

**致繁至简 · 极远及近**



7\*24小时数据中心在线运维  
和运营大数据分析

北京并行科技有限公司

<http://www.paratera.com>

# 陈健 Justin Chen



- 陈健现任**北京并行科技有限公司**CEO。并行科技 (<http://www.paratera.com>) 是专业高性能计算数据中心服务商。
- 2005-2010年在**英特尔**中国历任高性能计算架构师、资深性能优化工程师，负责中国大型HPC及前沿HPC项目系统架构设计、HPC系统的优化及高性能计算技术推广，研究方向为高性能计算机系统架构、并行计算、应用软件及系统调优。在2008年参与完成了中科院超算中心100万亿次超级计算机的设计，并最终实现实测性能超过100万亿次，全球排第19名、中国Top100第二名。于2007年完成胜利油田高性能计算机的性能优化设计，实现实测性能18.6万亿次，2007年中国排名第一。
- 2002-2005年在**联想集团**高性能服务器事业部工作，作为方案负责人参与了联想深腾万亿次、四万亿次超级计算机的研制。
- 2002年毕业于**清华大学**工程力学系，获流体力学博士学位，期间在荷兰TUDelft大学做访问学者一年，参与了荷兰国家计算中心SGI万亿次超级计算机的研制和测试

# ParaPlus全方位HPC服务



# 7\*24小时数据中心在线运维介绍

## ● 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

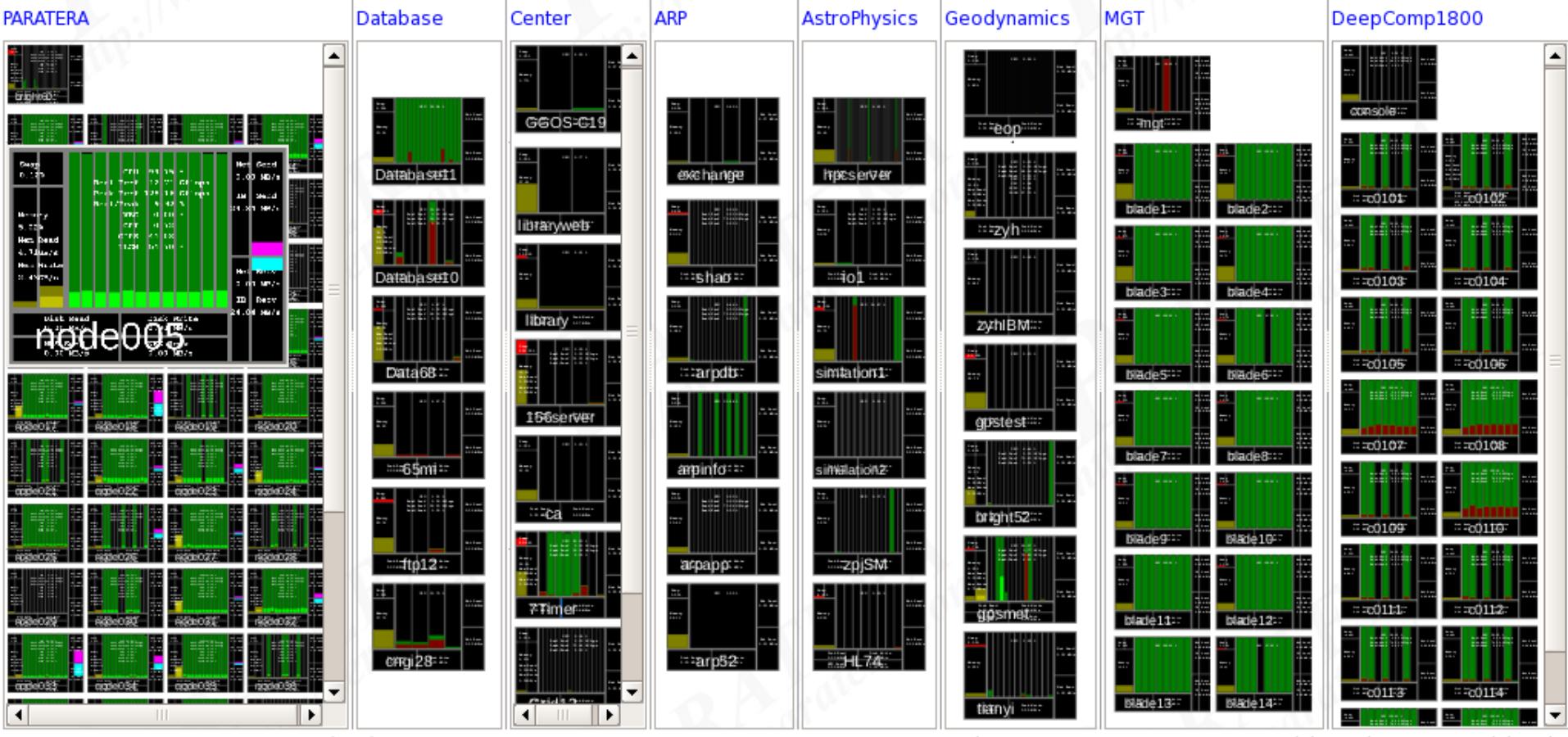
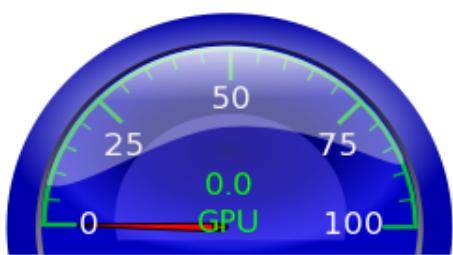
- 将现场运维服务转变为通过互联网的数据中心在线运维
- 减少购买运维软件
- 减少专业管理员的压力
- 减少购买上门运维服务

**致繁至简 · 极远及近**

# 某信息中心服务器状态



File Option View Window Help



# 并行科技在线运维监控中心上线



CCF YOCSEF 走进并行科技

ITS 在线运维

The Right Approach to HPC  
PARATERA 并行

多彩视界



PARATERA 并行

# OITS在线运维

## 让超算中心运营更高效，让研究人员更专注

### 用户端

用户自助全自动部署后台数据采集模块，**CPU**消耗低于**0.1%**，**4KB/s**数据流量，防火墙打开数据传输端口，只对并行科技**IP**开放特定端口

### 云端

1 - **Paramon**数据中心实时可视化，程序自动分析值守

2 - **ParaAlarm**手机**App**

对管理员，系统维护实现无人值守，只需要处理报警事件，由管理员、厂商或者并行科技远程主动修复故障，让用户体验稳定、可靠、好用的超算环境

对用户，实时报告作业运行情况，作业异常推送报警事件，作业结束推送消息（开发中）

3 - **ParaPortal**历史数据分析

4 - **ParaReport**自动周报、月报、年报

5 - **ParaDashboard**数据中心运营看板，可自定义，从底层数据直接生成，提供准确数据供领导决策

6 - **Paratune**基于秒级历史数据的故障和性能分析，快速发现并定位系统运行故障和瓶颈

7 - **Intel Software**系统级、代码级优化，高效利用**Intel**软件提升系统软硬件整体运行效率

# 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

# 在线运维手机版

 PPSN1001803 集群连接异常

**PARATERA 并行**

未读 已读

**62 告警信息**

集群名称 : PPSN200180  
告警时间 : 2014-09-14 11:11:11  
告警信息 : PPSN200180  
告警类型 : 头结点/登录节点不可用  
标记为已读 :

集群名称 : PPSN100180  
告警时间 : 2014-09-14 11:11:11  
告警信息 : PPSN100180  
告警类型 : 头结点/登录节点不可用

[查看所有未读告警信息 ➔](#)

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

 PARATERA 并行

告警信息 作业列表 告警作业

作业号	作业
10414	lxy
10401	liya
10409	srn
10417	lxy
10401	liyan
10414	lxy
10409	srn
10417	lxy
10419	liyan
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  
① 2014-08-05 12:00:00

作业号 作业

# 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

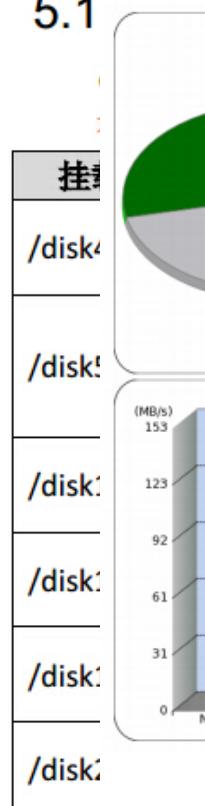
5.1

### 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	墙钟时间	计算用时	系统用时
	(单位:s)	(单位:s)	(单位:s)	(单位:s)
s	3587282.15	7195917.00	3151631.45	435636.97
h	1989081.68	470855.00	10.59	5.48
v	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
v	1587.98	1836.00	25.66	11.65
lxf	1.13	3.00	0.74	0.39

● 图 4 用户作业提交量统计图



# 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



The screenshot displays the ParaPortal web interface, specifically the 'Real-time Report' section for cluster monitoring. The left sidebar includes links for System Management, Asset Management, Equipment Monitoring, Statistical Analysis (Real-time Reports, Node Statistics, Cluster Statistics), Report Library, and Custom Reports. The main content area shows three line charts for a cluster named 'Tsinghua\_gonglab\_new' from March 18, 2014, to June 18, 2014.

- Top Navigation Bar:** Includes icons for system management, asset management, equipment monitoring, and statistical analysis. It also shows user information (0 notifications, 1 message, 0 emails) and a welcome message for 'System Administrator'.
- Breadcrumb:** Current location: Home > Statistical Analysis > Real-time Report > Cluster Statistics.
- Time Selection:** A dropdown menu allows selecting time intervals: 半小时 (Half Hour), 一小时 (Hour), 今天 (Today), 本周 (This Week), 本月 (This Month), 三个月 (Three Months) (highlighted with a blue border), 半年 (Six Months), and 一年 (One Year).
- Cluster Selection:** A dropdown menu set to 'Tsinghua\_gonglab\_new'.
- Chart 1: CPU Utilization Rate (%)** (Left): Shows a green line chart with values ranging from 0 to 100. The utilization rate remains high (around 90%) until June 10th, then drops sharply to around 60% and stabilizes. X-axis: Jun 4, Jun 6, Jun 8, Jun 10, Jun 12, Jun 14, Jun 16, Jun 18.
- Chart 2: Memory Usage Rate (%)** (Right): Shows two lines: yellow for memory usage and red for swap usage. Both are near zero until June 10th, then drop to near zero. X-axis: Jun 4, Jun 6, Jun 8, Jun 10, Jun 12, Jun 14, Jun 16, Jun 18.
- Chart 3: Network In/Out Rate** (Bottom Left): Shows a bar chart with values ranging from 0 to 1. The rate is consistently at 1 throughout the period. X-axis: Jun 4, Jun 6, Jun 8, Jun 10, Jun 12, Jun 14, Jun 16, Jun 18.
- Chart 4: Local Disk Read/Write Rate** (Bottom Right): Shows a bar chart with values ranging from 0 to 1. The rate is consistently at 1 throughout the period. X-axis: Jun 4, Jun 6, Jun 8, Jun 10, Jun 12, Jun 14, Jun 16, Jun 18.

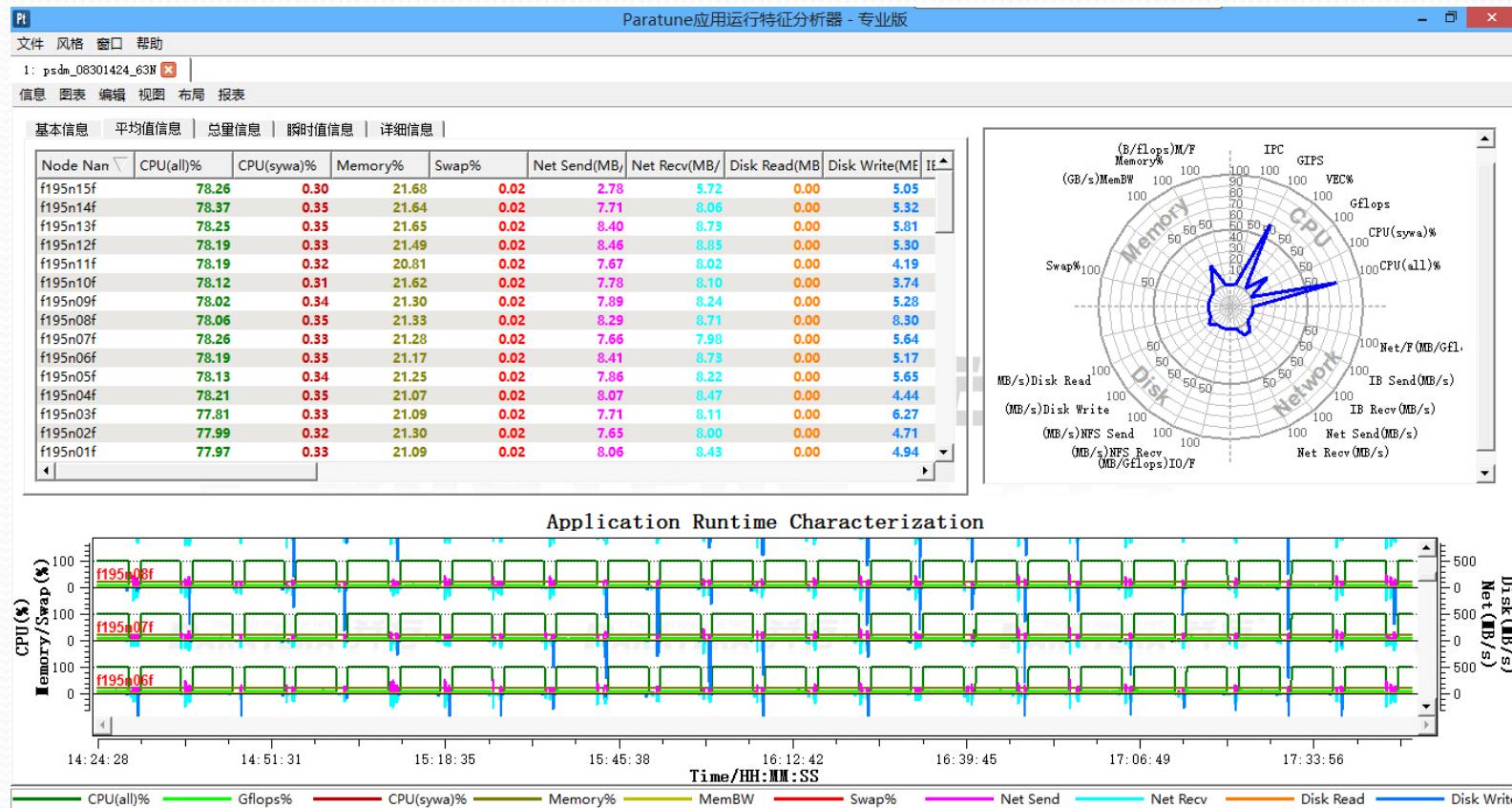
# 机群体检

- 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

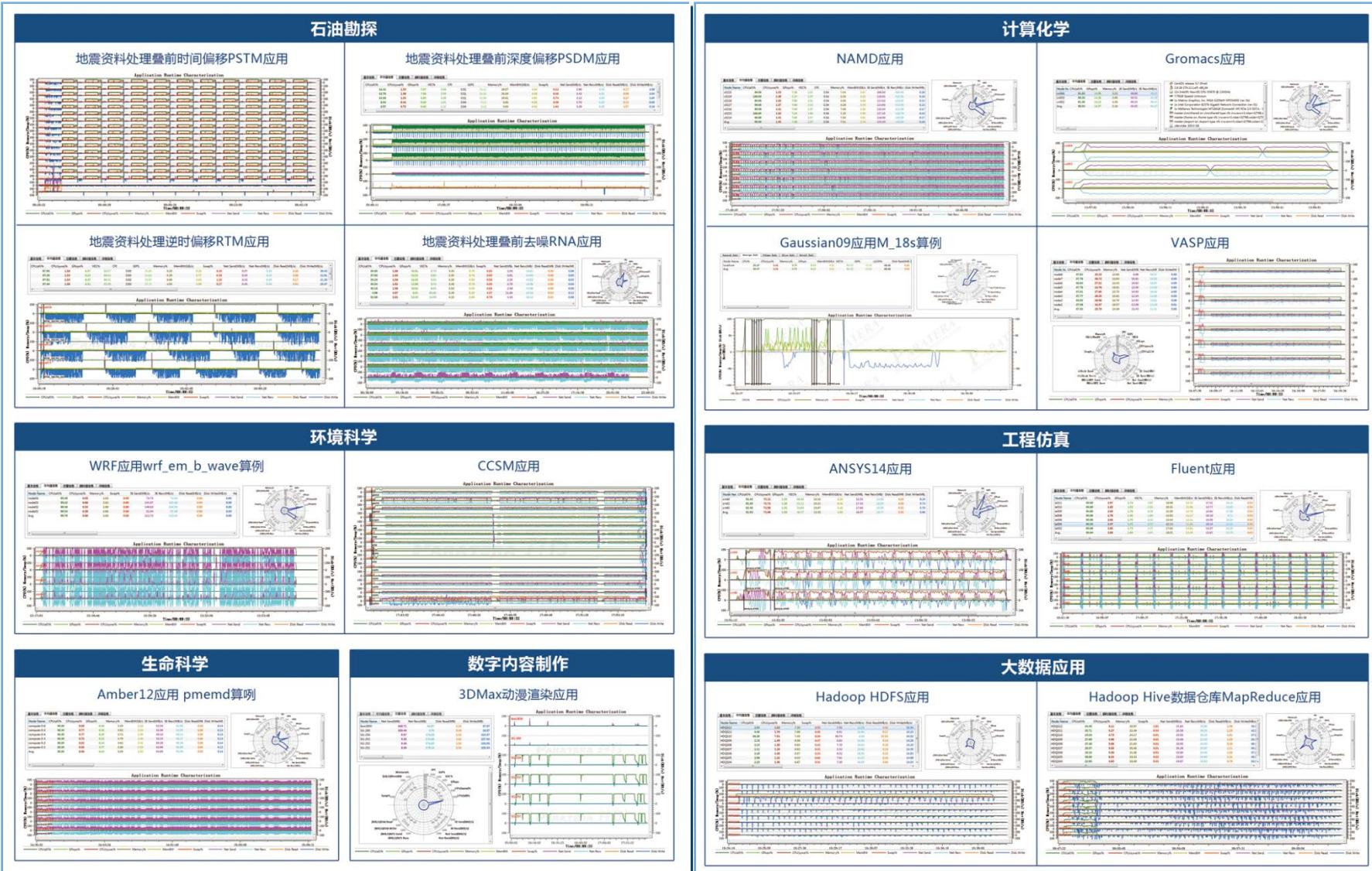
- 针对用户应用程序，利用Paratune提供详细分析建议，辅助用户提高程序性能及解决系统性能瓶颈



## 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

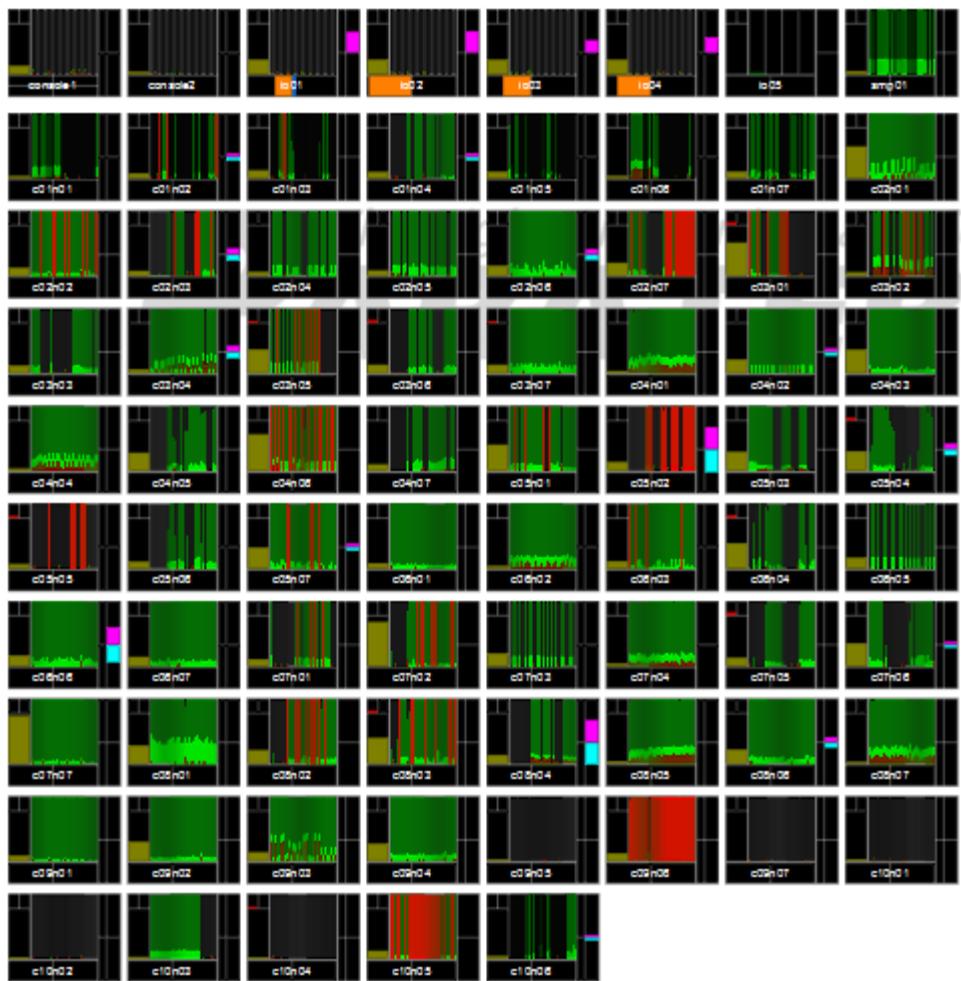
# 数据中心海量应用特征库



Performance Metrics for HPC Applications									
应用名称	时间占比	浮点峰值	平均浮点性能	内存读写平均带宽	本地磁盘平均IO	IB平均IO	CPU利用率峰值	CPU平均利用率	系统CPU平均占比
vasp	33.4971	55.3197	24.5348	14.7067	0.0609774	40.4301	100	90.462	0.572622
cesm	16.2914	7.81748	3.48189	4.72391	0.0511071	49.1703	99.9673	79.1891	0.745152
oceanM	10.9411	10.1355	9.2366	7.68755	0.00451452	165.7	100	99.349	0.373214
imp	8.00822	5.63389	3.88936	3.28481	0.0169285	34.3847	100	92.7057	0.17969
Gaussian	2.6964	98.4595	14.0791	7.00442	0.00820493	2.01839	100	81.6251	0.713142
lim	2.51249	4.15869	0.787213	0.414309	0.00602521	3.40864	16.7518	4.95766	0.367733
castep	2.42163	79.6827	29.0472	13.4754	0.676469	7.91233	100	92.1288	7.28084
sbspsm	2.20981	4.78455	1.92755	0.432583	0.00836993	1.23078	99.4139	72.5009	0.277058
swan	1.78265	1.48628	0.198289	0.146504	0.0030771	1.30165	87.7983	46.6239	0.460762
gromacs	1.44799	46.647	18.8783	7.49172	0.0261395	7.6496	100	99.068	0.189774
lodestar	1.44078	16.4365	14.7846	1.79186	0.00211625	0.295823	83.3294	78.3552	1.07824
wrf	1.43668	1.31363	0.355859	0.239231	0.0021978	3.58248	6.49849	3.38077	0.339132
fvcom_sph	0.9255	12.6312	11.3862	4.48049	0.0198671	15.4814	99.9995	95.2077	0.268714
fvcom	0.883585	12.7624	11.7904	4.36361	0.0148211	11.3722	100	98.3573	0.277469
fluent	0.748406	7.22421	3.49431	5.2175	0.00452232	26.0167	100	73.5896	0.899127
pim	0.467961	7.84933	1.13309	0.594526	0.00871249	2.9119	22.511	5.8948	0.798571



PARATERA Refresh(F5)



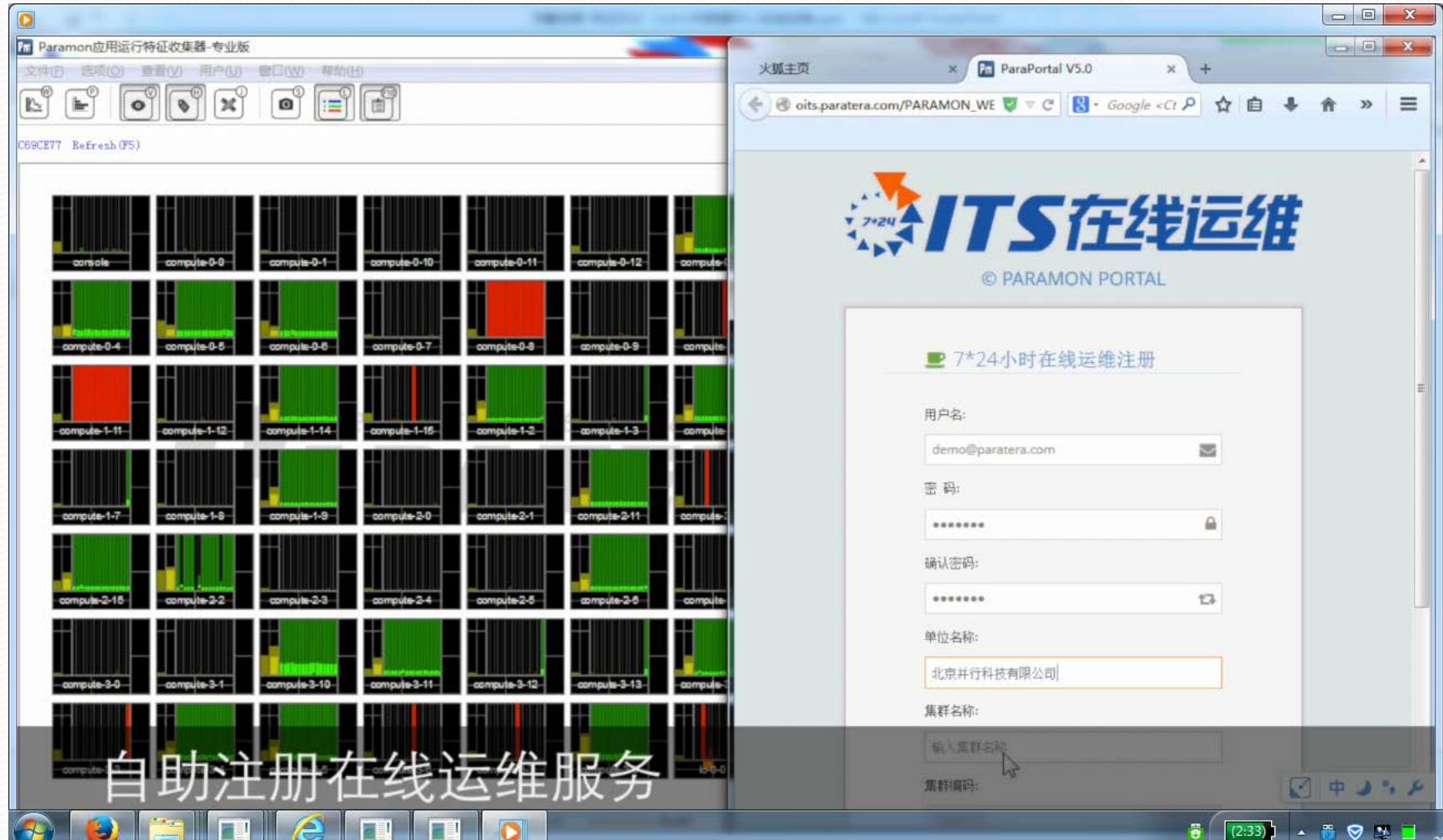
## 应用性能列表

应用名称	最大Gflops	平均Gflops
_ommain	2.111280	0.315993
ansys.e140	21.887600	10.330800
apops	14.218000	4.920450
athena	9.062360	8.661070
blastall	9.464830	1.306750
bowtie	1.114800	0.205367
cafemol	2.755240	1.459980
cam	0.899390	0.846457
castep	28.269400	2.494950
ccsm	4.406920	0.575910
cesm	7.975770	0.482326
charmm	11.141300	10.516800
clm	4.162160	2.338360
cmscan	8.138420	6.740860
cpl	0.797182	0.298690
craft	0.341591	0.218514
cc_color	4.621610	2.626170

可用资源 应用性能

CPU(all)% Gflops% CPU(sywa)% Memory% MemRW Swap% Net Send Net Recv Disk Read Disk Write

# 7\*24小时数据中心在线运维DEMO



自助注册在线运维服务

致繁至简 · 极远及近