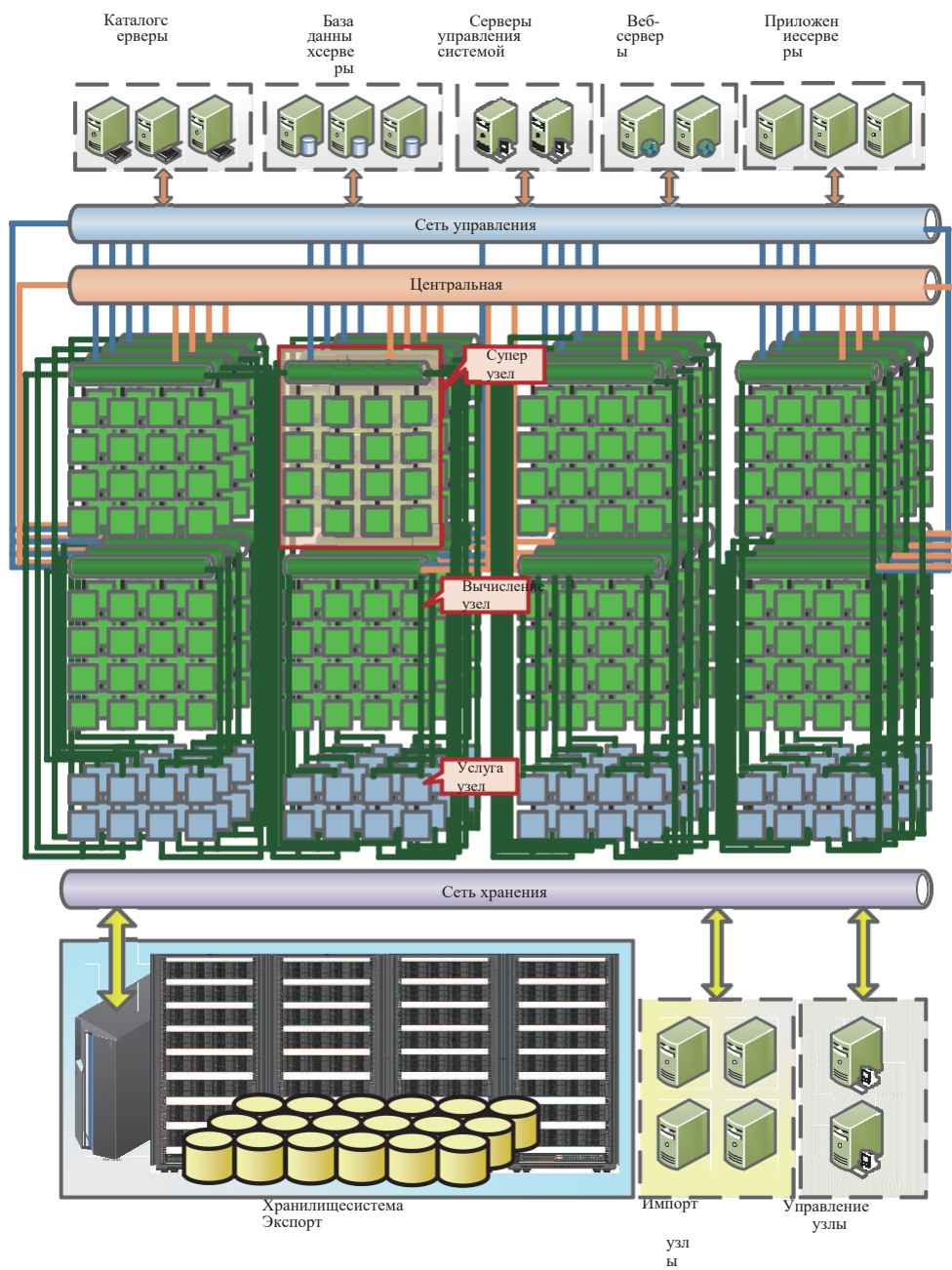


Рисунок 1. Общая архитектура системы Sunway TaihuLight.

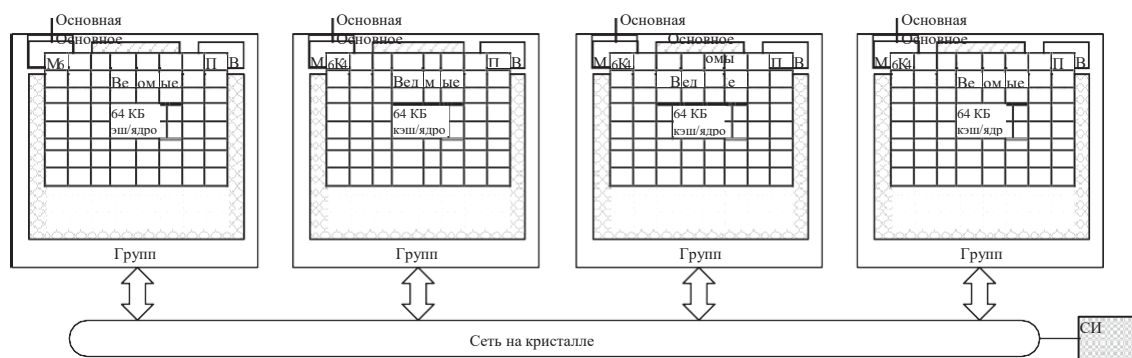


Рисунок 2. Общая архитектура нового процессора Sunway.

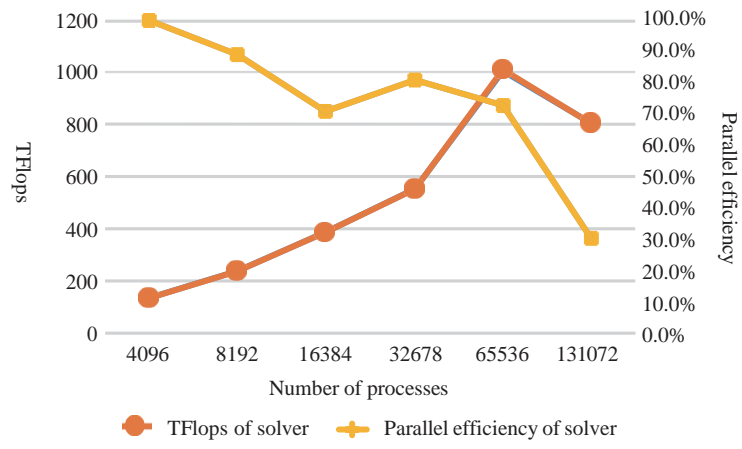


Figure 5 Strong scaling results of the non-hydrostatic dynamic solver on the Sunway TaihuLight supercomputer.

Table 3 Hardware, case configurations, and time to solutions (solver part, and the entire program) for strong-scaling tests

Parallelism	4096	8192	16384	32768	65536	131072
Process	NX	128	256	256	512	1024
	NY	32	32	64	64	128
NZ = 1						
Sockets	1024	2048	4096	8192	16384	32768
Cabinets	1	2	4	8	16	32
Total cells	$16384 \times 2048 \times 64 = 2.15 \text{ B}$					
Total unknowns	12.88 B					
Total solver time (h)	4.744	2.665	1.675	0.588	0.407	0.493
Total time (h)	4.771	2.679	1.682	0.591	0.409	0.494

the model and compare with reference results.

In the strong scaling tests, we fix the total problem size to be 2.15 billion mesh cells ($16384 \times 2048 \times 64$) and increase the number of computing nodes from 4096 processes to 131072 MPI processes. The test results are provided in Figure 5. Detailed configurations of the hardware resources (CGs, processors, and cabinets) and the simulation case (number of cells, number of unknowns, etc.) are shown in Table 3. From the figure, we observe that when we increase the number of processes from 4096 (1 cabinet) to 8192

(2 cabinets), the parallel efficiency of the solver part drops to 89%, due to the introduction of inter-cabinet communication. For the scale of 16384 (4 cabinets) to 65532 (16 cabinets), parallel efficiency persisted at around 70%, as the number of iterations remains almost the same for these different configurations. When the scale increases to 131072 processes (32 cabinets, 8.52 million cores in total), the efficiency drops to 30%, as the communication becomes the major overhead. The computational performance of the solver part changes from 130 TFlops to 806 TFlops (the highest performance, 1.01 PFlops, is achieved with the configuration of 65536 processes). Compared with the solver part, the performances for the entire program (including the I/O and initialization) are quite similar.

In the weak scaling tests, we observe better scalability when using 32 cabinets, with the sustained performance further improved to 1.5 PFlops.

As shown in Table 3, the time to solution of the solver part for the simulation of 1000 time steps reduces from 17080 s to 1775 s (the shortest time to solution, 1465 s, is achieved with the configuration of 65536 processes).

Архитектура и компоненты

Суперкомпьютер Sunway TaihuLight имеет уникальную архитектуру и конфигурацию, которая сделала его одним из самых мощных суперкомпьютеров в мире на момент его запуска в 2016 году. Вот основные характеристики его архитектуры и компонентов:

1. Процессоры (CPUs):

- Sunway SW26010: Этот процессор был специально разработан в Китае для Sunway TaihuLight. Он содержит 260 ядер и имеет высокую производительность в вычислениях с плавающей запятой. Каждый процессор имеет 4 SIMD (Single Instruction, Multiple Data) блока.

2. Сеть:

- Коммутатор: Sunway TaihuLight использует собственный коммутатор Sunway STN-A1000, разработанный специально для этой системы. Он обеспечивает высокую пропускную способность и низкую задержку для связи между узлами.

3. Память:

- Объем памяти: Суперкомпьютер имеет огромный объем оперативной памяти, что позволяет обрабатывать большие объемы данных и выполнять сложные вычисления.

- HBM (High Bandwidth Memory): Sunway TaihuLight также использует высокоскоростную память HBM для обработки данных с высокой пропускной способностью.

4. Система охлаждения:

- Водяное охлаждение: Для поддержания нормальной температуры в такой мощной системе Sunway TaihuLight использует водяное охлаждение.

5. Система хранения данных:

- Sunway TaihuLight также имеет большой объем хранилища данных для сохранения и обработки результатов вычислений.

6. Отказоустойчивость:

- Стратегии отказоустойчивости, вероятно, включают в себя чекпоинтинг и мониторинг состояния системы для быстрого обнаружения и реагирования на

ошибки.

7. Системное программное обеспечение:

- Sunway TaihuLight работает под управлением собственной операционной системы, разработанной в Китае, и использует специализированные библиотеки и средства разработки для параллельных вычислений.

Sunway TaihuLight был весьма инновационной системой, и его архитектура была разработана с учетом высокой производительности и эффективности при выполнении сложных параллельных программ. Он представляет собой яркий пример современной суперкомпьютерной архитектуры.

Средства и способы отказоустойчивого выполнения параллельных программ

Однако, в общем контексте суперкомпьютерных вычислений, можно выделить несколько ключевых методов и подходов к обеспечению отказоустойчивости, которые потенциально могут применяться в таких центрах.

1. Репликация данных и вычислений:

Этот метод включает создание копий данных и процессов на разных узлах суперкомпьютера. В случае сбоя одного узла, другой может продолжить выполнение задачи без потери данных или значительных задержек.

2. Чекпоинтинг и восстановление:

Процесс сохранения состояния программы в определенные моменты времени. Это позволяет восстановить выполнение программы с последнего сохранённого состояния в случае сбоя.

3. Изоляция и восстановление узлов:

Быстрое обнаружение и изоляция неисправных узлов, их замена или

восстановление без влияния на работу остальной части системы.

4. Программное обеспечение для управления отказами:

Использование специализированных программных решений для управления ресурсами суперкомпьютера, мониторинга состояния системы и оперативного реагирования на проблемы.

5. Алгоритмы отказоустойчивости в параллельных программах:

Разработка специальных алгоритмов, способных адаптироваться к изменениям в системе и поддерживать работу даже при частичных сбоях.

Эти методы и подходы являются стандартными в сфере высокопроизводительных вычислений и могут использоваться в различных суперкомпьютерных центрах, включая Национальный суперкомпьютерный центр в Уси. Однако, точной информации нет, нужно напрямую обращаться к источникам, связанным с центром.

Суперкомпьютеры часто имеют уникальные архитектуры и характеристики, но есть несколько известных систем, которые можно рассматривать как сравнимые с Sunway TaihuLight в некоторых аспектах, включая методы обеспечения отказоустойчивости. Например, суперкомпьютеры, такие как Summit и Sierra, разработанные IBM, могут предложить аналогичные подходы к отказоустойчивому выполнению параллельных программ.

1. Summit и Sierra:

Эти суперкомпьютеры, разработанные IBM и установленные в Oak Ridge National Laboratory (ORNL) и Lawrence Livermore National Laboratory (LLNL) соответственно, используют архитектуру, схожую с той, что используется в Sunway TaihuLight. Они предлагают расширенные возможности для отказоустойчивости и обработки больших объемов данных.

2. Методы отказоустойчивости:

- Чекпоинтинг и восстановление: Это стандартный подход в суперкомпьютерах, позволяющий сохранять состояние вычислительных задач в определенные моменты времени для последующего восстановления после сбоя.

- Распределенное хранение данных: Распределение данных по разным узлам системы уменьшает риск потери данных и повышает отказоустойчивость.

- Автоматическое обнаружение и устранение ошибок: Системы могут автоматически обнаруживать и корректировать ошибки, возникающие во время вычислений.

- Резервирование ресурсов: Некоторые суперкомпьютеры используют дополнительные ресурсы для обеспечения отказоустойчивости, включая резервные процессоры и память.

Эти методы являются общими для многих современных суперкомпьютеров и могут дать представление о том, какие подходы к отказоустойчивости могут использоваться в системах, аналогичных Sunway TaihuLight. Однако точной информации нет.

Проводя анализ статей о данном суперкомпьютере, можно прийти к выводу, что создатели не говорят о точных моделях отказоустойчивости вычислительной системы, хотя есть работы, которые показывают ошибки при работе суперкомпьютера, но не показывают способы их решения.

Список литературы

1.A Large-Scale Study of Failures on Petascale Supercomputers - Rui-Tao Liu¹ and Zuo-Ning Chen², Fellow, CCF

Received July 29, 2017; revised December 7, 2017.

2.A highly effective global surface wave numerical simulation with ultra-high resolution
Fangli Qiao, Wei Zhao, Xunqiang Yin, Xiaomeng Huang, Xin Liu, Qi Shu,Guansuo Wang, Zhenya Song, Xinfang Li, Haixing Liu, Guangwen Yang, Yeli Yuan
National Research Center of Parallel Computer Engineering and Technology,
Wuxi, 214083, P.R. China

3.Report on the Sunway TaihuLight System

University of Tennessee

Department of Electrical Engineering and Computer Science

Tech Report UT-EECS-16-742

4.The Sunway TaihuLight supercomputer:system and applications

Haohuan FU , Junfeng LIAO , Jinzhe YANG , Lanning WANG ,Zhenya SONG ,
Xiaomeng HUANG , Chao YANG , Wei XUE ,Fangfang LIU , Fangli QIAO , Wei
ZHAO , Xunqiang YIN , Chaofeng HOU, Chenglong ZHANG , Wei GE , Jian ZHANG ,
Yangang WANG ,Chunbo ZHOU & Guangwen YANG

Received May 27, 2016; accepted June 11, 2016; published online June 21, 2016