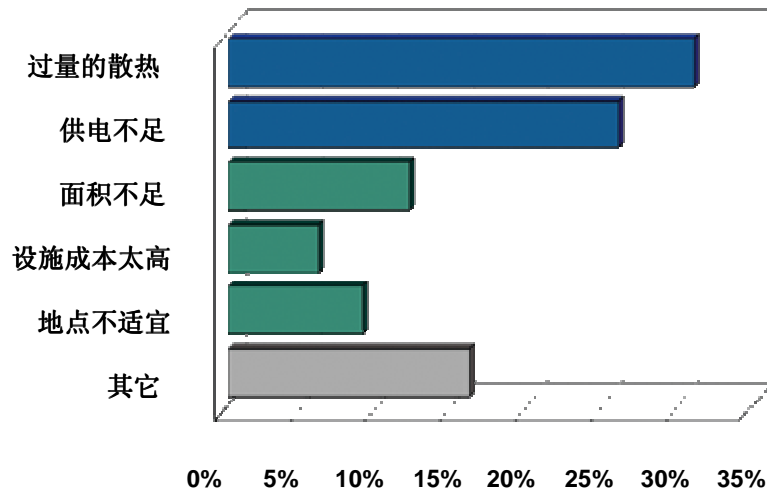# 为什么需要绿色数据中心？
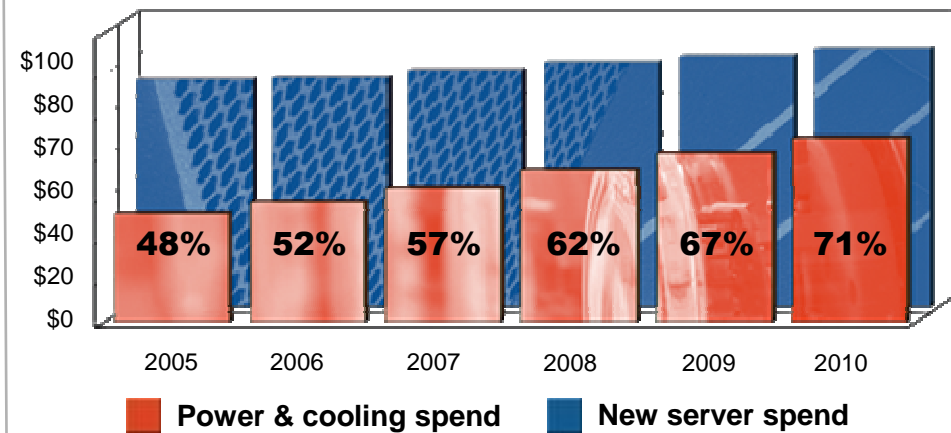
- 高耗电密度和迅速增长
- 每平方米能源消耗10倍甚至100倍于普通办公大楼
- 对供电系统的潜在影响巨大
- 2005年美国数据中心消耗约为450亿度电
- 以目前的发展速度，能源需求可能在5年内倍增

数据中心最大的设施问题是什么？ Gartner 2006

问题	百分比
过量的散热	~32%
供电不足	~27%
面积不足	~13%
设施成本太高	~8%
地点不适宜	~10%
其它	~18%

0% 5% 10% 15% 20% 25% 30% 35%

服务器安装的供电和冷却成本 ($US)

年份	Power & cooling spend
2005	48%
2006	52%
2007	57%
2008	62%
2009	67%
2010	71%

■ Power & cooling spend   ■ New server spend

IBM Power6[T] 拥有的节能减排功能

POWER6 芯片级功能

- **Power Reduction:** Monitor & reduce power to idle logic within cores
- **NAP Mode:** Power off inactive cores, restore power when needed
- **Thermal Tuning:** Sensors monitor & reduce power to overactive circuits
- **Virtualization:** Moving running UNIX and Linux operating system workloads from one POWER6 server to another.

System POWER6 服务器级功能

- **Enhanced System Design & Implementation::** Improved server Performance / WATT uplift over POWER5*.
- **EnergyScale I/O:** Powering off PCI slots not being used
- **Variable Fan Speed (10,500 – 5500 RPM):** Reduces power to fans (1/3 of total server power) by up to 45% based on ambient temperature**
- **Rear Door Heat Exchanger::** Cools exhaust air from 19 & 24" rack, removes up to 60% of the heat[#]
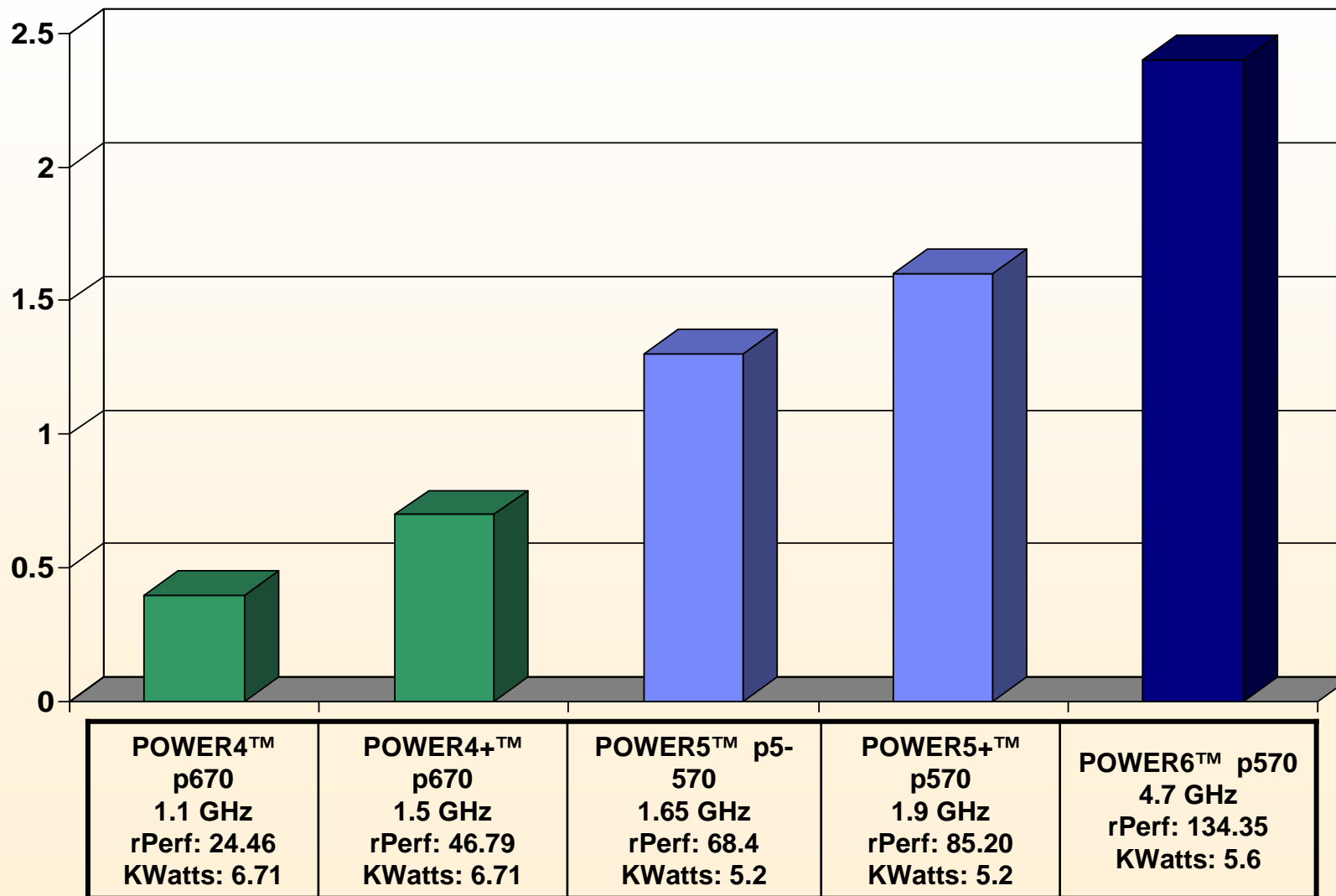
* Based on various SPEC benchmarks; IBM p570 POWER6 result to be submitted on 5/21/07
** Based on IBM internal measurements
[#] IBM press release; 05/10/07; http://www.ibm.com/press/us/en/pressrelease/21517.wss

IBM POWER Systems

Performance per Watt

rPerf per KWatt

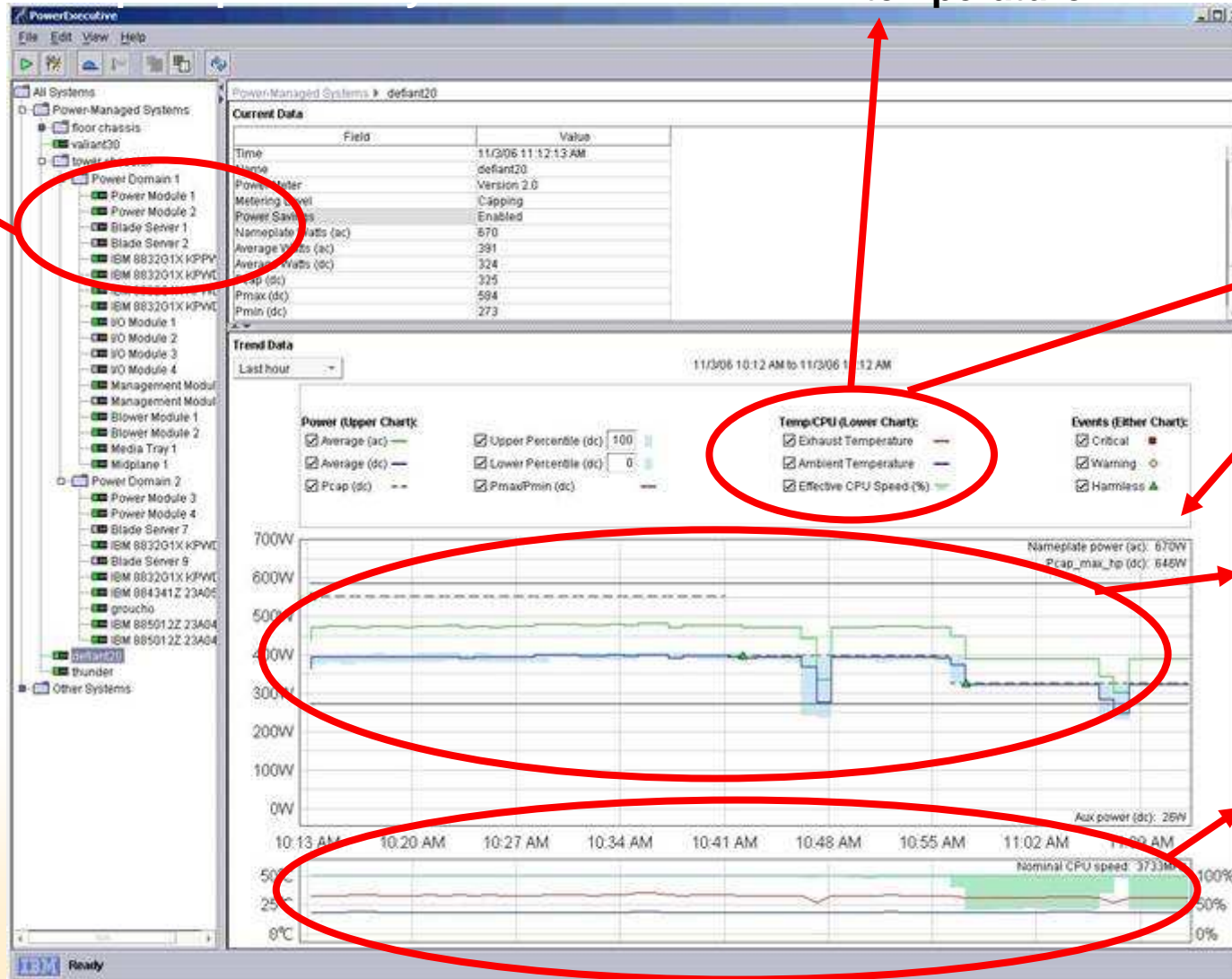| POWER4™ p670 1.1 GHz rPerf: 24.46 KWatts: 6.71 | POWER4+™ p670 1.5 GHz rPerf: 46.79 KWatts: 6.71 | POWER5™ p5-570 1.65 GHz rPerf: 68.4 KWatts: 5.2 | POWER5+™ p570 1.9 GHz rPerf: 85.20 KWatts: 5.2 | POWER6™ p570 4.7 GHz rPerf: 134.35 KWatts: 5.6 |

IBM POWER Systems

温度和功耗的管理
p520/p550服务器配置了TPMD (Thermal Power Management Device)

View inlet and exhaust temperature

Manage Power at the rack and server level

Track heat emitted

Compare rack actual power vs. Label Power

Trend power use over time

Trend temperature over time

IBM 先进的虚拟化技术

1967	1973	1987	2001	2004	2007	2008
IBM develops **Hypervisor** that would become VM on the mainframe	IBM announces first machines to do **Physical Partitioning**	IBM announces **LPAR on the mainframe**	IBM introduces **LPAR in POWER4 based systems with AIX / Linux**	**Advanced POWER Virtualization** POWER Hypervisor	**Partition Mobility** **AIX WPAR**	**PowerVM** **Lx86 Support**

"In our opinion, they [IBM POWER servers] bring mainframe-quality virtualization capabilities to the world of AIX."

- Ulrich Klenke, CIO, rku.it

January 2006

PowerVM on IBM System p servers

Timeline reference http://www.levenez.com/unix/history.html#01
Client quote source: rku.it case study published at http://www.ibm.com/software/success/cssdb.nsf/CS/JSTS-6KXPPG?OpenDocument&Site=eserverpseries

IBM Systems

IBM System p领先的虚拟化技术和功能

Advanced POWER Virtualization

分区负荷管理器
- 自动平衡对处理器和内存的请求

集成虚拟化管理器
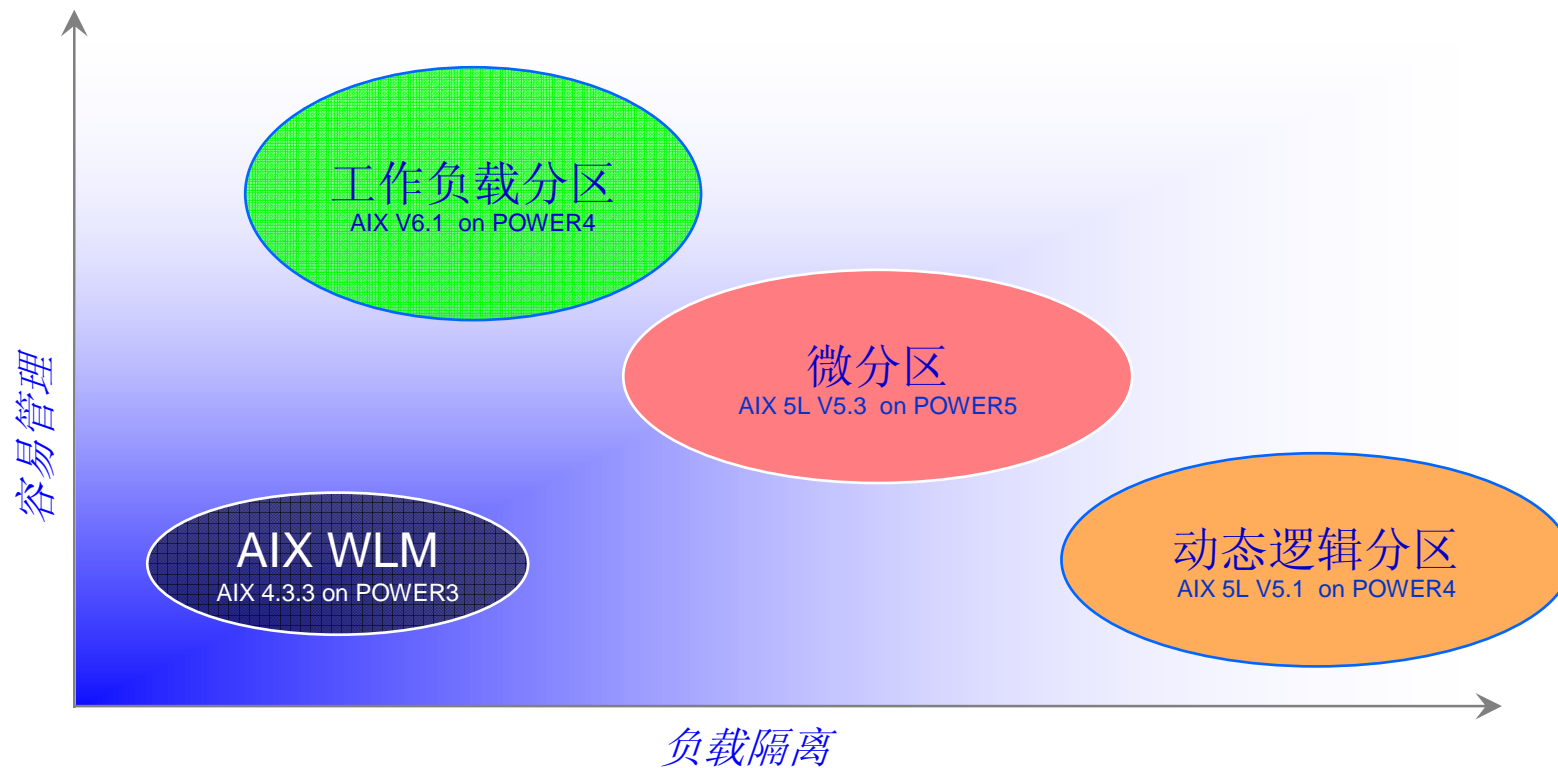- 不需要HMC（硬件控制台）便可管理逻辑分区

虚拟I/O服务器
- 简化以太网、SCSI和光纤通道连接

实时分区迁移
- 将逻辑分区在P6服务器之间动态迁移，不中断运行

Micro-Partitioning™ (POWER Hypervisor)
- 单个处理器可用于创建多至10个的微分区——单台服务器可创建多至254个
- 动态调整分区的大小（处理、内存等）而不中断运行
- 通过类似System z的PR/SM硬件微码实现——精简、优化、整合

业界领先的虚拟化技术，支持AIX和Linux操作系统
实现在硬件实用率和灵活性的显著提高

IBM POWER Systems

Power 服务器提供了灵活的资源管理方式

容易管理

负载隔离

工作负载分区
AIX V6.1 on POWER4

微分区
AIX 5L V5.3 on POWER5

AIX WLM
AIX 4.3.3 on POWER3

动态逻辑分区
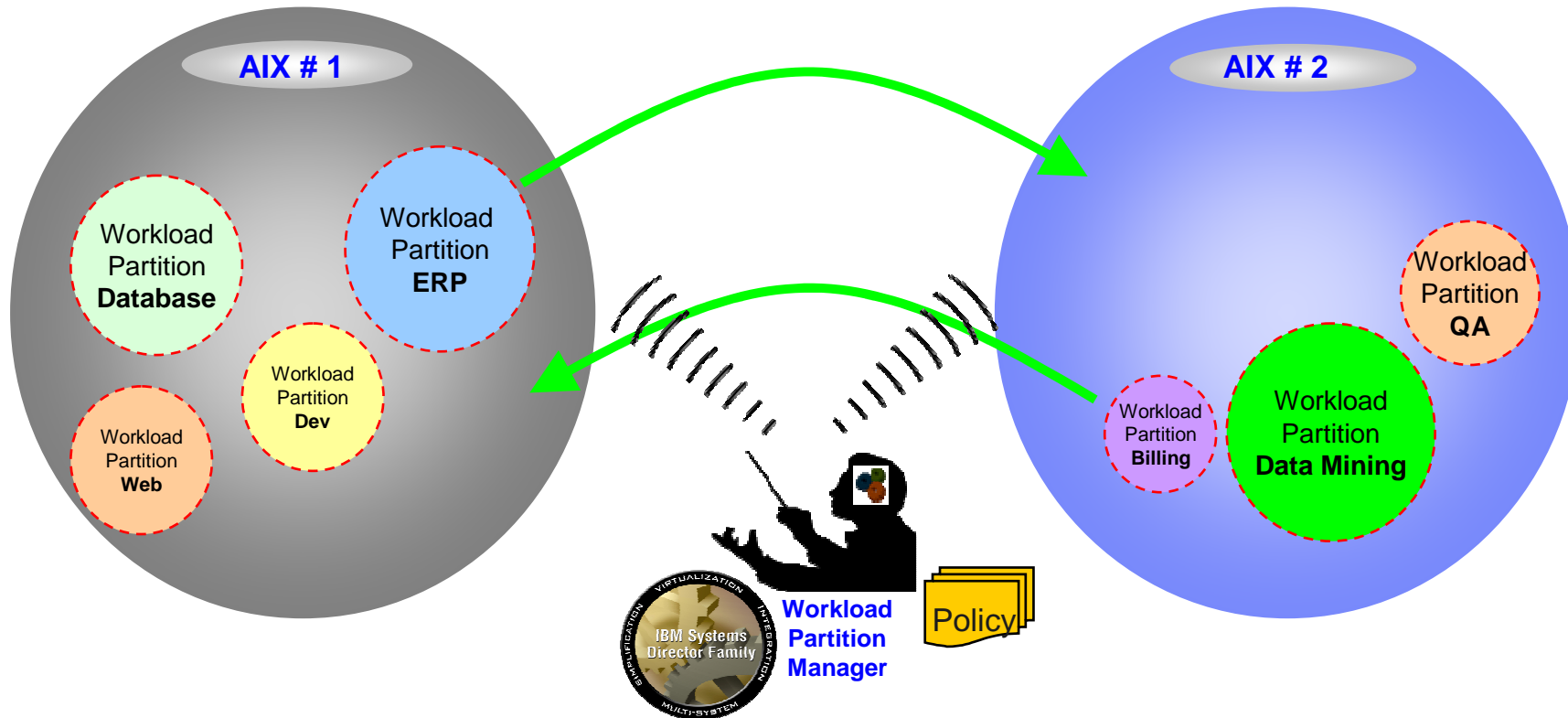AIX 5L V5.1 on POWER4

IBM Systems

AIX V6.1 基于工作负载的分区技术

Virtualized AIX OS environments within a single AIX image

- Partitioned system capacity
  - Each Workload Partition obtains a regulated share of system resources

- Two types of WPAR
  - System WPARs have separate security and appear like a completely separate OS
  - Application WPARs are manageability wrappers around a single application

- Resource controls for WPAR
  - CPU, memory, paging space, number of threads and number of processes

- Shared system resources
  - Operating System / Shared Library and Text
  - Processor / I/O Devices

AIX Image

Workload Partition A

Workload Partition B

Workload Partition D

Workload Partition C

Workload Partition E
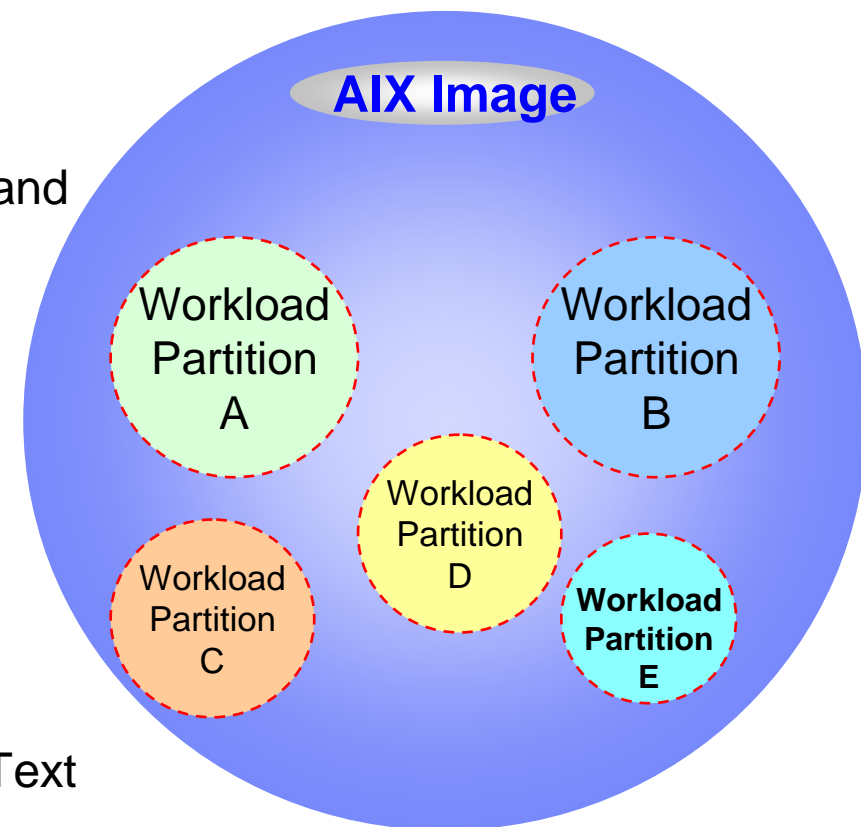
AIX V6.1 在线负载分区的迁移（Live Application Mobility）

The ability to move a Workload Partition from one server to another

AIX # 1

Workload Partition **Database**

Workload Partition **ERP**

Workload Partition **Dev**

Workload Partition **Web**

AIX # 2

Workload Partition **QA**

Workload Partition **Billing**

Workload Partition **Data Mining**

IBM Systems Director Family

Workload Partition Manager

Policy

Provides outage avoidance and multi-system workload balanci

Policy based automation can provide more efficient resource usage

IBM Systems

What does PowerVM Lx86 do?

Native Linux for POWER Binary

x86 Linux Applications

ISV Binaries   **User Binaries**

PowerVM Lx86

Front End

Optimizer

Back End

Linux OS (Red Hat or Novell SUSE)

POWER Processor

Power Systems platform
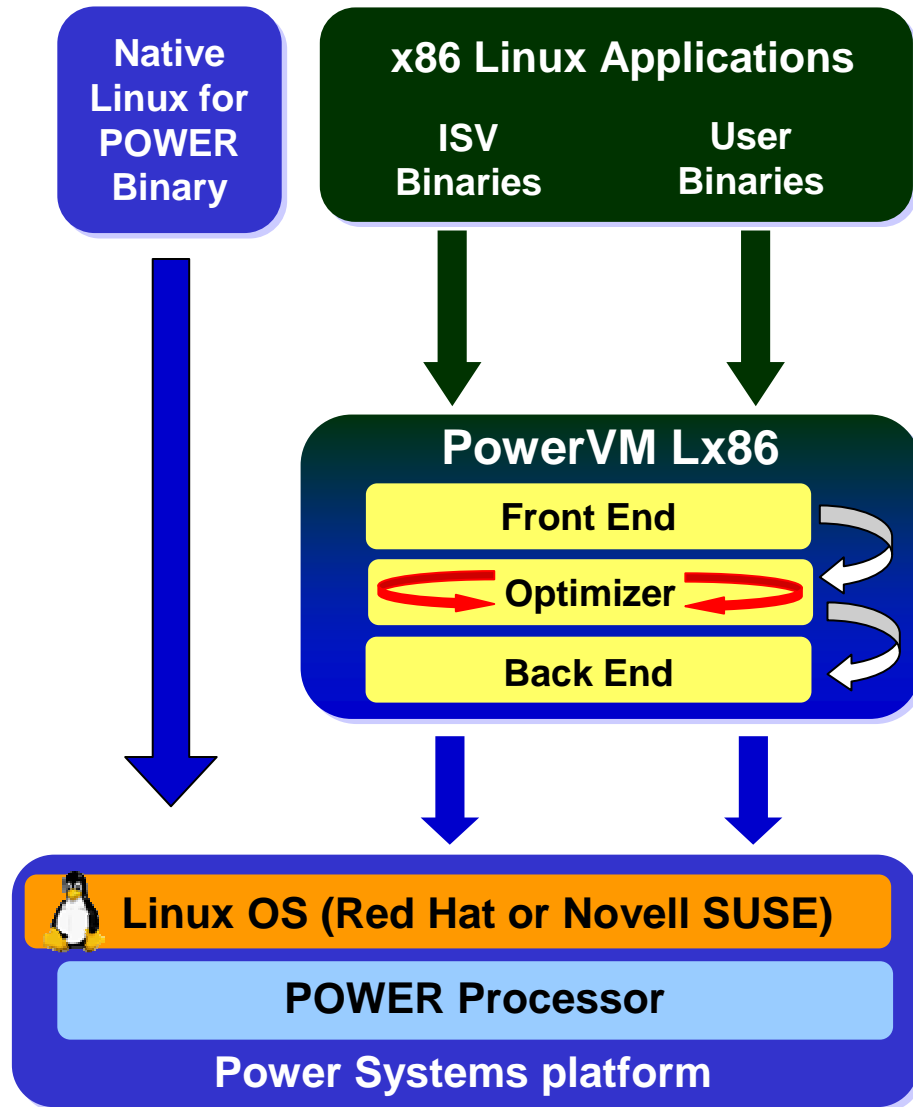
➢ Dynamically translates and maps x86 Linux instructions to POWER

➢ Translation process

✓ **Translates blocks of code into intermediate representation**

✓ **Performs optimizations**

✓ **Stores optimized, frequently used blocks of code in cache**

✓ **Handles Linux OS call mapping**

✓ **Encodes binary for target POWER processor platform**

➢ Best for certain applications and usage scenarios

✓ **Power architecture can provide many advantages**

✓ **But these make our architecture very different from x86 architecture**

✓ **Translation can be resource intensive**

IBM Systems

虚拟化技术提高服务器的使用率
服务器合并可以有效地节能

System 1	System 2	System 4
APP 1 / APP 2	APP 3 / APP 4	APP 7 / APP 8
10% busy 2KW	10% busy 2KW	10% busy 2KW

Advanced Virtualization — APP 1 / APP 2 / APP 3 / APP 4 / APP 5 / APP 6 / APP 7 / APP 8

70% busy
4KW
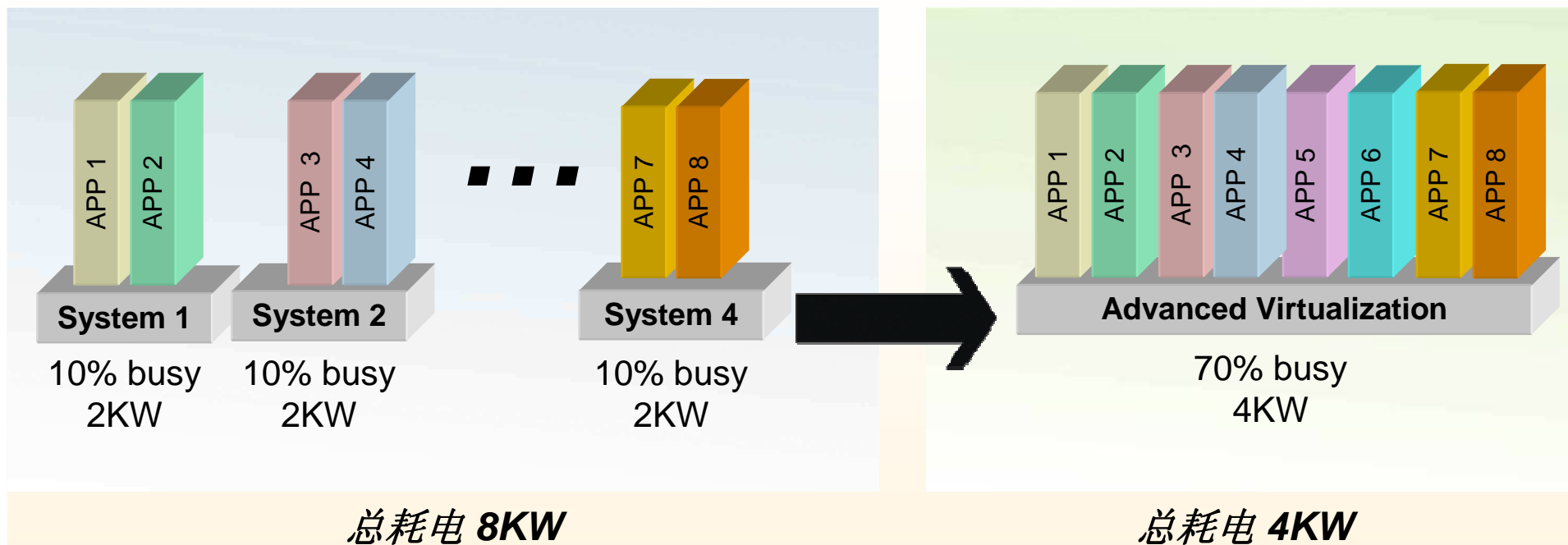
总耗电 8KW   *总耗电 4KW*

使用虚拟化技术进行服务器合并是一个非常有效的节约能源的工具

IBM POWER Systems

Power Blades – 具有与高端服务器同样的虚拟化技术
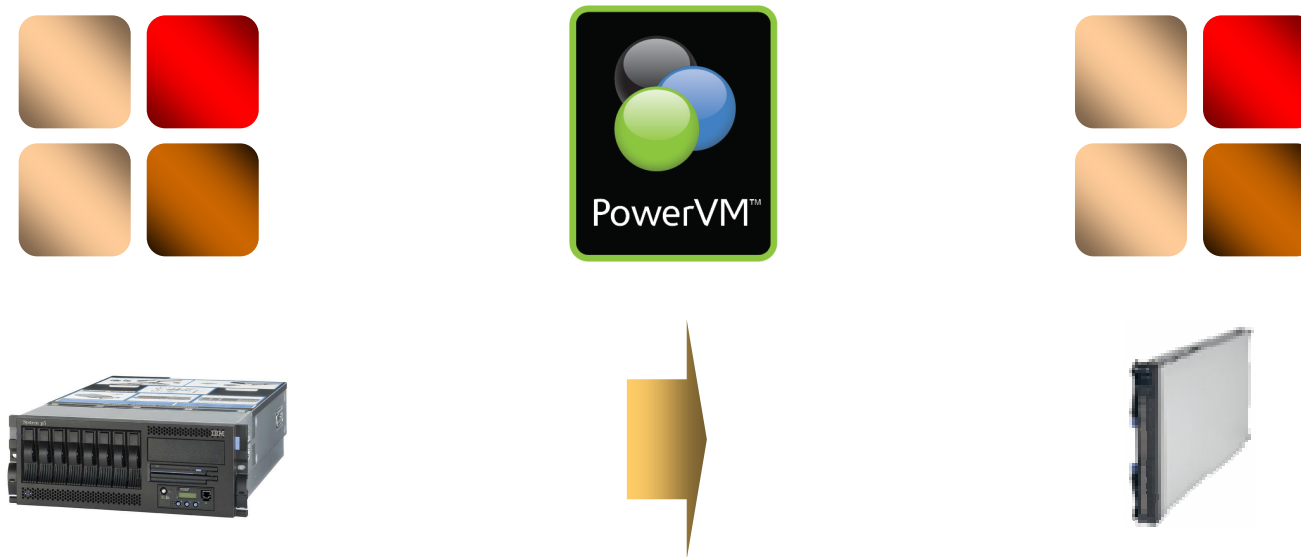
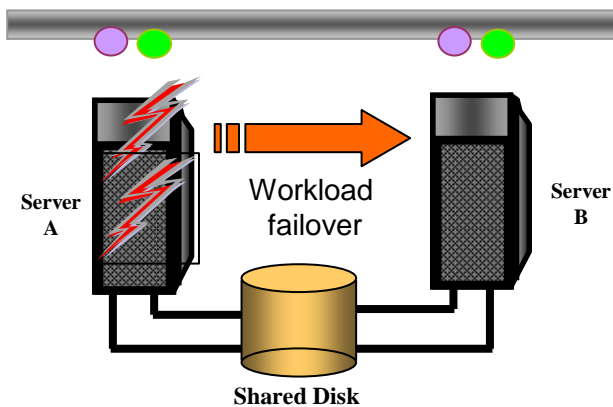**p5 SLE**   **PowerVM**   **JS**

✓	**Micro-Partitioning™ (10 LPARs / CPU core)**	✓
✓	**Virtual I/O Server**	✓
✓	**Integrated Virtualization Manager**	✓
✓	**Lx86 (formerly System p AVE)**	✓
X	**Live Partition Mobility**	✓

IBM Systems

HACMP V5.4 – 高可用性管理软件

Leadership AIX High Availability and Disaster Recovery Product

Available for Linux!

Server A   Workload failover   Server B

Shared Disk

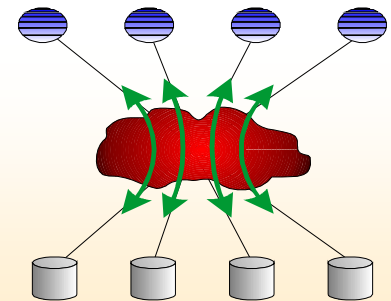**HACMP**

▪**Protect your critical business applications through reliable monitoring, failure detection and automated recovery of business applications**

▪ **New Linux Support**

–SLES9 and RHEL4 support

▪ **Ease of Use Enhancements**

–Configure an HACMP cluster or upgrade HACMP on a node without disrupting the target application

–Fast Failover Detection through enhanced AIX integration improves failover time

–Recognize Application and Resource presence

–Resource Dependency Graph

–Adjustable preferences

GPFS: 高性能并行文件系统

- GPFS™ 允许多台服务器并行地读写同样的文件系统

- 多路径访问和避免单点故障，保证数据的高可靠性
  - 大量的商业客户在关键业务中使用GPFS

- 数据管理功能保证了数据管理的效率和方便性

- Multi-cluster可以在多个地点共享同样的数据

- 并行的文件访问只需要保存一份文件，隔离了多份文件的弊病

- 最快速的文件访问性能，大型计算环境都在使用GPFS

- 商业应用同样可以使用GPFS并行文件系统
  - Digital media
  - Medical imaging
  - Life sciences
  - Financial analysis
  - – Business intelligence
  - – Engineering design
  - – Geographic information systems
  - – Data sharing