

高性能计算应用特征剖析



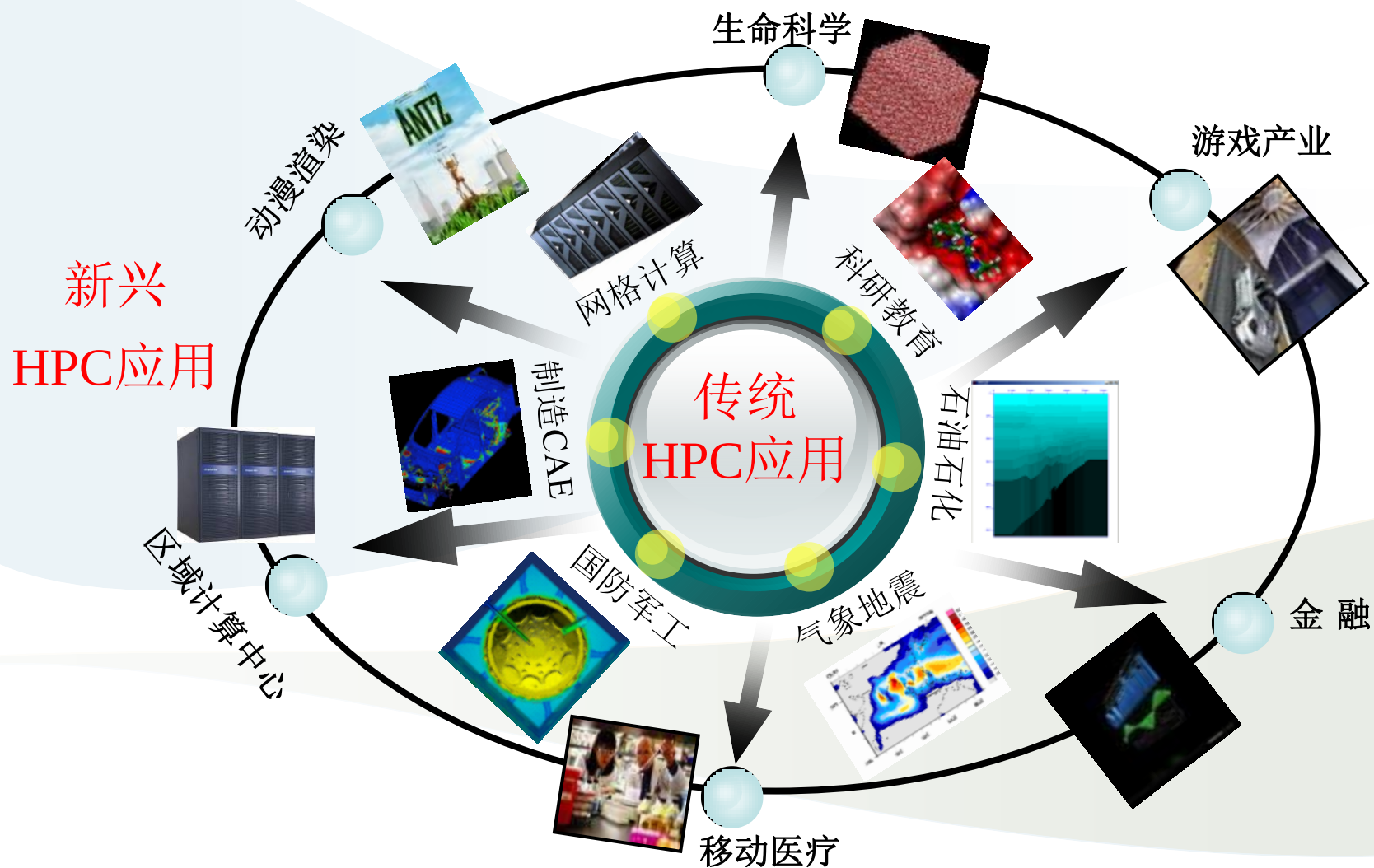
inspur 浪潮

刘性能服务器产品部

应用支持处

刘羽 博士

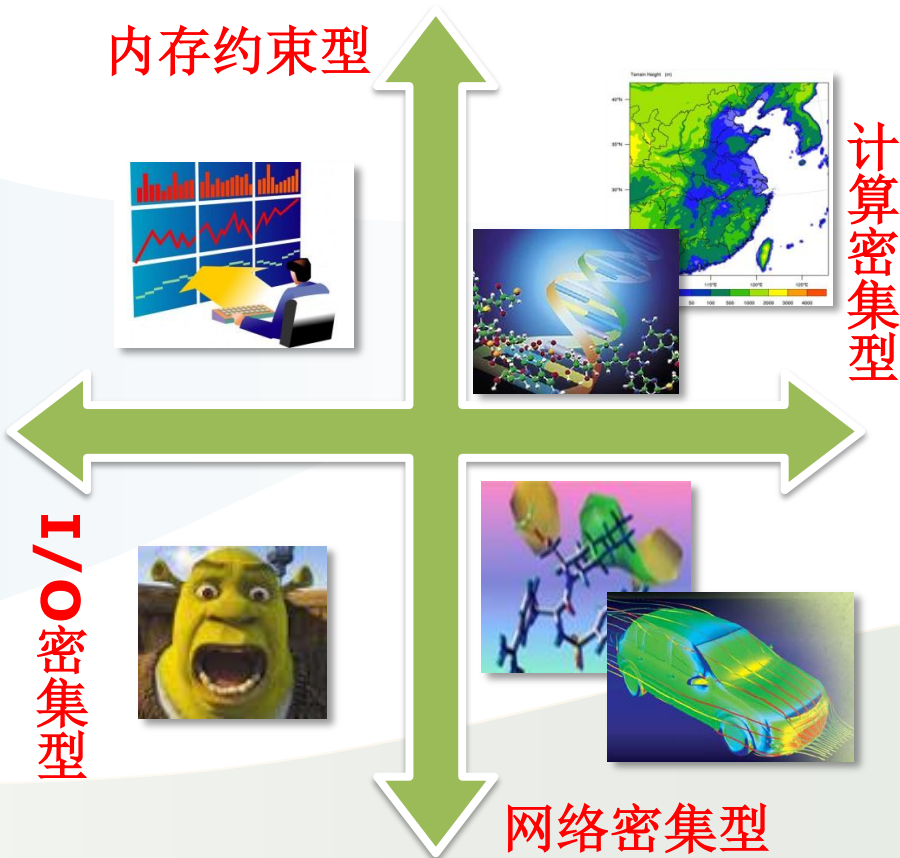
需求各异的高性能计算领域



浪潮HPC特征分析

HPC应用需求按硬件主要组成划分:

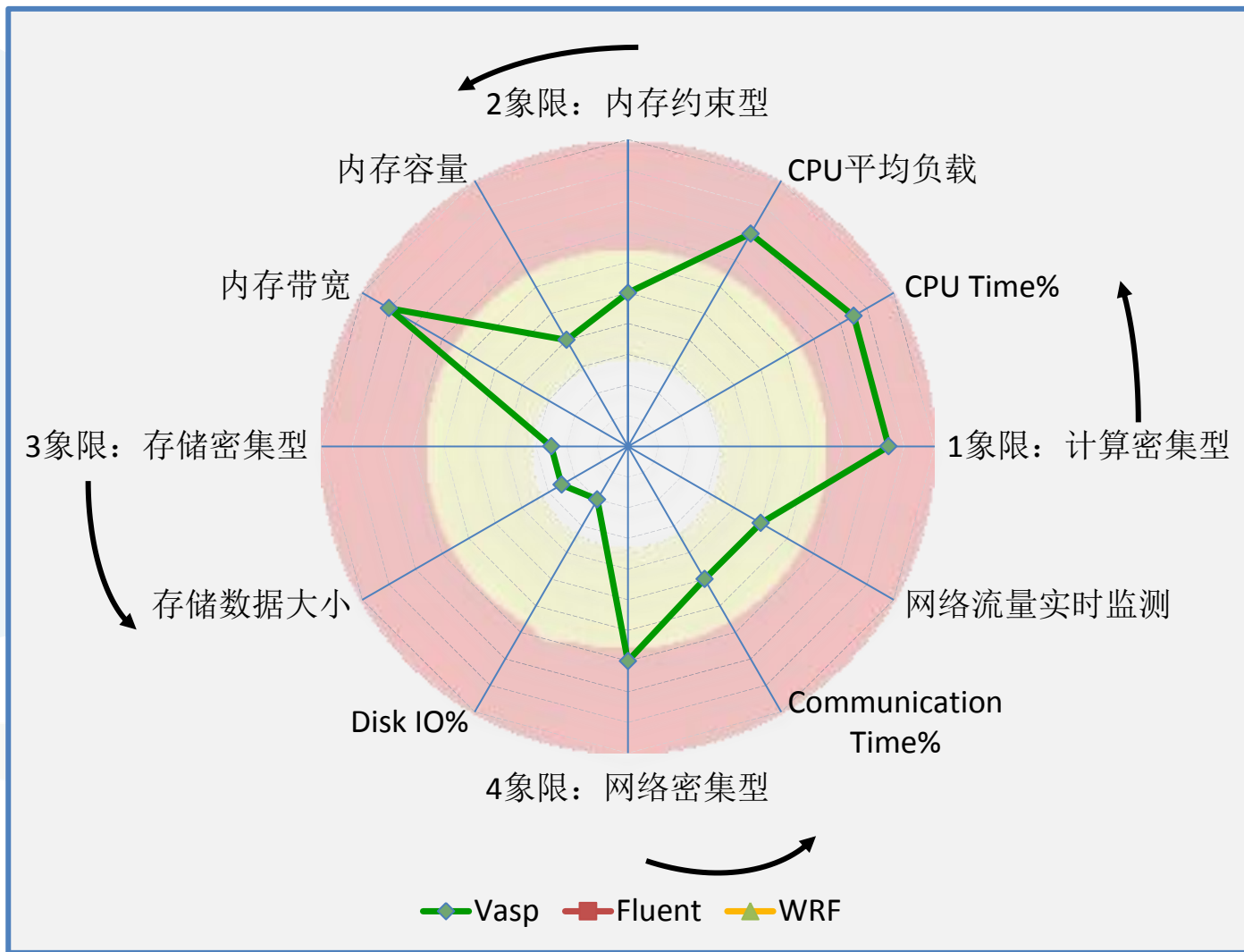
- CPU
- 内存
- 存储
- 互联



平衡多样的计算需求，构建**可扩展**的高性能计算系统。

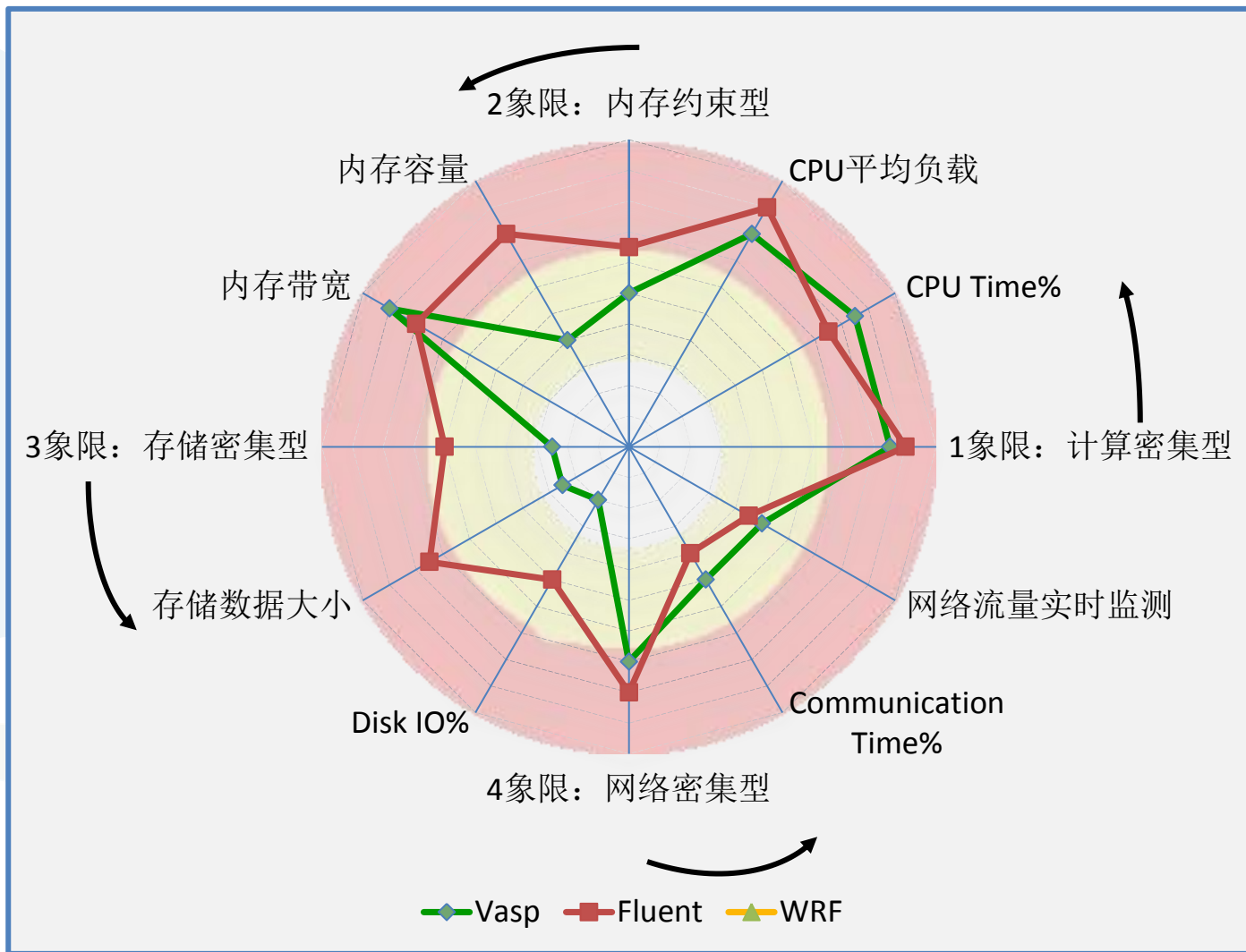
高性能行业应用分析

生命科学、
计算化学、
CAE、
数值气象等
行业应用
分析图。



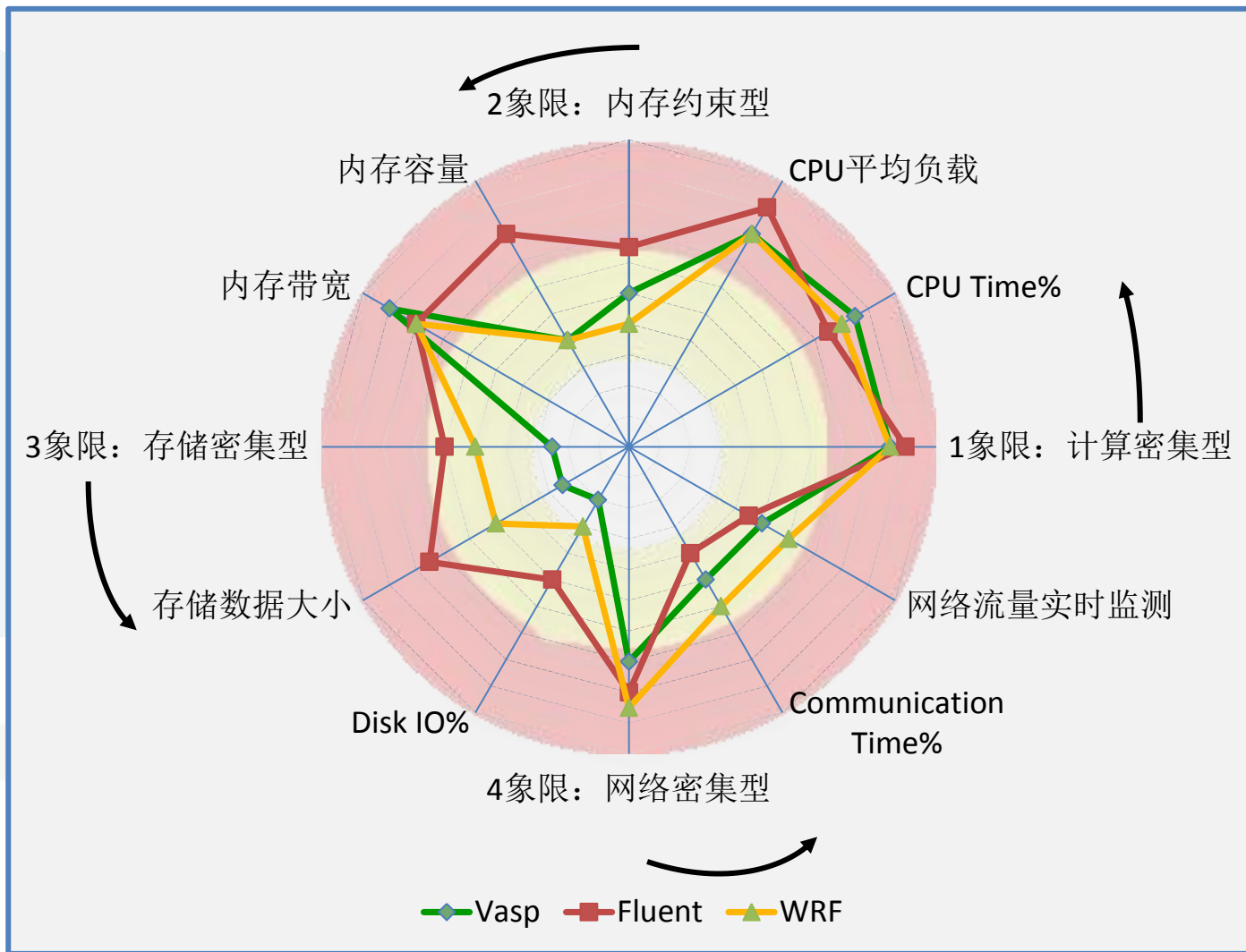
高性能行业应用分析

生命科学、
计算化学、
CAE、
数值气象等
行业应用
分析图。



高性能行业应用分析

生命科学、
计算化学、
CAE、
数值气象等
行业应用
分析图。



实例剖析

Fluent

Gaussian VASP

计算机辅助工程（CAE）概述

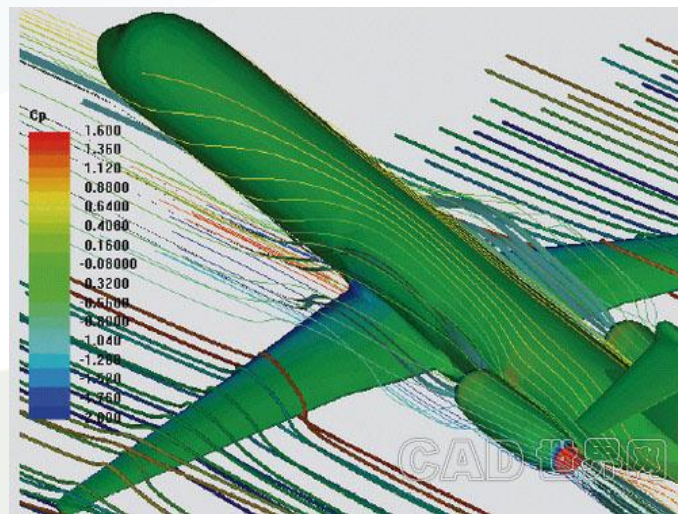
- 计算机辅助工程（**CAE**, **Computer Aided Engineering**）

就是用计算机辅助求解工程和产品的强度、刚度、屈曲稳定性、动力响应、热传导、弹塑性等力学性能以及性能的优化设计等方法。

- **CAE应用与高性能计算**

- 更快的速度, Performance
- 更高的精度, Accuracy
- 更大的规模, Scale

❓ 如何选配提供最佳应用性能的高性能计算平台



计算机辅助工程（CAE）概述

CAE处理流程：

◆前处理——

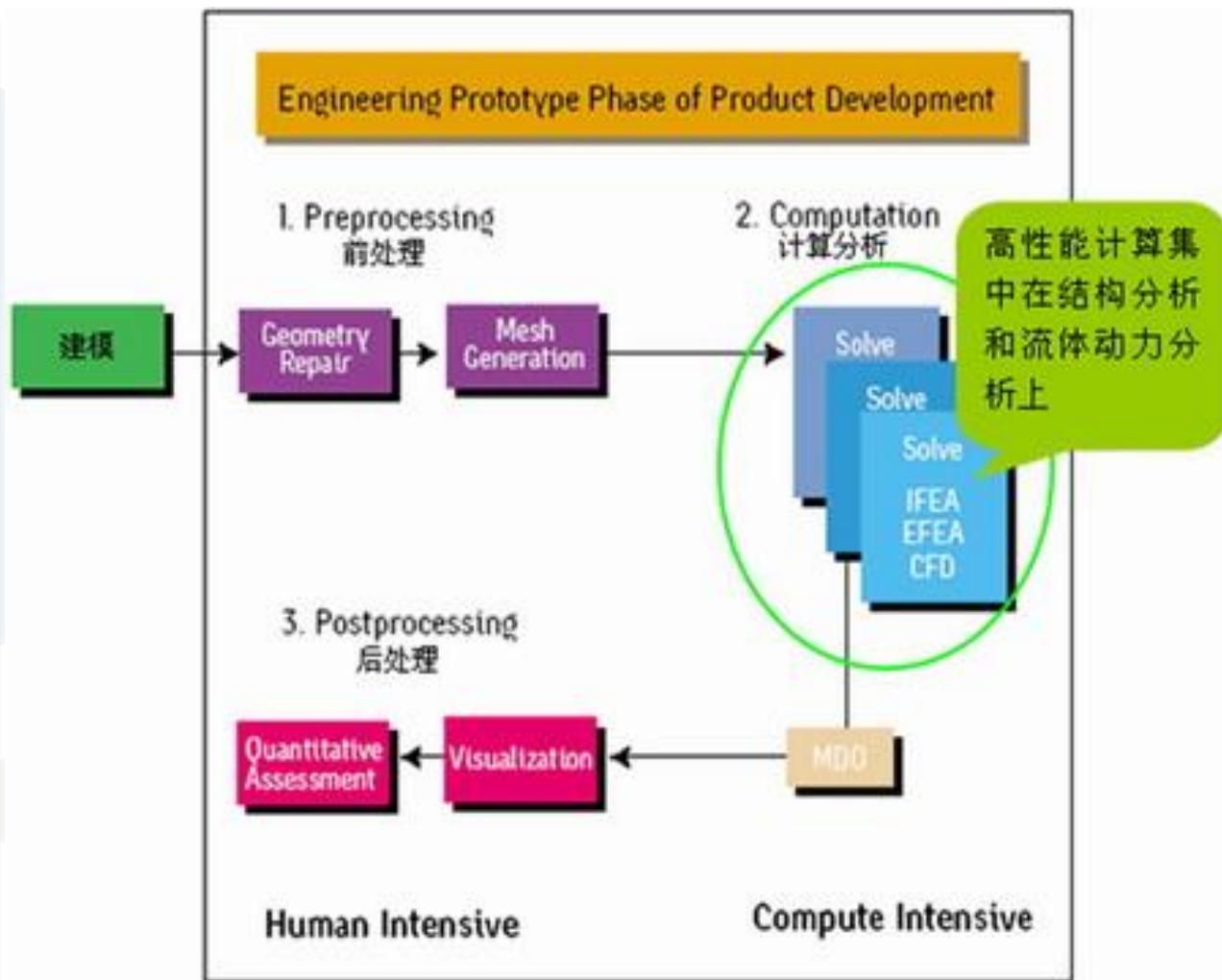
建立问题的几何模型、进行网格划分、建立用于计算分析的数值模型、确定模型的边界条件和初始条件等

◆求解计算——

是对所建立的数值模型进行求解，经常需要求解大型的线性方程组

◆后处理——

以图形化的方式对所得的计算结果进行检查



计算机辅助工程（CAE）概述

	前处理	计算	后处理
硬件配置需求	高端显卡 高分辨率显示器 高端处理器 内存容量 网络一般无特殊要求 IO一般无特殊要求	显卡无特殊要求 显示器无特殊要求 高端处理器 内存容量、内存主频 低延迟高带宽网络 IO可能成为瓶颈	高端显卡 高分辨率显示器 高端处理器 内存容量 网络一般无特殊要求 IO一般无特殊要求
主要使用的OS	Windows	Linux	Windows
人机交互	需要	一般不需要	需要
主要软件	GridGen,ICEM, HyperMesh	Ansys,Abaqus,Nast ran,Fluent, CFX, 以 及各种专用软件	Tecplot, Ensight

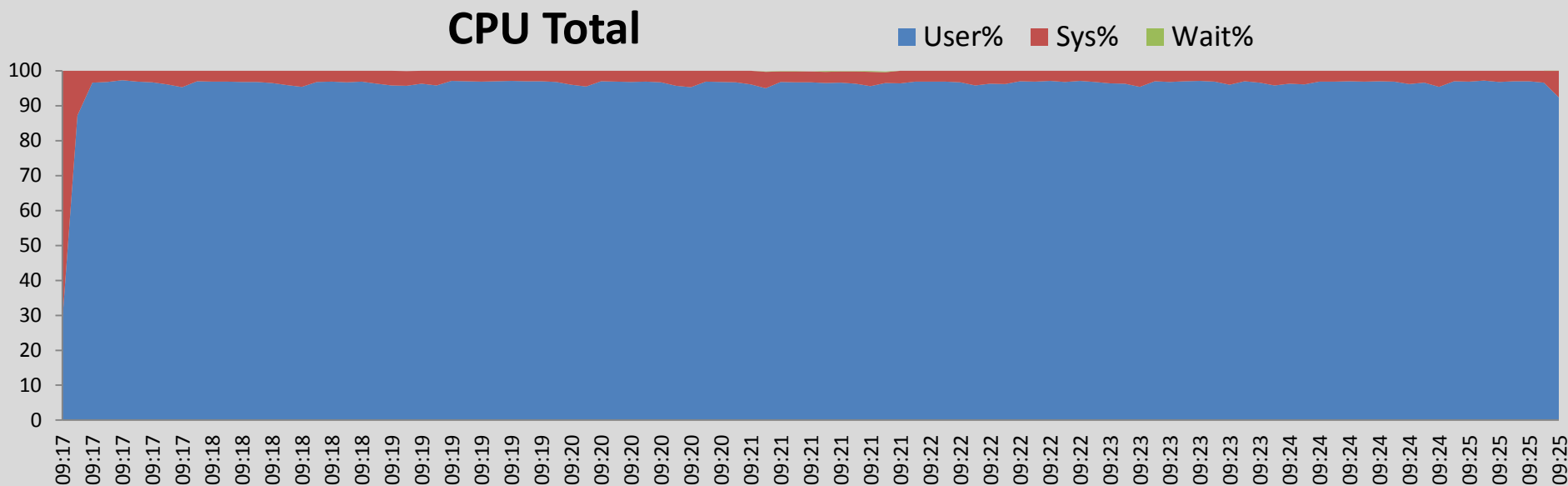
CAE应用分类

分类	软件	并行方式	可扩展性	特点
静态隐式有限元 (IFEA Static)	ABAQUS	Pthreads	<100p	内存占用多 磁盘IO大 进程通信量大 扩展性一般
动态隐式有限元 (IFEA Dynamic)	ANSYS	OpenMP, MPI		
	NASTRAN	Pthreads, MPI		
显式有限元 (EFEA)	LS-DYNA	OpenMP, MPI	~500p	对内存、磁盘IO和通信延迟的要求一般 扩展性较好
	PAM-CRAS	OpenMP, MPI		
	RADIOSS	OpenMP, MPI		
计算流体力学 (CFD) 非结构	FLUENT	MPI	~1000p	计算量大 内存占用多 进程通信量大，通信频繁 扩展性好
	STAR-CD	MPI		
	PowerFLOW	OpenMP, MPI		
计算流体力学 (CFD) 结构	OVERFLOW	MPI	~2000p	

Fluent案例分析

组成	选配方案	考虑角度
Workload	Truck_14M Truck_111M	算例规模、算例设置
FLuent	FLuent	并行方式
机型	计算节点: NX560T (刀片) I/O节点: NF5280M2 (机架)	根据项目需求选择机型
CPU	Intel Xeon X5675/X5650	CPU频率、Cache、QPI频率、 Turbo配置、NUMA配置
内存	48GB DDR3 1333MHz 24GB DDR3 1333MHz	内存容量、内存频率、 DIMM/Socket
存储/文件系统	IO节点, NFS	500G3、1000G3 NFS、Lustre
网络	Infiniband	Infiniband、10G Eth
OS	RHEL5.5	OS版本
MPI	Intel MPI	MPI软件选择、版本选择、 参数选项

Fluent案例分析——CPU需求

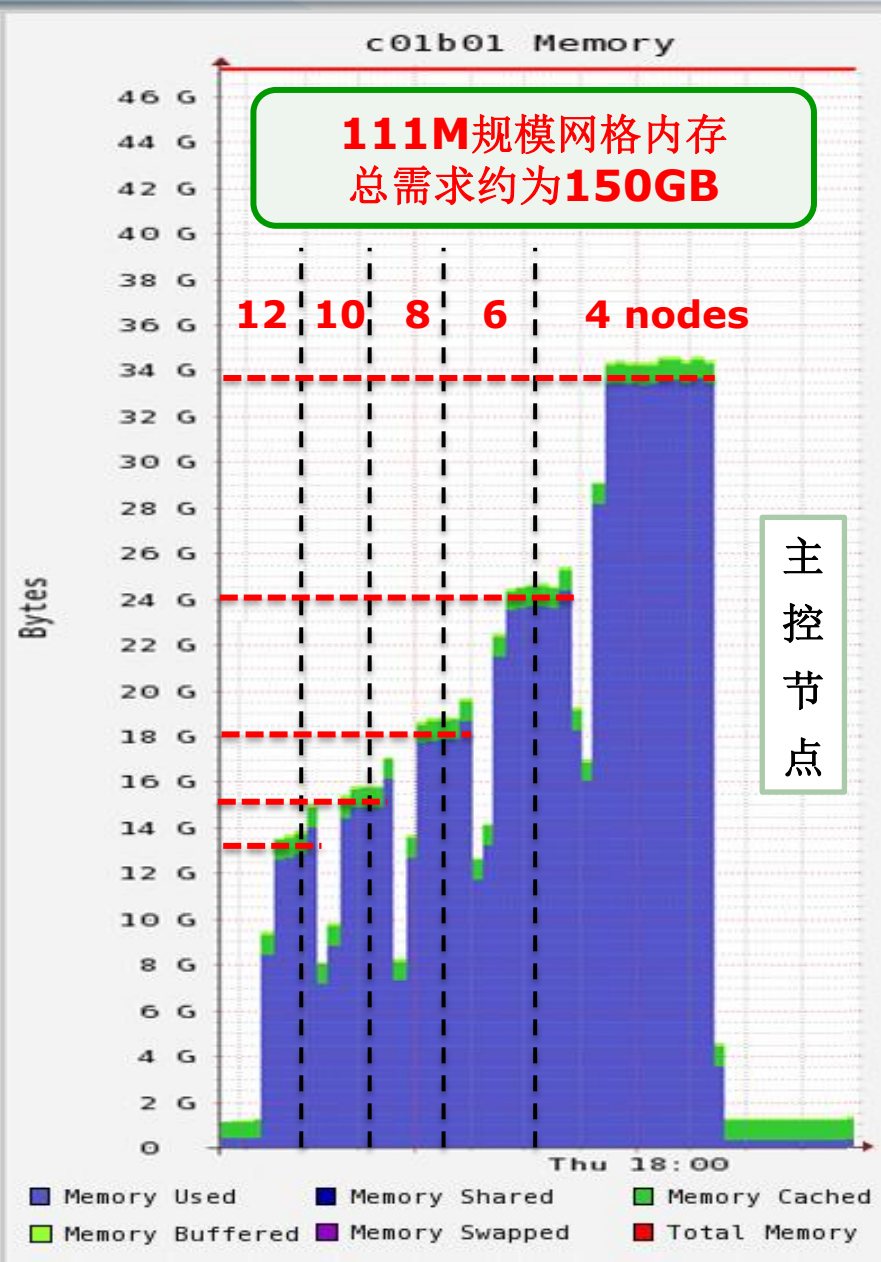


- Fluent运行中，CPU利用率一直为100%，其中系统内核使用率仅占4%左右。
- Fluent具有较好的CPU Frequency Scalability，性能与主频提升成正比。

Inspur NX560T
1node: 48GB,1TB
Fluent
truck_poly_14m

CPU	Rpeak	比例	Fluent时间	性能比例
X5650	127.68	1	954.85	1
X5675	146.88	提升15%	841.28	提升13.5%

Fluent案例分析——内存需求

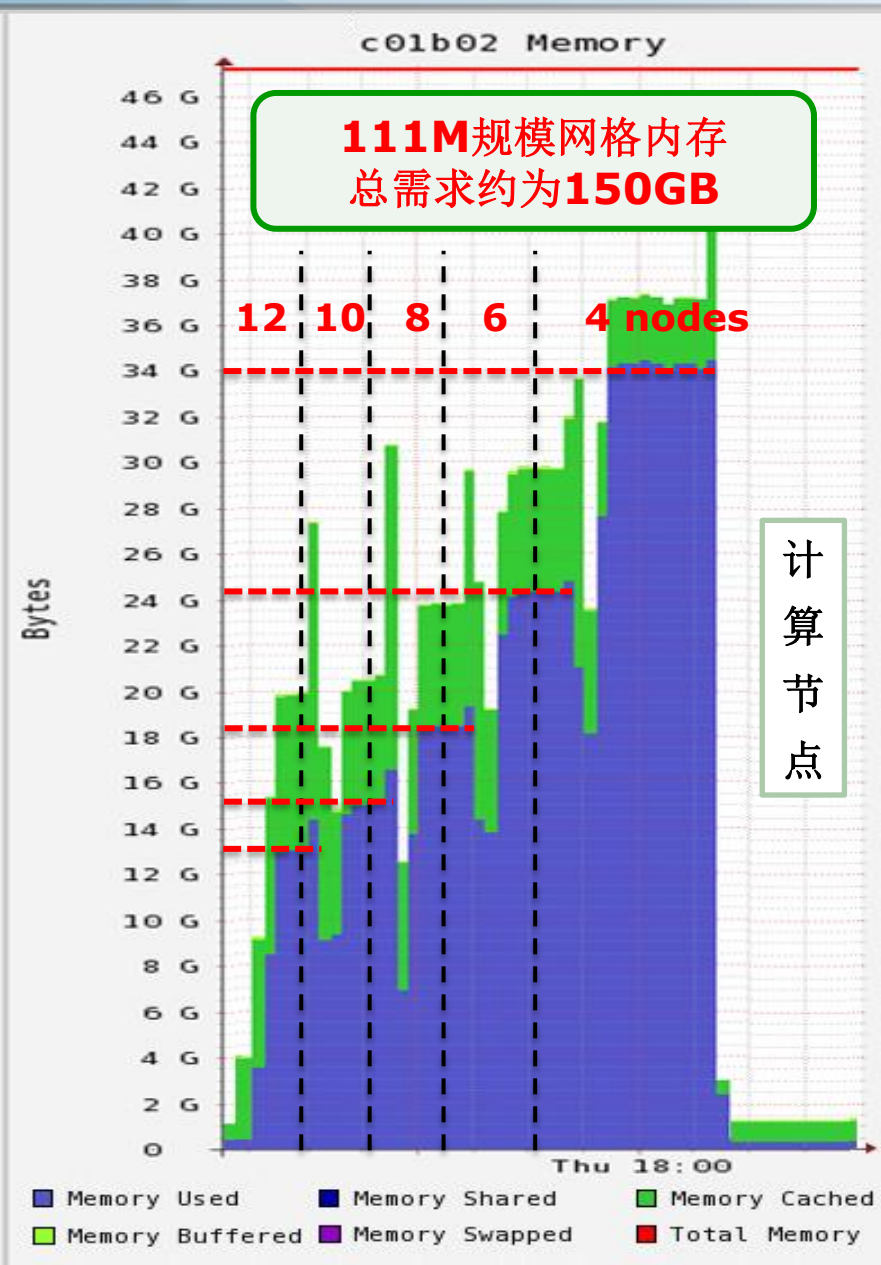


- 网格规模确定了计算所需内存，无论并行多少节点内存总需求是一定的。

需求公式 $\approx 1.5\text{GB}/\text{百万网格}$

- 并行更多的节点，可对每节点内存容量需求减小。
- 主控节点与所有计算节点的内存使用大小相同，甚至计算节点需要更大的内存用于Memory Cache。

Fluent案例分析——内存需求



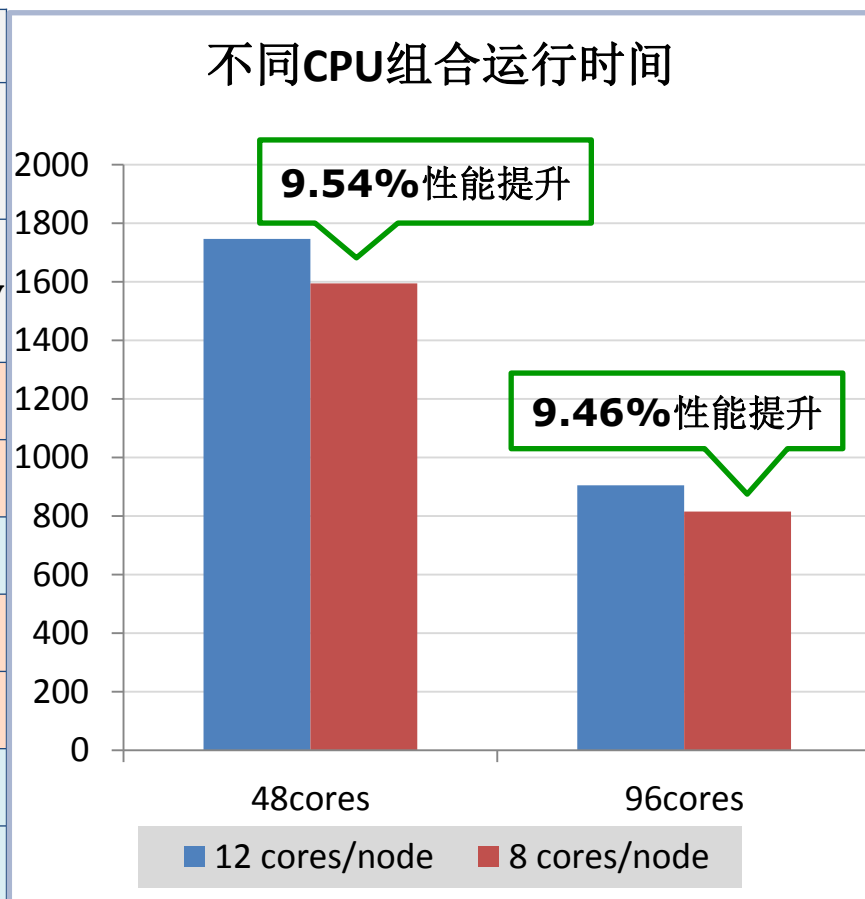
- 网格规模确定了计算所需内存，无论并行多少节点内存总需求是一定的。

需求公式 $\approx 1.5\text{GB}/\text{百万网格}$

- 并行更多的节点，可对每节点内存容量需求减小。
- 主控节点与所有计算节点的内存使用大小相同，甚至计算节点需要更大的内存用于Memory Cache。

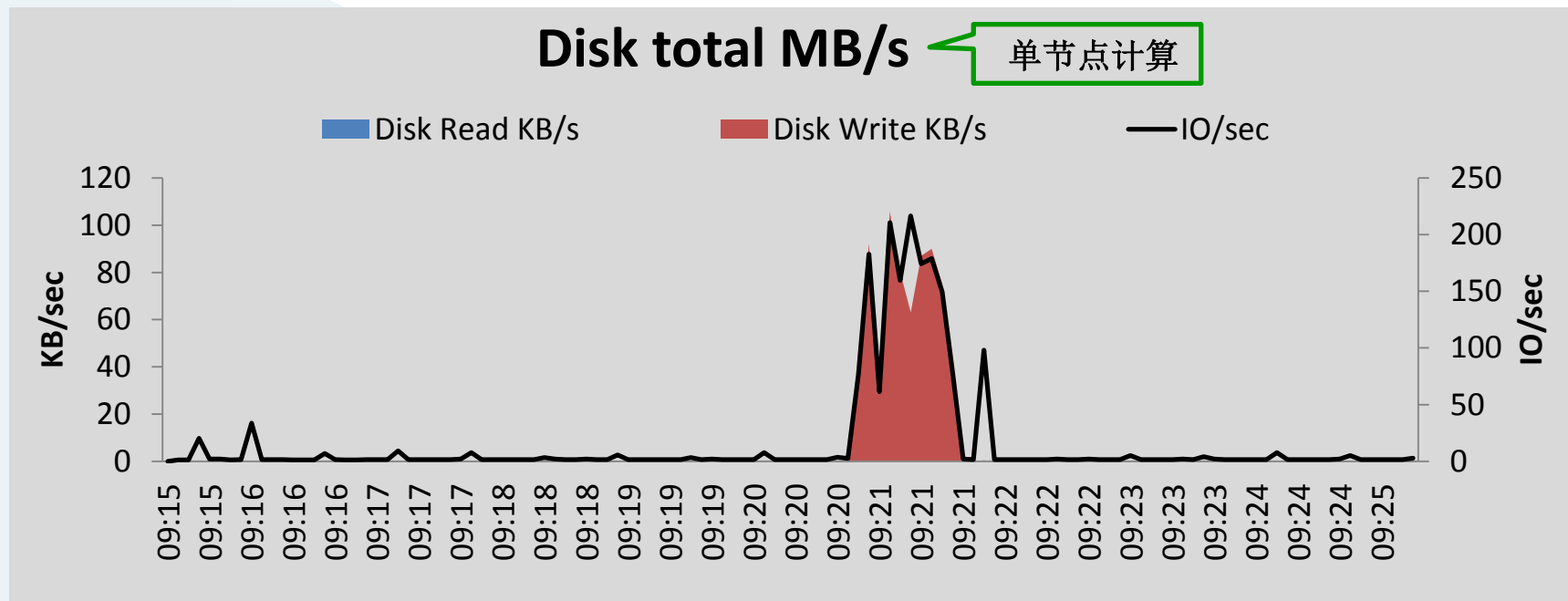
Fluent案例分析——内存带宽

NX560T (X5675, 48GB, IB QDR)					
Machines		Fluent V13: truck_111m (1亿1千1百万网格)			
#of Nodes	#of Cores	Total wall time(s)	标准线性加速比	实测 Speed up	efficiency
4	48	1746.67	1	1.00	100.00%
6	48	1594.45	1	1.10	109.55%
6	72	1209.67	1.5	1.44	96.26%
8	96	904.72	2	1.93	96.53%
12	96	815.46	2	2.14	107.10%
10	120	728.85	2.5	2.40	95.86%
12	144	635.63	3	2.75	91.60%



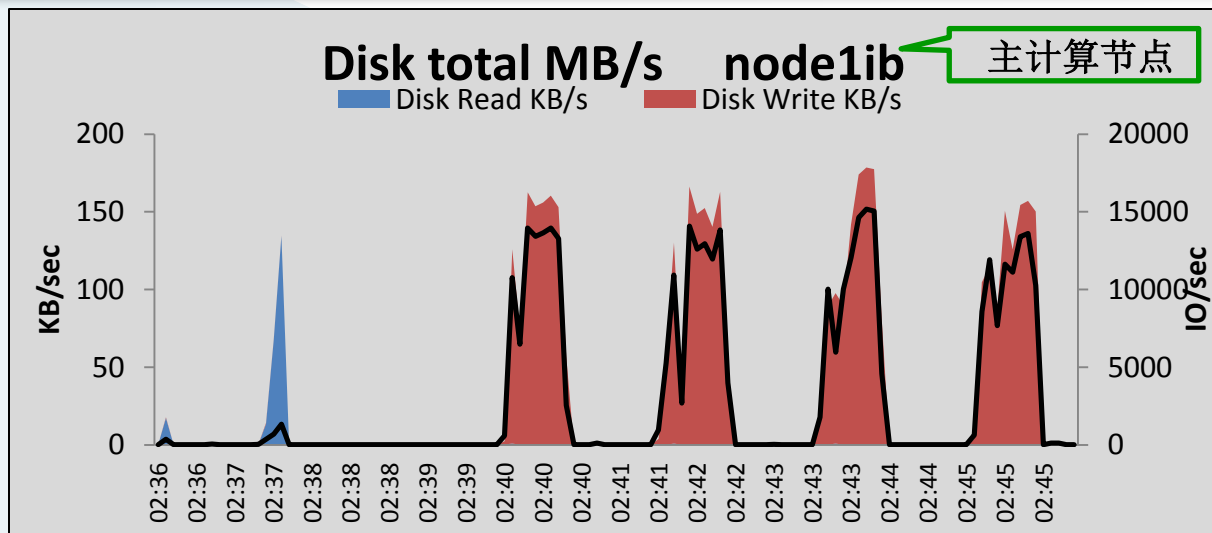
使用节点内更少的核心，降低了内存带宽的竞争

Fluent案例分析——IO



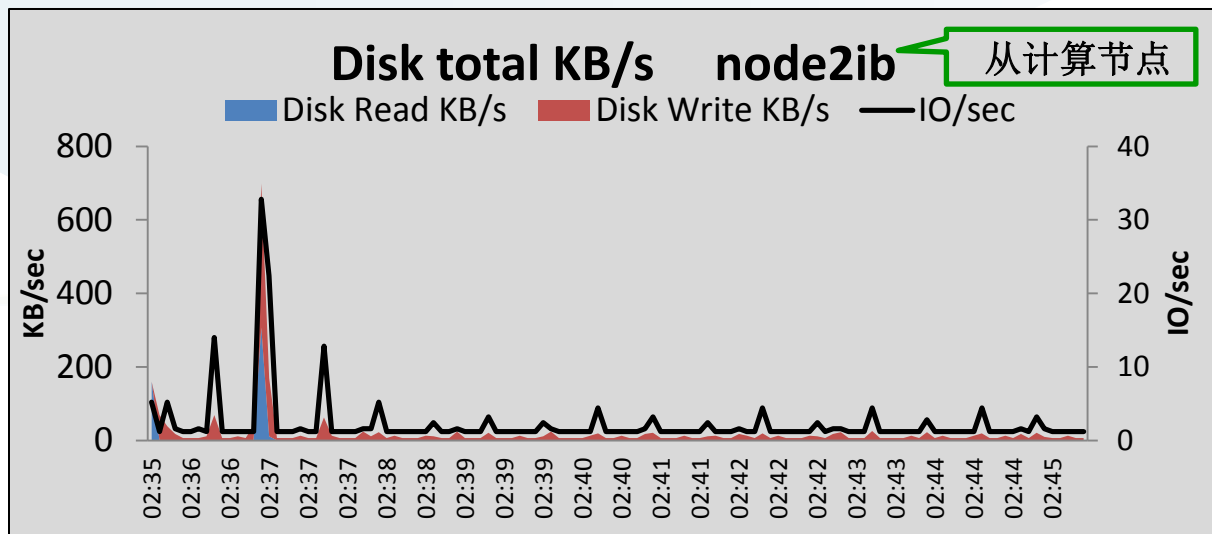
单节点并行： 计算过程很少的读写，生成CAS/DAT数据时即图中大部分红色区域，单硬盘写速率平均为43MB/s，写的最高性能达到了87.9 MB/s。

Fluent案例分析——IO



四节点并行**Fluent**计算

- 多点并行运行的**Fluent**在 write cas/dat时的IO压力都在主节点上



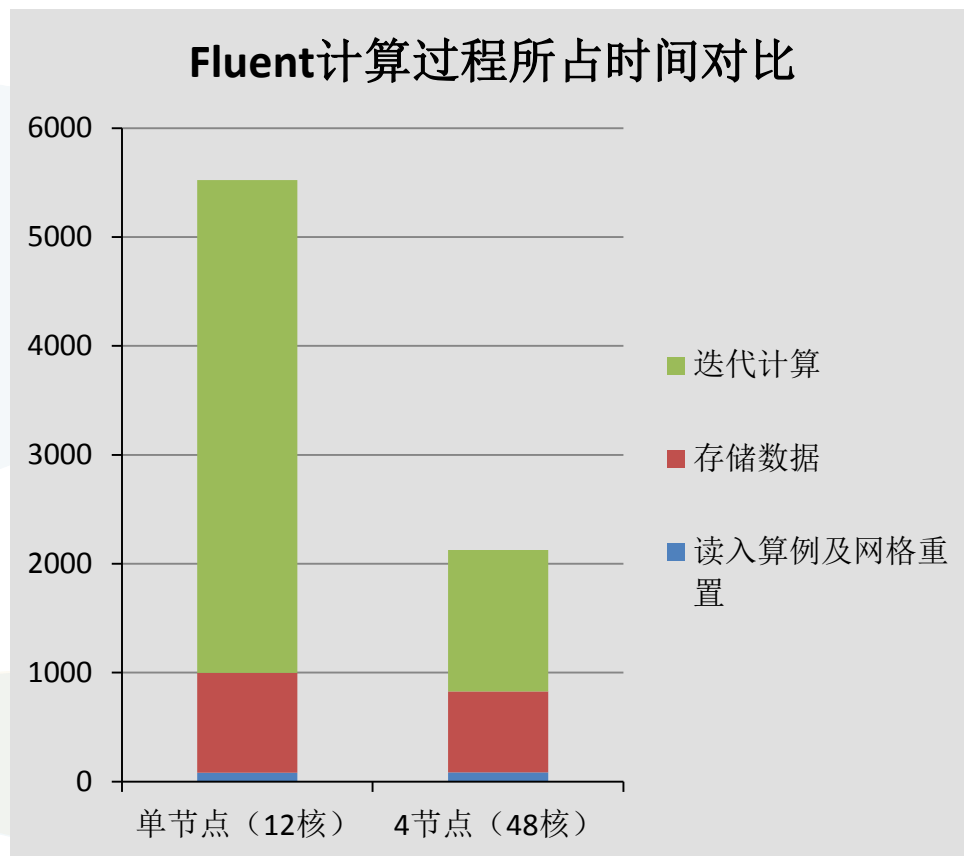
- 从节点的IO压力主要在启动**Fluent**读入算例时

Fluent案例分析——IO

计算平台：NX560T（X5650，24GB，IB QDR）；Workload：660万网格

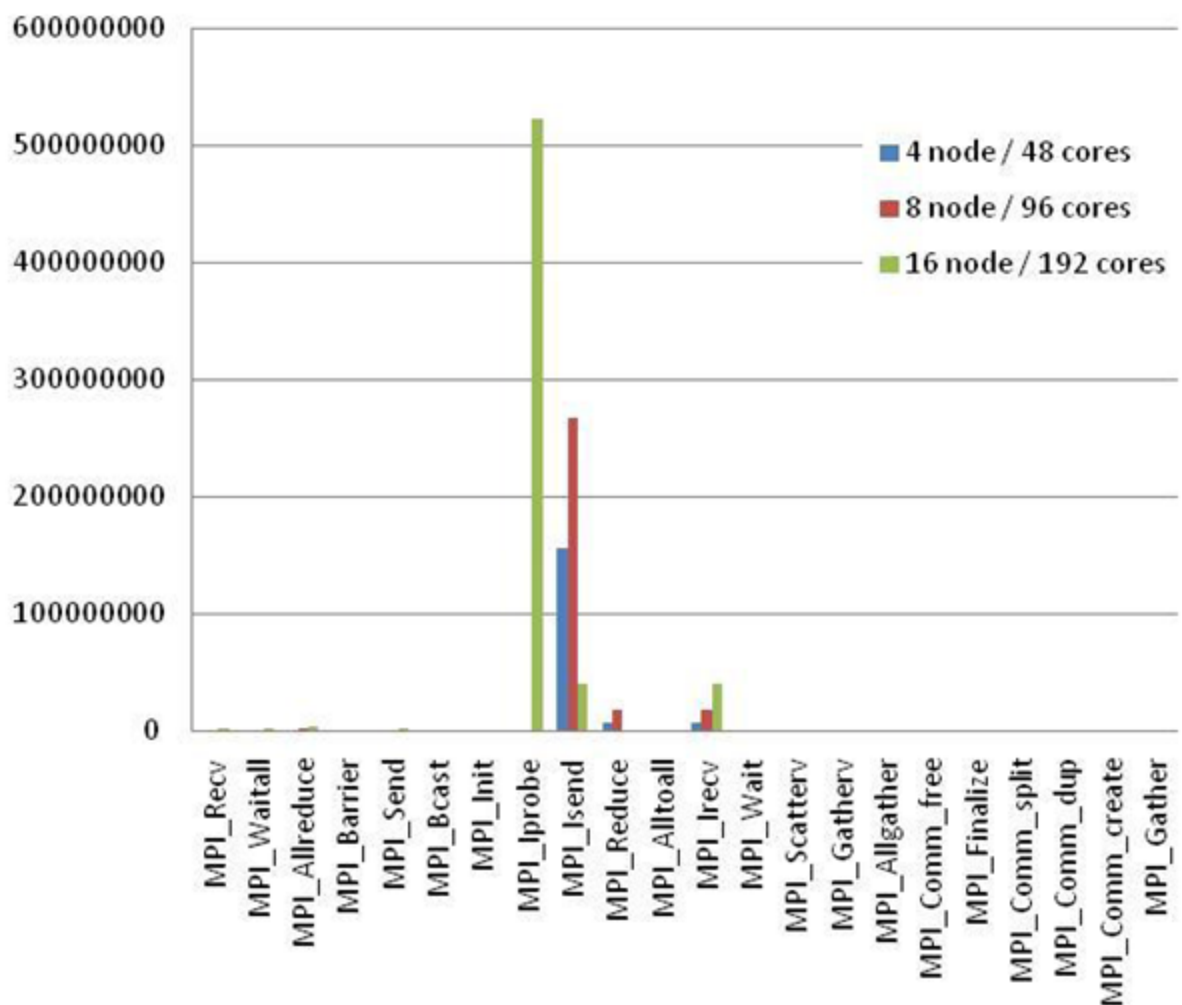
	单节点	4节点
总时间	5569	2088
迭代时间	4525	1296
存储时间	914	745
IO%	16%	36%

Fluent在多点并行情况下计算有着近似线性的加速比，所以由于总时间缩短，存储所占的时间比例会变大。



Fluent案例分析——MPI函数

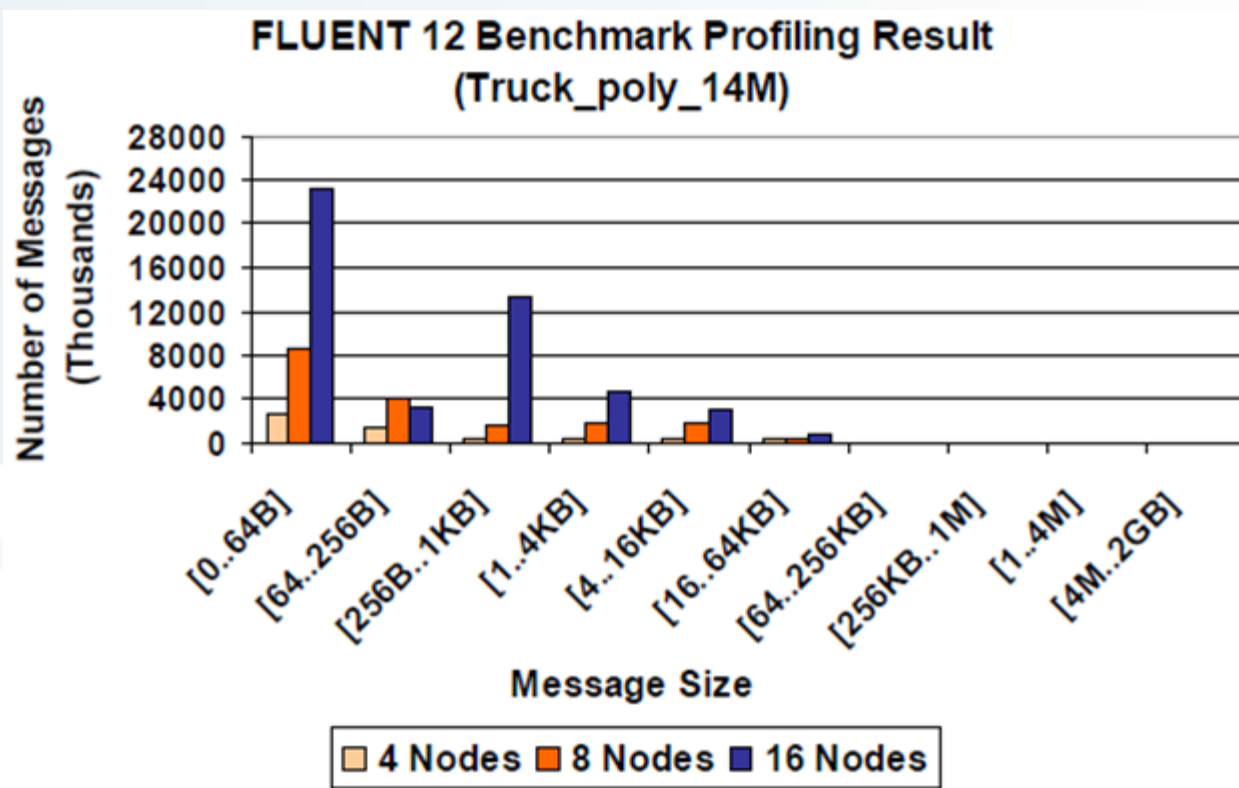
Truck_111m MPI Routines



- 调用最多的几个函数为：
MPI_Iprobe,
MPI_Isend,
MPI_Irecv,
MPI_Allreduce
- 运行时间比例最大的几个函数：
MPI_Recv,
MPI_Allreduce,
MPI_Waitall

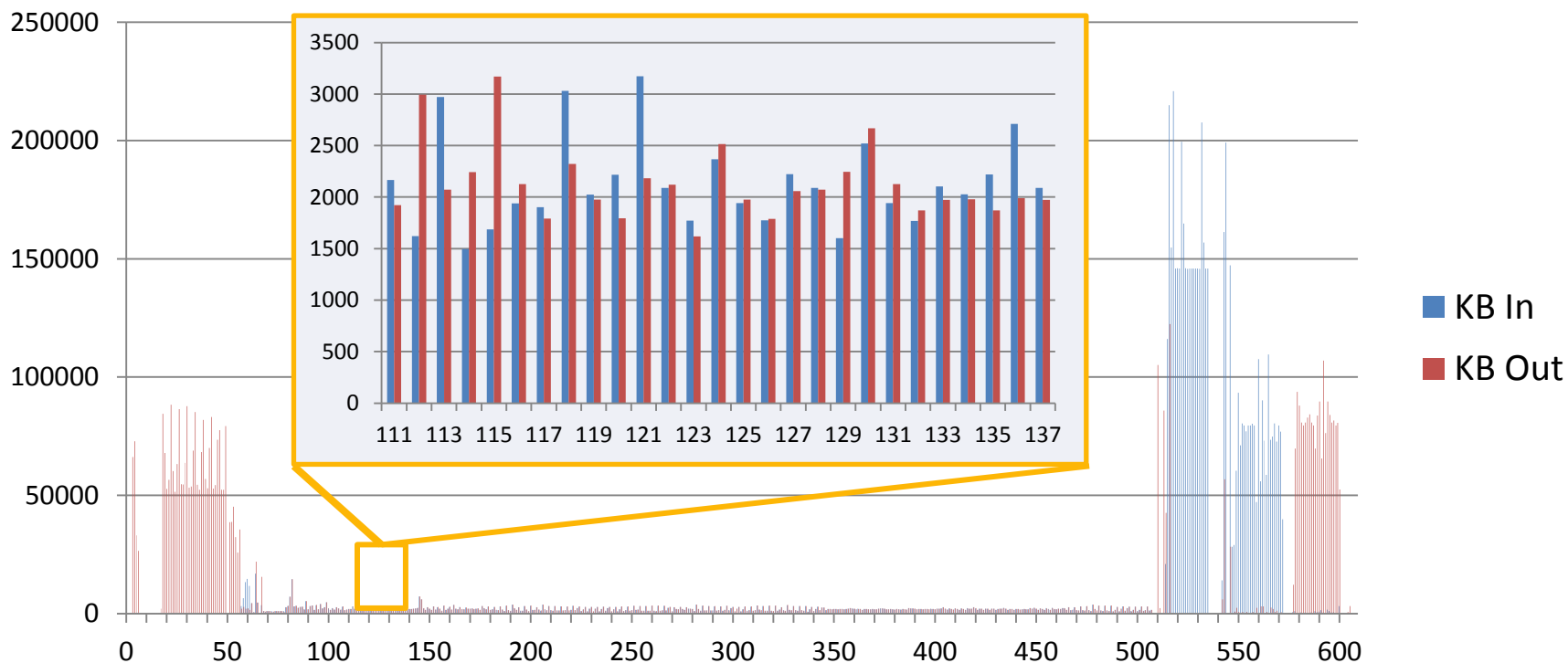
Fluent案例分析——MPI函数

- 大多的MPI消息传递数据大小在256B-1KB之内
- 典型的MPI同步消息传递数据小于64B
- MPI传递的消息数目随计算节点增加而增加



Fluent案例分析——网络监测

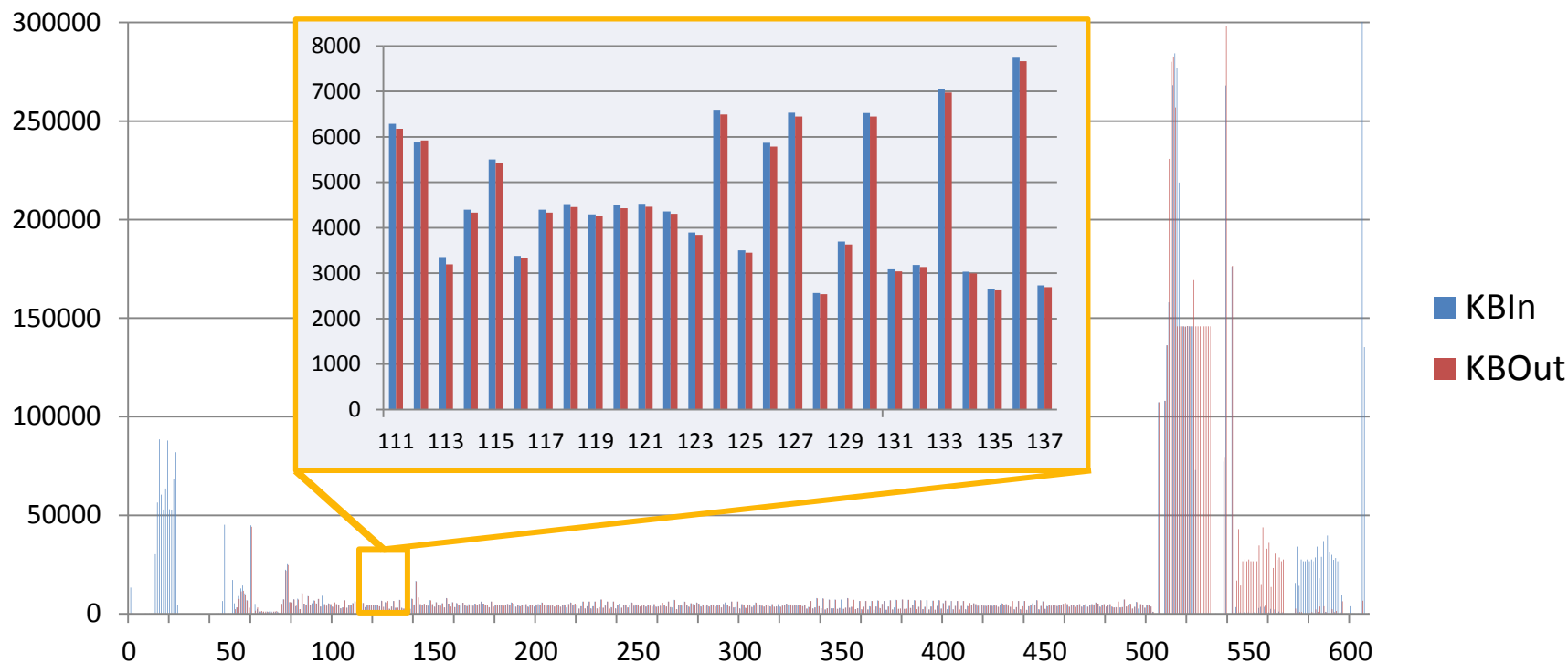
Fluent_14M_Infiniband 流量监测 (主节点)



- Fluent求解最大的网络压力出现在计算的开始和结束阶段。
- Fluent求解对网络的延迟特性(latency)有很高的要求。
- 主节点与从节点的网络进出是相对的，但对网络的要求是一致的。

Fluent案例分析——网络监测

Fluent_14M_Infiniband 流量监测 (从节点)

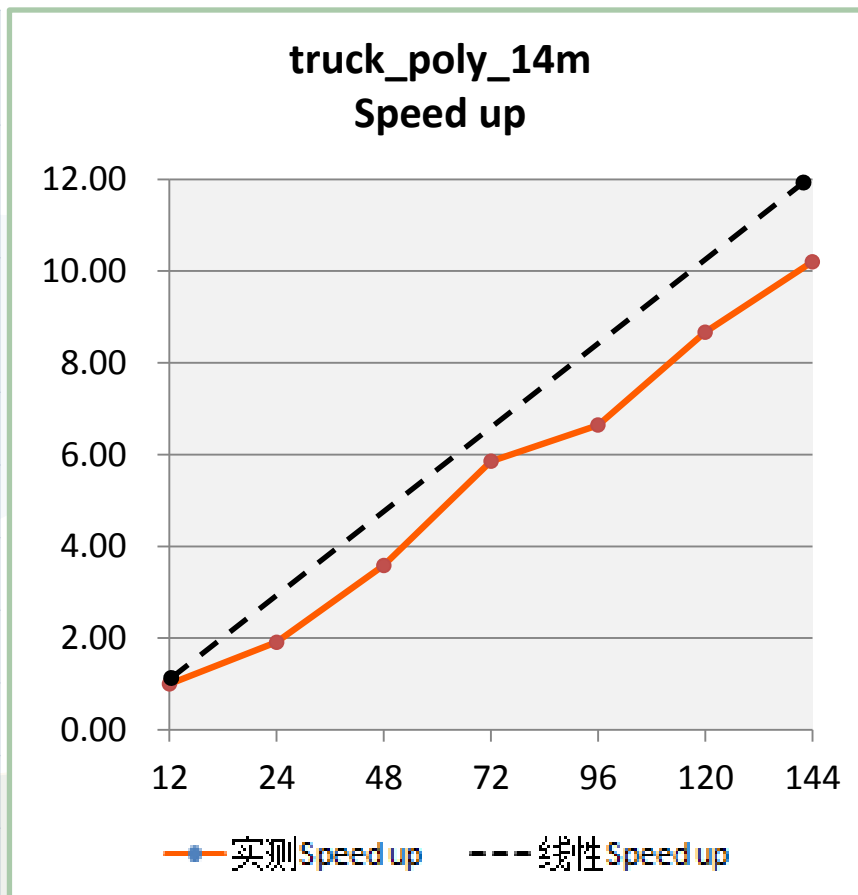


- Fluent求解最大的网络压力出现在计算的开始和结束阶段。
- Fluent求解对网络的延迟特性(latency)有很高的要求。
- 主节点与从节点的网络进出是相对的，但对网络的要求是一致的。

Fluent案例分析——加速比

NX560T (X5675, 48GB, IB QDR)

Machines		truck_poly_14m (一千四百万网格)		
#of Nodes	#of Cores	Total wall time(s)	Speed up	Efficiency
1	12	841.28	1.00	100.00%
2	24	440.61	1.91	95.47%
4	48	234.82	3.58	89.57%
6	72	143.72	5.85	97.56%
8	96	126.62	6.64	83.05%
10	120	97.05	8.67	86.69%
12	144	82.44	10.20	85.04%



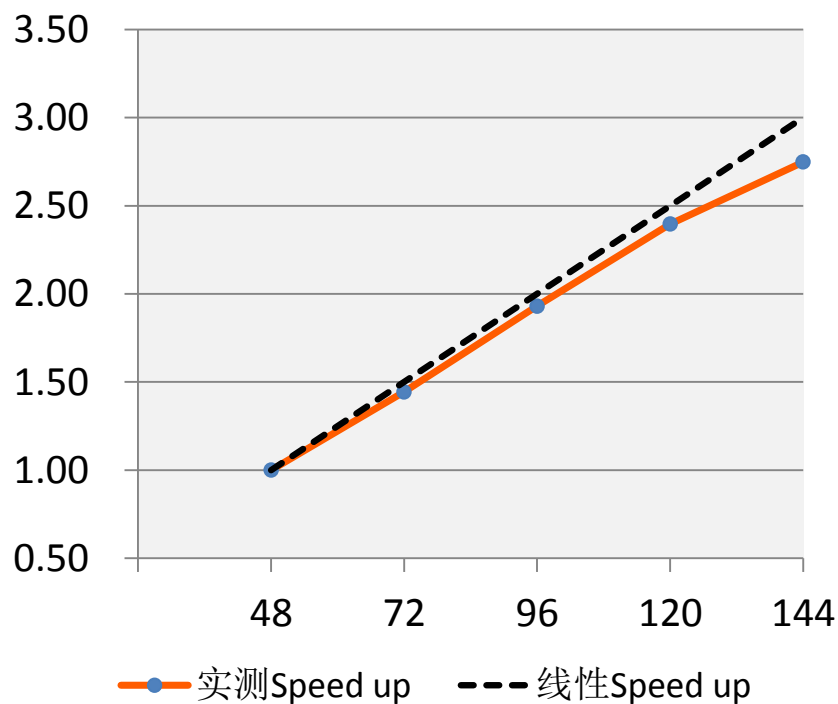
Fluent在Inspur平台上有着近似线性的加速比

Fluent案例分析——加速比

NX560T (X5675, 48GB, IB QDR)

Machines		Fluent V13: truck_111m (一亿一千一百万网格)			
#of Nodes	#of Cores	Total wall time(s)	标准线性加速比	Speed up	Efficiency
4	48	1746.67	1	1.00	100.00%
6	72	1209.67	1.5	1.44	96.26%
8	96	904.72	2	1.93	96.53%
10	120	728.85	2.5	2.40	95.86%
12	144	635.63	3	2.75	91.60%

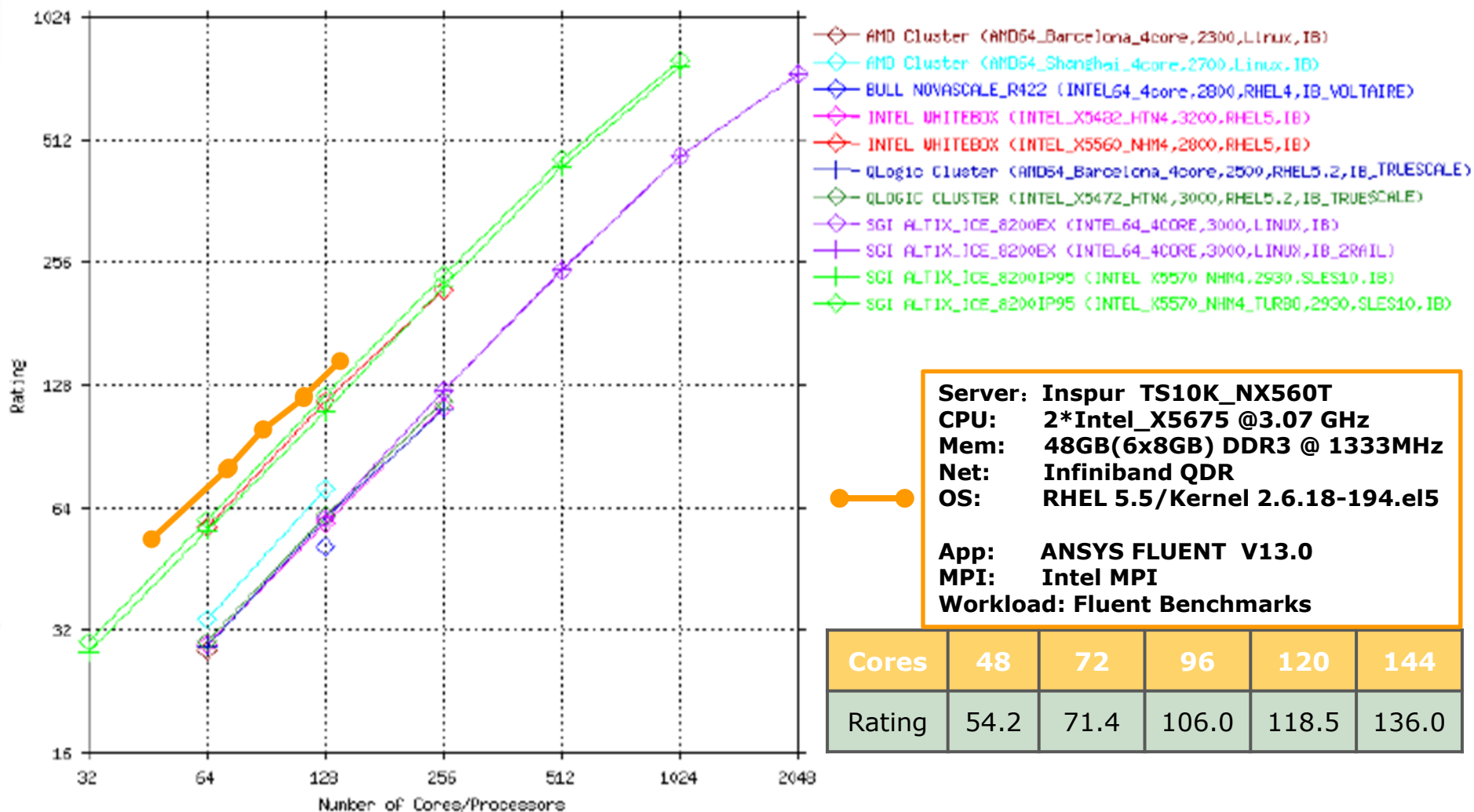
Fluent truck_111m
Speed up



Fluent在Inspur平台上有着近似线性的加速比

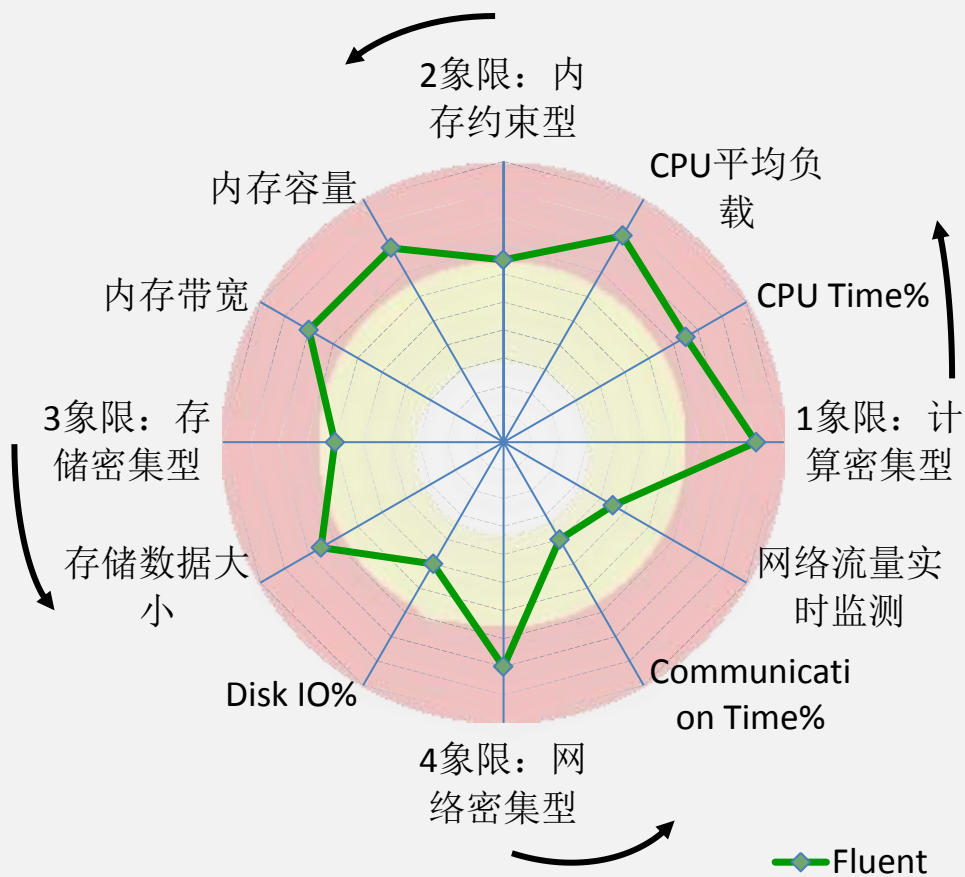
Fluent案例分析——综合性能

FLUENT 13.0 beta Parallel Performance of TRUCK_111M



Fluent应用特点小结

Fluent应用特征雷达图

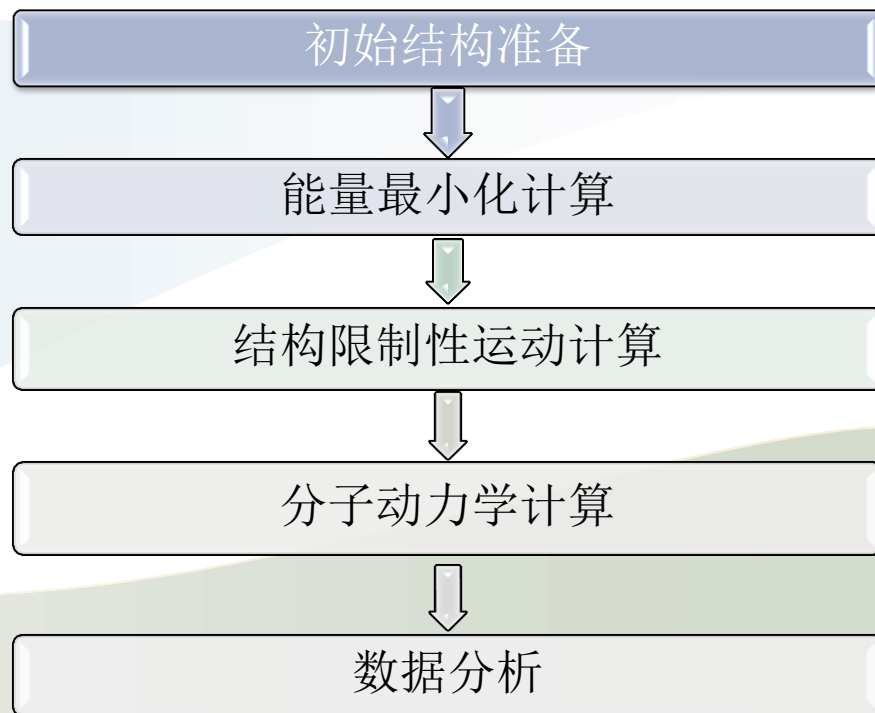


- **计算密集型应用**，对CPU主频有较高要求
- **内存**：按算例规模及节点规模配置相应内存大小；使用更多的节点资源以减小进程间的内存带宽竞争
- **存储**：大规模算例保存结果对IO有较高要求
- **网络密集型应用**，大量的进程间消息传递，对网络延迟有较高要求
- **扩展性**：很好，在上千个进程上也有较好加速比

分子动力学软件--NAMD

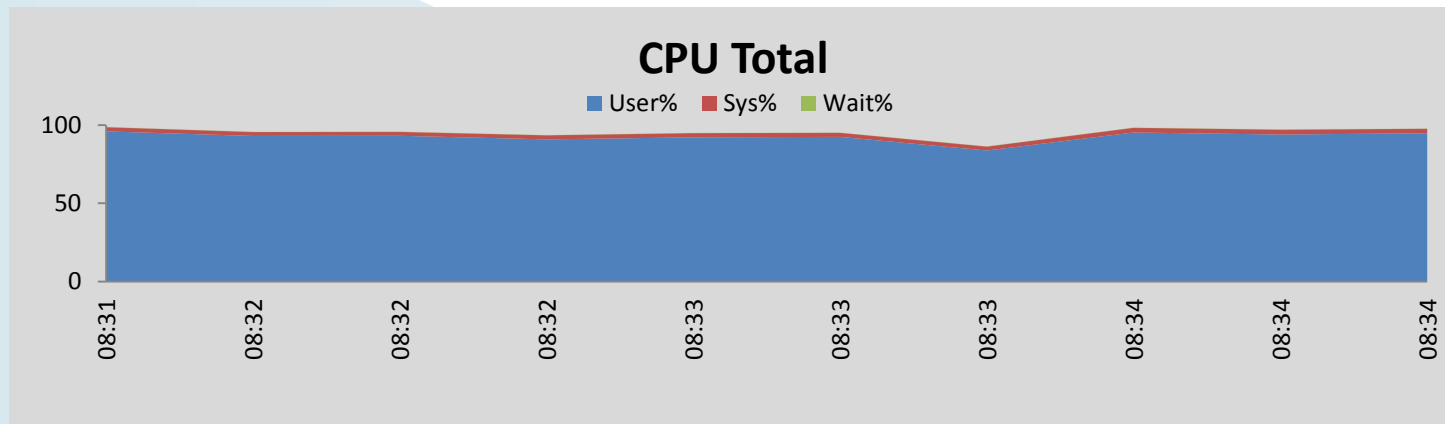
- NAMD

是一款并行分子动力学模拟软件，主要用来研究大分子。它是由Illinois 大学的理论生物物理研究小组开发。NAMD 可以在数百个处理器上进行高性能并行计算。常用力场包括Amber、CHARMM 和Dreiding。



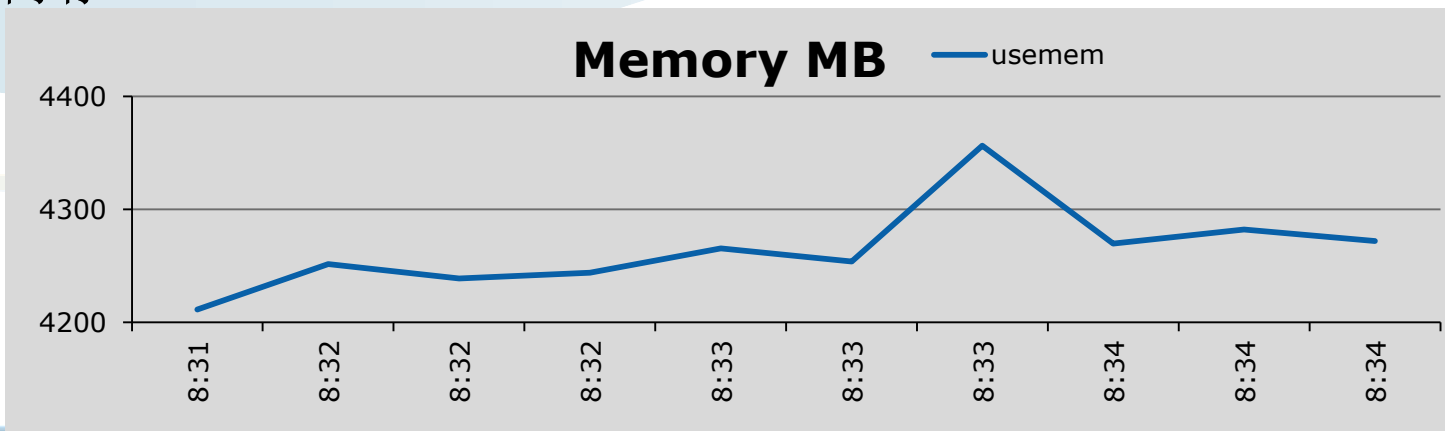
NAMD案例分析

- CPU



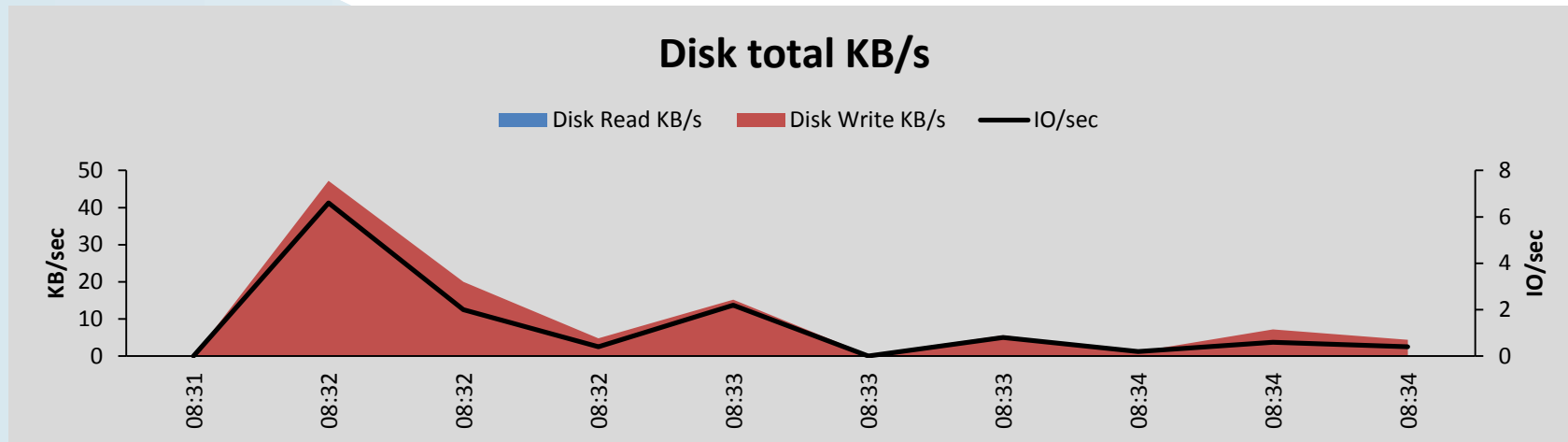
CPU
利用率较
高

- 内存

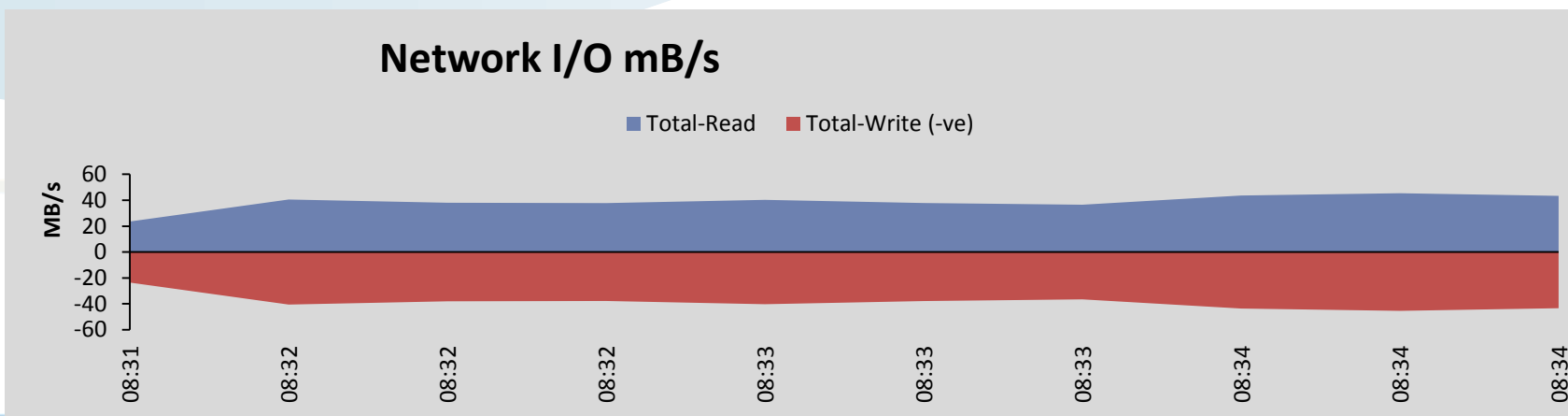


NAMD案例分析

• 磁盘IO



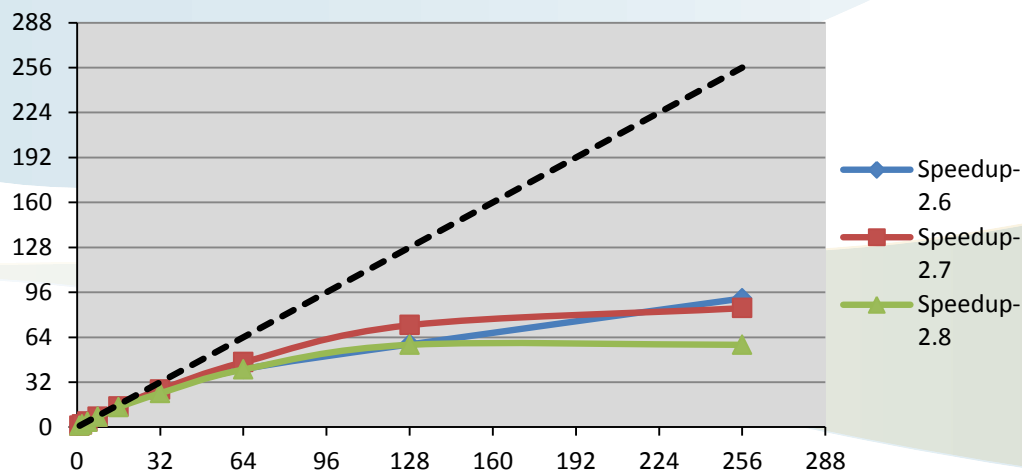
• 网络IO



NAMD案例分析

• 扩展性

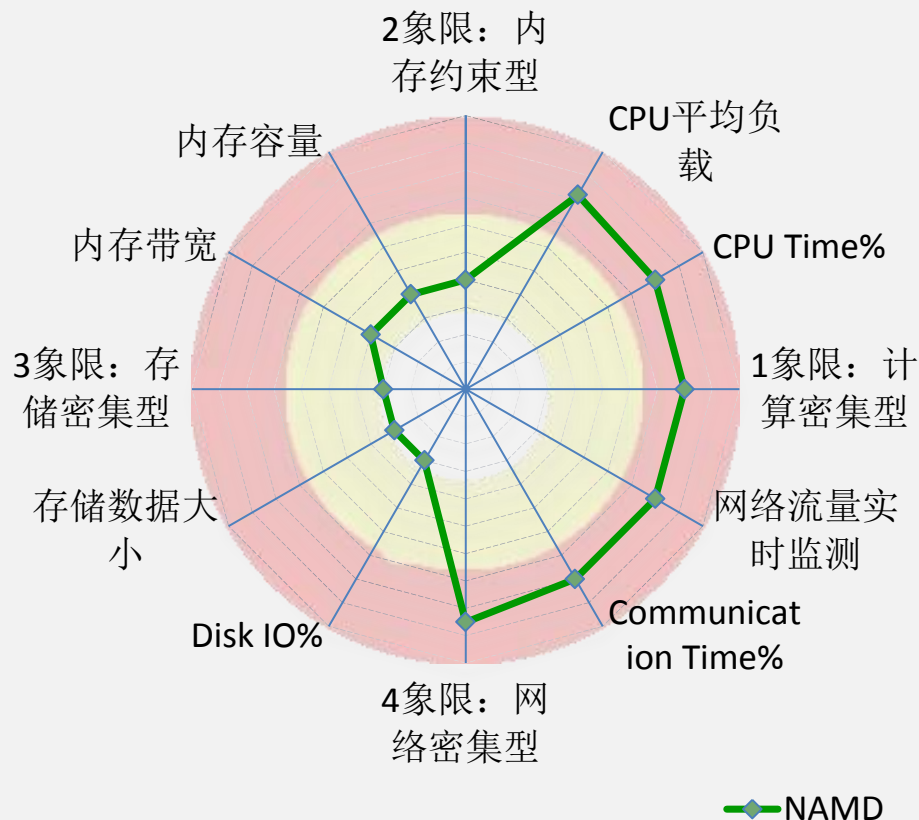
NAMD 2.6			NAMD 2.7		NAMD 2.8	
Procs	Time	Speedup-2.6	Time	Speedup-2.7	Time	Speedup-2.8
1	<u>8.24</u>	1	<u>5.08</u>	1	<u>4.1</u>	1
2	<u>4.26</u>	1.93	<u>2.66</u>	1.91	<u>2.16</u>	1.9
4	<u>2.19</u>	3.76	<u>1.35</u>	3.76	<u>1.11</u>	3.69
8	<u>1.12</u>	7.36	<u>0.69</u>	7.36	<u>0.57</u>	7.19
16	<u>0.57</u>	14.5	<u>0.35</u>	14.5	<u>0.29</u>	14.1
32	<u>0.32</u>	25.8	<u>0.19</u>	26.7	<u>0.17</u>	24.1
64	<u>0.2</u>	41.2	<u>0.11</u>	46.2	<u>0.1</u>	41
126	<u>0.14</u>	58.9	<u>0.07</u>	72.6	<u>0.07</u>	58.6
252	<u>0.09</u>	91.6	<u>0.06</u>	84.7	<u>0.07</u>	58.6



较高的扩展性，运行100个进程时也具较高的效率，是其他分子动力学不可比拟的

NAMD应用特征小结

NAMD应用特征雷达图



计算密集型应用：CPU利用率高，高主频性能提升明显

内存：占用量很小，单节点**4G**内存基本能满足计算需求

磁盘IO：压力不大

网络密集型：高速INFINIBAND网络下能够获得较高的计算能力

扩展性：较好

量子化学软件-Gaussian

- **Gaussian**是做半经验计算和从头计算使用最广泛的量子化学软件，可以用来预测气相和液相条件下，分子和化学反应的许多性质。如，分子能量和结构，过渡态的能量和结构，化学键以及反应能量，分子轨道，极化率和超极化率，热力学性质，反应路径等。
- **Gaussian**并行方式：
 - ▶ 节点内:OpenMP，并行效率高
 - ▶ 节点间:TPC-Linda，并行效率低
 - ▶ 根据任务计算量选择并行方式

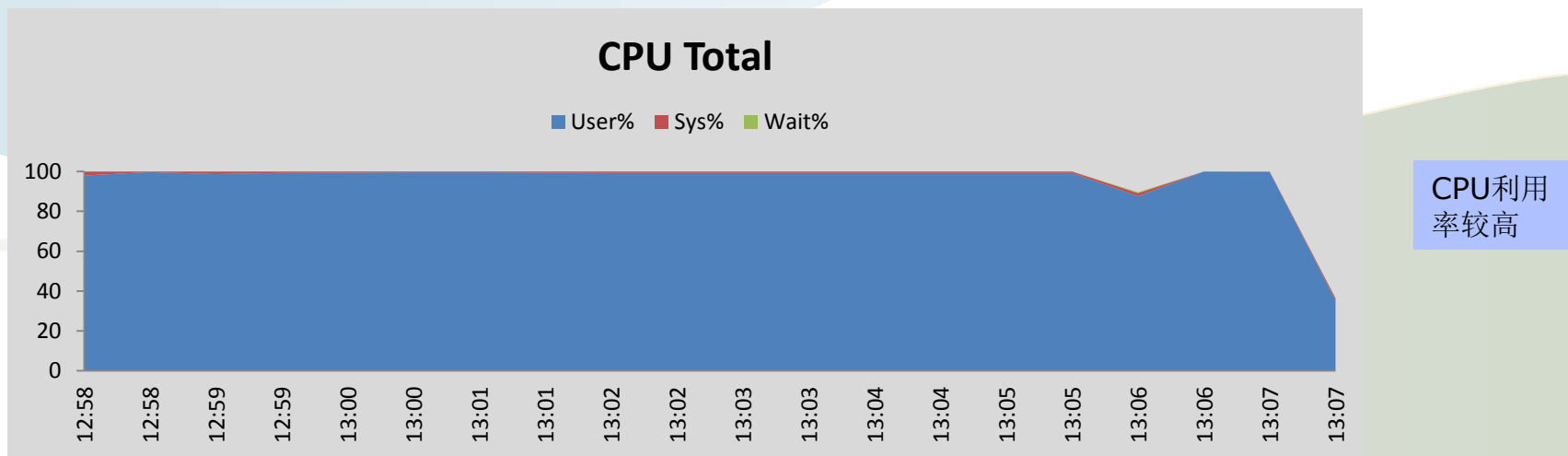
Linda是专门为Gaussian设计，实现G03分布式并行的必需程序

GAUSSIAN案例分析

- 并行方式

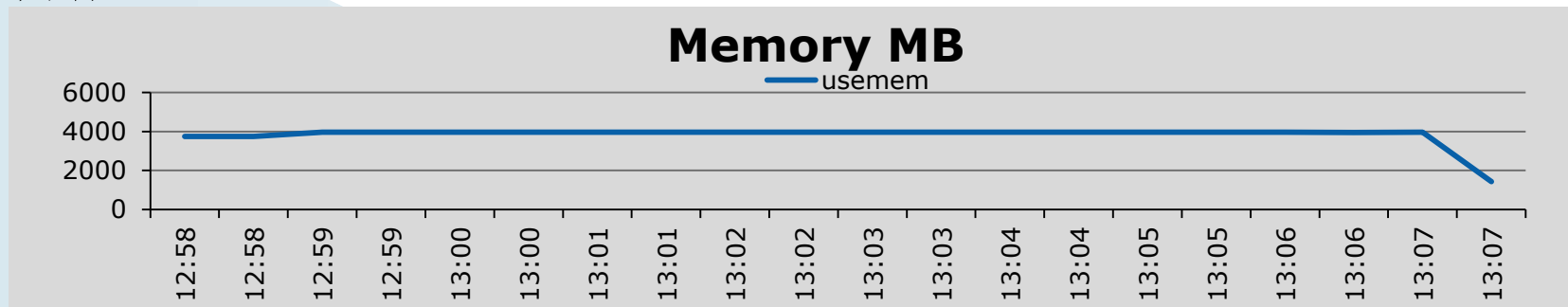
机型	CPU	内存	并行方式	Machines		run time(s)
				#of Nodes	#of Cores	
NX560T	E5620	6*4G	OPENMP	1	8	579
			LINDA	2	4	892

- CPU使用情况



GAUSSIAN案例分析

- 内存

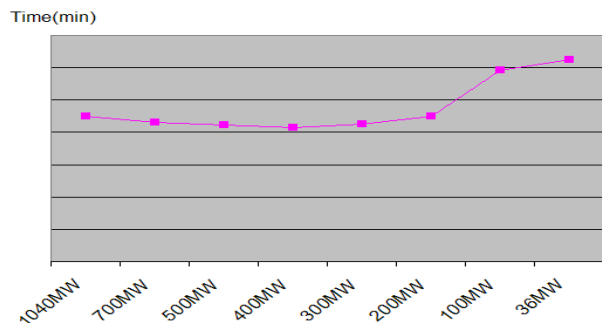


不同的分子结构需要设置不同memory大小

$$\text{Memory} = M + 2N_B^2 \quad (\text{单位是8字节的位})$$

其中 N_B 是计算中使用的基函数数量， M 是和工作任务类型有关的最小值

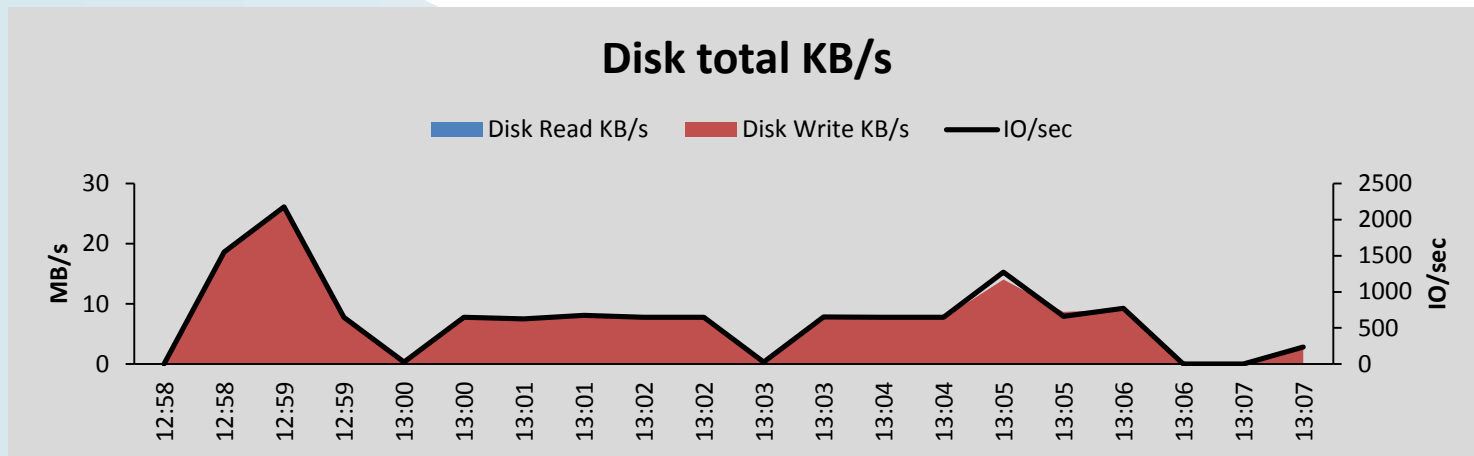
注意1 MW = 8MB,表中的值是32位计算机系统的；对64位系统需要增加一倍。



机型	CPU	内存	Machines		run time(s)	备注
			#of Nodes	#of Cores		
NX560T	E5620	6*4G	2	4	892	满核运算时内存带宽竞争激烈
			2	8	952	

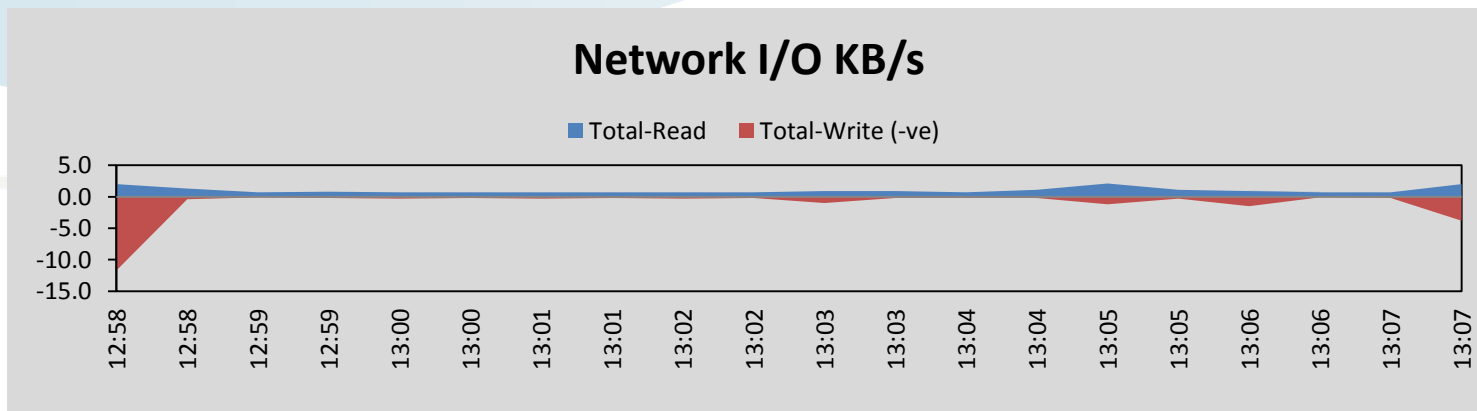
GAUSSIAN案例分析

• 磁盘IO



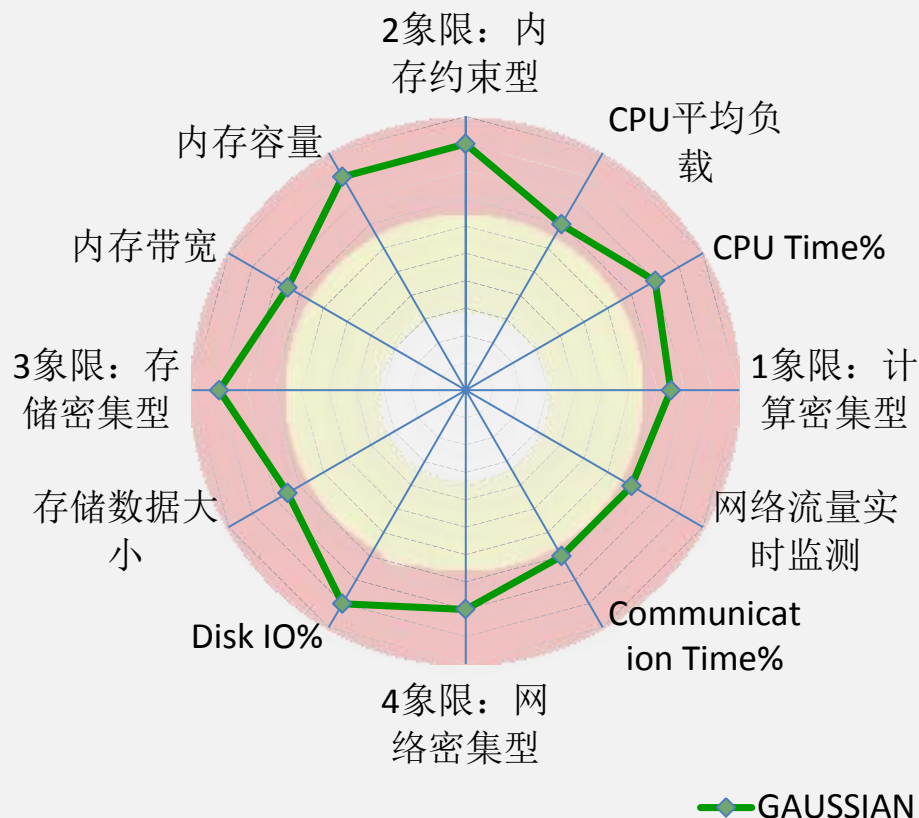
起始阶段磁
盘读写速率
为20M/S

• 网络IO



Gaussian应用特征小结

GAUSSIAN应用特征雷达图

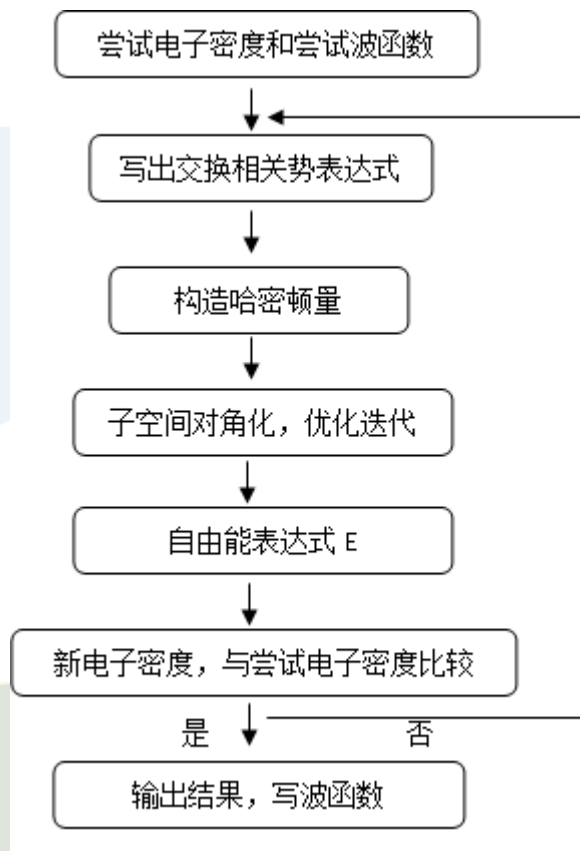


- **计算密集型：** Gaussian 计算对计算能力的需求是很大，主频高对性能影响明显；
- **内存容量约束型：** 由于目前大多数计算都要消耗比较大的内存，且如果采用Linda方式计算内存需要更大；
- **IO密集型：** 需要相对较大的本地硬盘空间；对于有着大规模计算需求而且又希望保留临时文件的用户就有必要配置专门的存储系统；
- **网络：** 通常千兆以太网可以满足一般计算要求；Infiniband能提供一种更高带宽和延迟的高速网络选择。

材料科学软件--VASP

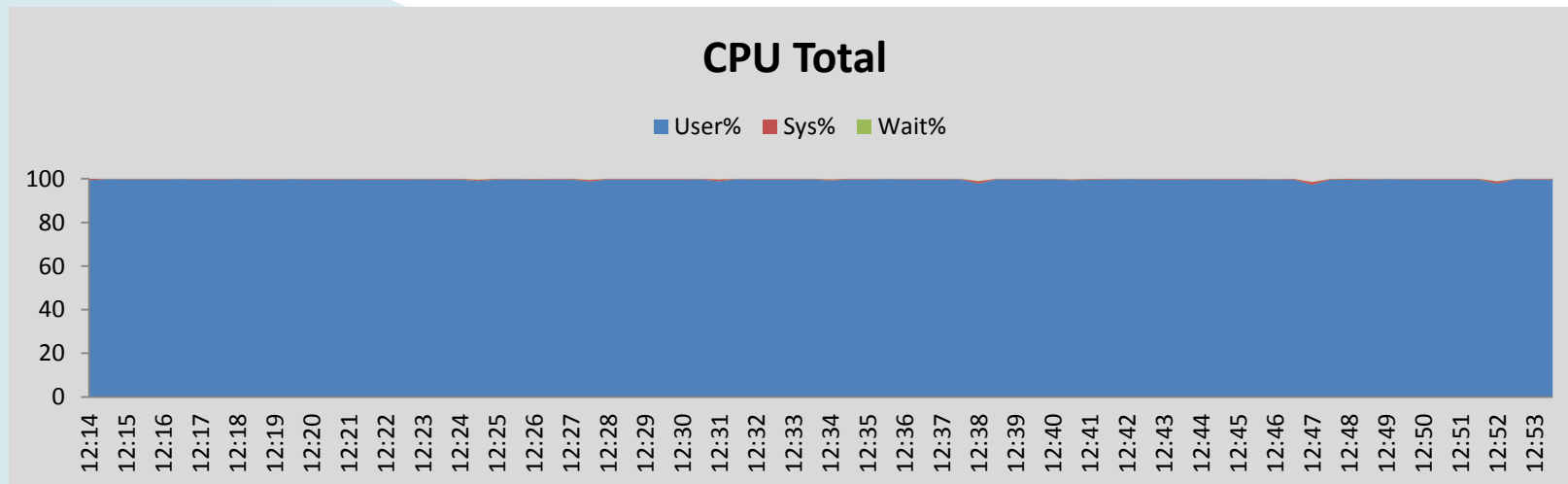
• VASP

它是基于密度泛函理论下并利用平面波势方法进行从头分子动力学和第一原理电子结构计算的软件包；可以计算材料的结构参数(键长,键角,晶格常数, 原子位置等)和构型；计算材料的状态方程和力学性质；计算材料的电子结构；光学、磁学和晶格动力学性质等



VASP案例分析

- CPU

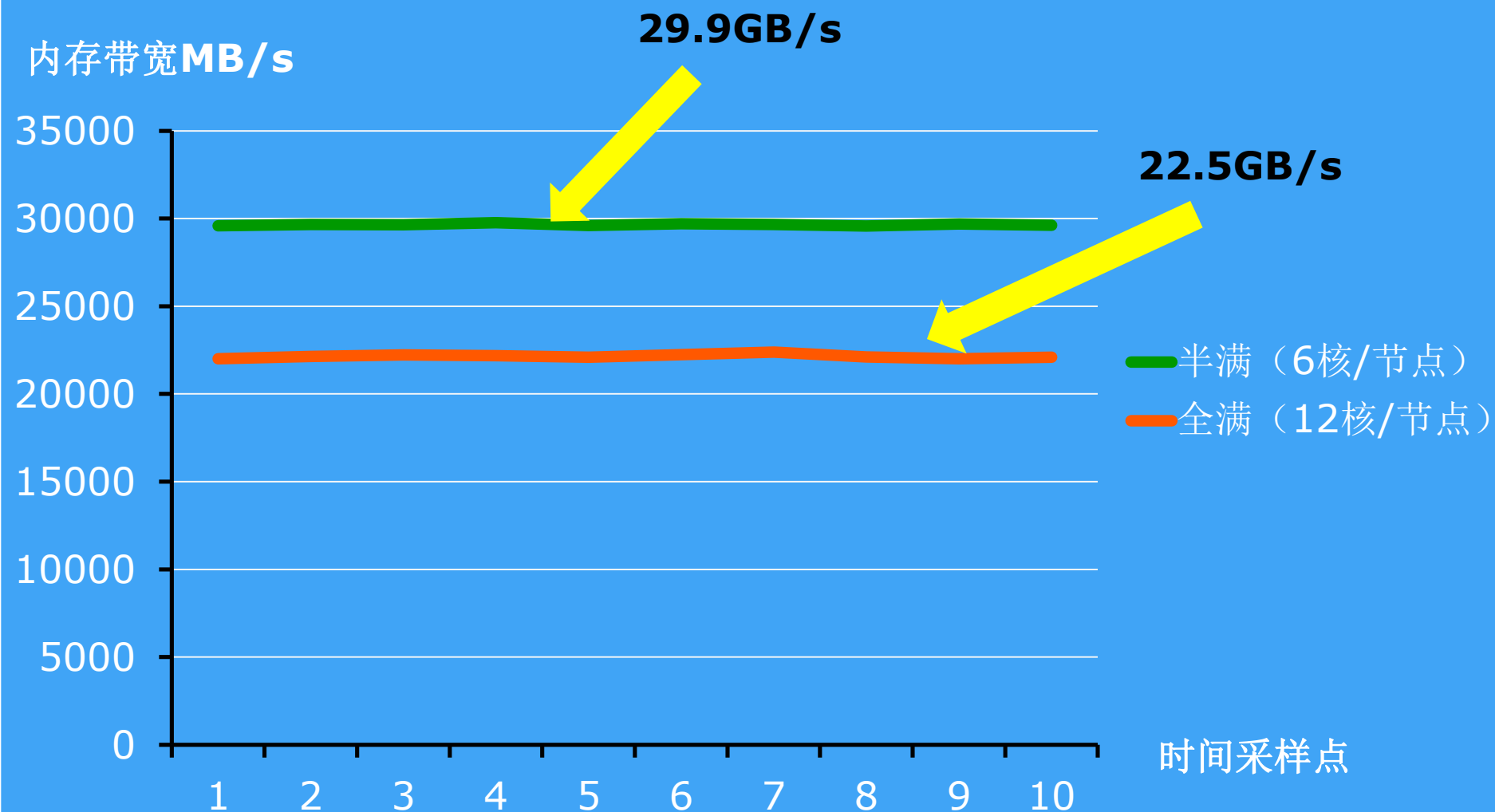


VASP运行中，CPU利用率一直为100%，其中系统内核使用率仅占2%左右。

VASP具有较好的CPU Frequency Scalability，性能与主频提升成正比。

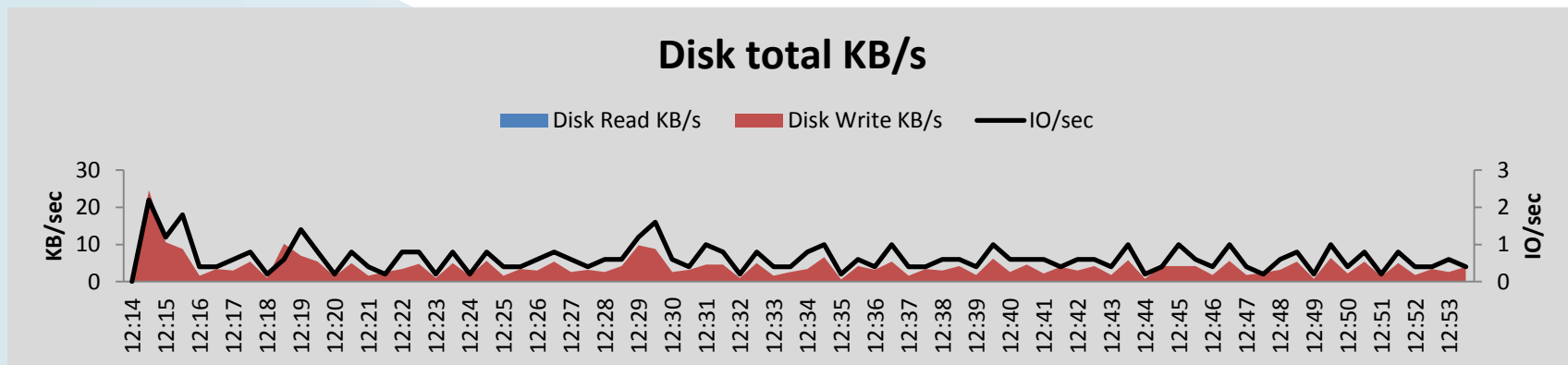
机型	CPU	Rpeak	比例	Time(s)	性能比例
NX560T	X5650	127.68	1	2727.651	1
NX560T	X5675	146.88	提升15%	2374.399	提升12%

VASP案例分析



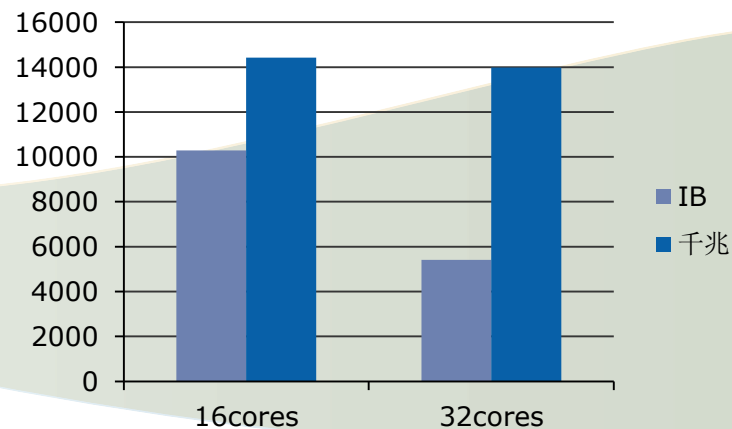
VASP案例分析

• 磁盘



• 网络

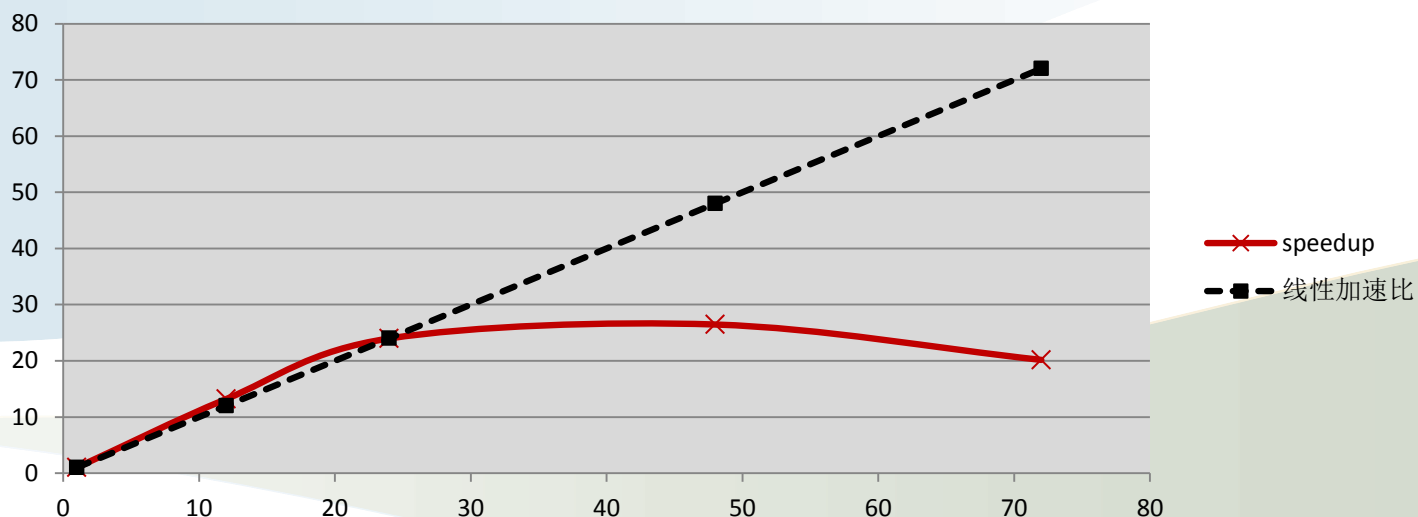
网络	Machines		Elapsed time (s)	备注
	nodes	cores		
IB	2	16	10289.254	IB网络较千兆网络 约有20-60%性能 提升；千兆网络 下，随着节点数的 增多，计算性能大 幅下降
	4	32	5415.084	
千兆	2	16	14423.560	
	4	32	13982.96	



VASP案例分析

- 扩展性

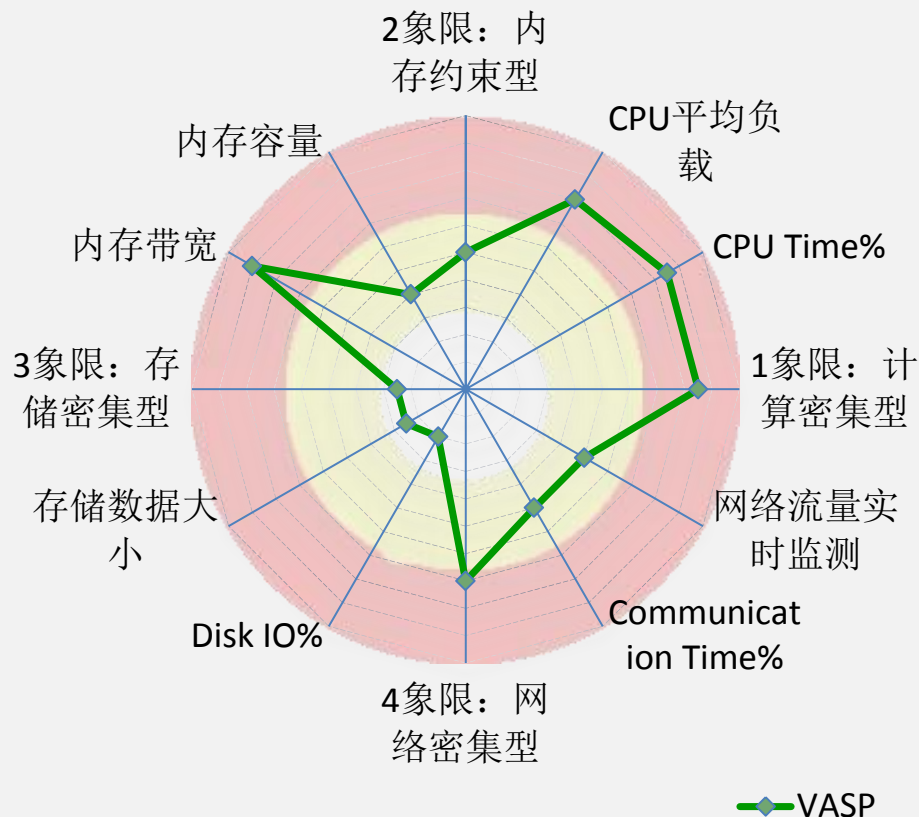
算例	IONS (粒子数目)	NBANDS (能带数目)	内存需求 (MB)	Elapsed Time(s)				
				Nodes=1 cores=1	Nodes=1 cores=12	Nodes=2 cores=24	Nodes=4 cores=48	Nodes=6 cores=72
光学材料	50	476	4000	36018	2729.986	1502	1361	1786



超过6节点加速比开始下降

VASP应用特征小结

VASP应用特征雷达图



计算密集型：主频高对性能影响明显

内存带宽敏感型：高内存带宽需求；一般情况下**48G**内存能够满足大部分应用的需要

磁盘IO：开销不大

网络：高速INFINIBAND网络的可保证较高的并行效率以及加速比。

典型HPC应用特征需求小结

应用领域	典型应用	CPU	内存容量	内存带宽	存储	网络	扩展性
CFD	Fluent						
分子动力学	NAMD						
量子化学	GAUSSIAN						
材料科学	VASP						

注： 信号满格表示对此项需求很高

谢 谢！



inspur 浪潮