



定制服务器的 前世今生

讲师介绍

提升服务器利用率

降低采购成本

提升单机性能
和每人民币性能



刘明生

新浪-研发-运维部

mingsheng@staff.sina.com.cn

微博：@明生78



? 思维导向

生意

500亿\$

2012年服务器行业收入

公司	2013年Q1销售 额(千美元)	2013年Q1 市场份额%	2012年Q1销售 额(千美元)	2012年Q1 市场份额%	同比 增长%
IBM	3016060	25.5	3490477	28.0	-13.6
惠普	2959030	25.0	3455760	27.8	-14.4
戴尔	2124462	18.0	1857579	14.9	14.4
富士通	583239	4.9	626722	5.0	-6.9
甲骨文	538542	4.6	739826	5.9	-27.2
其它	2604390	22.0	2273724	18.3	14.5
共计	11825724	100.0	12444088	100.0	-5.0

3426 亿\$

2012年软件行业收入

全球前五名软件供应商，2012年新兴市场及全球收入、全球收入份额及同比增幅（收入单位：百万美元）

供应商	2012年新 兴市场收入	2011年新 兴市场收入	2012年全 球市场收入	来自新兴市 场的收入占 全球收入的 比例	2012年新 兴市场增幅
Microsoft	\$11,072	\$10,658	\$58,454	18.9%	3.9%
IBM	\$4,444	\$4,068	\$29,129	15.3%	9.2%
Oracle	\$6,150	\$5,516	\$27,826	22.1%	11.5%
SAP	\$3,905	\$3,425	\$16,988	23.0%	14.0%
Symantec	\$849	\$813	\$6,423	13.2%	4.5%
其他	\$22,751	\$21,407	\$203,818	11.2%	6.3%
所有供应商	\$49,172	\$45,886	\$342,638	14.4%	7.2%

来源：IDC 全球半年度软件市场跟踪报告，2013年4月

? 合作基础

我要去做**软件**
我要去做**服务**

IBM
DELL
HP

我是专业做硬件制造的
您的需求就是我的使命

ODM

能否把服务器外壳去掉
能否用SD卡启动OS
能否用消费级SATA硬盘替代企业级SAS
能否将内部线缆做一些改动

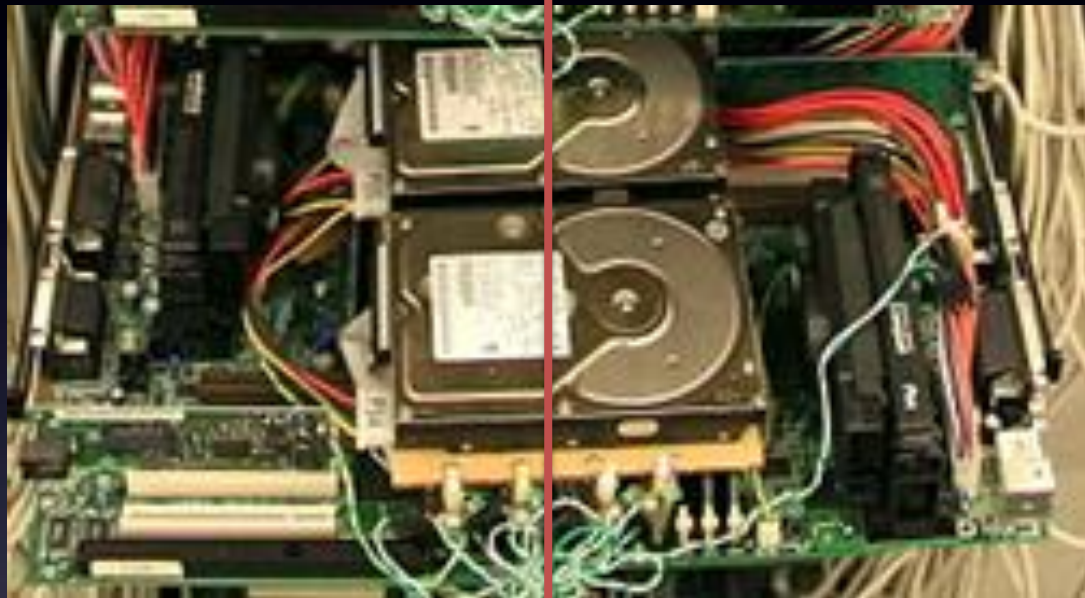
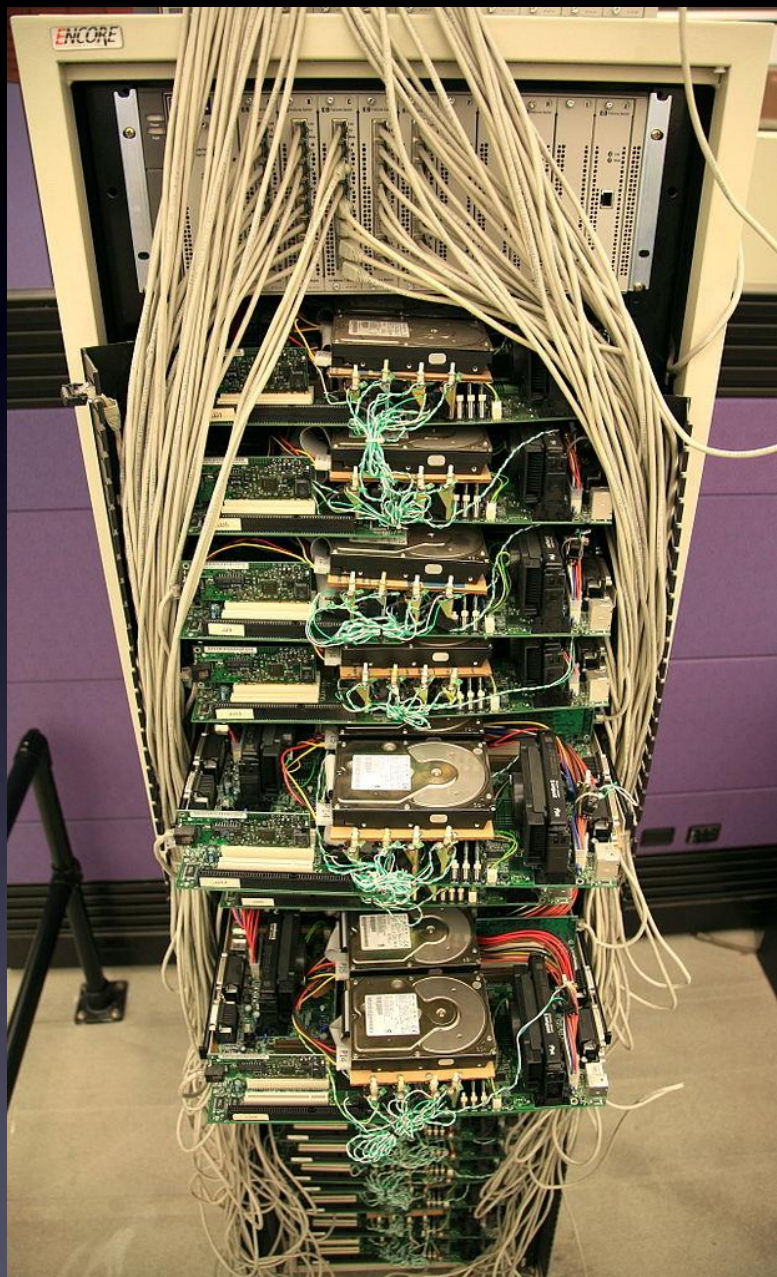
Facebook

Avnet、ZT系统和Hyve Solutions(母公司为业务流程服务企业Synnex)等，QCT(母公司为ODM厂商Quanta广达，戴尔服务器的主要代工厂)和Wiwynn(母公司为ODM厂商Wistron)

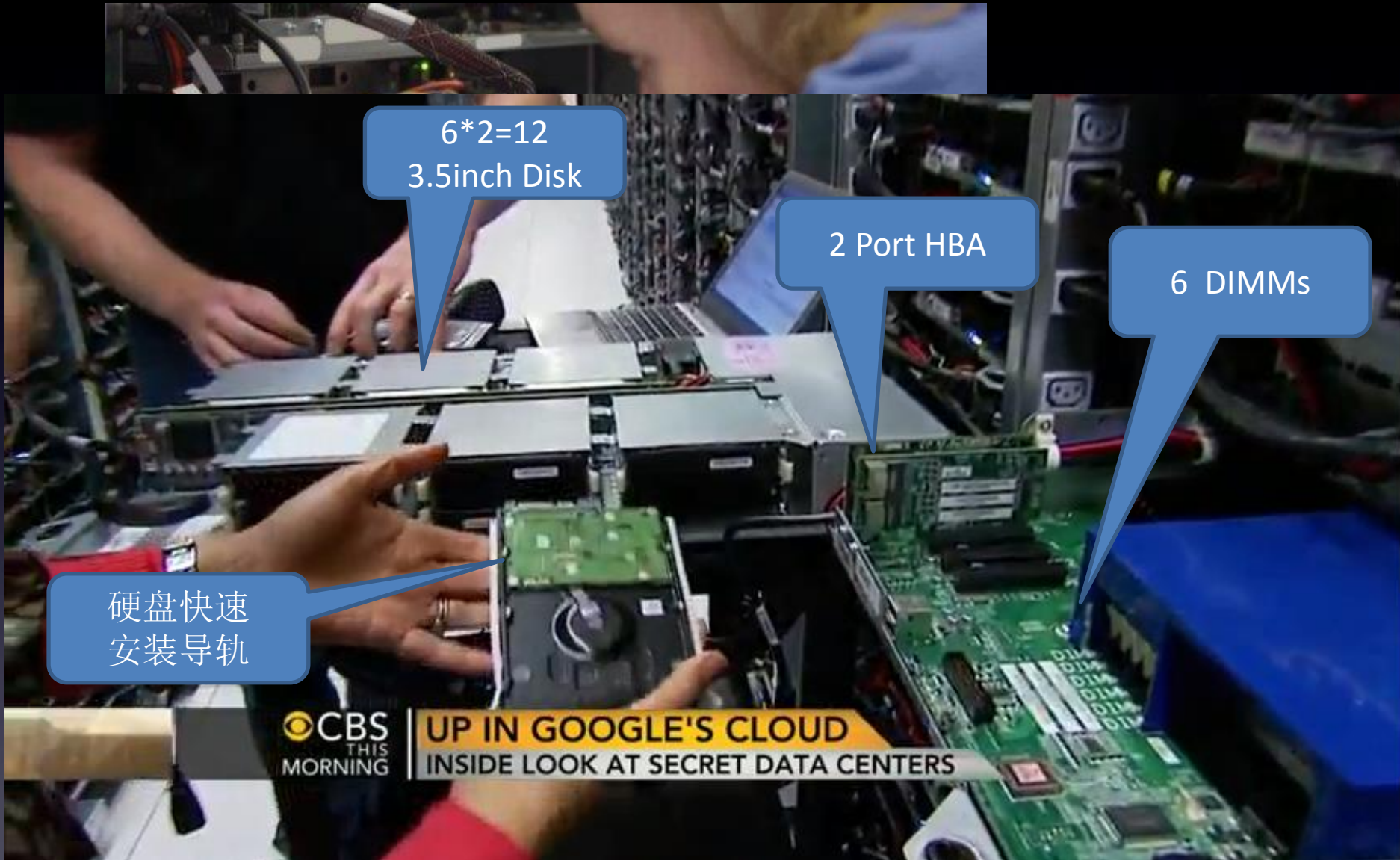
Google Video

[播放...](#)

二、Google 定制服务器的发展历程-第一代



二、Google 定制服务器的发展历程--第三代



二、Google 定制服务器的发展历程—第三代-2

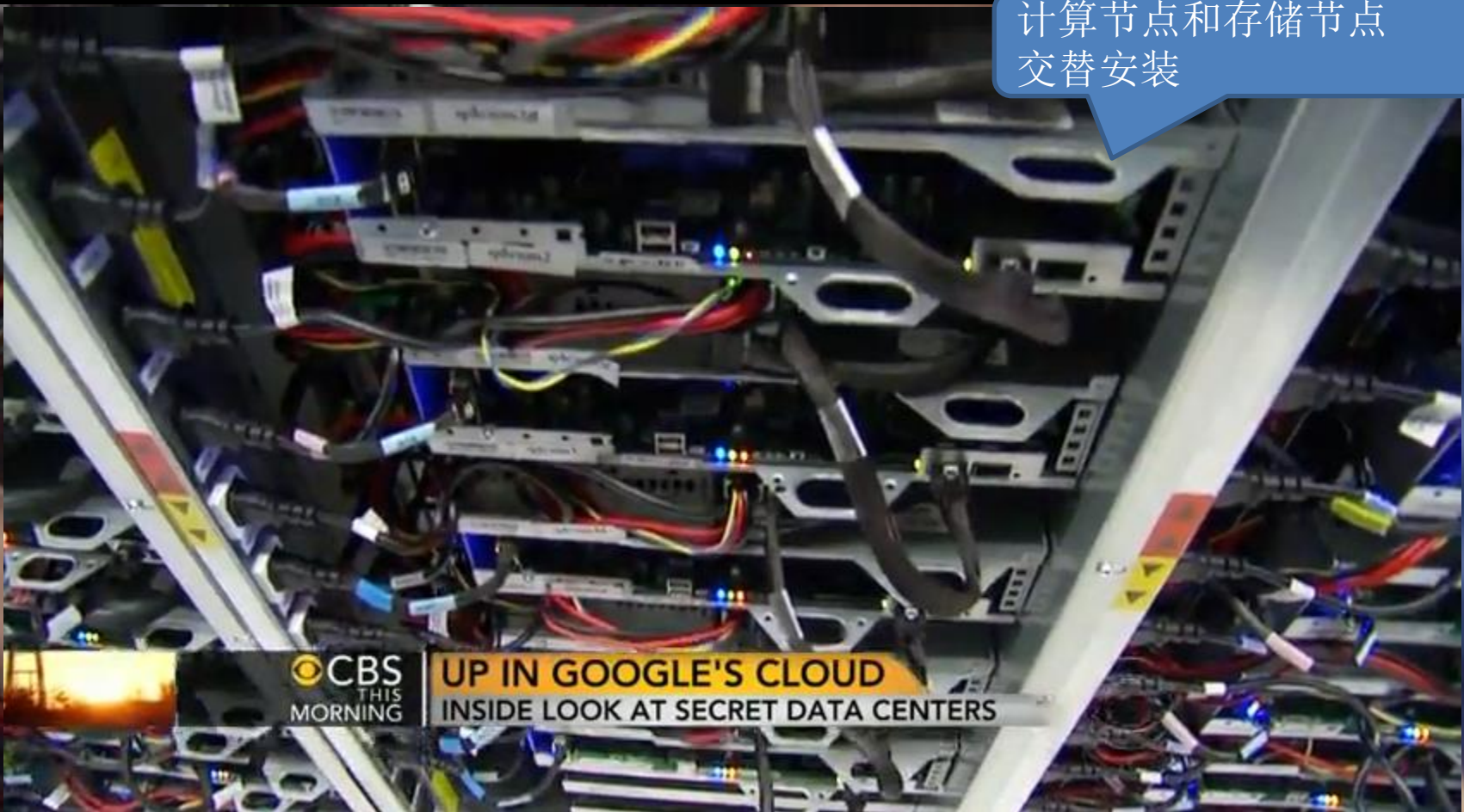


硬盘2 2 反向安装
抵消震动

计算和存储节点
按需组合

二、Google 定制服务器的发展历程

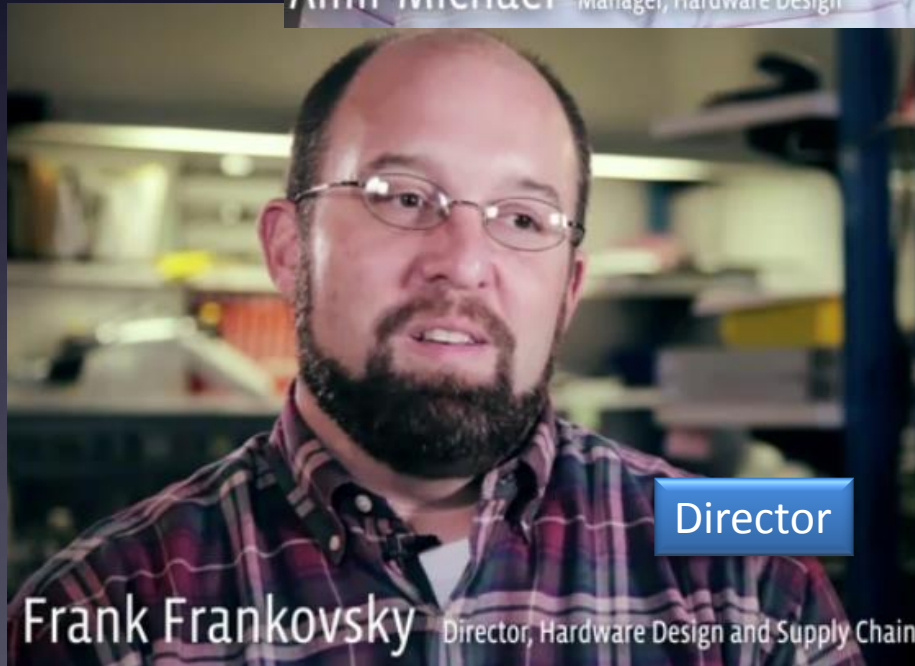
计算节点和存储节点
交替安装



Facebook

定制服务器

四、Facebook 定制服务器的发展历程-第一代 Server



四、Facebook 定制服务器的发展历程

---背后的能量



Mark Zuckerberg

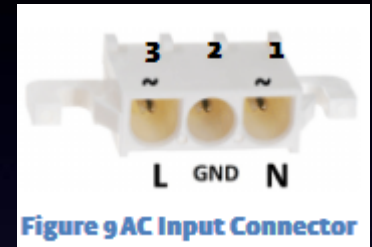
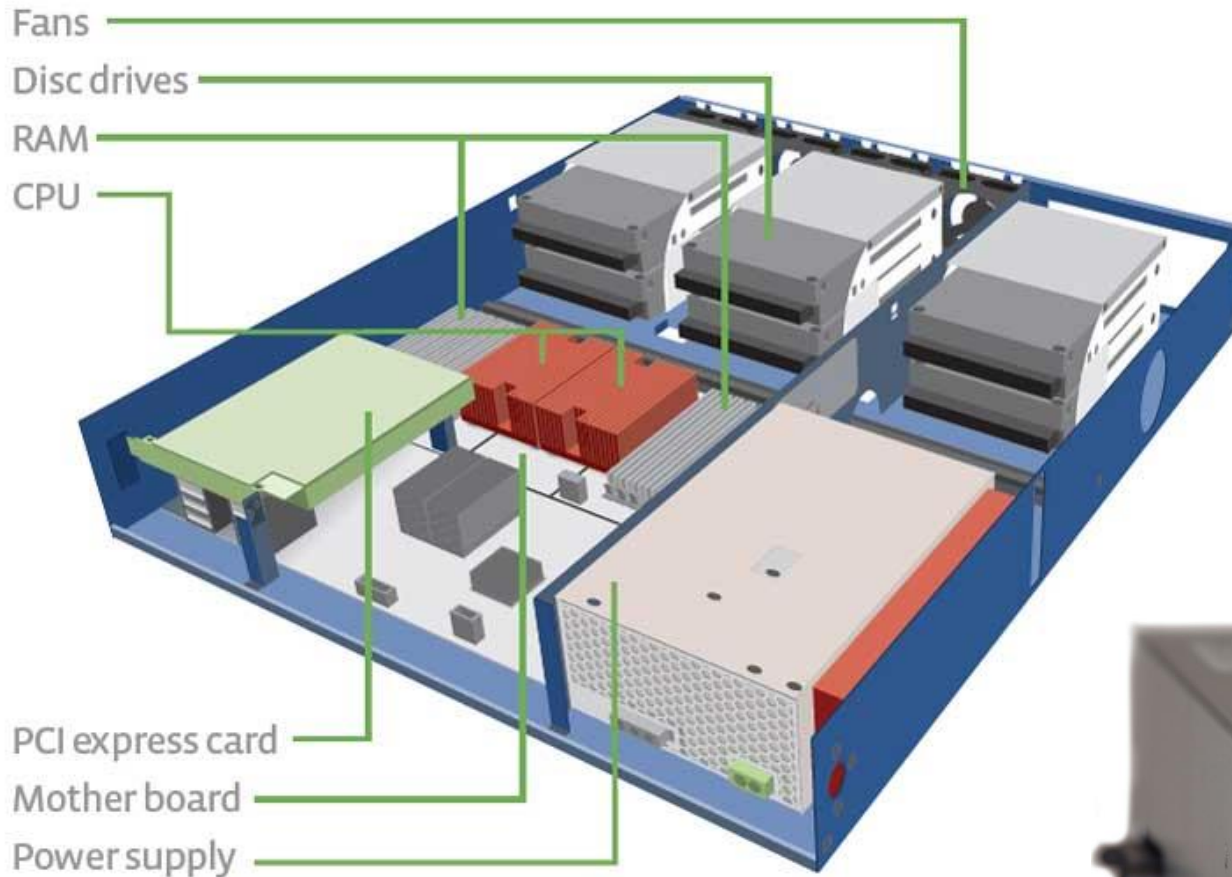


四、Facebook 定制服务器的发展历程-第一代 Server



四、Facebook 定制服务器的发展历程-第一代 Server

Open Compute Project Server



The nominal AC input voltage is 277VAC RMS (200 – 277 VAC).

The nominal DC input voltage is 48VDC.



四、Facebook 定制服务器的发展历程-第二代

Web v1



Web v2



届英特尔

Google和Facebook的定制化共性

- 1、达到成本节约的目的
- 2、无机箱盖，只有托盘
- 3、共享散热
- 4、共享供电
- 5、本机存储有限
使用HBA卡+SAS线缆连接扩展存储
- 6、规格统一，可互换

五、百度、阿里、腾讯的天蝎计划

第二版

2012年发布 新版 1U 双路 8盘 刀片



五、百度、阿里、腾讯的天蝎计划



为什么空闲很多位置？

天蝎需求 $40U \times 2\text{刀片} \times 200\text{瓦} = 16000\text{瓦/机柜}$
国内机房常规供应
 3500瓦/机柜



五、百度、阿里、腾讯的天蝎计划 生产环境





接地气的新浪定制服务器 Sina Cube



成果展示

需求来自？

评估标准


粗陋定制服务器生存基础

我们的做法

产品展示

成果展示

成本节约-采购

2*intel E5-2620v2 +4*4G RDIMM ECC +1*300G 2.5 SAS	12U 定制服务器 平均到每刀	对比品牌 A 
满载	$728/6=122$	181
待机	$410/6=69$	115
待机降幅		40%
满载降幅		32.6%
半载降幅		35.1%

节约 21.3%的设备采购费用

成本节约-机架

1万/秒 并发任务

每秒任务能力 1.1628 个—基准php benchmark
机架费用并非实际成交价

	定制服务器	传统服务器
半载功耗（VA）	96	148
需要采购设备（台）	8600	8600
13A 机架数量（个）	297	430
3年机架费用（万元）	5346	7740
3年节约费用（万元）		2384

节约 30.8%的机架费用

成本节约-空调

空调每调高1度，可以解决6%~8%的制冷电费

定制服务器的CPU温度，待机和满载均比友商低6度（24度进风）

在局部区域：sina可以将进风口温度提升到29度（机械硬盘）

在局部区域：sina可以将进风口温度提升到32度（SSD）

节约 30%的空调电费
在PUE=2的机房相当于15%的整体电费。

需求来自？

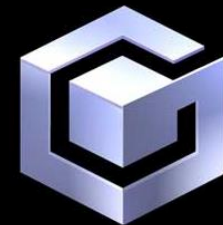


老板？

Or 成本

Or 年轻躁动的心

评估标准



在满足业务单机最低性能时
，每人民币性能要优于当前采购设备

具体某个业务的性能指标,在满足
99.99% 延迟低于50ms 时
Mysql=QPS, CDN=hits/s

每人民币性能=

设备采购费用
+ 机架租用费用
+ 网络二层端口及带宽费用
+ 运维人员费用

粗陋定制服务器生存基础：

业务稳定性**不依赖于** 单机/单机架/**单机房**



想法来自？



共享

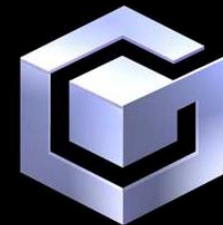


我们的做法？-step1

产品技术特点：

- 0、框架8年使用寿命（满足intel 换3代）
减少66%电源，70%风扇成本
- 1、超短距散热
- 2、非标准散热片（1.5U 高）提升散热效率
- 3、PWM（风扇型号）替代BMC 管理散热系统，
降低开发成本
- 4、无中板设计，降低设计复杂度
- 5、航空插头替代 铜排BUS,降低框架制造精度
- 6、框架设计，内置理线装置
- 7、自由并柜，高密度部署

我们的做法？-step2



用户的真实需求 → **计算能力
存储能力
稳定性**

增加共享 → **共享风扇和电源**

提高效率 → **增大散热片面积、
缩短散热通道距离
提高电源效率**



我们的做法？-step3

共享，可以节约

Server 采购成本21.3%

电源 **12** → **3**

风扇 **30** → **9**

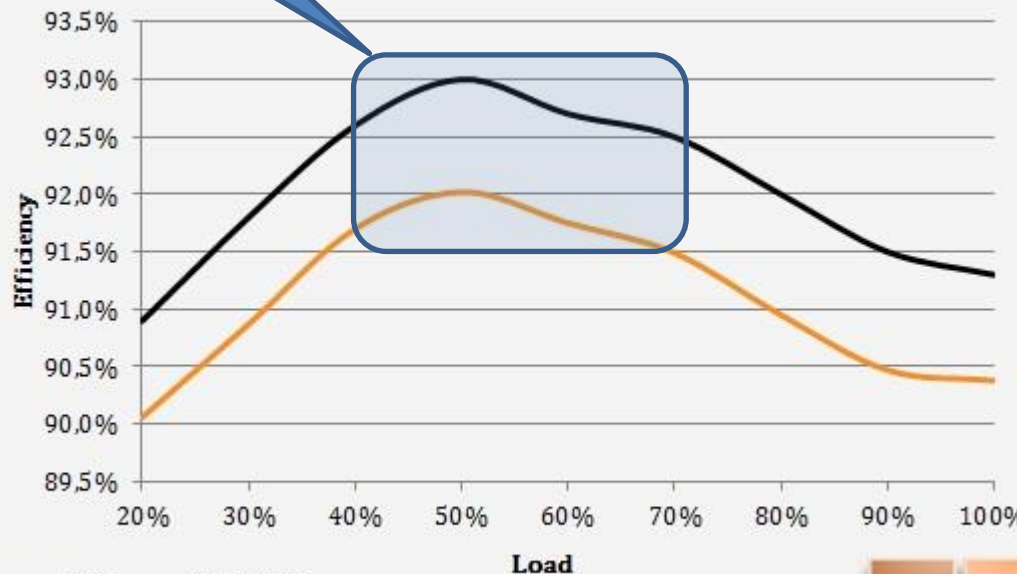


我们的做法？-step4

电源半载的好处

甜点区，
效率高

Efficiency

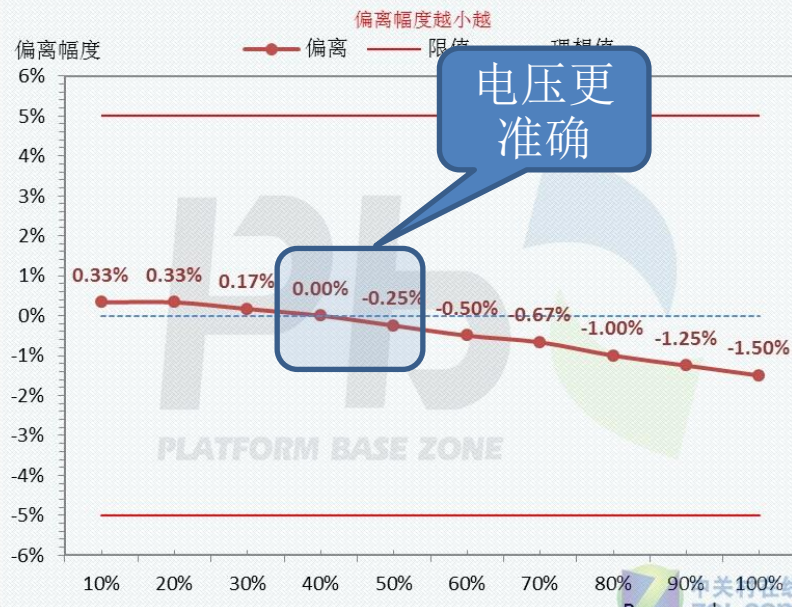


— Efficiency @ 110VAC

— Efficiency @ 230VAC

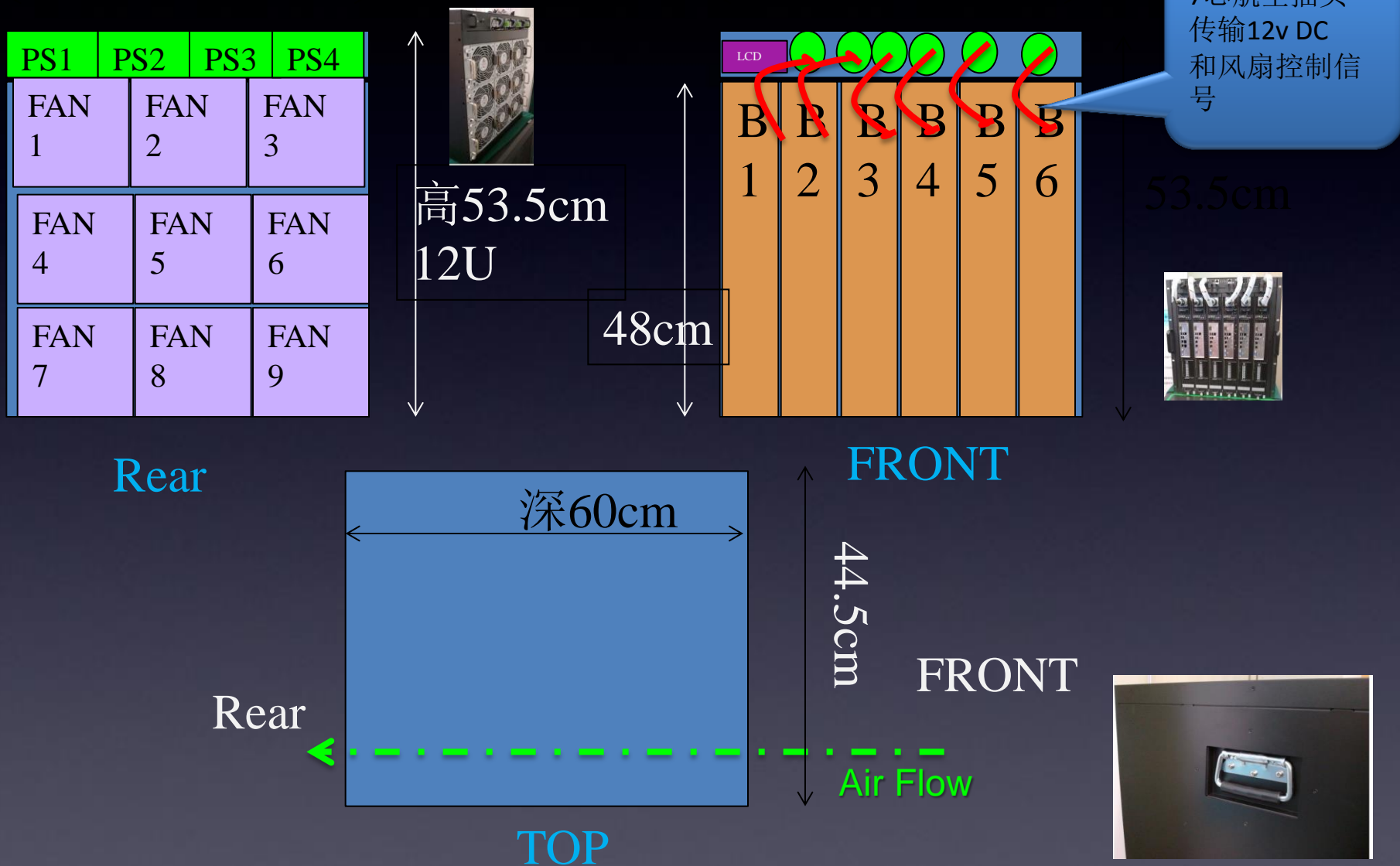
					
	白牌	铜牌	银牌	金牌	白金牌
负载	转换效率				
20%	80%	82%	85%	87%	90%
50%	80%	85%	88%	89%	92%
100%	80%	82%	85%	87%	89%

+12V1电压偏离曲线



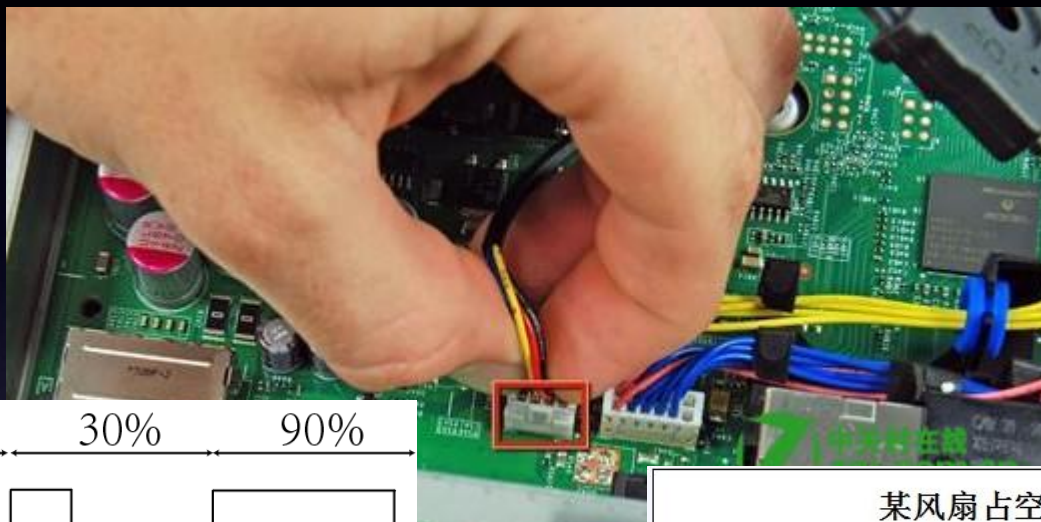
电压更
准确

新浪定制服务器-外观



散热系统原理：

- 1、未使用BMC
- 2、使用主板上FAN口的风扇调速信号
- 3、6个node的FAN信号输入控制板，决策后输出9路FAN控制信号到风扇墙

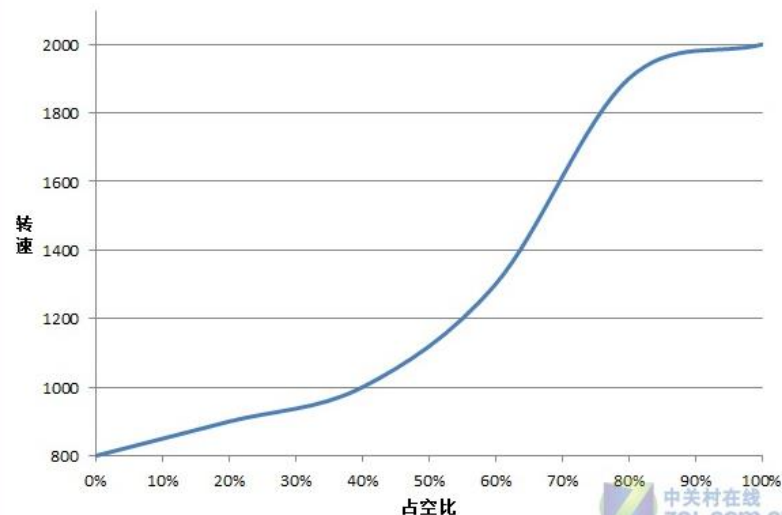


50% 10% 30% 90%

PWM波形图

电容积分后波形图

某风扇占空比与转速示意



某风扇占空比示意

集群管理机控制整体策略：

- 1、半夜风扇降速
- 2、IDC局部热点自动解决等

集群管理机



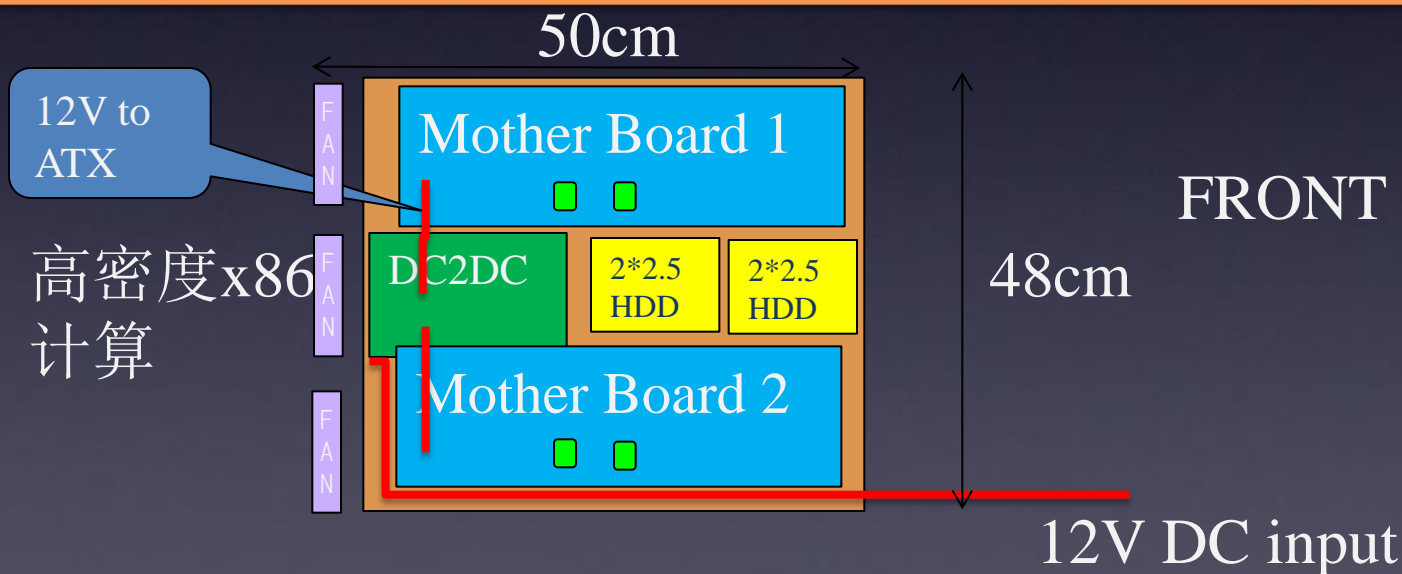
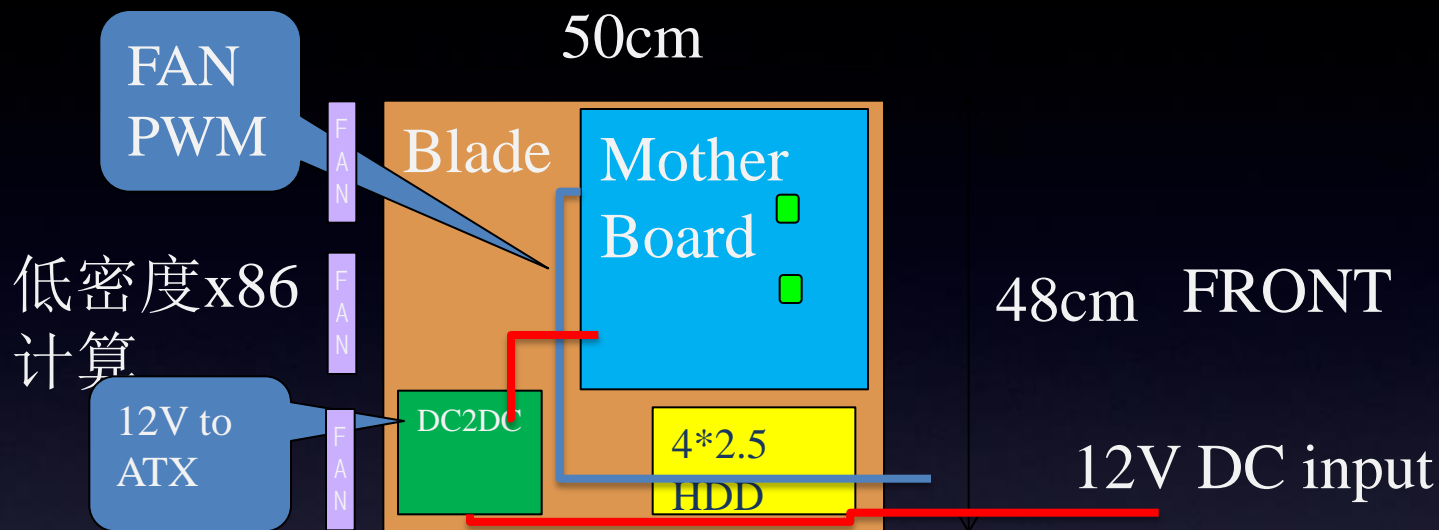
本机“生命维生系统”，
风扇转速更新时间 $<1s$

本机“生命维生系统”，
风扇转速更新时间 $<1s$

本机“生命维生系统”，
风扇转速更新时间 $<1s$

新浪定制服务器-布局-计算

动态前端
/虚拟化



新浪定制服务器-部署

电源线×6

1Gbps*2/4 or 10Gbps*2

13A Rack

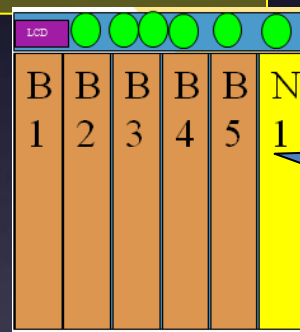
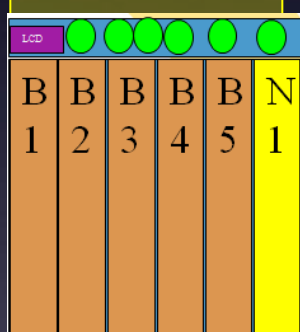
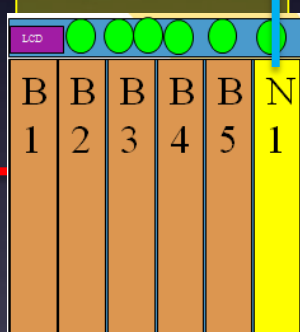
22 server +1 switch

满载2800w

(ivybridge)

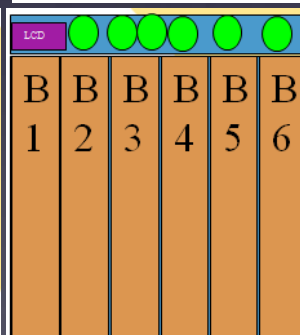
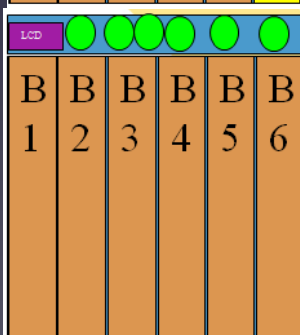
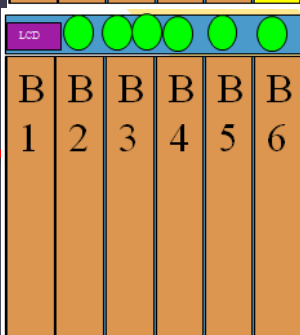
透明挡
风帘
隔绝冷
热通道

3



5*高密刀
+1 交换刀

3



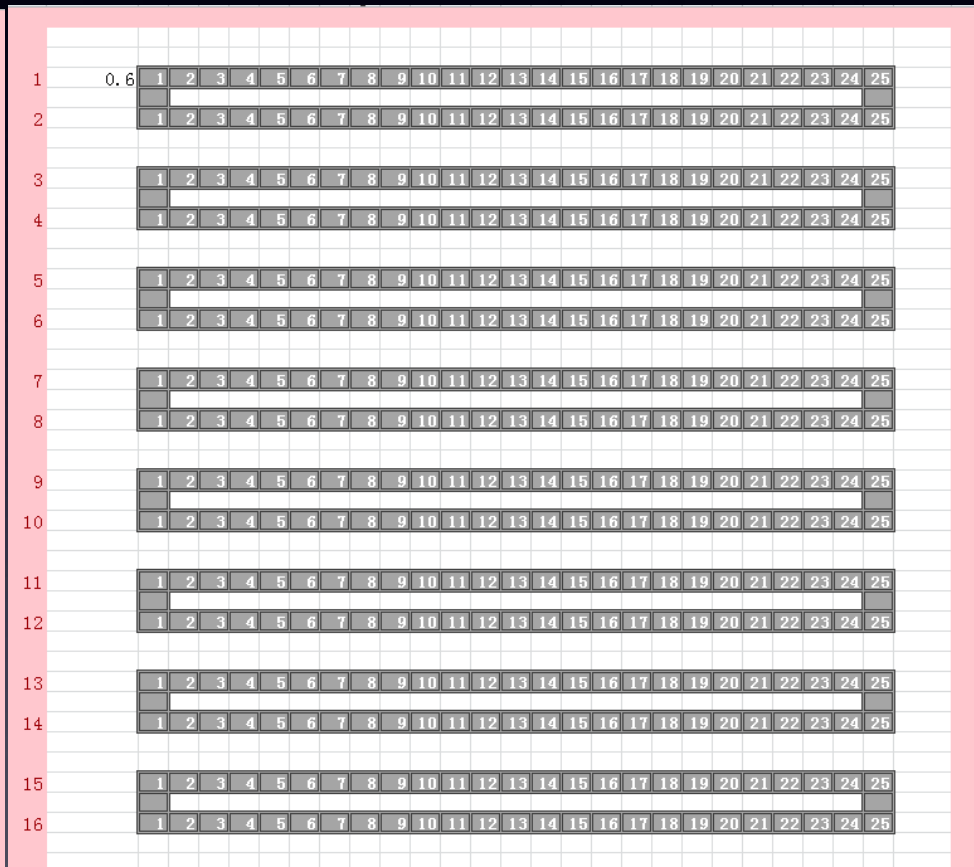
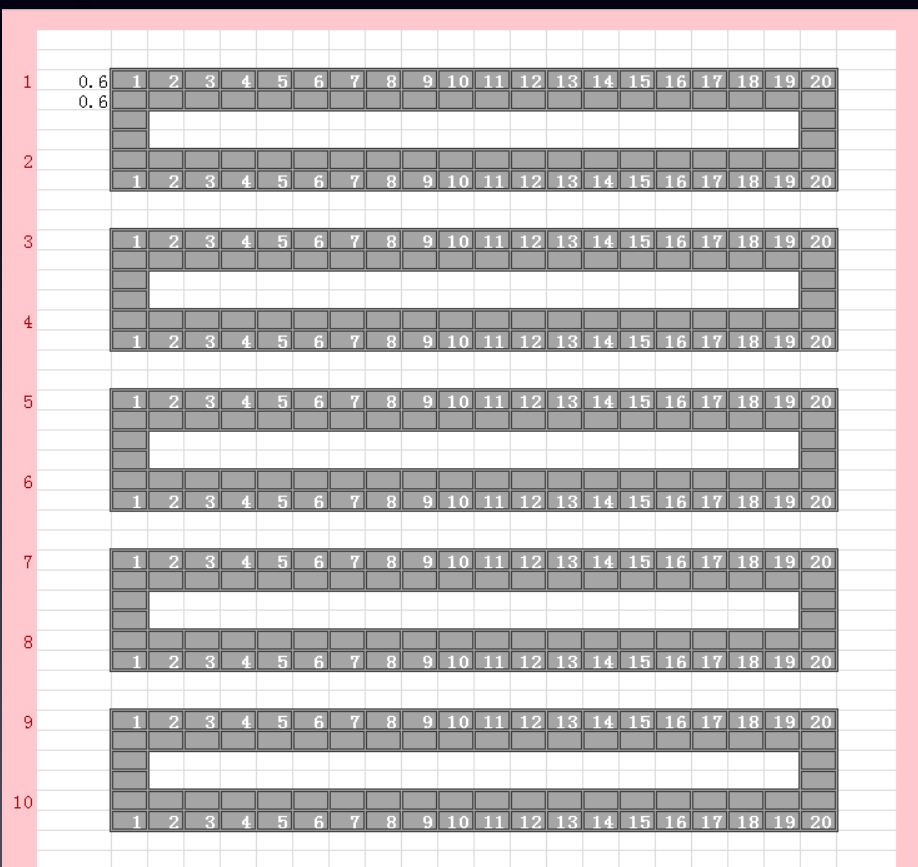
6*高密刀

新浪定制服务器-部署-无机柜

无机柜部署的优势:

机柜部署密度比整机架方案高100%

天蝎 20列*10行 vs 新浪 25列*16行=200 rack vs 400 rack



整机柜占地: 120cm*60cm 40U

新浪定制占地 :60 cm*48cm 48U

无机柜部署的优势：

机柜部署密度比整机架方案高100%

天蝎 20列*10行 vs 新浪 25列*16行=200 rack vs 400 rack

设备部署密度比整机架方案高20%

天蝎 200 rack *40U*2 node/U=16000 node

新浪 400 rack *48U*1 node/U=19200 node

Q&A