

A futuristic digital interface with a glowing globe and data points. The scene is set against a dark blue background with a grid of light blue lines. A central globe is surrounded by various data points and lines, suggesting a complex network or data analysis. The overall aesthetic is high-tech and digital.

应用案例

第10章 大数据分析应用案例



1 让出行更加通畅——智能交通

2 让生活更加健康——智慧医疗

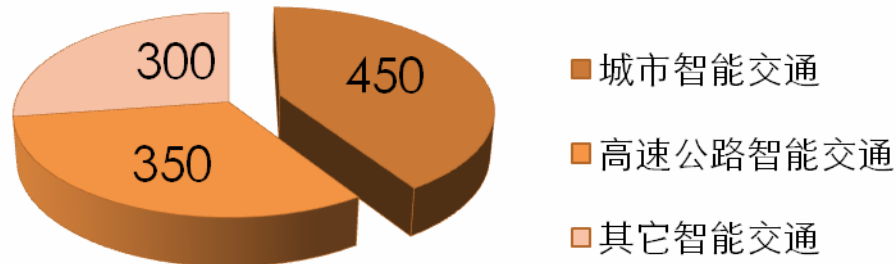
10.1 让出行更加通畅——智能交通

1. 智能交通是国家“十二五”重点规划

“十二五”规划中明确指出，将大力推进交通信息化建设，大力发展智能交通，提升交通运输的现代化水平。

“十二五”期间，国家将对智能交通投入超过**1000亿元**，其中城市智能交通投入达到**450亿元**，预计到2015-2020年，智能交通给相关行业带来的商机将超过**1000亿元**。

国家“十二五”规划智能交通投入



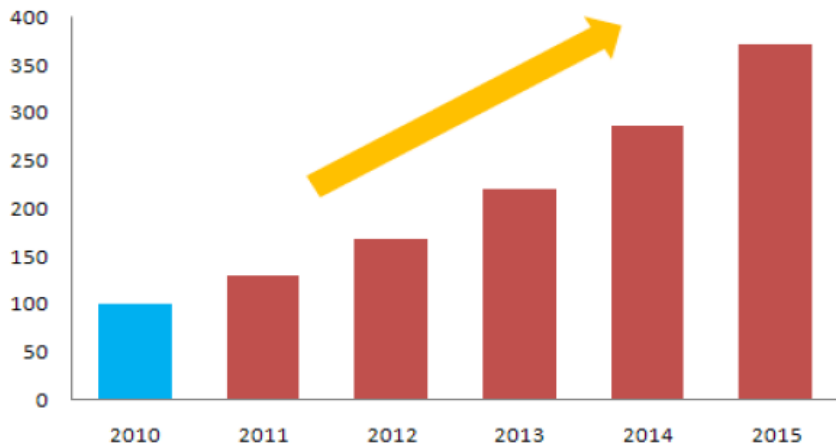
单位：亿元

数据来源：《交通运输“十二五”发展规划》

“十二五”规划智能交通投入

- ü 预计未来5年内，我国将在200个以上的大中型城市建立城市交通指挥中心。
- ü 其中，城市智能交通投资约450亿元人民币，高速公路智能交通系统投资约350亿元人民币，其它智能交通系统投资大概300亿元人民币。

城市智能交通市场规模预测



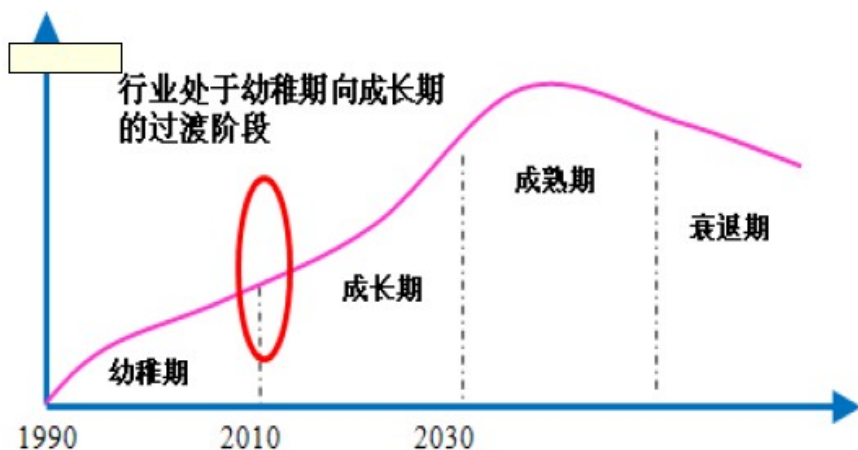
ρ 城市智能交通行业“十二五”期间预计平均增速30%，

ρ 东部沿海城市的智能交通系统已经初具规模，智能交通建设除硬件系统进一步提高覆盖和性能升级外逐步向软件系统升级和信息服务方向过渡。

ρ 中西部地区是城市化进展最快的地区，新城市区的出现带来了对智能交通系统的强烈需求，普遍采取的“一步到位”的建设方式催生了大批千万级整体解决方案项目的出现。

ρ 行业整体处于幼稚期向成长期的过渡阶段行业需求增长明显。

城市智能交通市场规模预测



2. 智能交通与大数据



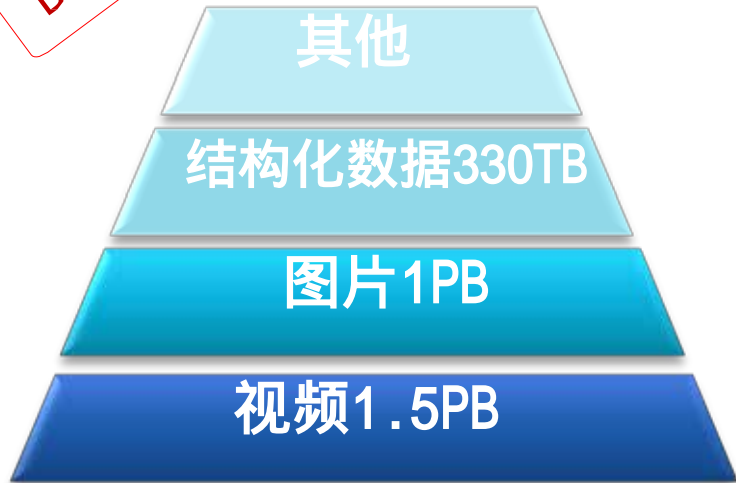
大数据技术能够应用于交通卡口视频和图片数据的实时采集、存储、分析、分类、查询，能够提供比以往更快、更精确地分析、预测交通状况，并对能够影响复杂交通的潜在因素进行新的洞察。

不断增长的数据带来的挑战

挑战

如何存储大体量的数据
并从中获取想要的信息

Big Data



数据存储

- 能够满足横向扩展要求，存储PB级别的数据
- 在大型智能交通应用中提供分布式数据中心
- 满足数据快速容灾要求

系统吞吐

- 图片和视频的并行传输带来了更高的系统I/O要求
- 系统的I/O必须可横向扩展

实时处理

- 实时统计、监报告警
- 互动搜索

模式识别

- 对大量历史数据的分析处理要求
- 需要一个并行处理数据的计算框架

大数据处理的需求和特点

新一代系统的建设需求

扩展性

- 增量式的、几乎无限的扩展

可用性

- 要求系统总是在线运行

灵活性

- 灵活可动态改变的数据模型

传统平台

大数据平台

扩展性

纵向扩展

横向扩展

分布式

资源集中

计算和存储分布

可用性

单份数据

数据复制

一致性

- 数据最终一致

传统数据存储方案分析

关系数据库集群+NAS存储方案

- 将卡口文本、图片数据分别存储：用关系数据库集群（多节点）存储卡口文本数据，并支持对数据进行检索统计；用NAS集中存储（多节点）技术来存储海量图片数据。
- 文本数据存储：使用 Share Nothing 架构的智能存储层，将数据库查询下移分布到存储层的各个节点（具备计算能力），然后再将查询中间结果汇总到数据库节点。
- 图像数据存储：摄像头拍摄的图像经工控机处理后，图片与文本数据上传至中心系统，中心系统的图像处理服务器经过计算处理后，将图像以文件的方式存储到磁盘阵列上；然后图像处理服务器再将请求发送给中心业务应用服务器，由业务应用执行相应的业务逻辑处理，将元数据以及图像的文件位置信息记录到关系数据库中。

系统架构



优劣势分析

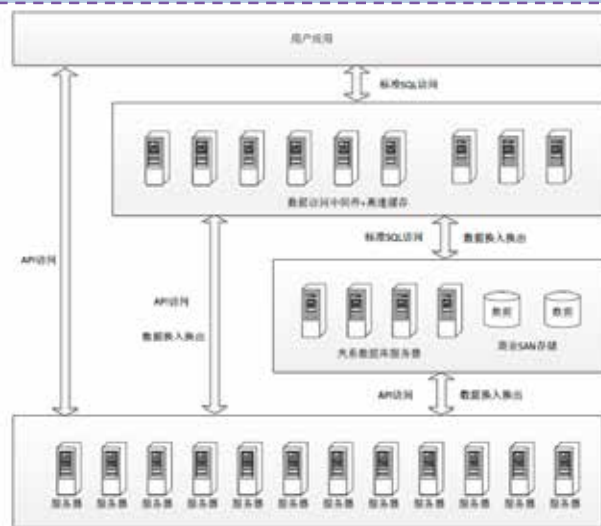
- 优势：1、用基于传统关系型数据库的OLAP和数据挖掘应用相对比较成熟2、上层的兼容性好。
- 劣势：1、集群规模受限（关系型数据库集群的扩展能力有限）；2、需要购买NAS作为图片存储，价格昂贵，NAS存储的扩展在PB已经出现瓶颈。

大数据智能存储分析方案分析

基于Hadoop的海量数据存储方案

- 采用基于Hadoop的海量数据存储方案，同时提供针对不同应用的数据实时分析和查询功能。
- 文本数据存储：文本类数据主要包括车牌、日期、地点、颜色、车型、违法行为分析结果、交通事件识别结果等。采取分布式索引技术，实现海量数据情况下的全文检索在内的简单索引以及支持分布式检索集群检索技术；同时支持关系型索引，实现标准SQL 在内的复杂查询需求和面向应用 的各类优化。
- 图像数据存储：图像类数据主要指 50-500KB 的图片文件数据（非结构化数据），采用基于 Hadoop/HBase/Lucene 的海量数据对象存储技术，采用完全扁平化的对象存储结构，抛弃了传统文件系统的目录结构。

系统架构



优劣势分析

- 优势：1、系统可扩展性强，2、集中I/O和性能可线性扩充，3、数据冗余，4、不停机扩容，5、跨数据中心大表，便于数据的统一管理
- 劣势：技术应用尚待时间的检验

构建智能交通一站式云平台



技术创新点

海量非结构化数据存储技术

视频、图片智能分析分布式计算调度及框架

分布式、高性能数据库（在线实时查询）

图像处理、分析、分类技术

基于语义的分布式视频搜索技术

多维度数据叠加描述技术



创新点

3. 大数据智能交通应用介绍

智能交通云平台

海量数据存储



智能交通系统实时监控城市的交通状态，将各个路口的车辆实时抓拍数据传输到大数据云平台进行长期数据存储

实时数据查询



通过大数据云平台可实时掌控任一车辆的行驶，运行轨迹、分析车辆是否违章

智能数据分析



大数据云平台对海量的交通信息进行比对、分析和预测，为车辆出行提供最优路径规划、交通管理服务、刑侦分析、模式识别

即席查询

- 在1PB的过车和图片数据中查询任意车辆在任意时间点的过车信息，系统在1S内返回查询结果。

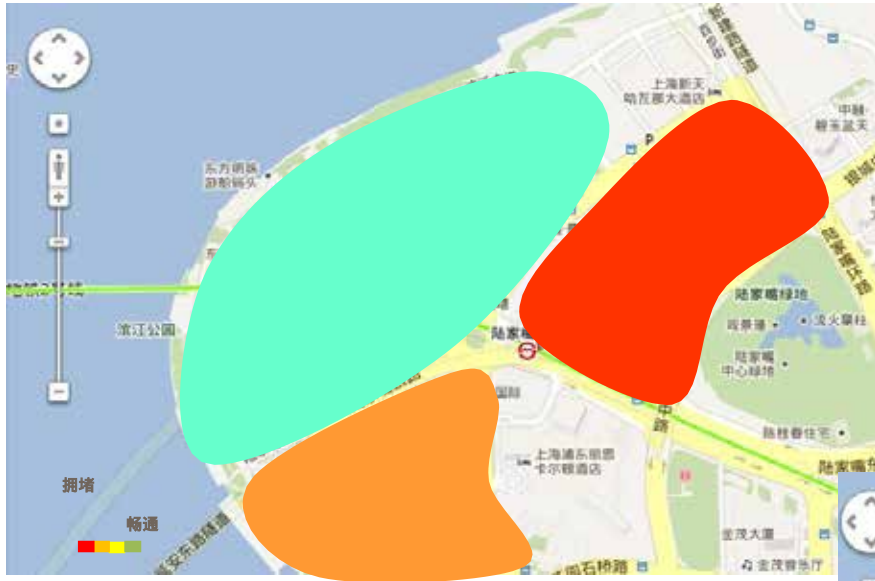
卡口信息
查询

车牌信息
查询

区域信息
查询

The screenshot displays the '逸迅智能交通云' (E-Speed Smart Traffic Cloud) interface. The top navigation bar includes '交通管理' (Traffic Management) and '统计分析' (Statistical Analysis). The search criteria are set to '沪A (上海)' (Shanghai), '车牌' (License Plate) 'D12345', and '时间' (Time) '2013-05-21'. A calendar shows the date '2013年4月' (April 2013) with the 16th highlighted. The search results list 9 entries for license plate '沪A D1234', all recorded at '通过闸口 S1195' (Through Gate S1195) on '2013-02-28 16:48:11'. The details for the first entry show a 'SUV : 阿斯顿·马丁' (Aston Martin SUV) with a '正常状态' (Normal Status). A map view shows the vehicle's location on a highway, with a callout box displaying a photo of the vehicle, its license plate '沪A D123456', and details: '卡口: B1195B1195', '速度: 24', '行驶方向: 自北向南', and '时间: 2013-02-28 16:48:11'. The map also shows a speed limit of 40 km/h.

路况监控



道路拥堵状态展现

- 根据两个卡口之间的平均车速（两卡口即时车速的平均值），计算当前道路的拥堵状态
- 通过红、橙、黄、绿四种颜色标识拥堵状态

区域拥堵状态展现

- 根据卡口过车的实时速率成区域测绘
- 通过红、橙、黄、绿四种颜色标识拥堵状态



车辆远程监控调度

货车调度：

n 为货运企业提供车辆位置、车载状况、运送情况等的信息服务

n 对车辆进行实时调度提高运输效率。



安全监控：

n 使政府监管部门对危险品运输进行安全健康和管理

n 掌握危险品运输车辆的数量、位置、路线，及时发现异常情况并协助快速施救

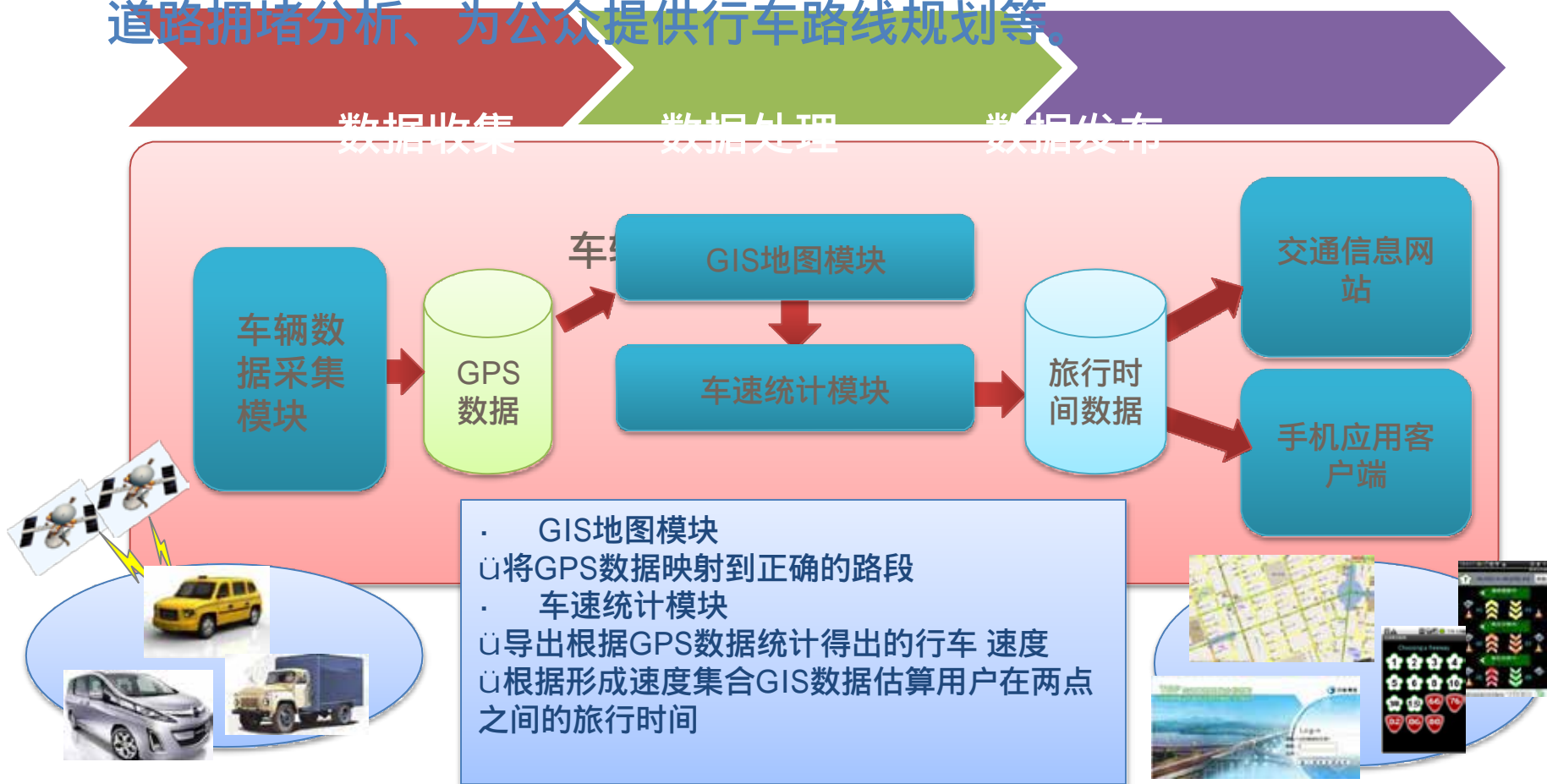
碰撞分析

- 分析要求：分析在某两个特定的时间点在某区域A和区域B都出现的车辆，通过分布式计算框架，实时的进行大量过车数据比对。
- 例如上海最近发生多起盗窃，刑警侦查后确认为同一伙盗贼所为，A区案件一般发生在早上，B区案件一般发生在下午，警察在A、B两区确定以某点为中心的半径2公里的区域为案件多发区，进行碰撞分析，分析结果输出上午在A区域内，下午在B区域内出现的车辆

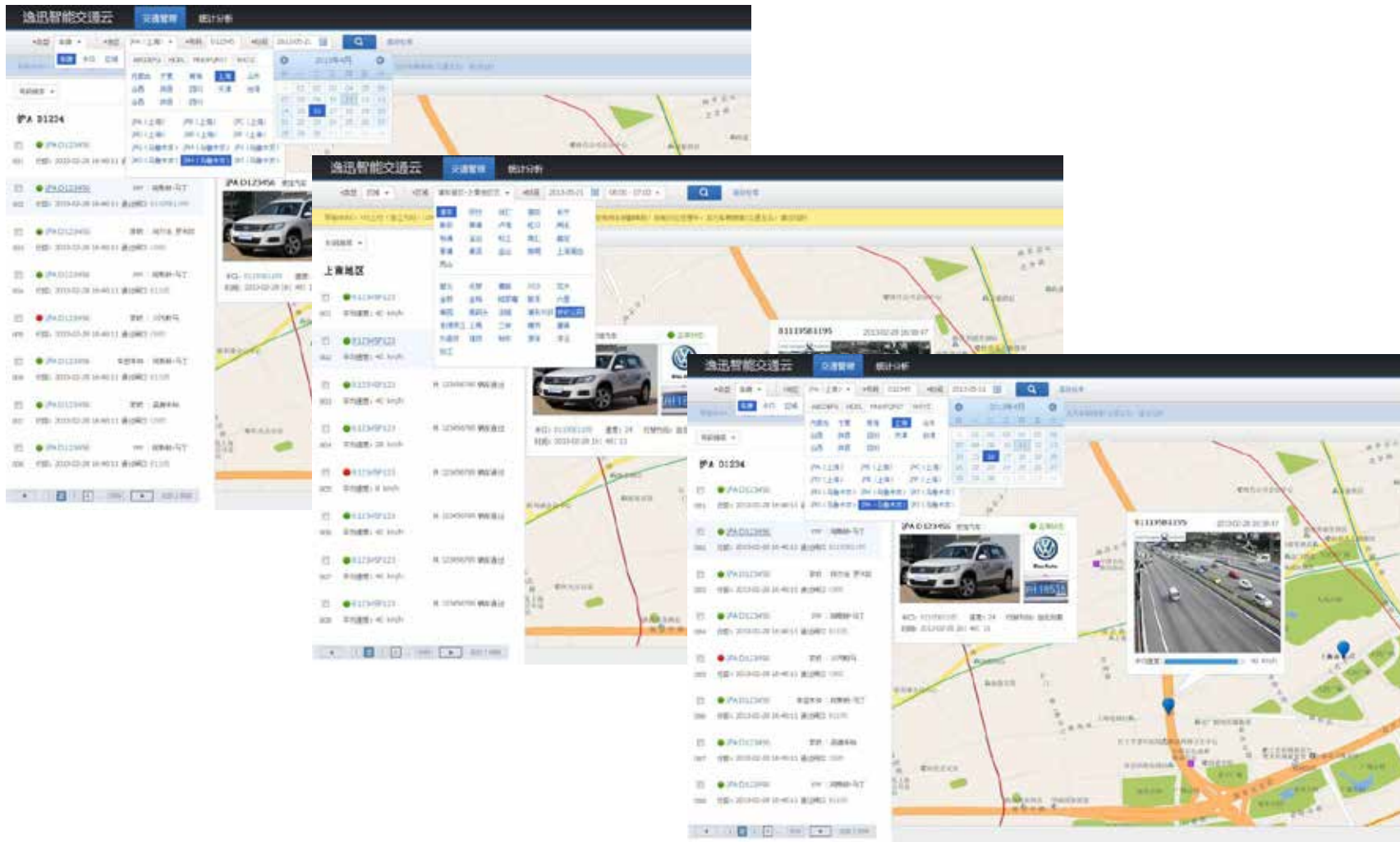


旅行时间分析



基于车辆GPS数据的旅行时间分析系统，能够用于进行城市道路拥堵分析、为公众提供行车路线规划等。



智能交通卡口应用展示



智能交通大数据一体机

预装智能交通云平台软件	型号	8800S
	机框	最多支持14块可插拔的刀片 支持两个可热插拔的万兆以太网交换机 可支持InfiniBand 7U (12.1 “ x 18.5 ” x 29 “) 四个热插拔2500W电源，N 1冗余
Processor Blade 	计算刀片	双路 E5-2650 v2 可升级至4路 128G DDR3 (16X240-pin DDR3 DIMM) 2*2T SATA (2 x 2.5 “ Hot-swap HDD Bays)
Storage Blade 	存储刀片	单路/双路 E5-2600 v2 16 DDR3 (16X240-pin DDR3 DIMM) 6*2T SATA (6x 2.5 “ Hot-swap HDD Bays)

• 高密度

最高14刀

• 高可用

关键部件冗余

• 扩展性

可堆叠

• 灵活性

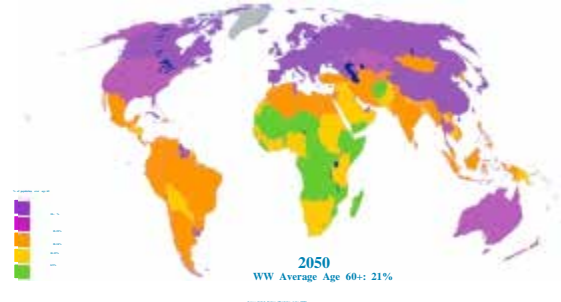
计算和存储刀片按需配置

10.2 让生活更加健康——智慧医疗

趋势分析：我们正处在医疗行业的一个重要转折点



医疗费用在不断上升
GDP的占比非常高



全球老龄化
平均年龄60 +
：目前的10%，到
2050年将达到20%

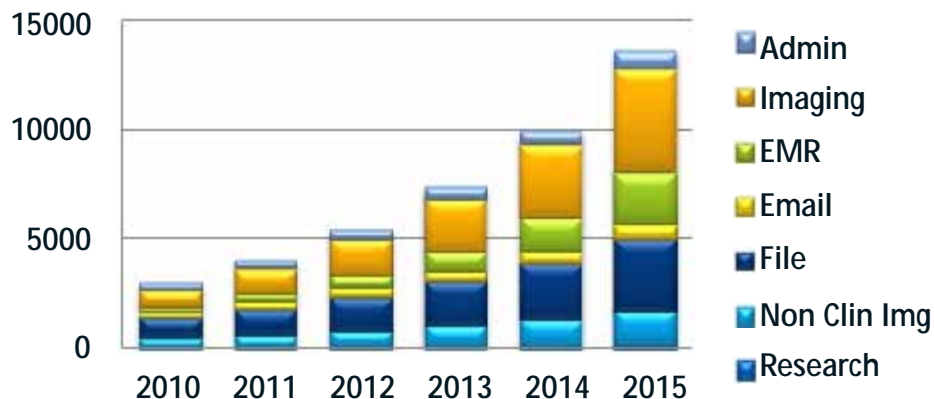


以美国为例：医疗大数据的价值
3千亿美元/年，相当于每年生成总
值增长0.7%

趋势分析:我们正处在医疗行业的一个重要转折点

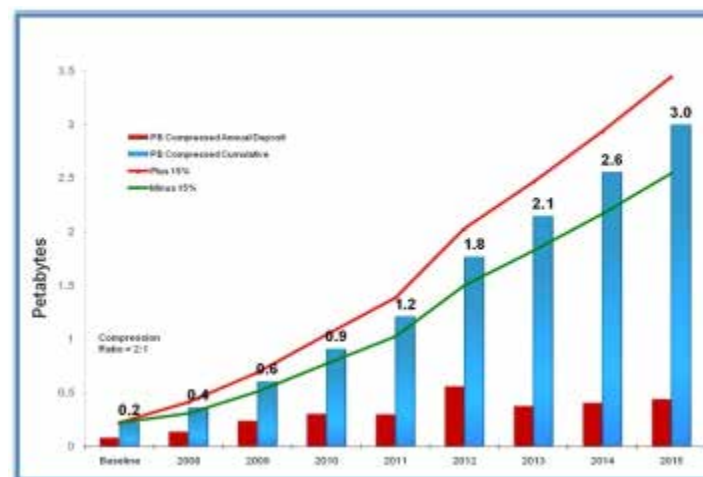
存储的增长

医疗服务产生的数据总量(PB)



医疗影像归档

一个医疗系统案例的数据



到2020年，医疗数据将急剧增长到35 Zetabytes，相当于2009年数据量的44倍增长

Source: McKinsey Global Institute Analysis

ESG Research Report 2011 – North American Health Care Provider Market Size and Forecast

1. 医疗大数据简介

数据来源包括哪些？

1. 制药企业/生命科学

2. 临床决策支持 & 其他临床应用（包括诊断相关的影像信息）

3. 费用报销, 利用率 和 欺诈监管

4. 患者行为/社交网络

我们如何利用大数据创造价值？（示例）

1. 个体化医疗

2. 临床决策支持

3. 欺诈监测得以加强

4. 由生活方式和行为引发的疾病分析

2. 医疗大数据相关解决方案

健康信息服务

基础医疗服务

个人健康管理

老龄社会

新兴的医疗服务
应用

临床决策支持

个体化医疗

肿瘤基因组学

数据分析及
视觉化处理

类SQL的检索

机器学习

医疗影像分析

数据处理/
管理

医疗记录

基因数据

医疗影像

分布式平台

存储优化

安全和隐私

影像数据处理加速

大数据的挑战不仅来自于数据量的增长... 需要新技术的支持

数据量

检验结果, 费用数据, 影像, 设备产生的感应数据, 基因数据等

类型

- 结构化数据, 遵循标准的数据标准(如,HL7)
- 非结构化数据, 如口述、手写、照片、影像等

价值

基于现有数据库中的数据进行分析, 来支持不同类型的业务: 如费用及报销、患者病史、归档影像分析、实时临床决策支持(数据分析)

速度

- 实时数据分析, 而非传统的批量处理分析
- 数据以流的方式进入系统, 进行抽取和分析
- 对于实时运行中的每个时间节点产生影响, 而不是事后处理

在传统的解决方案之上, 引入新的数据及分析模型和技术,
实时有效的商业价值

3. 大数据在中国医疗行业中的应用模式

- 药品研发
- 基因测序
- 分布式计算加快基因测序计算效率

1. 制药企业/ 生命科学

- 2. 临床决策支持 & 其他临床应用 (包括诊断相关的影像信息)

- 临床数据比对
- 临床决策支持

- 公共卫生实时统计分析
- 新农合基金数据分析
- 基本药物临床应用分析

3. 费用报销, 利用率和欺诈监管

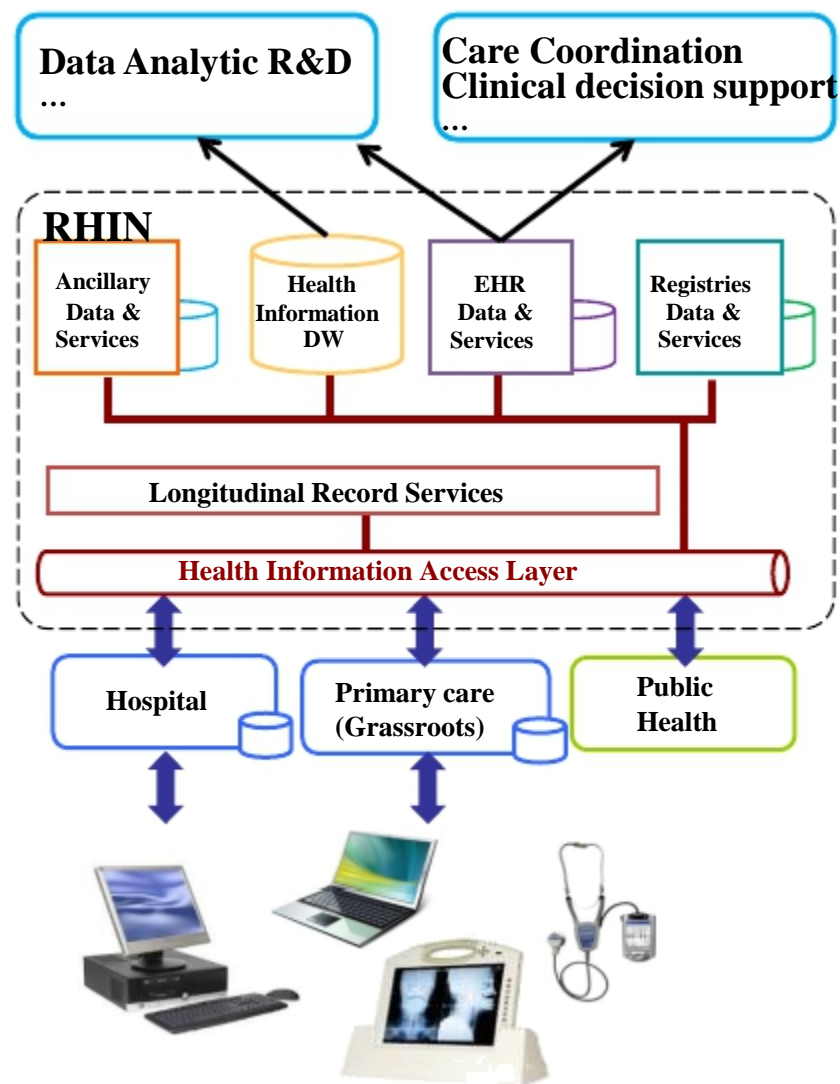
4. 患者行为/ 社交网络

- 远程监控
- 人口统计学分析
- 了解病人就诊行为

案例分享: Regional Health Info Network – China

Real-time Clinical Decision Support

- 实时的医疗数据处理（电子健康档案，医疗影像数据），支持医疗协同、临床决策支持和公共卫生管理
- 采用 Hadoop* (HBase*/Hive*)来实现医疗数据分析和处理
- 未来将扩展到不同领域、不同区域/地区（包括数据交换、处理和分析）
- 与本地的软件厂商及OEM厂商进行了广泛合作
- 技术挑战
 - Hadoop (HBase/Hive)与传统关系型数据库如何有效结合
 - 大数据在区域卫生信息平台中的切实可行应用场景



区域医疗及基层医疗信息系统大数据解决方案(Hadoop*)

