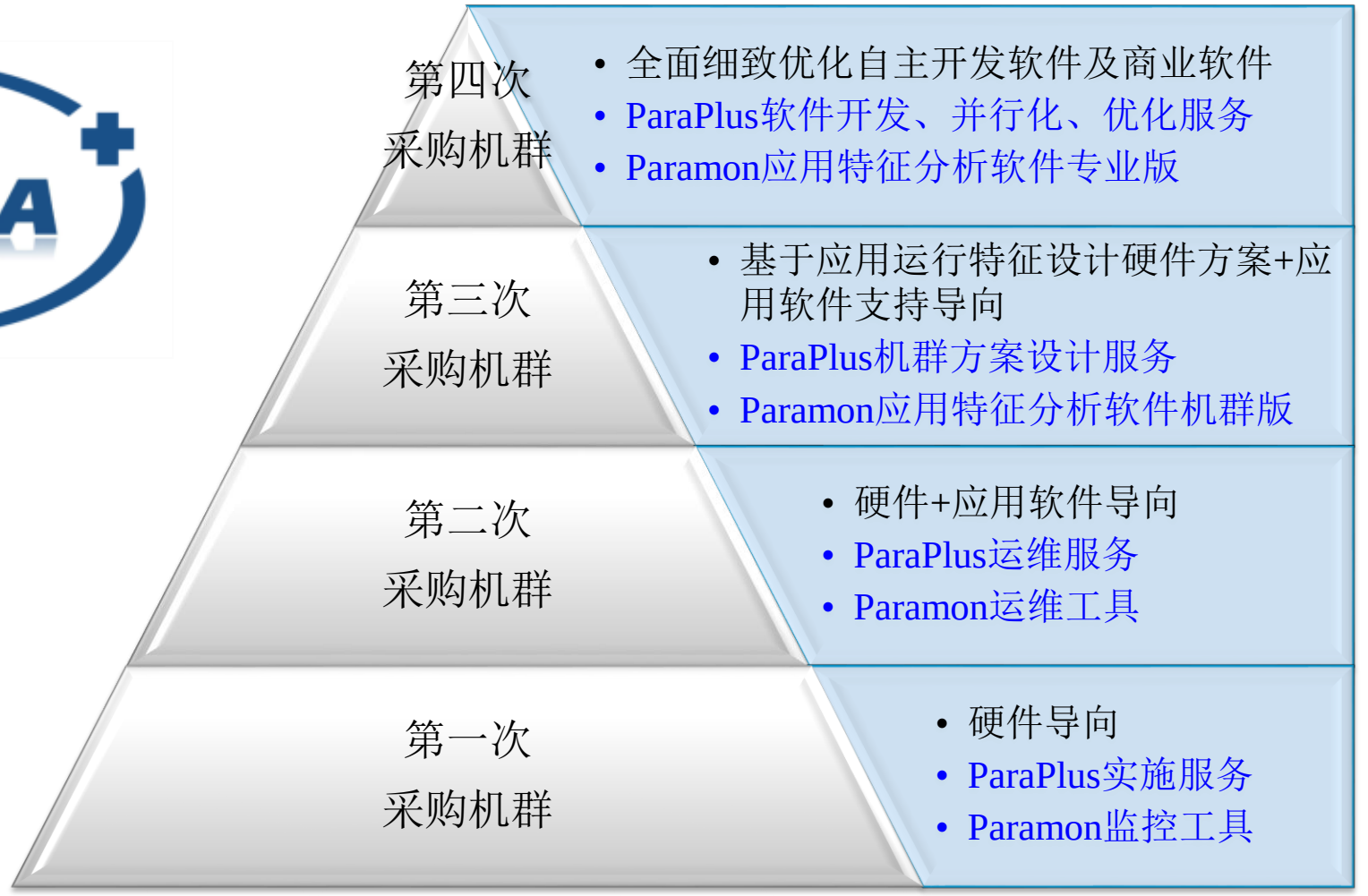
ParaPlus高性能计算数据中心建设和7*24小时在线运维

北京并行科技有限公司
http://www.paratera.com

议题目录

- ParaPlus HPC 专业服务和Lustre
- Paratera OITS 在线运维服务
- Intel与并行的广泛合作

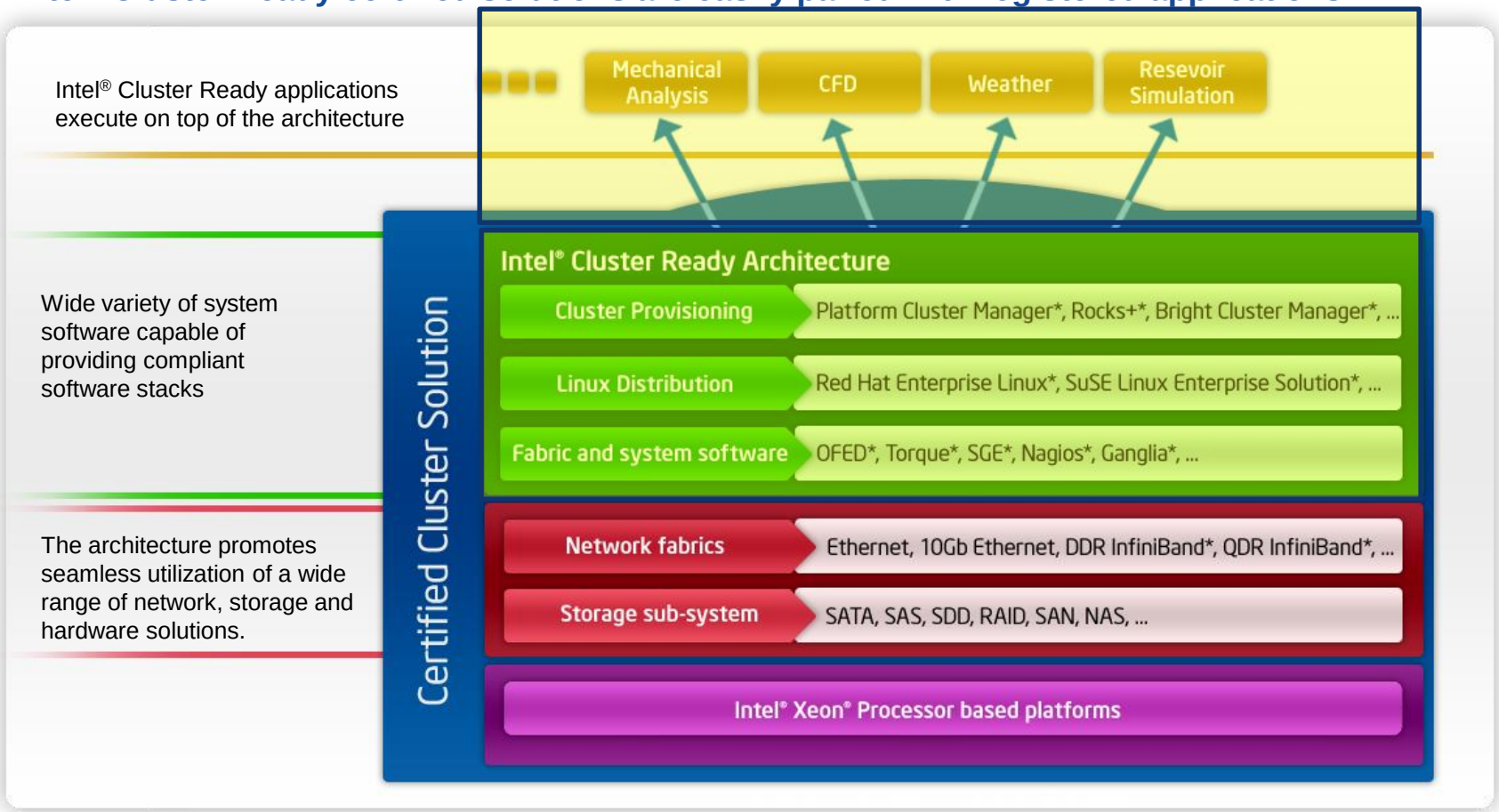
机群用户发展历程


第四次
采购机群
• 全面细致优化自主开发软件及商业软件
• ParaPlus软件开发、并行化、优化服务
• Paramon应用特征分析软件专业版
第三次
采购机群
• 基于应用运行特征设计硬件方案+应用软件支持导向
• ParaPlus机群方案设计服务
• Paramon应用特征分析软件机群版
第二次
采购机群
• 硬件+应用软件导向
• ParaPlus运维服务
• Paramon运维工具
第一次
采购机群
• 硬件导向
• ParaPlus实施服务
• Paramon监控工具

基于Intel® Cluster Ready的HPC方案

One Architecture Meant To Maximize Your Product's Potential Reach

Intel® Cluster Ready certified solutions are easily paired with registered applications



The Right Answer in HPC PARATERA 并行

ParaPlus全方位HPC服务

ParaPlus高性能计算解决方案

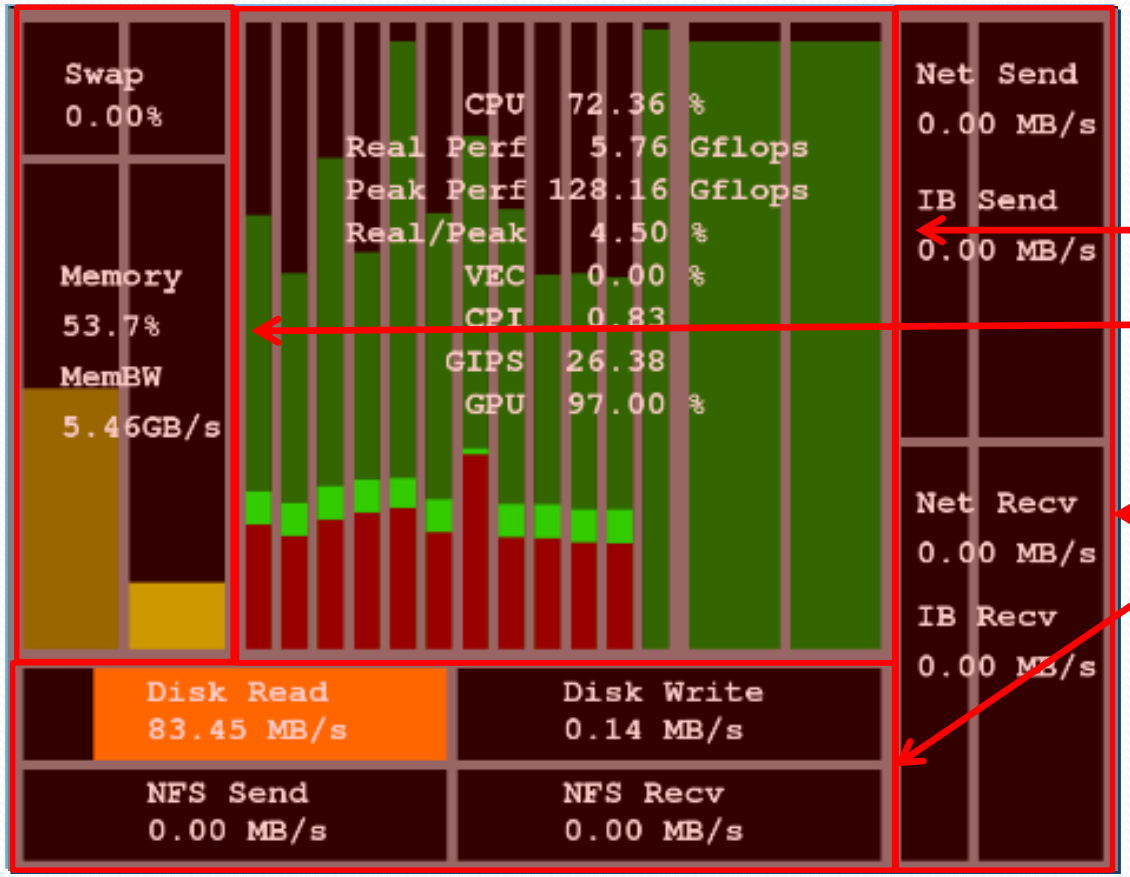
Lustre 并行文件系统

并行科技的 ParaPlus 高性能计算解决方案中并行文件系统采用 Lustre 并行文件系统：

- Lustre 的服务器和存储可以来自不同的厂商；
- 用户可以逐步的动态的扩充文件系统；
  - 通过增加数据服务器以提升聚合带宽
  - 通过增加数据服务器以提升系统整体容量
- Lustre 是基于 GPL 的开源文件系统，还可以从英特尔购买专业的技术支持（并行科技是 Intel Enterprise Edition for Lustre 的授权服务提供商）；

- 以色列Paradigm 地球物理公司在自己的开发测试集群全面采用 Lustre 并行文件系统（总容量 300TB）
- 某研究院 500+ 节点大型高性能计算中心全面采用 Lustre 并行文件系统（总容量 400+ TB）

Paramon指标可视化



关键指标颜色	
CPU USED%/MIC%/GPU%	
CPU Gflops (微架构数据)	
CPU SYS%	
Memory%	
Memory BandWidth (GB/s)	
SWAP%	
Net Send(MB/s)	
Net Recv(MB/s)	
Disk Read(MB/s)	
Disk Write(MB/s)	

Data Center PARATERA 并行 ®

Paramon软件目前采集的数据

基本性能数据模块

- 系统级性能数据
  - CPU USED%利用率
  - CPU SYS%利用率
  - Memory%利用率
  - SWAP%利用率
  - Ethernet Send速率
  - Ethernet Recv速率
  - Disk Read速率
  - Disk Write速率
- 微架构级性能数据
  - Gflops浮点计算性能
  - Memory BandWidth速率
  - CPI，代码执行效率
  - VEC%，向量化比例
  - GIPS，每秒执行指令总数
  - LLCM%，Last Level Cache MISS百分比

机群管理模块

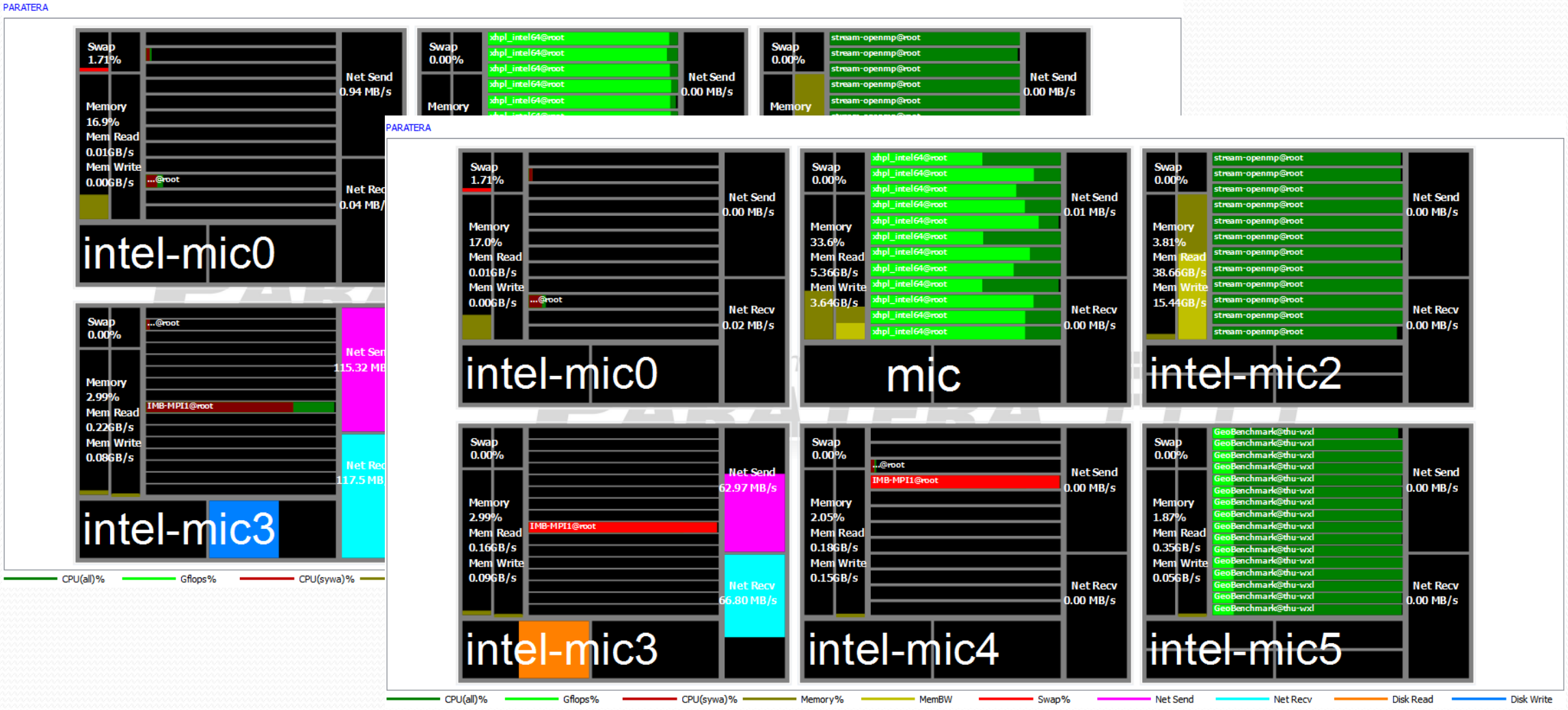
- 远程登录管理
- 远程操作管理
  - 批量节点远程关机/重启
  - 批量节点清除SWAP
  - 自定义节点操作
- 进程管理
- 并行命令管理
- 机器数据
  - 系统硬件配置数据
  - 系统软件配置数据
  - 系统日志
  - 用户信息数据
  - 作业数据
  - License数据

扩展模块

- 网络模块
  - InfiniBand 网络收发速率
- 文件系统模块
  - NFS读写速率
- IO统计模块
  - tps, await, util%
- GPU管理模块
  - NVIDIA GPU%
- MIC数据
  - Intel Xeon Phi MIC%
  - CPI / VPU
- 能耗数据
  - 节点POWER%利用率
- 进程数据

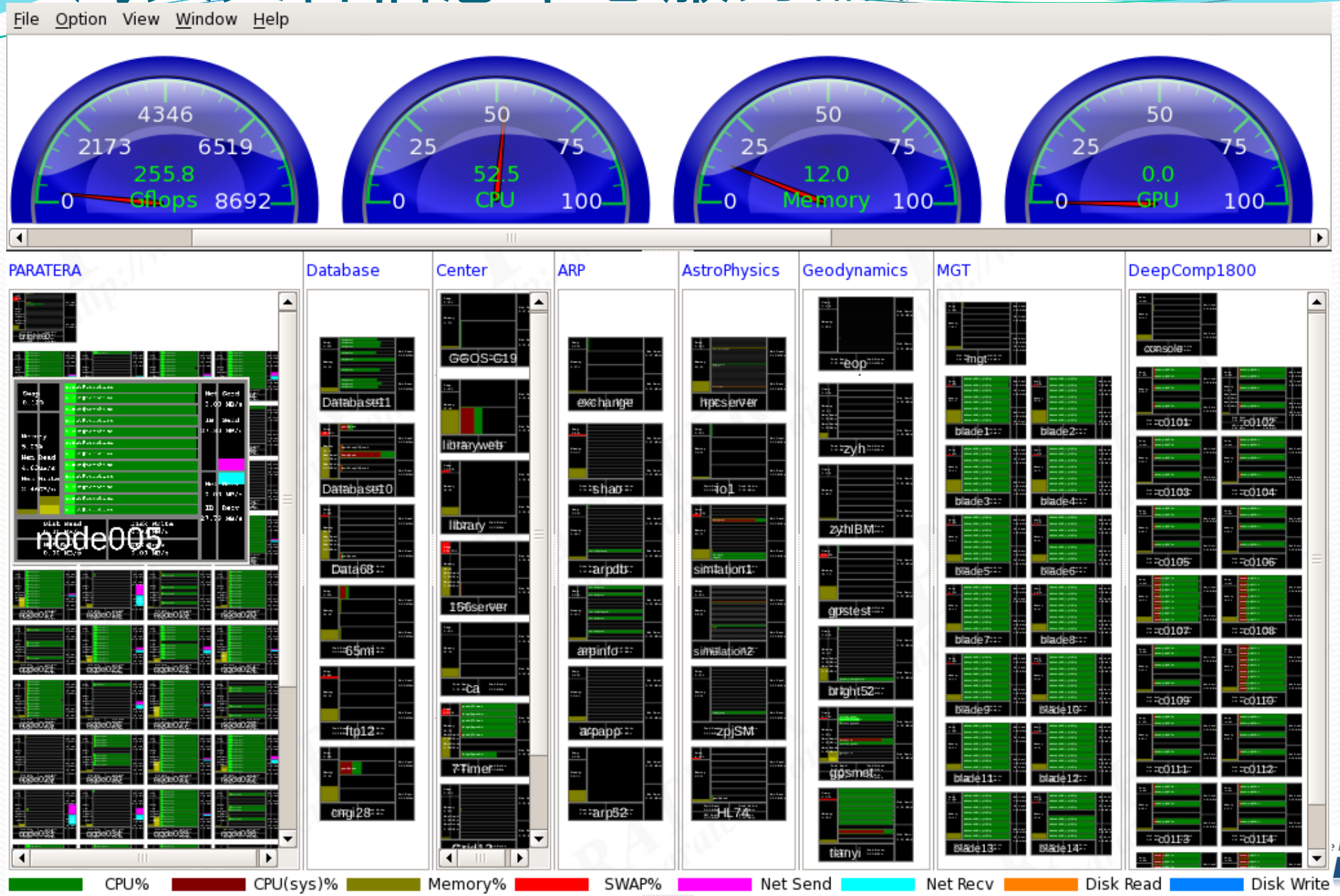
独立于硬件平台，支持上千节点混合机群

Paramon数据采集服务和故障报警



The Right Answer in Data Center
PARATERA 并行 ®

上海天文台信息中心服务器状态



Right Answer in Data Center
RATERA 并行

并行科技支持嫦娥三号项目

- 测量程序的并行化、优化、算法改进
  - 测量处理卫星位置由最初的5分钟处理完成最终达到5秒完成
- 提供高可用实时机群软件系统
- 提供现场技术保障，软件异常Paramon可以在第一时间报警

ParaPlus典型客户

国家电网
STATE GRID
中国电力科学研究院
CHINA ELECTRIC POWER RESEARCH INSTITUTE

中国石油

中国航天

中国移动通信
CHINA MOBILE

中国科学院
CHINESE ACADEMY OF SCIENCES

工业和信息化部电信研究院
China Academy of Telecommunication Research of MIIT

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

The Right Answer in Data Center
PARATERA 并行 ®

在线运维基本介绍

- **传统运维服务**
  - 现状：中国存在大量的中小数据中心，缺少专业的运维团队，缺少专业数据中心管理软件，面对越来越复杂的系统，急待解决日益复杂的管理问题
  - 现有解决方案：购买运维软件、雇佣高级管理员、购买驻场、巡检、上门运维服务等
- **创新7*24小时在线运维**
  - 将现场运维服务转变为通过互联网的数据中心在线运维
  - 减少购买运维软件
  - 减少专业管理员的压力
  - 减少购买上门运维服务

ITS在线运维 7*24
The Right Answer in Data Center PARATERA 并行®

并行科技在线运维监控中心上线

The Right Answer in Data Center
PARATERA 并行 ®

在线运维手机版

PPSN1001803 集群连接异常
PARATERA 并行
62
1
未读
已读
集群名称：PPSN2001803
告警时间：2014-09-14 17
告警信息：PPSN2001803
告警类型：头结点/登录节点
标记为已读：
集群名称：PPSN1001805
告警时间：2014-09-14 17
告警信息：PPSN1001805
告警类型：头结点/登录节点
62 告警信息
PPSN2001806 集群连接异常 201...
PPSN1001807 集群连接异常 201...
PPSN1001805 集群连接异常 201...
PPSN2001803 集群连接异常 201...
查看所有未读告警信息
记为已读
集群名称：PPSN1001805
告警时间：2014-09-14 17:00:00
告警信息：PPSN1001805 集群连接异常
告警类型：头结点/登录节点不可用
31
告警信息
作业列表
告警作业
10414 lxy
10401 liyan
10409 srn
10417 lxy
10419 liyan
作业号
10128 lxy
10124 lxy
10414 lxy
10401 liyang
10409 srn
10417 lxy
10419 liyang
1条作业任务消息
清除全部
作业10128异常结束
2014-08-05 12:00:00

专业的运维报告（周报、月报、年报）

5 5.2 6 用户作业统计

以下 以下 3

5.1

6.2 按用户使用资源统计

● 表 4 用户使用计算资源量统计

用户	CPU 总用时 (单位:s)	墙钟时间 (单位:s)	计算用时 (单位:s)	系统用时 (单位:s)
s	3587282.15	7195917.00	3151631.45	435636.97
h	1989081.68	470855.00	10.59	5.48
w	1479373.58	2079026.00	1262.48	433.88
x	1419856.98	3512936.00	1188657.76	231199.22
e	1299268.78	3341648.00	1081466.99	217801.79
g	921021.03	327534.00	8.06	3.20
y	510172.55	242735.00	389.86	105.70
lj	367516.61	371560.00	67.86	30.39
w	1587.98	1836.00	25.66	11.65
lxf	1.13	3.00	0.74	0.39

挂载

/disk4

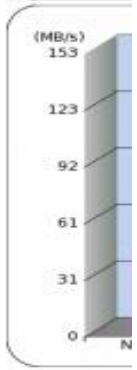
/disk5

/disk1

/disk1

/disk1

/disk2



● 图 4 用户作业提交量统计图

The Right Answer in Data Center PARATERA 并行 ®

丰富灵活的报表

- 提供ParaPortal网站服务，通过独有用户名及登录密码，自主查询机群运行的历史数据信息，生成精美的统计分析报告，为科学决策提供依据

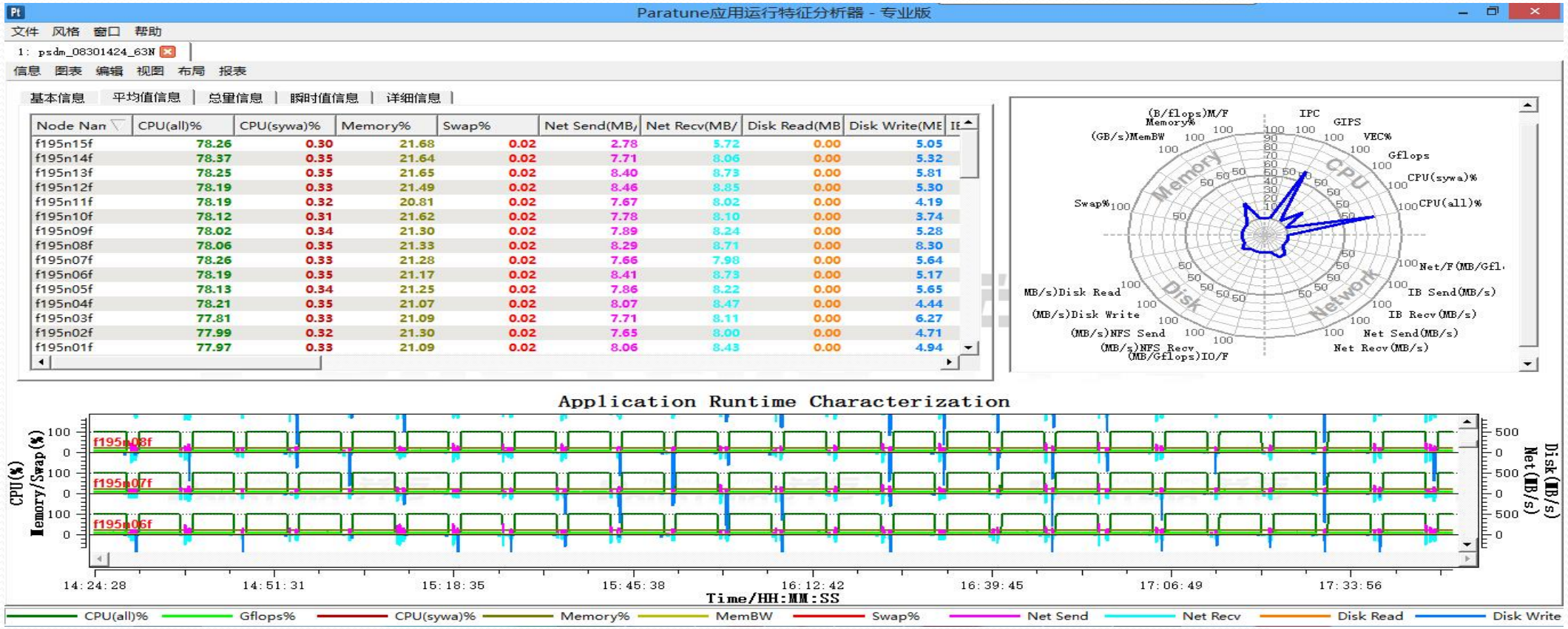
机群体检

- 定期或在运行大作业前为机群进行健康体检，筛除问题节点，避免问题节点成为瓶颈影响机群的整体运行效率

主机名	CPU浮点计算量 (Gflops)	内存带宽 (MB/s)	磁盘读 (KB/s)	磁盘写 (KB/s)	网络延迟 (us)	网络带宽 (Mb/s)
compute-0-2	47.98	6139.7	85847.01	95274.38	6.1	950.93
compute-0-3	48.50	6140.8	87392.21	96975.98	6.4	1054.13
compute-0-4	48.99	6140.5	68487.62	97102.92	6.37	955.43
compute-0-5	49.61	6141.3	60563.05	72687.54	6.37	955.43
compute-0-6	50.12	6141.2	63623.41	100535.08	6.44	1057.92
compute-0-7	50.65	6140.1	85116.3	95632.05	6.44	1057.92
compute-0-8	51.13	6140.1	90434.68	135852.28	6.06	663.94
compute-0-9	51.80	6141.1	80075.04	96527.44	6.06	663.94
compute-0-10	53.13	6140.5	80822.13	95560.77	6.51	957.34
compute-0-11	52.28	6140.6	79778.6	97122.47	6.72	956.65
compute-0-12	52.87	6140.5	92863.99	100410.74	6.72	956.65
compute-0-13	48.47	6140.9	73418.14	95177.05	6.51	957.34

应用程序优化

- 针对用户应用程序，利用Paratune提供详细分析建议，辅助用户提高程序性能及解决系统性能瓶颈



运营大数据分析、挖掘、可视化

- DEMO：大赛比赛用机数据采集结果分析
- DEMO：并行应用优化大赛应用运行特征记录文件
- DEMO：VASP应用运行问题
- DEMO：Hadoop数据加载性能问题
- DEMO：大规模应用运行中有慢节点
- DEMO：多种程序混合运行
- DEMO：……

Paramon online live Demo

- 基于>500种已采集应用运行特征数据
- 性能数据挖掘：
  - Gflops%比例低，浮点计算效率低，需要优化
  - CPI高，需要提高代码效率
  - LLCM%高，修改访存模式，减少对内存带宽压力
  - X87%比例高，辅助编译器提高SSE指令使用率提高性能
  - VEC%比例低，将标量SSE指令转为向量SSE指令
  - E5平台上，SSE%高，AVX%为0：重编译后性能加倍
  - MIC平台上，VPU比例低，需要提高512位FMA指令使用
  - ……

用户收益

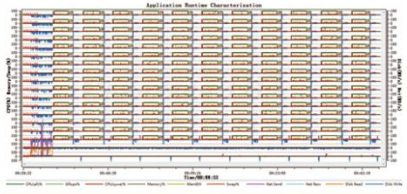
- 通过并行科技的7*24小时监控中心实时查看大量数据中心的运行情况，补充用户晚上和周末监控人员的不足，为用户实现全天候的保驾护航
- 变被动响应为远程主动发现问题，并有实时历史数据可以查看、分析，在得到用户授权的情况下可以主动修复故障，将服务响应时间极大幅度的减少
- 长时间的在线运维监控可以自动生成周报、月报、年报等统计分析报表，满足用户对数据中心的宏观把控
- 提供最终用户关心的业务运行和资源使用统计报告
- 提供领导关心的HPC数据中心运营看板

数据中心海量应用特征库

ITS在线运维

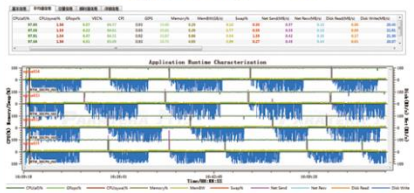
石油勘探

地震资料处理叠前时间偏移PSTM应用

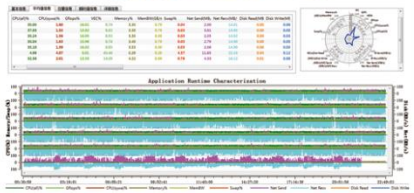


地震资料处理叠前深度偏移PSDM应用

地震资料处理逆时偏移RTM应用



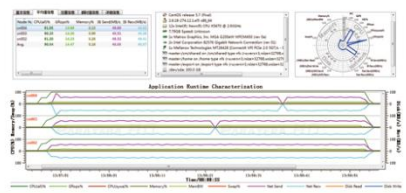
地震资料处理叠前去噪RNA应用



计算化学

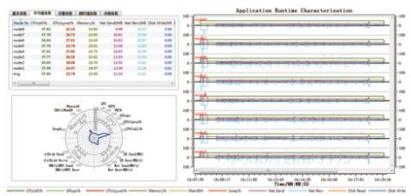
NAMD应用

Gromacs应用



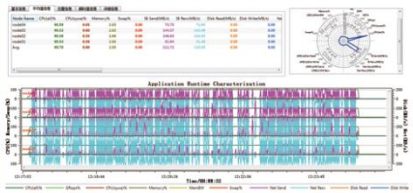
Gaussian09应用M_18s算例

VASP应用



环境科学

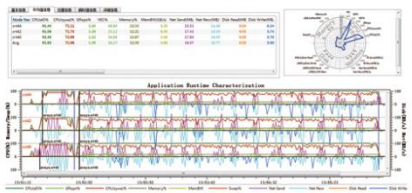
WRF应用wrf_em_b_wave算例



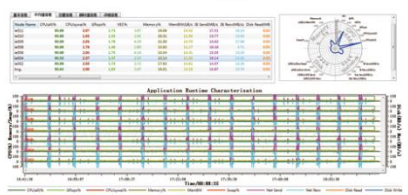
CCSM应用

工程仿真

ANSYS14应用

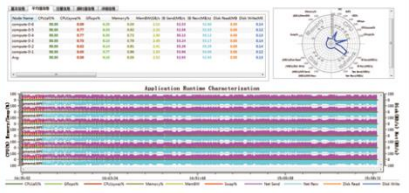


Fluent应用



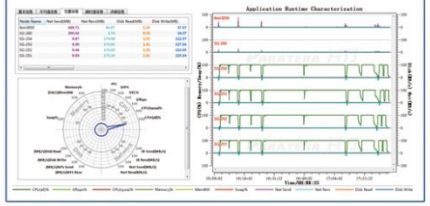
生命科学

Amber12应用 pmemd算例



数字内容制作

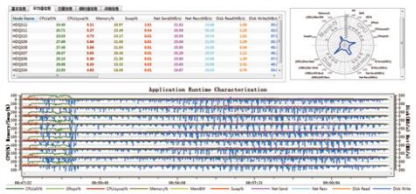
3DMax动漫渲染应用



大数据应用

Hadoop HDFS应用

Hadoop Hive数据仓库MapReduce应用

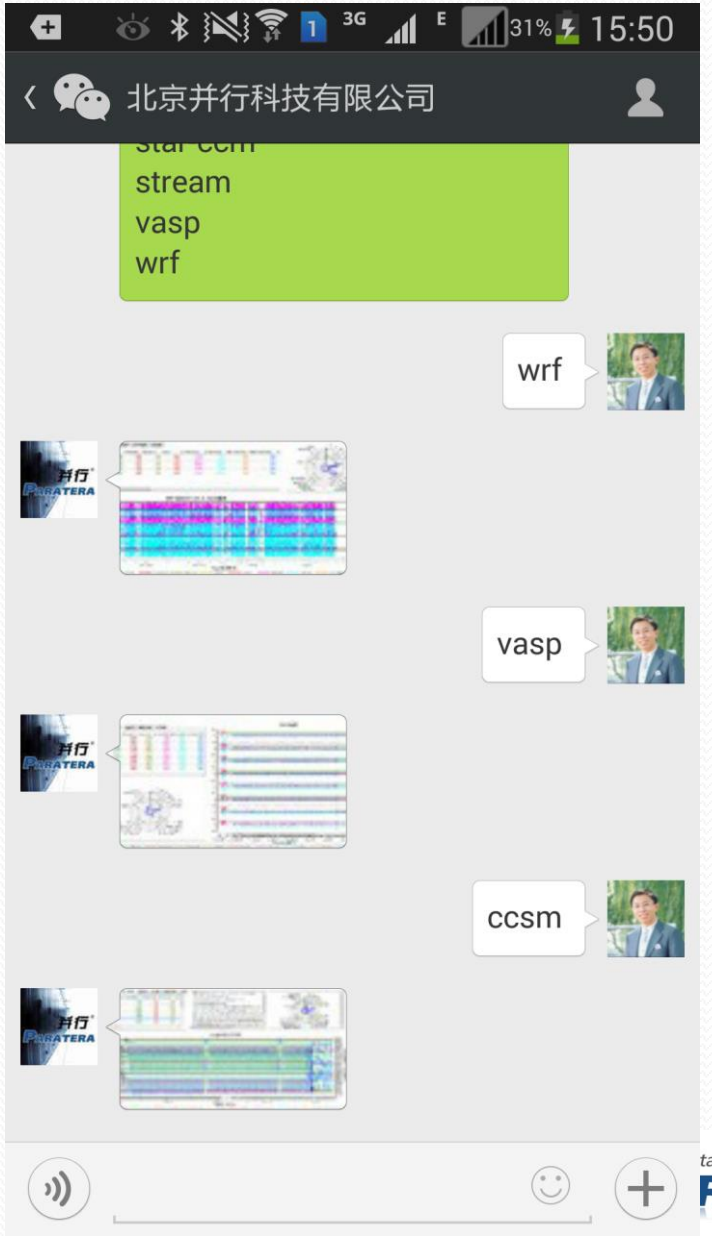


The Right Answer in Data Center PARATERA 并行®

并行科技微信公共号
—应用运行特征服务

ITS在线运维

(查找公共号”并行”)



与Intel携手Intel Oil&Gas HPC Seminar 2014

全国高性能计算学术年会/并行应用大赛

The Right Answer in Data Center
PARATERA 并行 ®

携手并行，共赢未来！

北京并行科技有限公司

http://www.paratera.com