

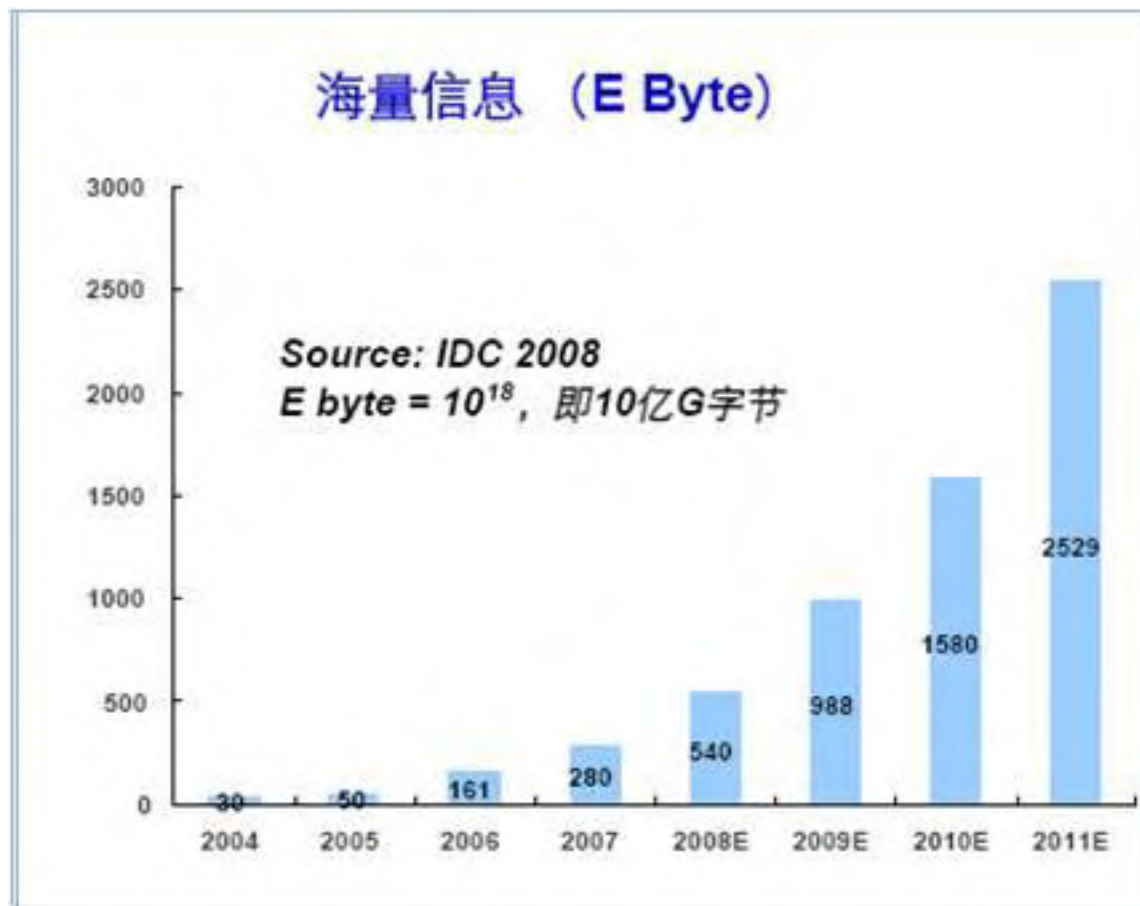
大数据 云计算 及应用实践

刘 鹏 微信: chinagloud



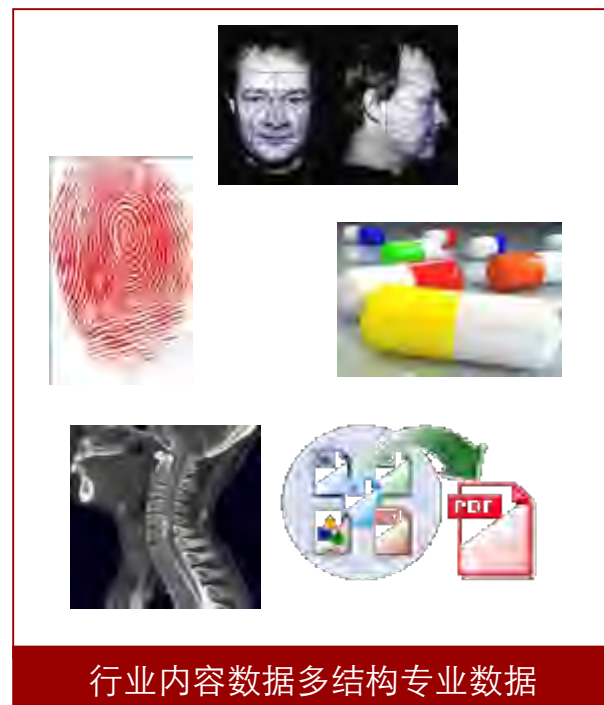
大数据

大数据现象



- 信息爆炸: 2006年, 全球产生161 EB 的数据, 印成书是地球到太阳距离的10倍; 2007年, 全球产生280 EB 数据, 全世界平均每人45G; 而人类历史5000年的文字记载只有5EB

大数据产生原因



马云：大数据赋予我们洞察未来的能力

马云成功预测2008 年经济危机

- “2008 年初,阿里巴巴平台上整个买家询盘数急剧下滑, 欧美对中国采购在下滑。海关是卖了货, 出去以后再获得数据; 我们提前半年时间从询盘上推断出世界贸易发生了变化了。”
- 通常而言, 买家在采购商品前, 会比较多家供应商的产品, 反映到阿里巴巴网站统计数据中, 就是查询点击的数量和购买点击的数量会保持一个相对的数值, 综合各个维度的数据可建立用户行为模型。因为数据样本巨大, 保证用户行为模型的准确性。因此在这个案例中, 询盘数据的下降, 自然导致买盘的下降。



人类从依靠自身判断做决定到依靠数据做决定的转变, 也是大数据作出的最大贡献之一。——《大数据时代》

禽流感 and 板蓝根

板蓝根: 成交指数

2013.03.01 - 2013.04.08



地域细分 从2013-03-01到2013-04-08, 38天来通过搜索购买 板蓝根 的消费者

喜好度(TGI)

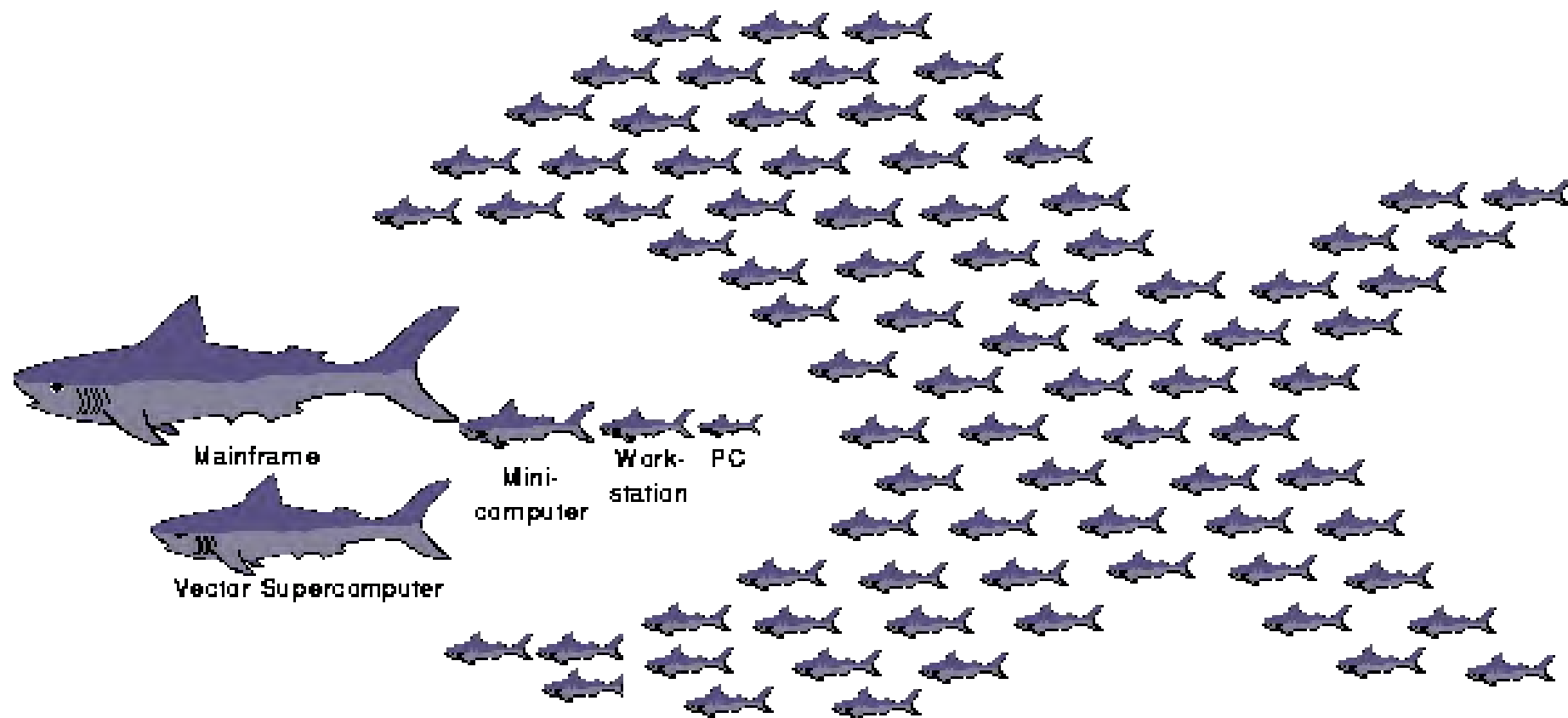
人群占比





云计算

云计算的核心思想



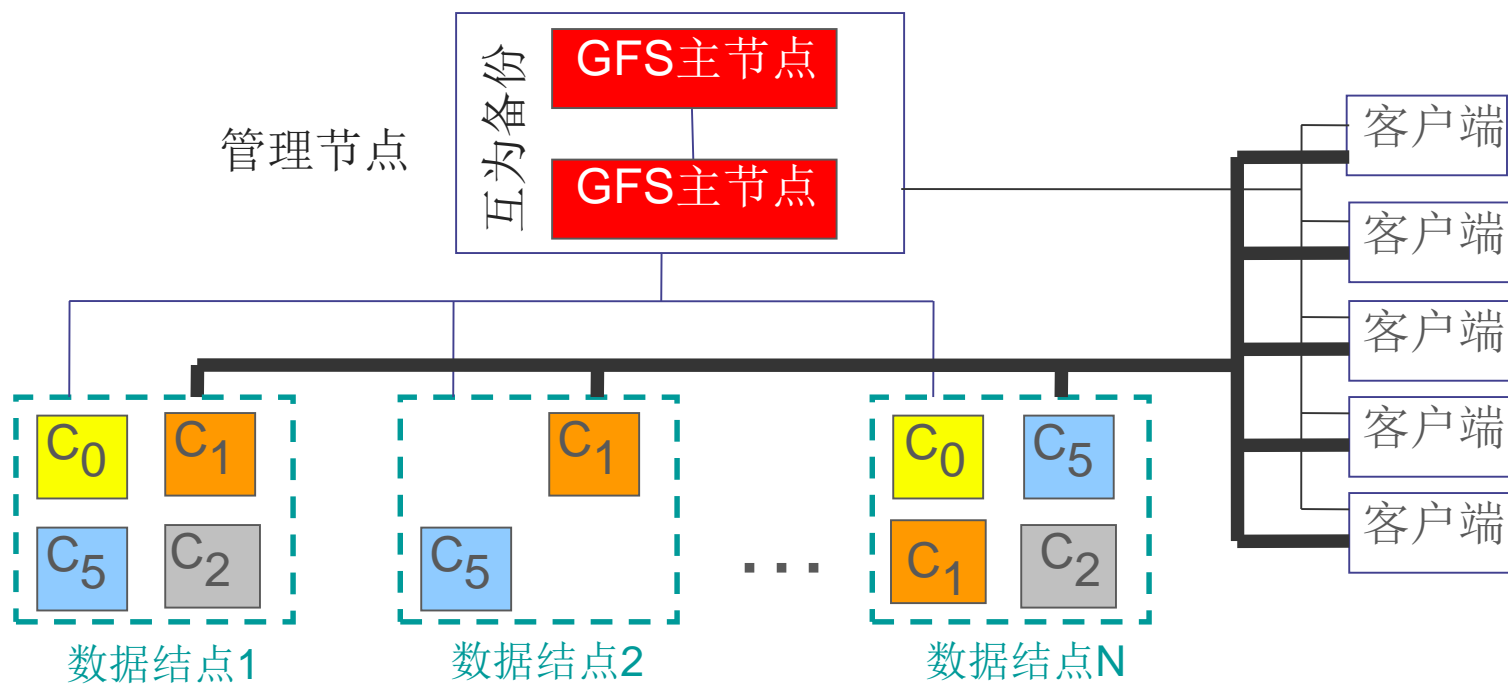
NOW

Google云计算关键技术



- Google文件系统GFS (Google File System)
- 并行数据处理MapReduce
- 结构化数据表BigTable
- 分布式锁管理Chubby

Google文件系统(GFS)

















绿色和平组织的报告

- 由于数据中心和远程通信网络的快速增长，使得用电量激增，从而对气候变暖产生极大的负面影响。
- 到2020年，数据中心将消耗的电力资源，比法国、德国、加拿大和巴西目前消耗的电力总量还要多。

绿色云计算



Atlantic Wind Connection(大西洋风力传输, AWC)计划。总成本约50亿美元, Google将提供该计划37.5%的资金



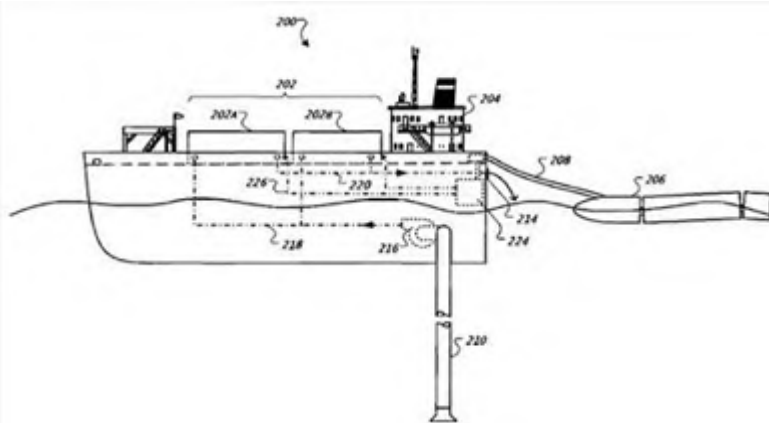
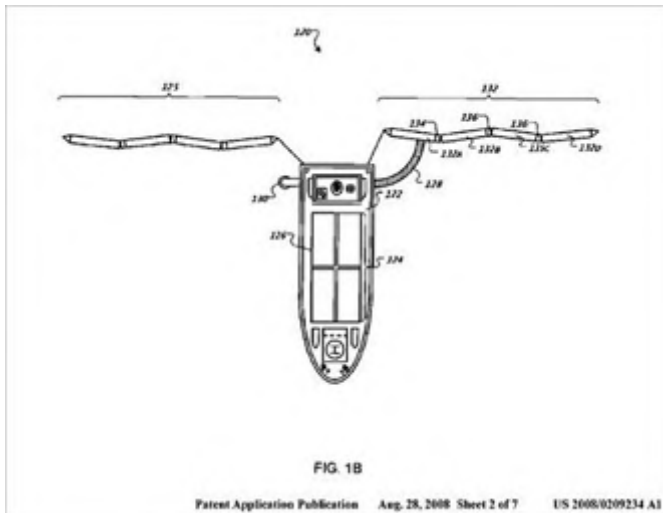
Intel太阳能面板阵列充分吸太阳能可以产生10千瓦的电力, 为数据中心提供部分的电力。



Driver



Google海上数据中心







云计算给用户体验带来的革命

小小终端=超级计算机

更加智能的服务

如影随行的私人数据中心

服务器永远不会塞车

90% Off 的价格

信息随手可得

计算智能将大幅提升



Google 翻译

源语言: 中文 目标语言: 英语 翻译

无论是2010年还是2011年,云计算都被评为十大战略技术之首。Google云计算平台规模已经超过200多万台计算机,其企业用户数量超过300万家。今年微软公司将有90%的研发投入用于云计算,华为公司将有70%的研发投入用于云计算,全世界的互联网企业和电信企业都在向云计算进军,云计算技术同时成为学术界的研究热点领域。

允许输入拉丁字符的拼音

将中文译成英语

Whether it is 2010 or 2011, cloud computing have been the first top ten strategic technologies. The size of Google's cloud computing platform has more than 200 million computers, the number of over three million corporate users. This year Microsoft will have 90% of R & D investment for cloud computing, there will be 70% of Huawei's R & D investment for cloud computing, the world's Internet companies and telecommunications companies to enter the cloud computing, cloud computing technology at the same time as academic industry research focus areas.

Google 翻译

源语言: 中文 目标语言: 阿拉伯语 翻译

无论是2010年还是2011年,云计算都被评为十大战略技术之首。Google云计算平台规模已经超过200多万台计算机,其企业用户数量超过300万家。今年微软公司将有90%的研发投入用于云计算,华为公司将有70%的研发投入用于云计算,全世界的互联网企业和电信企业都在向云计算进军,云计算技术同时成为学术界的研究热点领域。

允许输入拉丁字符的拼音

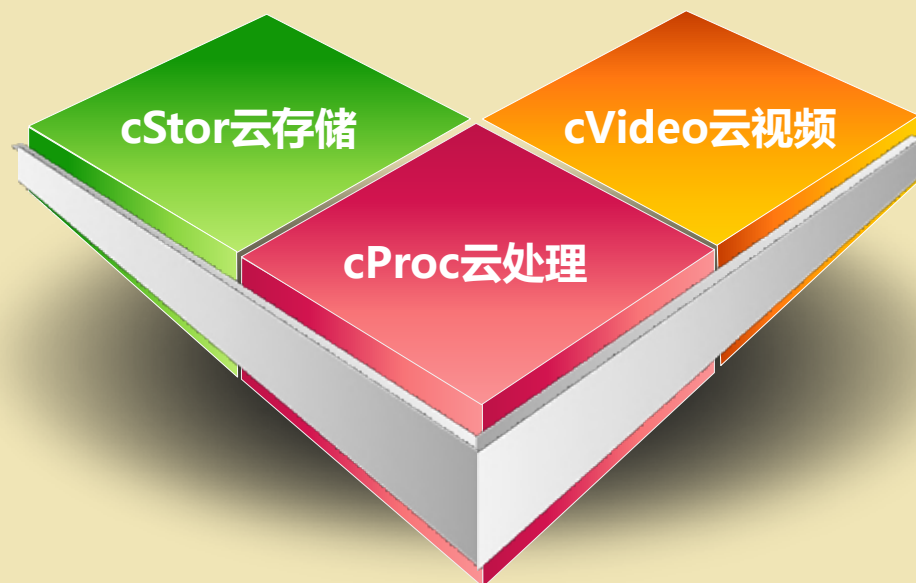
将中文译成阿拉伯语

سواء كان عام 2010 أو عام 2011 ، وقد الحوسبة السحابية أول المشورة التقنيات الاستراتيجية حجم غوغل سحابة منصة الحوسبة لديها أكثر من 200 مليون جهاز كمبيوتر ، وعدد أكثر من ثلاثة ملايين مستخدمى الشركات. هذا العام ان مايكروسوفت ستطرح 90 ٪ من R & D الاستثمار لالحوسبة السحابية ، سيكون هناك 70 ٪ من R هوائي والتطوير الاستثمار من أجل الحوسبة السحابية ، وشركات الإنترنت في العالم وشركات الاتصالات السلكية واللاسلكية لدخول الحوسبة السحابية الحوسبة السحابية ، والتكنولوجيا ، في نفس الوقت الذي الأكاديمية مجالات الصناعة التركيز على البحوث.



云计算技术

三条大数据产品线



CSTOR云存储系统

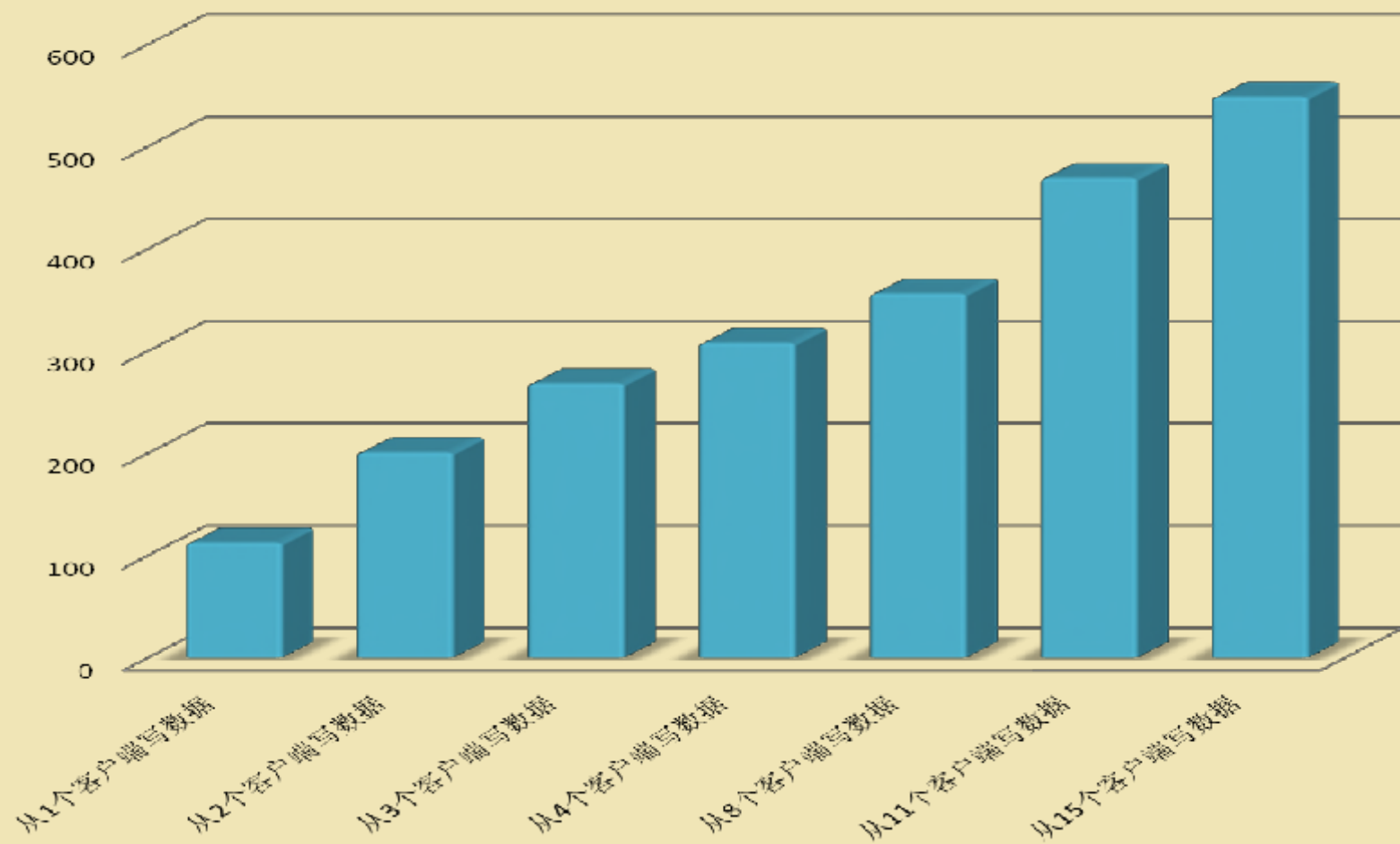


CSTOR云存储管理界面



CSTOR性能

往15个存储节点写数据(MB/s)



全球首款64位低功耗服务器



主要参数指标:

产品型号	技术规格
外形规格	mini ITX (170mm×170mm)
处理器	Intel Micro Server S1269 , TDP 11.7W
内存	DDR3,8GB ECC,SODIMM,1333MHz
存储接口	支持16个SATA 2.0/3.0盘位
网络模块	2路千兆以太网口
操作系统	Linux Fedora 17 (Core 3.8.4)
管理系统	支持IPMI、iKVM、vMedia (可选配)
电源系统	ATX电源接口，单板功耗25W
工作环境	工作温度：10℃~40℃，每小时变化<10℃ 相对湿度：8%~90% (非凝露、非结霜)
功耗	25W

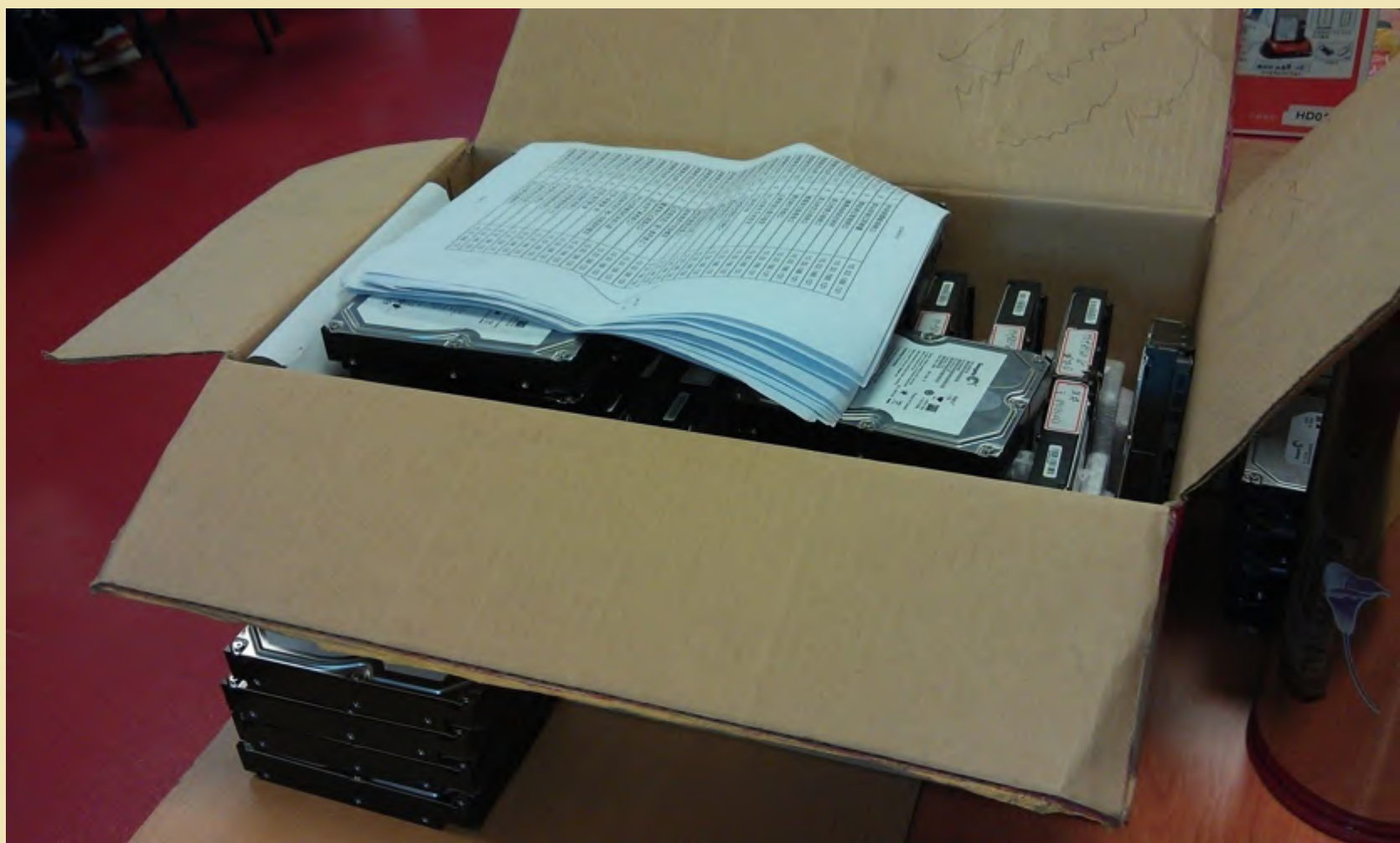
CSTOR A8000云存储系统



助力南京市公安局



助力南京市公安局



部署于南京政务数据中心

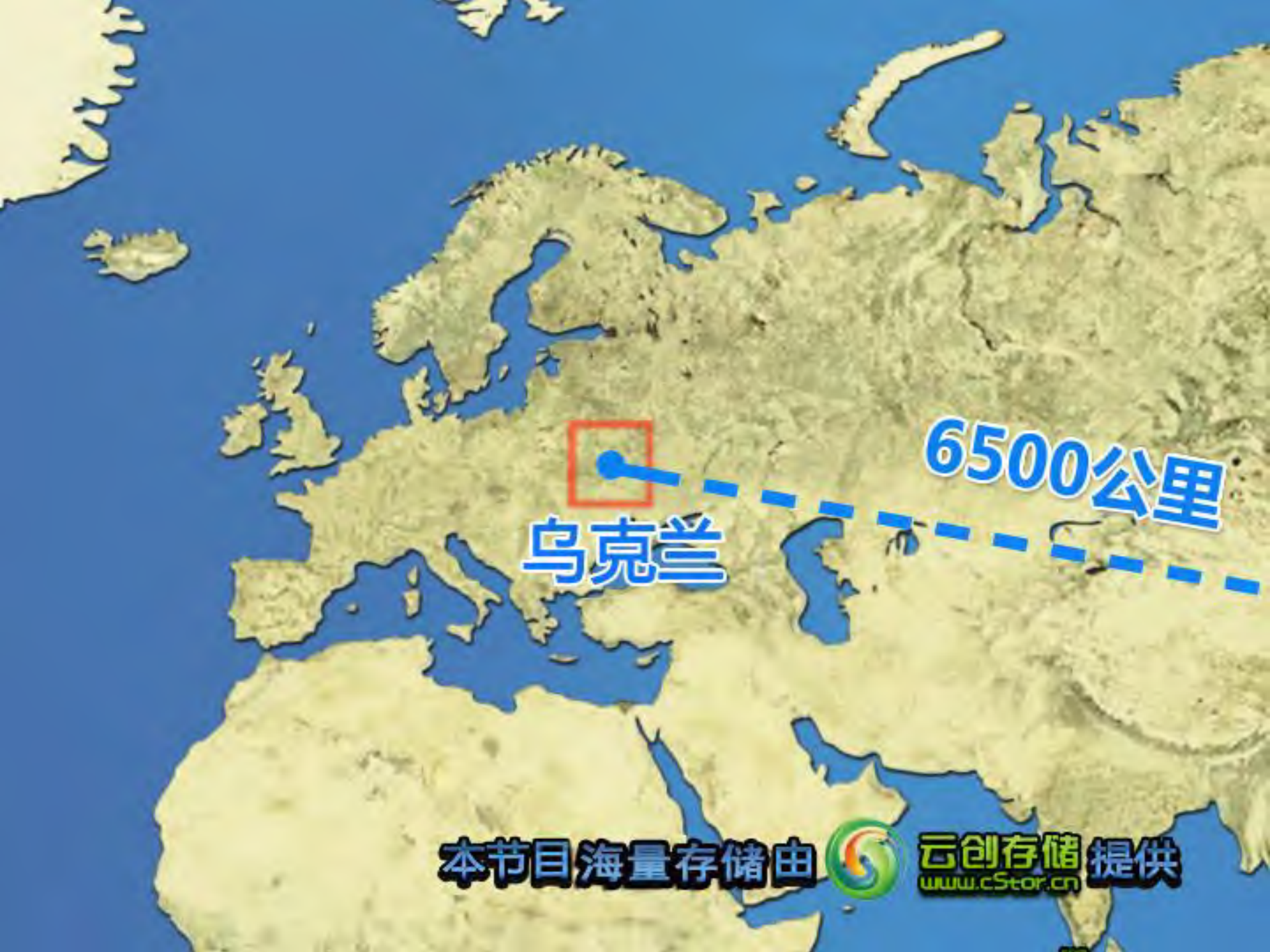


助力镇江、兰州等智慧城市建设



部署于中科院





乌克兰

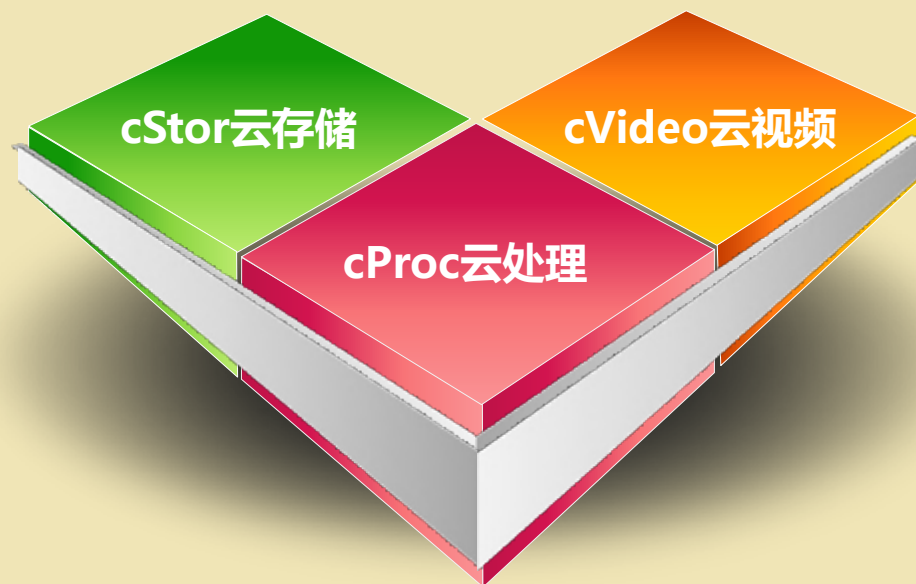
6500公里

本节目海量存储由  云创存储 提供
www.cStor.cn

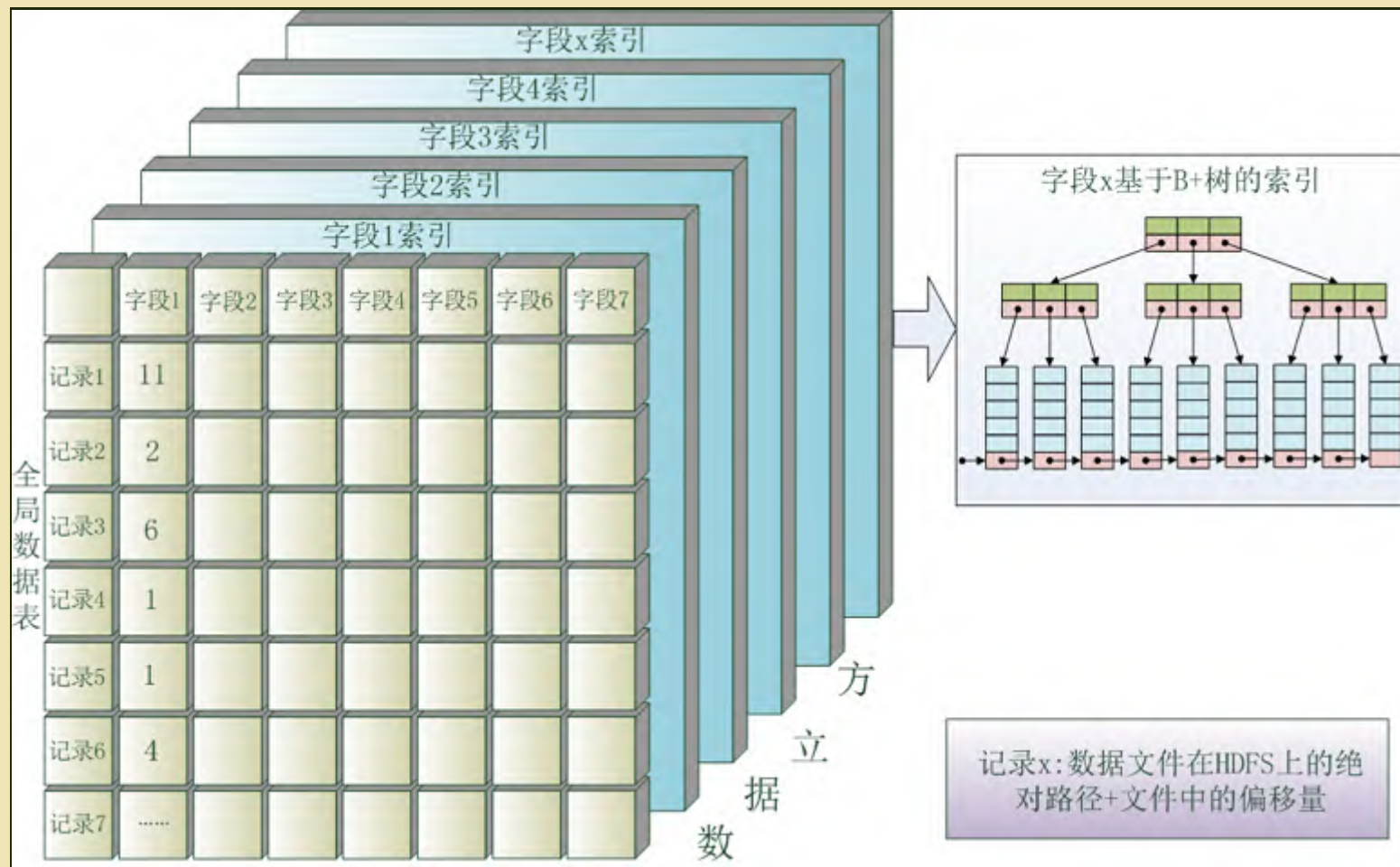
MINICLOUD迷你云



三条大数据产品线

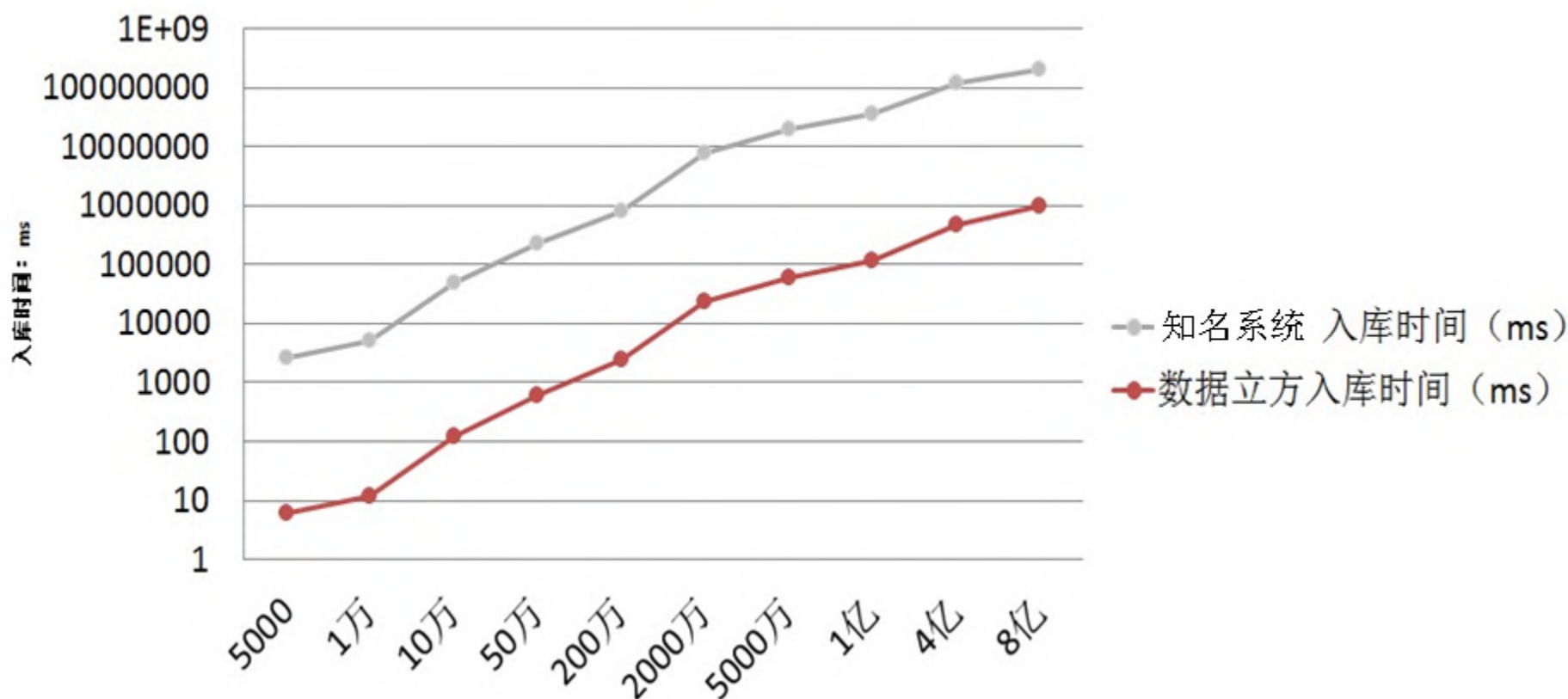


数据立方(DATACUBE)



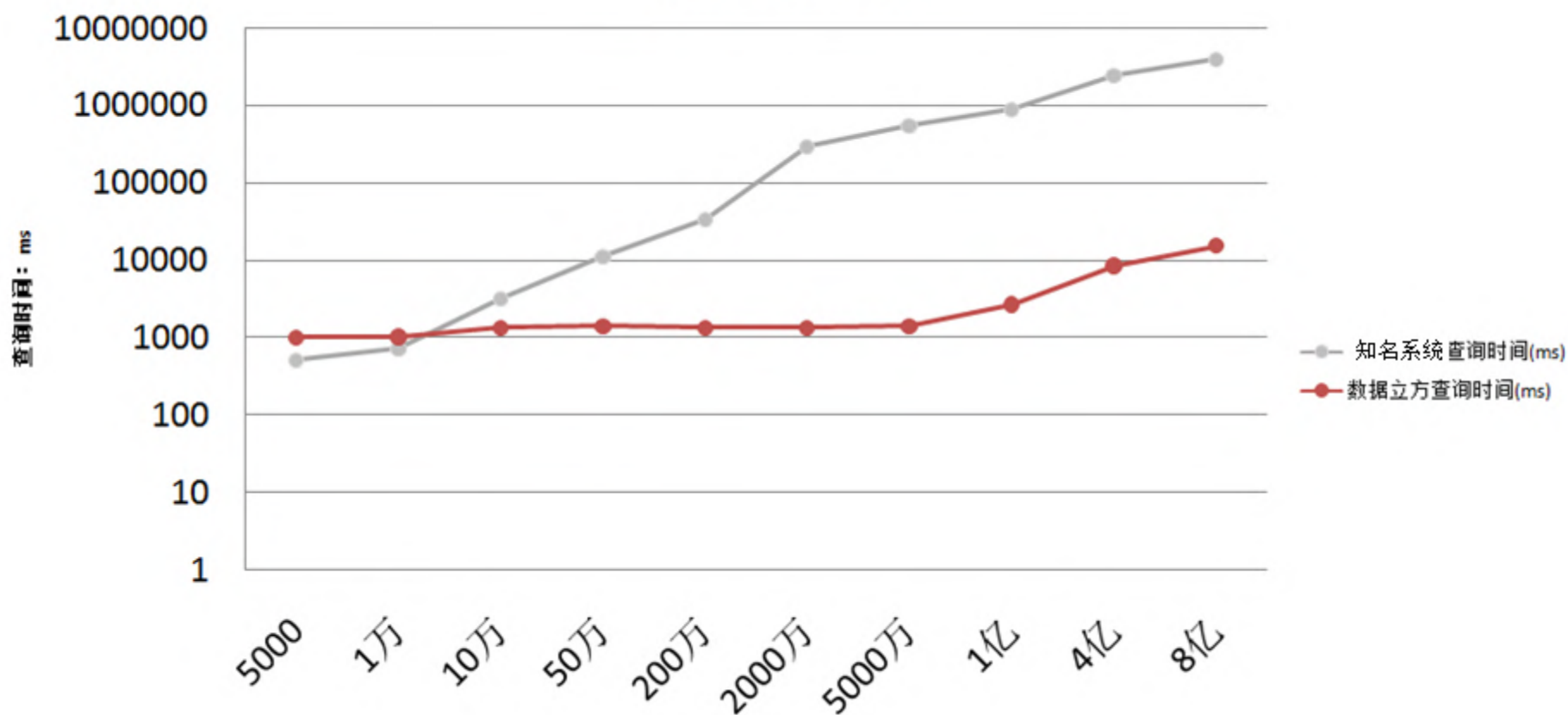
入库性能

入库性能对比图

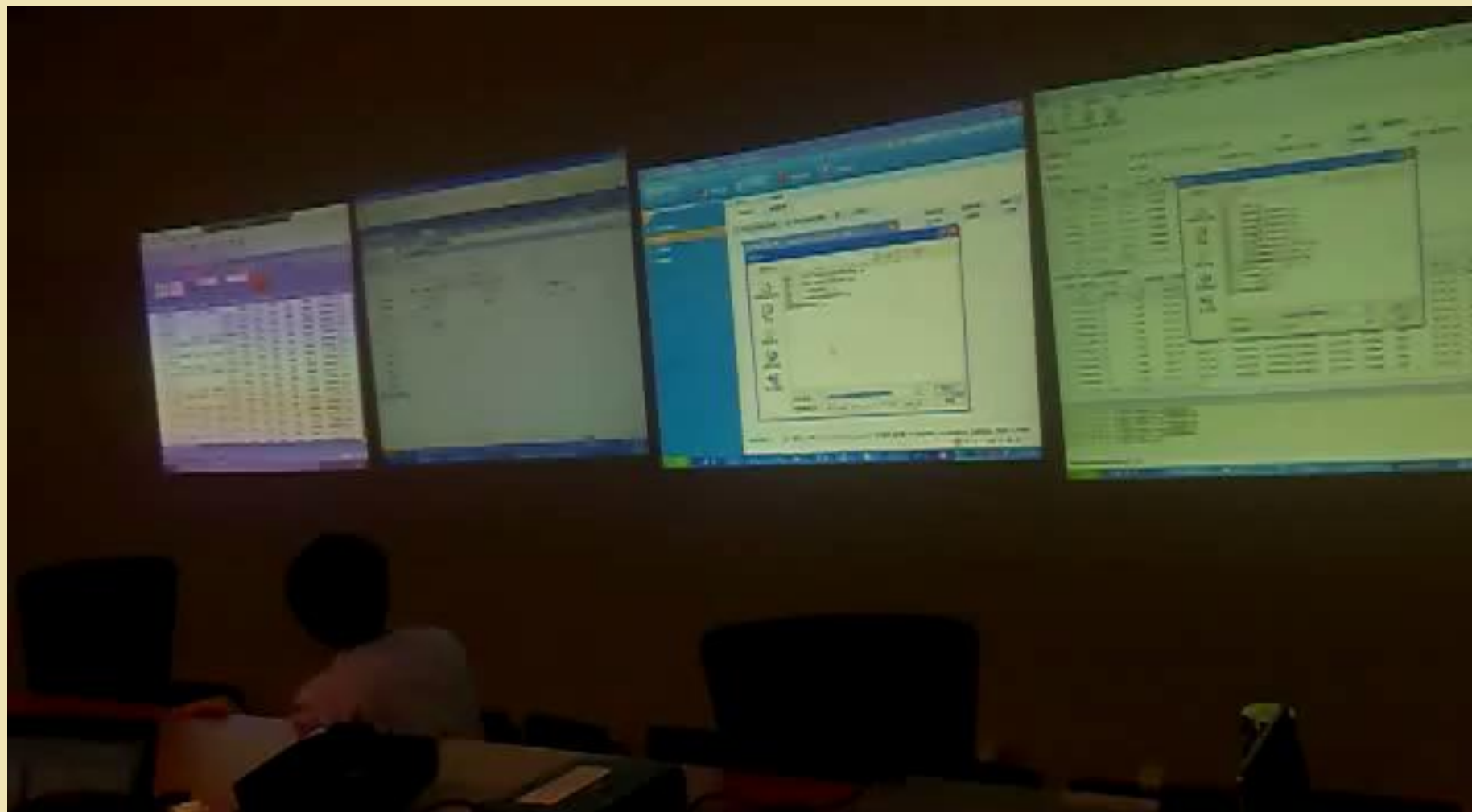


查询性能

查询性能对比图



中国移动比赛现场



部署到江苏移动



智慧交通云平台



智能交通云处理演示



历史总记录数

当前入库速度

2 1 5 6 6 4 6 2 5 5 0 条

7 8 5 6 条/s

车辆通过记录

卡口号	车牌号	通过时间
2	苏B30521	2012-4-25 10:34:14
16	苏B45821	2012-4-25 10:34:14
17	苏A95874	2012-4-25 10:34:14
18	苏B05634	2012-4-25 10:34:14
19	苏G11254	2012-4-25 10:34:14
20	苏A36841	2012-4-25 10:34:14
1	苏B75412	2012-4-25 10:34:14
2	京B30584	2012-4-25 10:34:14
3	苏A25684	2012-4-25 10:34:14
4	苏E32869	2012-4-25 10:34:14
5	苏C56894	2012-4-25 10:34:14

视频监控信息



导出本页 导出全部

卡口信息列表

序号	卡口名称	号牌号码	号牌种类	通过时间	通过方向	卡口详细地址	所属辖区	号牌照片
1	和平路红军街	冀A6C336	小型车牌	2013-10-19 15:07:27	东向西	和平路红军街	null	
2	西二环建国路南	冀A6C336	小型车牌	2013-10-19 15:23:32	北向南	西二环建国路南	null	
3	中华大街新华路北	冀A6C336	小型车牌	2013-10-21 17:55:49	南向北	中华大街新华路北	null	
4	中华大街工农路	冀A6C336	小型车牌	2013-10-23 14:25:34	南向北	中华大街工农路	null	
5	槐安路时光街东	冀A6C336	小型车牌	2013-10-23 20:23:15	西向东	槐安路时光街东	null	
6	中华大街工农路	冀A6C336	小型车牌	2013-10-24 09:59:29	南向北	中华大街工农路	null	
7	中华大街工农路	冀A6C336	小型车牌	2013-10-25 11:51:09	北向南	中华大街工农路	null	
8	中华大街新华路北	冀A6C336	小型车牌	2013-10-25 12:00:48	南向北	中华大街新华路北	null	
9	裕华路青园街	冀A6C336	小型车牌	2013-10-29 09:23:03	东向西	裕华路青园街	null	

http://10.25.42.241/p2/2013/10/29

文件(F) 编辑(E) 查看(V) 收藏(A) 工具

后 地址(D) http://10.25.42.241/p2/2013/10/29



完毕

查看照片 查看照片

查看照片 查看照片

查看照片 查看照片

每页显示 10 记录 记录共 59 条记录 记录分[6 / 6]页 首页 上一页 下一页 尾页 转到 页

正在下载图片 http://127.0.0.1:8080/Hebei/css/kk_cls/welcome_bg.jpg

Intel-云创数据立方云计算一体机



英特尔最新的硬件平台和管理技术

- 英特尔E5家族系列CPU
- 英特尔服务器组件
- DataCenterManager能耗管理

云创存储自主知识产权的软件架构

- 分布式文件系统cStor/HDFS
- 分布式数据库DataCube
- 任务执行引擎TaskMaster
- 数据立方云计算一体机管理系统

大数据处理利器，极具突破性的整柜解决方案



集群管理

设备监控

系统监控

磁盘监控

数据管理

表管理

数据入库

在线查询

对外接口

文件管理

任务管理

告警管理

告警配置

告警信息

告警日志

故障管理



集群管理

设备监控

系统监控

磁盘监控

数据管理

表管理

数据入库

在线查询

对外接口

文件管理

任务管理

告警管理 3

告警配置

告警信息

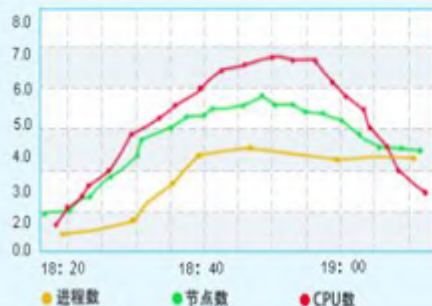
告警日志

故障管理

🔍 查询

时间: 最近 从 到
节点: IP 监控类型

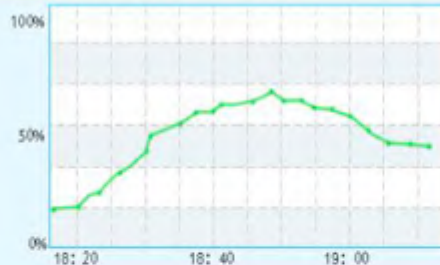
集群负载



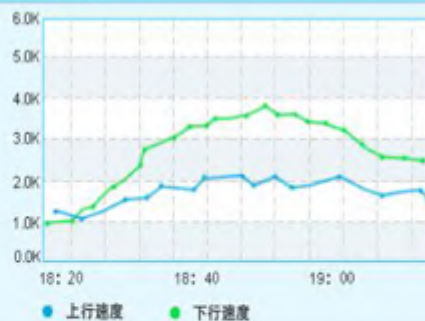
集群内存



集群CPU



集群网络



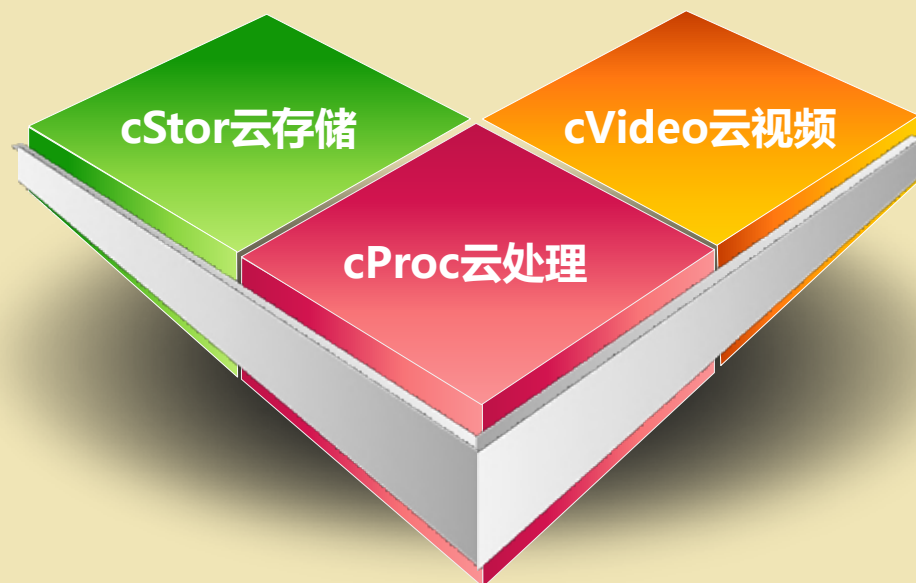


数据立方云计算一体机

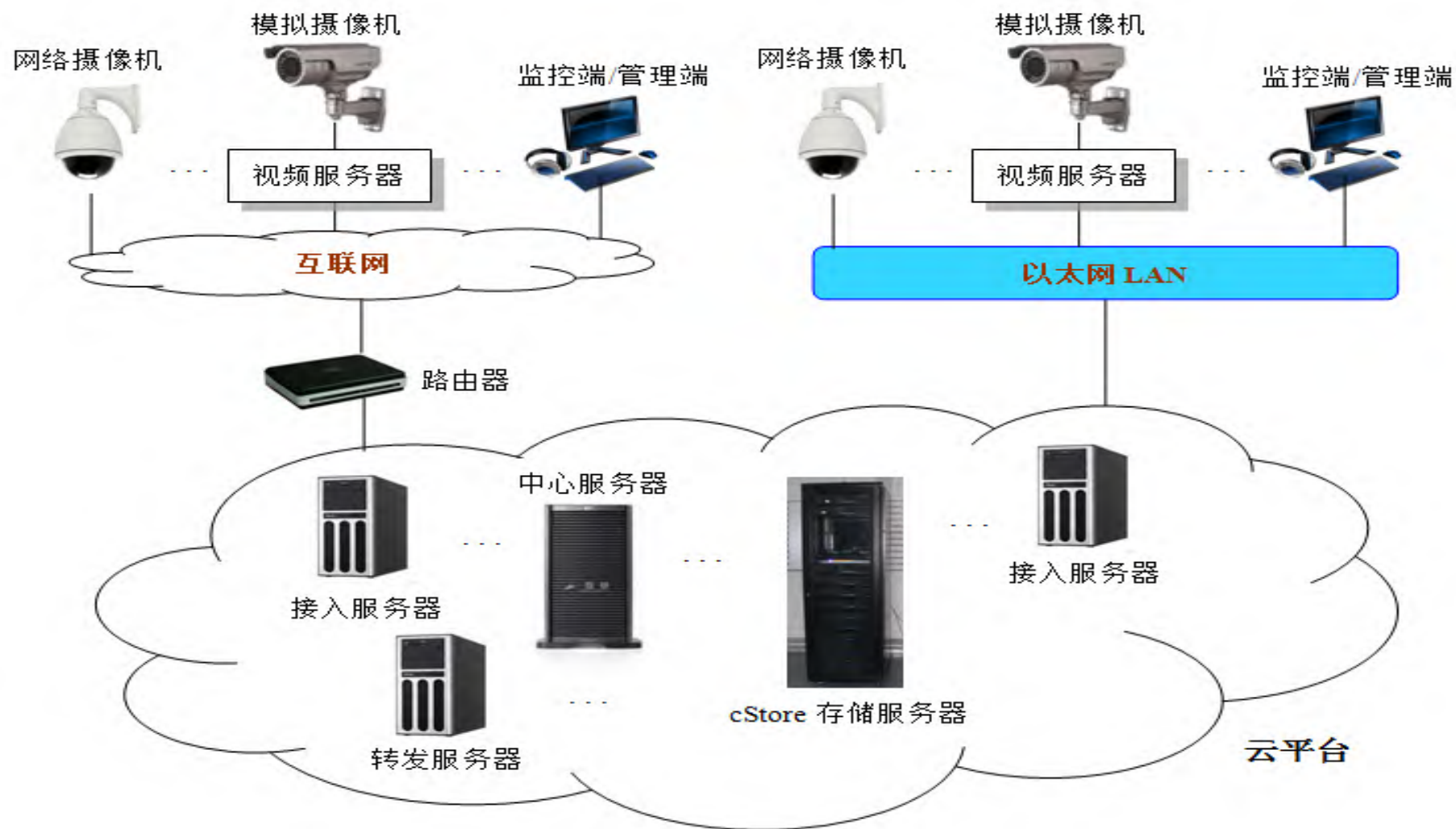
DataXBase大数据库机



三条大数据产品线



城市级别的视频监控



CVIDEO智能分析



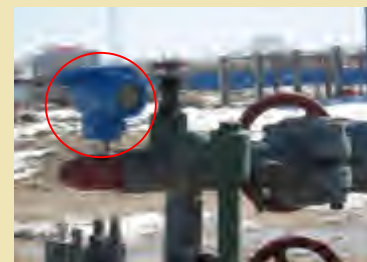
人或车的越界检测



人流量检测



可疑遗留物检测



物品保全侦测



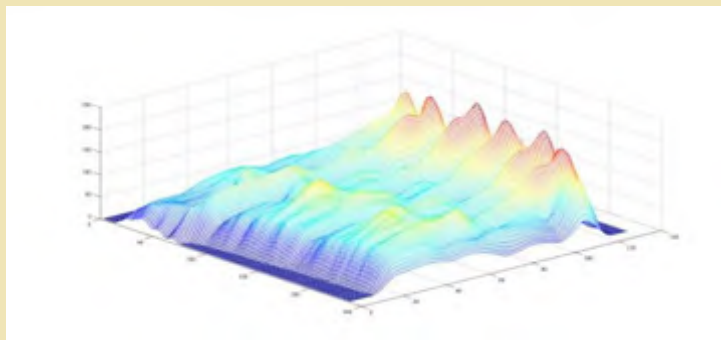
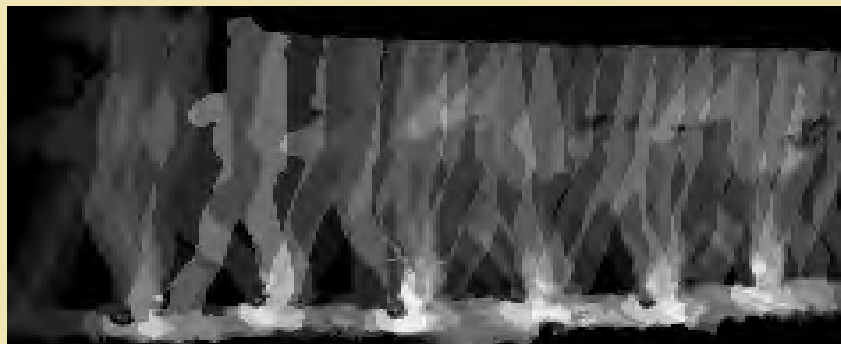
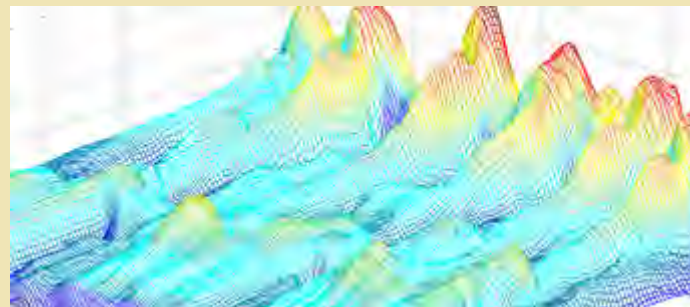
车辆逆行检测



人脸识别

CVIDEO智能分析

■ 视频DNA技术



目标运动视频DNA序列生成

视频DNA可视化图谱



第二届夏季青年奥林匹克运动会 主运行中心

Main Operations Centre (MOC) of Nanjing 2014 Summer Youth Olympic Games







场馆设施遗产

Legacy of Venues and Facilities

2014

赛事遗产 Legacy of the Games

与聪明人一起做精彩的事!

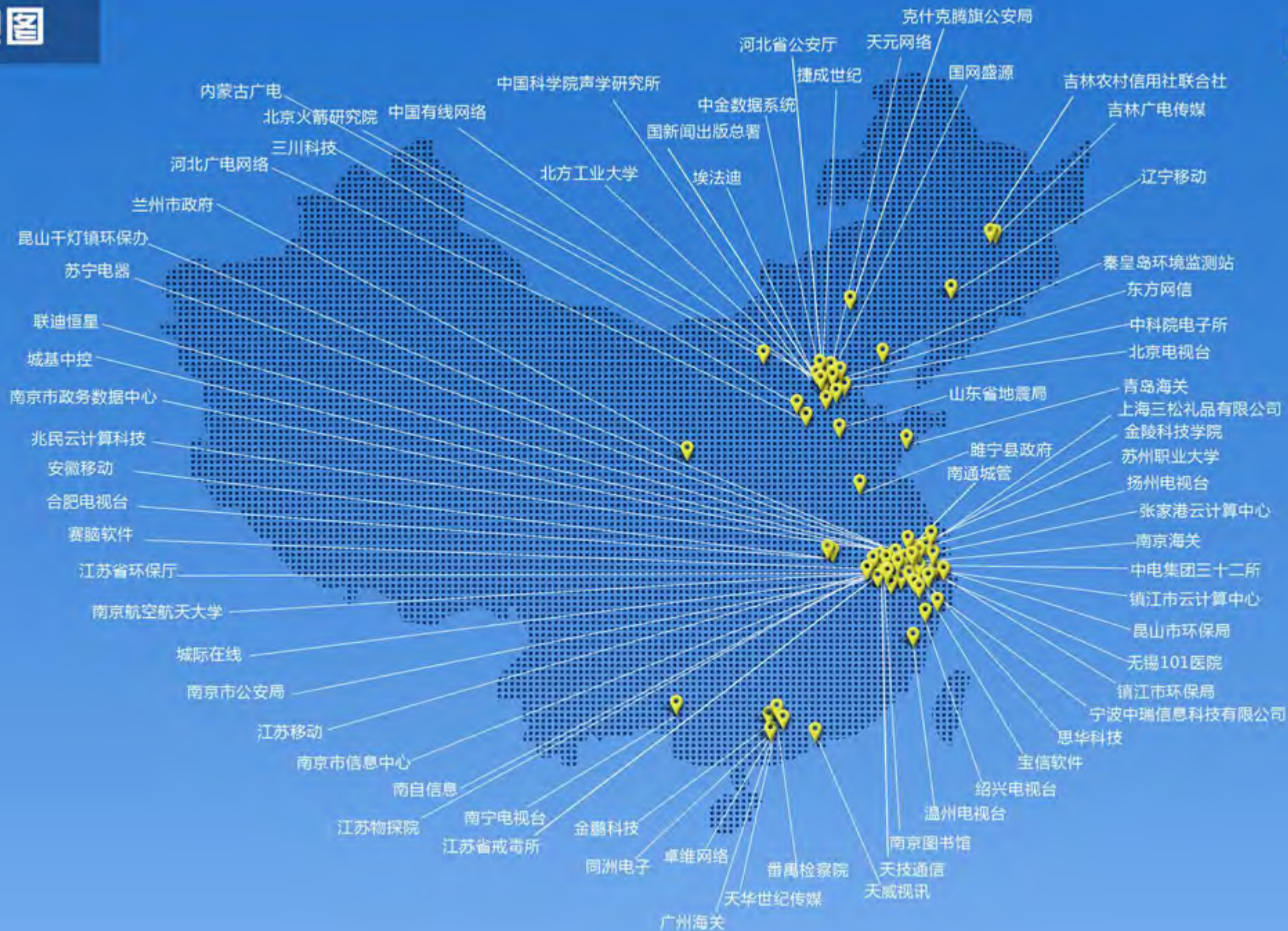
Cre

rt.

云助青奥
创引未来

赠：南京云创存储科技有限公司

南京青奥会主运行中心技术支持保障组
二〇一四年八月



关于公布2012年度第二批江苏省科技型中小企业名单的通知

发布时间：2012-12-13 作者：中小企业科技创新处 文章类型：原创

[【加入收藏】](#)

[【关闭窗口】](#)

[【打印此文】](#)

苏中小科技〔2012〕1013号

各市、县（市）经信委（中小企业局）：

为加快培育创新型企业集群，促进中小企业创新转型，根据《江苏省科技型中小企业评审推荐办法》，经企业自愿申请，各市、县（市）推荐和专家评审，确认南京云创存储科技有限公司等629家企业为2012年度第二批江苏省科技型中小企业，现予公布。

希望被评定的省科技型中小企业，进一步加大研发投入，完善创新体系，增强创新能力，为全省中小企业加快转型升级起到示范带动作用。各级中小企业管理部门要加强对省科技型中小企业的跟踪管理、指导和服务，优先推荐其中报国家和省扶持中小企业发展专项资金项目。

附件： 2012年度第二批江苏省科技型中小企业名单

» PM2.5监测云

背景



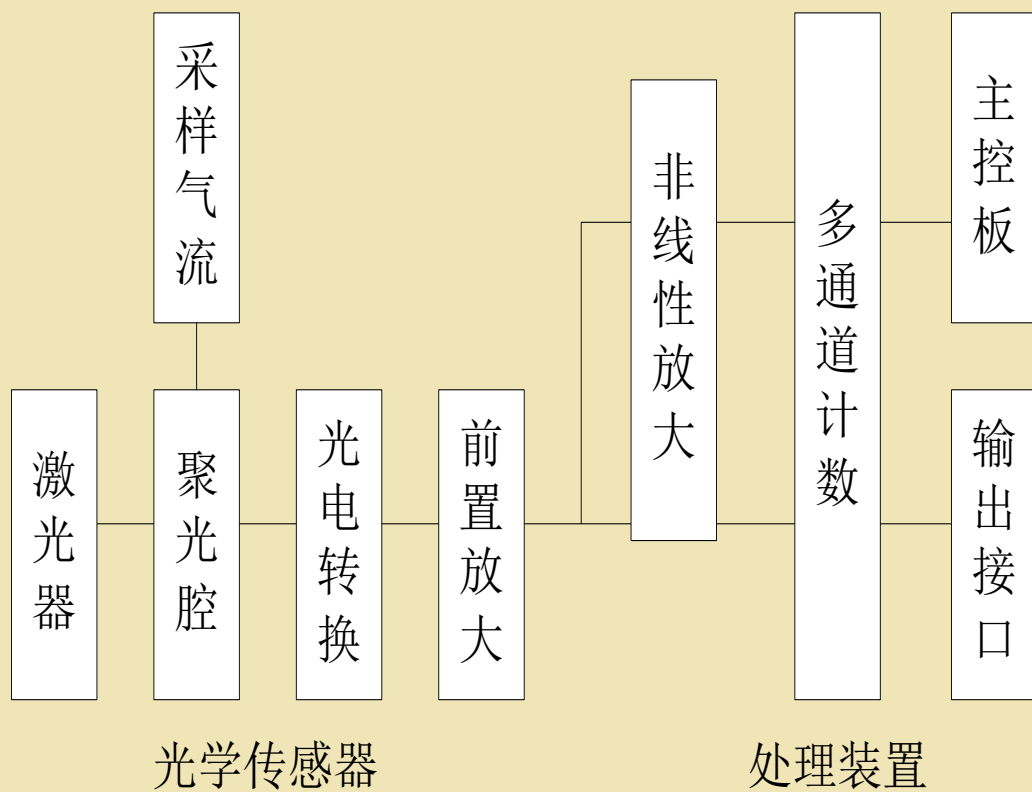
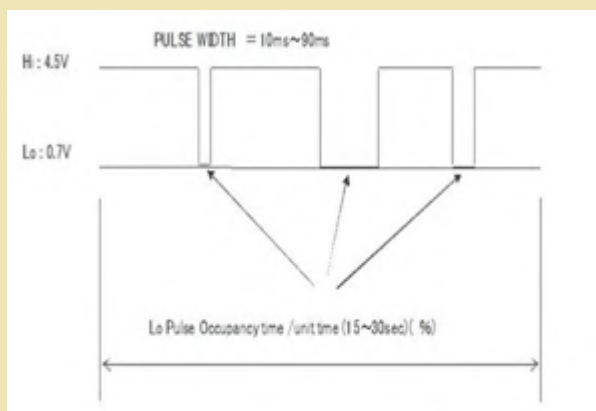
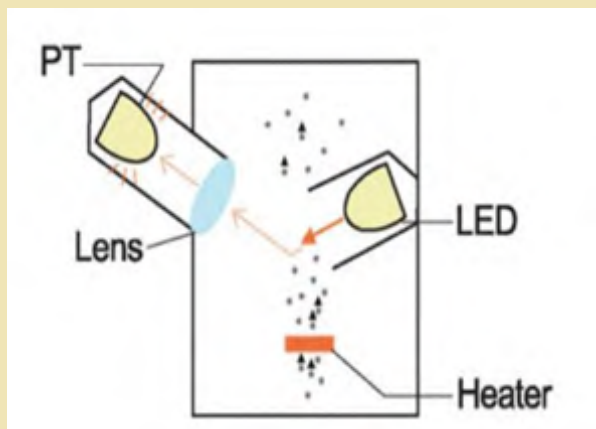
现有**PM2.5**监测设备非常昂贵，需要专门机房安放，因而城市里只部署了非常少的站点，无法对空气污染进行精确实时监测和分析。



监测节点



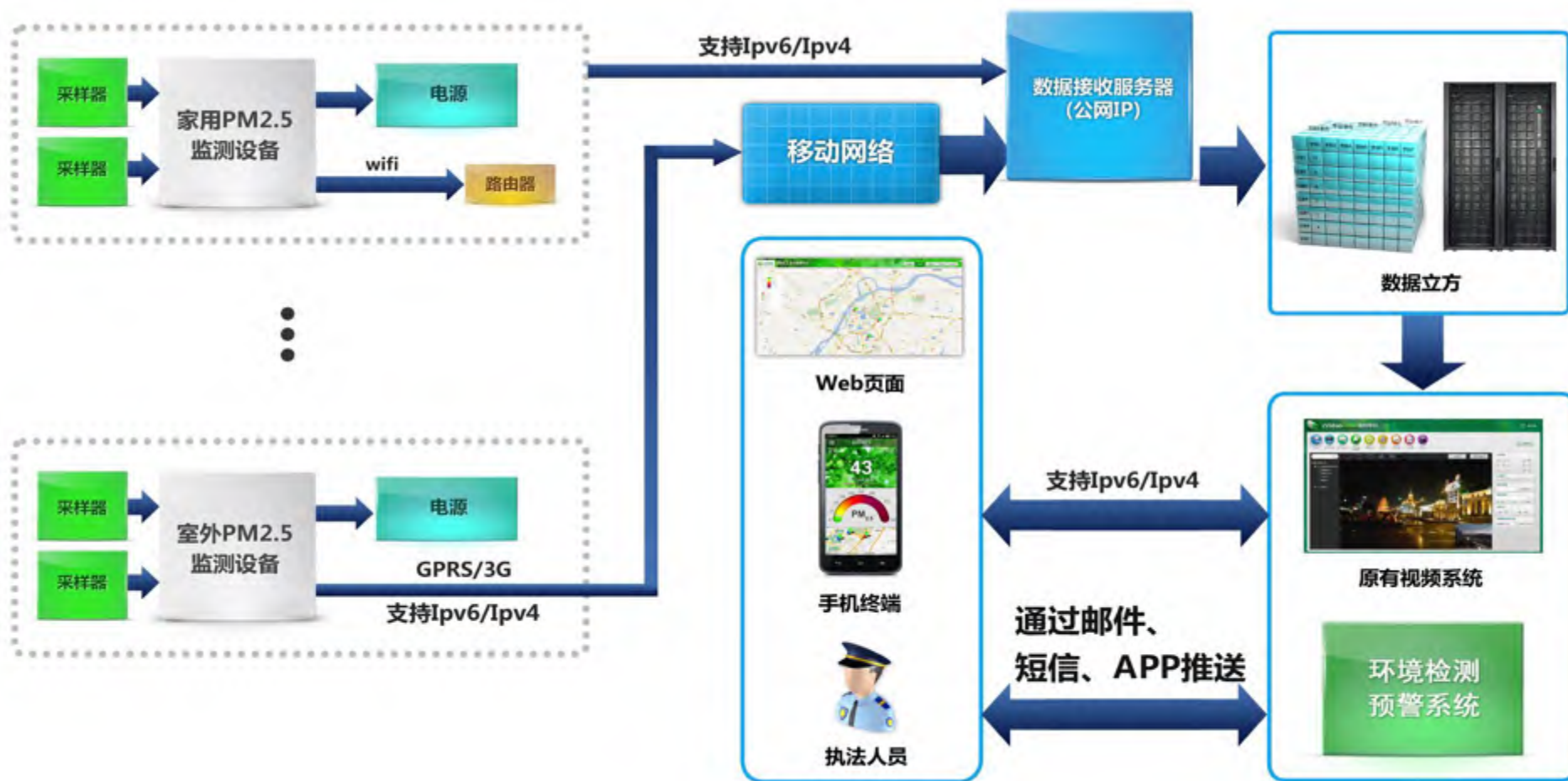
传感器



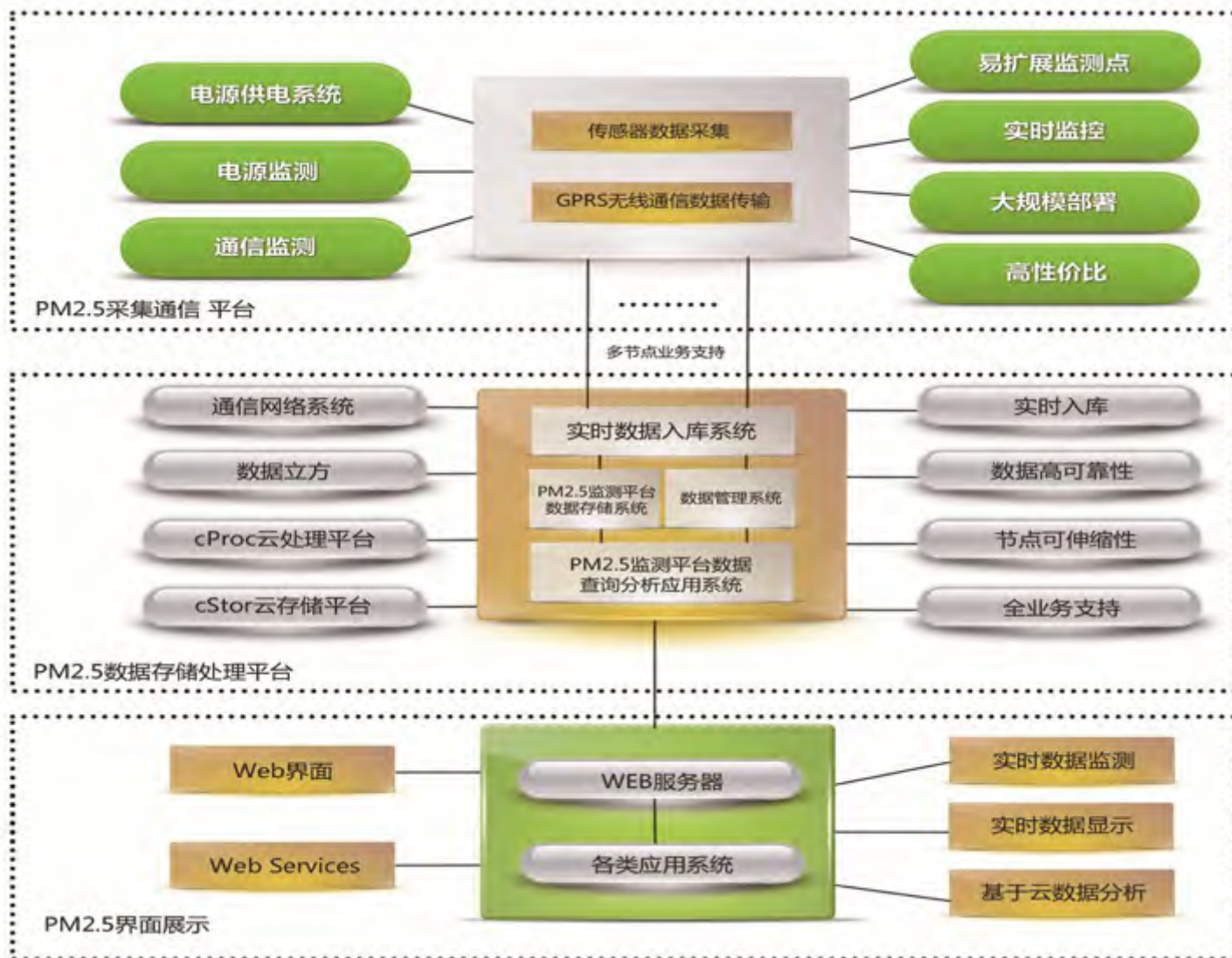
空气质量指标连续监测仪



平台



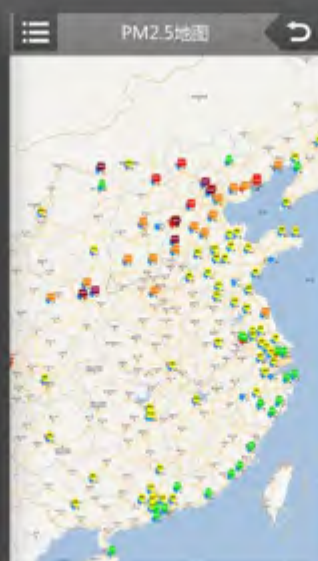
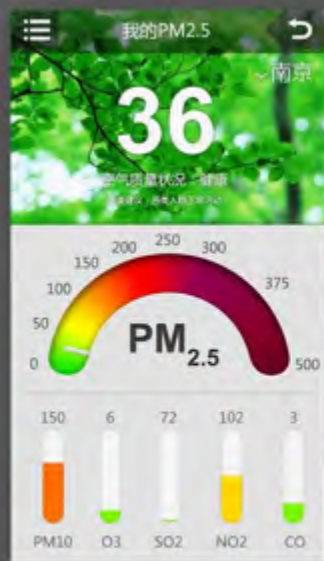
架构



界面



APP



实时排名

空气质量最差城市TOP10

排名	城市	指数	状况
1	石家庄	405	危险
2	秦皇岛	385	危险
3	石家庄	405	危险
4	秦皇岛	385	危险
5	石家庄	405	危险
6	秦皇岛	385	危险
7	石家庄	405	危险
8	秦皇岛	385	危险
9	石家庄	405	危险
10	秦皇岛	385	危险

空气质量最佳城市TOP10

排名	城市	指数	状况
1	南京	8	优
2	唐山	15	优



部署

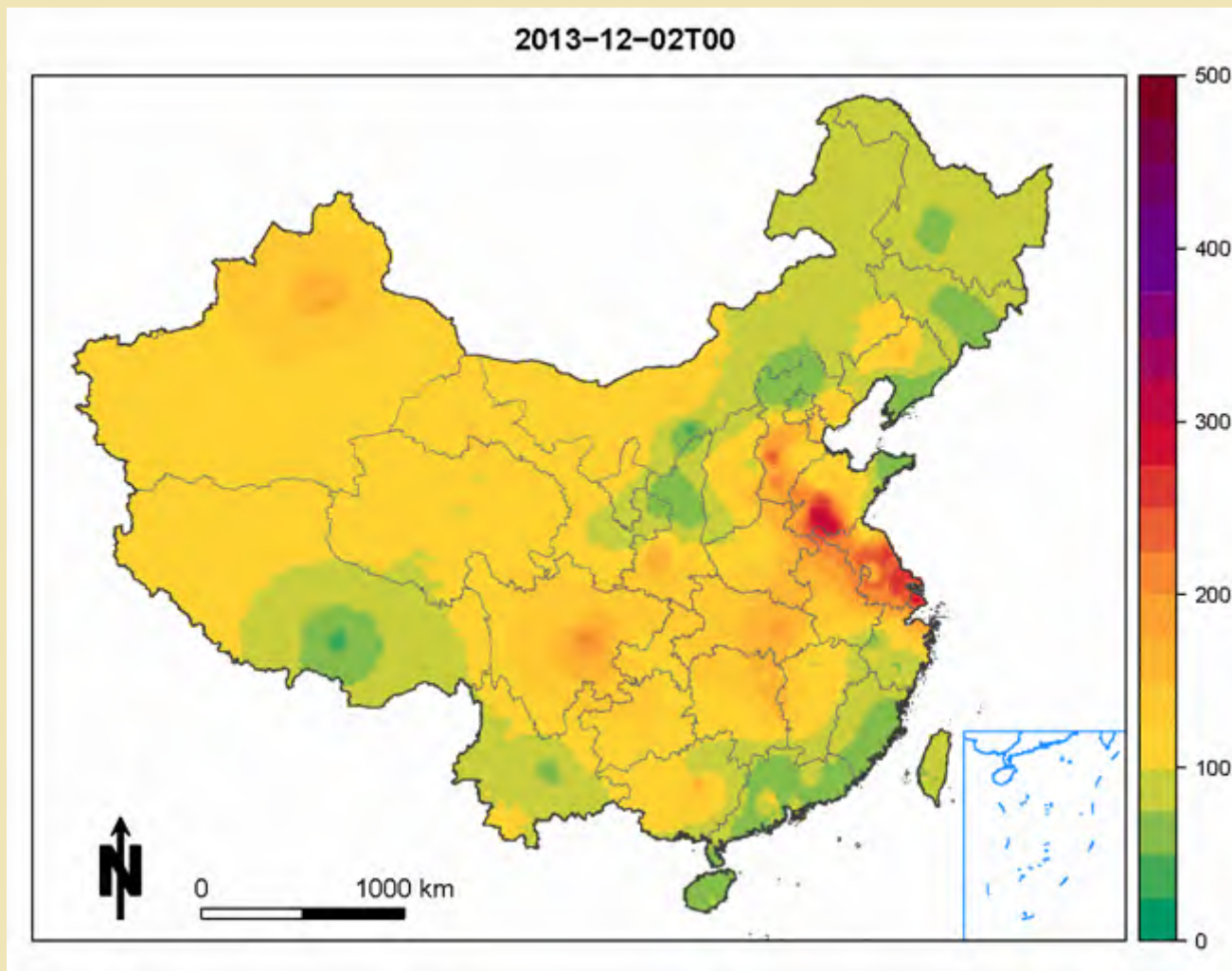


[illegible]

精度



趋势分析



对比

序号	对比内容	云创存储PM2.5监测平台	传统PM2.5在线监测仪器
1	灵活性	设备小巧，可灵活放置在室外，很灵活	设备庞大，需要专门房间进行放置仪器，不灵活
2	部署方式	PM2.5监测云节点是设备箱形式，可在立杆或者墙体挂载即可	需要机架进行放置仪器，同时对房间有高要求
3	部署规模	基于云计算平台，价格低廉，可大规模部署	需要专门房间进行处理，价格昂贵不可能进行大规模部署
4	采样时间	秒级采样，迅速	小时采样
5	检测方式	在线监测，连续监测	在线监测，连续监测
6	参考价位	1万人民币一个PM2.5监测云节点	10万-100万美元部署一个监测点
7	采样原理	光散射，自动分析方法	微量振荡天平法、 β 射线法，自动分析方法
8	设备维护	定期进行清洁(一般半年进行一次)，方便	须自动或手动置换滤网膜，或定期设备检修和维护
9	测试精度	比传统PM2.5在线监测仪器略低，可完全反映污染情况	测试准精确度高
10	数据存储和处理	云计算和云存储，可以进行大数据处理，百万节点数据处理	单个测试节点进行传统数据库的存储和处理
11	数据分析	结合分布式计算和云存储框架，可以进行海量数据的挖掘	传统关系型数据库，当数据量过大时无法有效的对历史数据分析
12	环境动态展示	利用并行计算和云存储的优势，可以实时的动态展示任何时段的污染情况	传统关系型数据库，只能展示某一时刻的污染情况
13	污染预警	利用并行计算和云存储的优势，通过大量的监测节点以及污染分布算法，可以有效的进行污染源预警	无法预警
14	对外服务接口	提供污染历史数据下载	提供历史污染数据下载
15	动态扩展	可以在不影响服务的情况下进行后台扩展	扩展复杂，需要停止服务



地震大数据

山东大数据应用实践

- 为解决山东区域中心地震数据存储、管理与服务问题，2013年12月，山东省地震局利用“省十二五”资金，启动了“地震数据处理平台”工程项目。
- 项目目标是采用大数据技术，实现山东台网历史数据（测震、强震、前兆等）和实时产出数据的在线综合存储、管理和服务，为进一步数据的分析和挖掘，提供数据资源和计算资源。

硬件软件环境

- 硬件：18个节点（X86服务器）组成的（存储能力180TB）区域中心地震大数据平台
- 软件：大数据平台支撑软件（分布式云存储，分布式云处理，云计算数据库）+应用软件

支持100PB量级数据存储，100GBps数据流实时索引，SQL形式HIVE访问，API多语言访问，Sqoop、Flum等ETL工具访问，支持JDBC通用接口访问等。

数据

- 历史连续波形数据（2002-2012）
SEED卷：2008-2012光盘文件
“九五”台网格式：2002-2007
- LISS实时波形 LISS服务器
- 台网、台站、仪器参数管理系统

应用界面



■ 引用自 王方建PPT 2014.12

数据查看



数据查看部分可实时查询各台站分量的历史波形数据。

■ 引用自 王方建PPT 2014.12

数据统计



数据统计部分可实时统计出数据库中数据总量及每个台站及台站下面每个分量的数据量。对于实时数据可统计出前一分钟、前一小时的入库量。

■ 引用自 王方建PPT 2014.12

数据管理

山东地震数据管理与服务云平台

分布式文件系统管理

上传文件 新建文件夹 path: / 全选 下载

5 * 记录数 搜索:

名称	类型	大小	块大小	修改时间	允许权限	用户	用户组	操作
<input type="checkbox"/> Desert.jpg	文件	826.11 KB	64 MB	2014-11-11 15:24	rw-r--r--	hadoop	supergroup	删除
<input type="checkbox"/> hbase	文件夹			2014-11-24 18:50	rw-r-xr-x	hadoop	supergroup	删除
<input type="checkbox"/> historyData	文件夹			2014-08-13 18:25	rw-r-xr-x	hadoop	supergroup	删除
<input type="checkbox"/> img	文件夹			2014-12-05 18:32	rw-r-xr-x	Mr.Yang	supergroup	删除
<input type="checkbox"/> user	文件夹			2014-08-12 16:21	rw-r-----	hadoop	supergroup	删除

显示 1 - 5 来自 5 的记录

云创存储

版权所有 © 南京云创存储科技有限公司 2011-2013. 保留一切权利。(苏ICP备11060547号-1)
云创存储-领先的云存储,云计算产品供应商

本系统中所有数据都是存储在分布式文件系统中，虚拟成完整的目录结构。

■ 引用自 王方建PPT 2014.12

台网、台站、仪器参数管理

山东地震数据管理与服务云平台

台网信息管理

新增

5 records

序号	台网编号	台网名称	经度	纬度	海拔高度	台网类型	备注	编辑	删除
1	25	江苏鲁地中心	118.0	33.0	120.0	25	江苏鲁地中心	修改	删除
2	27	山东鲁地中心	118.0	33.0	120.0	50	山东鲁地中心	修改	删除

显示 1 - 2 条记录 2 页/2 页

山东地震数据管理与服务云平台

100000 10.0

首页

新建数据

数据管理

数据管理

数据管理

数据管理

数据管理

数据管理

数据管理

数据管理

数据管理

台站信息管理

新增

新增

5

5 records

重置

序号	台站编号	台站名称	台站类型	隶属台网中心	经度	纬度	海拔高度	备注	编辑	删除
1	17001	泰山	TA	山东鲁地中心	117.1213	36.11	209.8		修改	删除
2	17002	烟台	YTA	山东鲁地中心	121.8907	37.525	37.09		修改	删除
3	17003	聊城	LCM	山东鲁地中心	115.637	36.4111	35.72		修改	删除
4	17004	德州	TDH	山东鲁地中心	118.4458	34.7021	59.44		修改	删除
5	17005	滨州	BHZ	山东鲁地中心	118.0	38.0	390.13		修改	删除
6	17006	菏泽	HZT	山东鲁地中心	115.61	35.12	61.28		修改	删除
7	17007	东营	DOY	山东鲁地中心	118.5	37.5	51.81		修改	删除
8	17008	潍坊	WFT	山东鲁地中心	119.12	36.7028	50.01		修改	删除
9	17009	威海	WCH	山东鲁地中心	122.0	37.0	61.0		修改	删除
10	17010	青岛	QHD	山东鲁地中心	120.34	37.5113	47.82		修改	删除

显示 1 - 10 条记录 6 页/6 页

1 2 3 4 5 6

山东地震数据管理与服务云平台

版权所有 山东地震数据管理与服务云平台 2019-2021 鲁ICP备19011111号 鲁公网安备37010502000000号 鲁

山东地震数据管理与服务云平台

测点信息管理

新增

5 records

测点编号	测点名称	测点类型	所属台网	经度	纬度	海拔高度(m)	测点地址	测点描述	创建时间	编辑	删除
75	烟台测点	LDH	烟台	118.48	36.41	209.8	42	42	2014-12-05 15:34:21	修改	删除
77	烟台测点	LDH	烟台	118.48	36.41	209.8	41	41	2014-12-05 15:31:48	修改	删除
78	烟台测点	LDH	烟台	118.48	36.41	209.8	40	40	2014-12-05 15:30:11	修改	删除
80	烟台测点	LDH	烟台	118.48	36.41	209.8	38	38	2014-12-05 15:21:00	修改	删除
81	泰山测点	LDH	泰山	118.11	36.11	209.8	37	37	2014-12-05 15:22:46	修改	删除

显示 1 - 5 条记录 18 页/18 页

山东地震数据管理与服务云平台

仪器管理

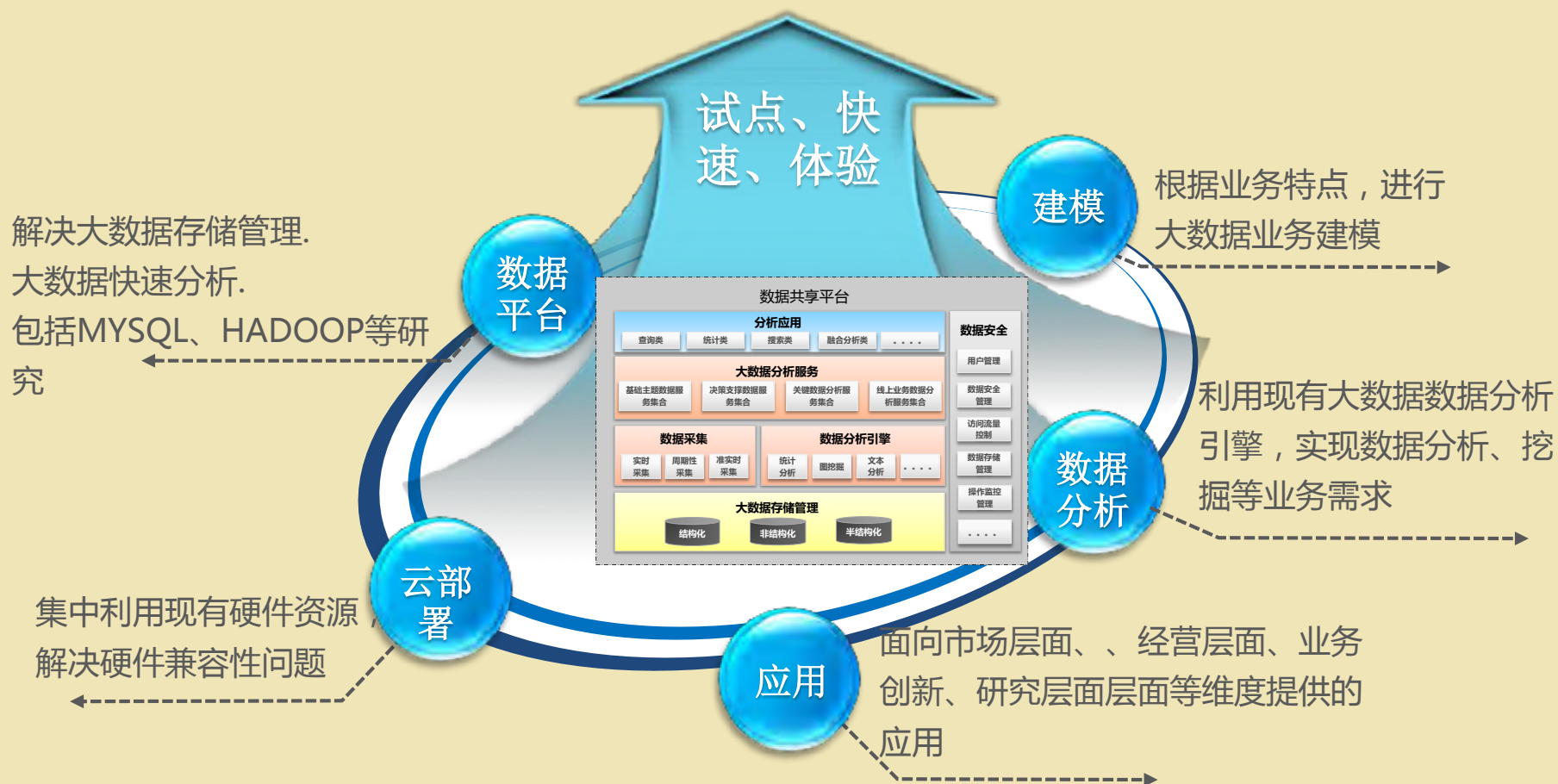
新增

5 records

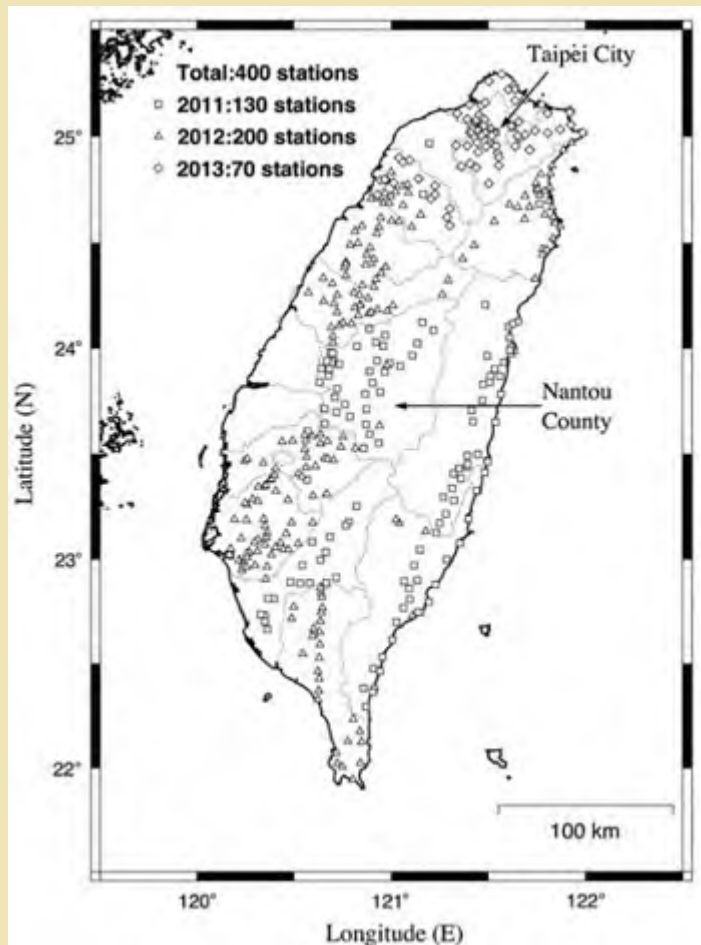
仪器编号	仪器名称	仪器类型	参数	分辨率	方向	状态	所属台网	创建时间	编辑	删除
23001	测震仪1	测震	77	77	HWZ	停用	烟台测点	2014-12-05 15:30:28	修改	删除
23002	测震仪2	测震	88	88	HWZ	停用	烟台测点	2014-12-05 15:34:48	修改	删除
23003	测震仪3	测震	88	88	HWZ	停用	烟台测点	2014-12-05 15:35:01	修改	删除
23004	测震仪4	测震	88	88	HWZ	停用	烟台测点	2014-12-05 15:36:15	修改	删除
23005	测震仪5	测震	75	75	HWZ	停用	泰山测点	2014-12-05 15:37:04	修改	删除

显示 1 - 5 条记录 18 页/18 页

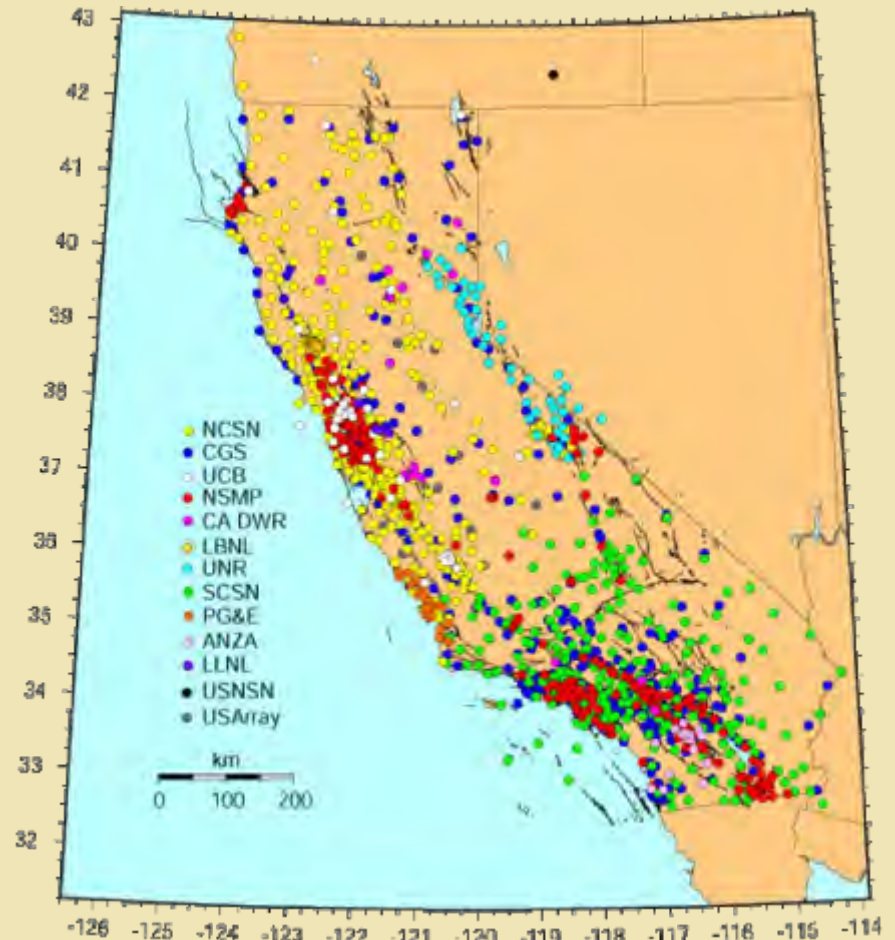
大数据应用要素



密集地震台网



台湾Palert密集地震预警网

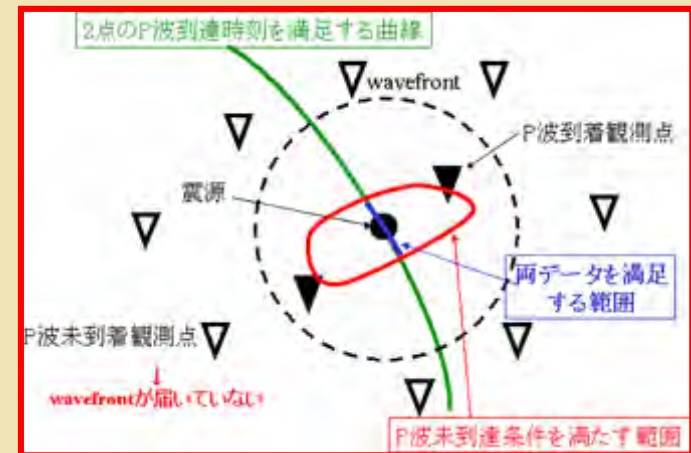


美国加州的地震动报警ShakeAlert网

从“复杂算法”到“简单算法”



小数据时代，在数据的限制无法突破的情形下，数据处理算法的研究越来越深入，发明的算法越来越复杂。



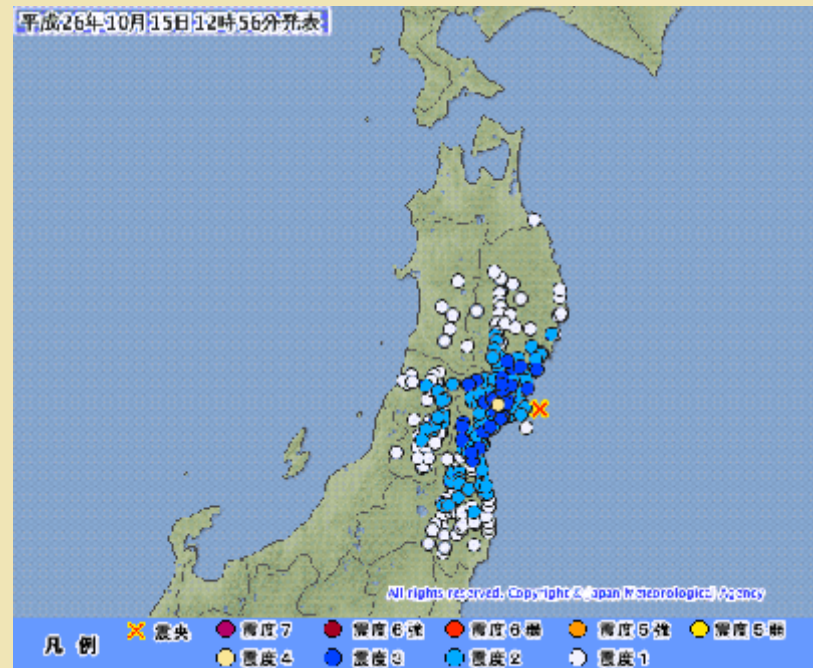
Hi-net観測点の処理方法
“着未着法”

当数据量以指数级扩张时，原来在小数量级的数据中表现很差的简单算法，准确率会大幅提高；大数据的简单算法比小数据的复杂算法更有效。

“审慎的决策” 到 “快速的决策”

小数据时代，通过收集和分析数据来验证这种假设；如果有一些数据有问题，就影响原有假设，决策与行动是审慎的。

大数据时代，我们不再受限于传统的方式，我们得到的是“是什么”，这就足以做出快速的决策与行动。因为机会和价值很快就会被刷新。



2014年10月15日日本宫城县金海4.5级地震
震后3分钟发表

“预警” 到 “预报”

密集地震观测网不仅仅是地震（测震）观测网，而且是“前兆网”

密集的温度网，将价格为百元的灵敏度为小数点后一位的地温计密集布设

密集的气体网，将极便宜的气体探测器密集布设

密集的地磁网，将一般的地磁仪密集布设

密集的压力网，将可以密集布设的应力应变仪密集布设

地震观测网将拥有数万、数十万传感器观测点。仅密集地震观测网，就将地震学带进大数据时代。更不要说3+1体系的其他巨量数据。

谢谢!

刘鹏微信

chinagloud

