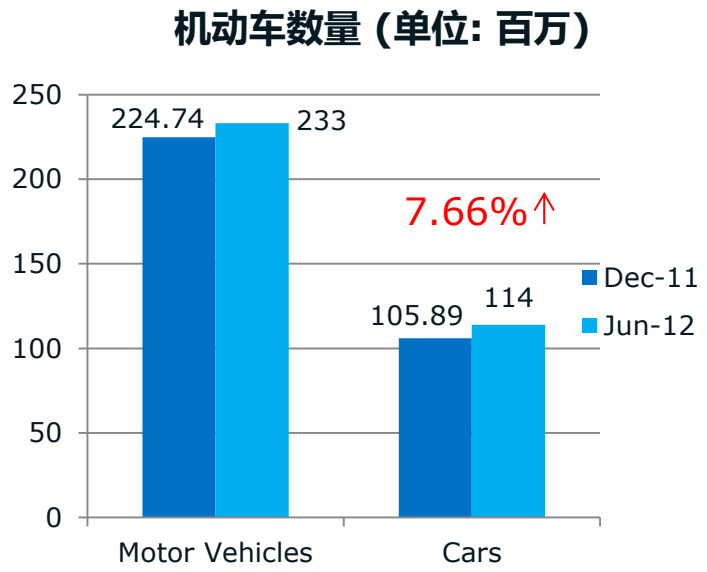
Intel Hadoop 发行版案例

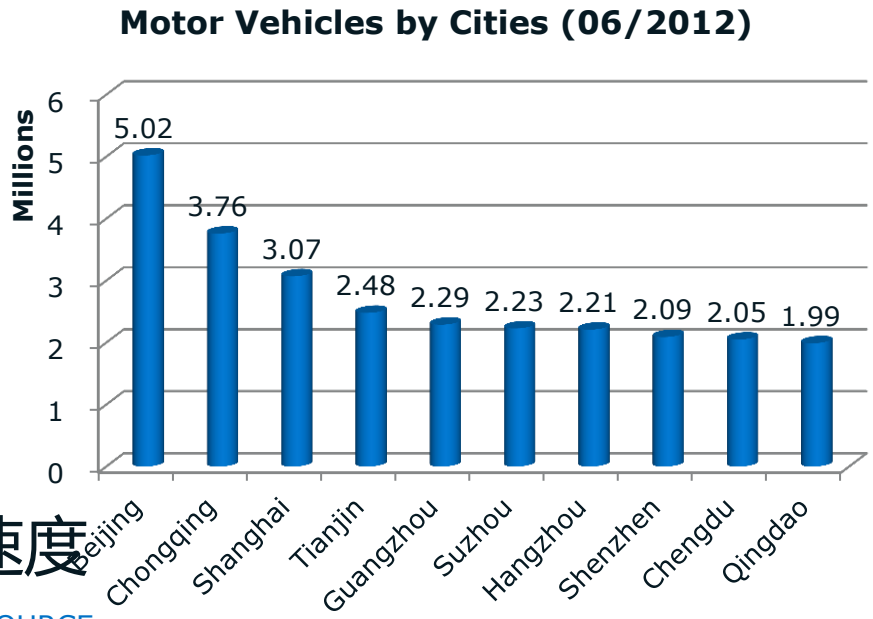
案例分享一
多个地级市智能交通系统大数据

动机

车辆统计和增长率

- 233M 机动车 (年增长率. 3.67%)
- 114M 轿车(年增长率. 7.66%)
- 103M 摩托车





挑战

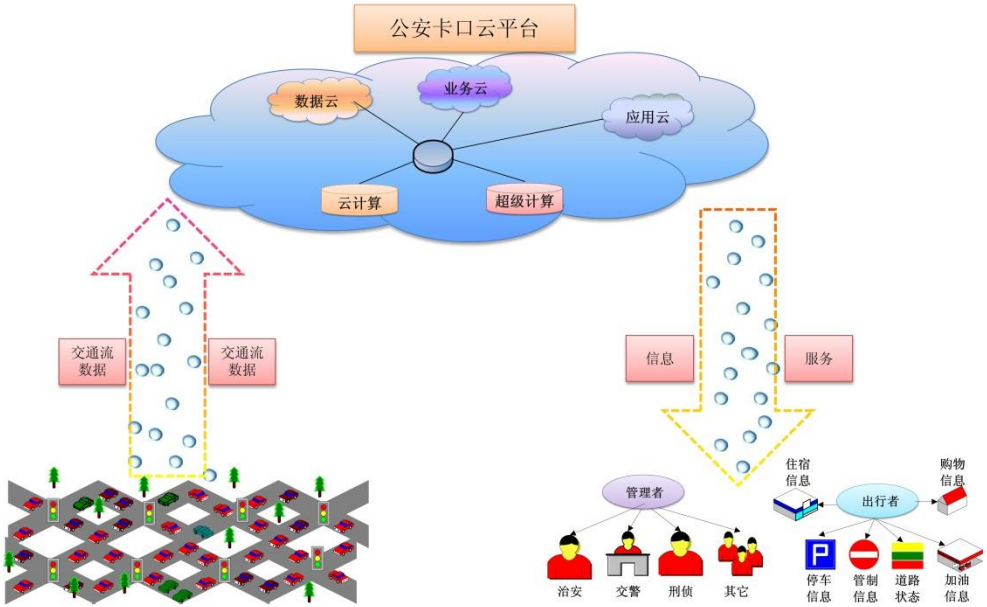
- 交通拥堵已成常态
- 道路建设不能赶上机动车增长速度
  - 每年增长 <5%

SOURCE:
http://www.mps.gov.cn/n16/n1252/n1837/n2557/3327565.html

- 现行管理方案取得效果，但不足够

智能交通系统目标

- 交通管理
  - 强制交通规则 (例如, 限速)
  - 运输计划支持
  - 按需交通控制
  - 交通情况研究
- 旅客信息系统
  - 实时路况
    - 畅通&堵塞
    - 历史照/摄相影像 & 统计
  - 出行时间信息
    - 不同出行方式
    - 前瞻性的出行计划
- 商用车辆信息
  - 商用车辆管理，跟踪，调度
- 公共安全



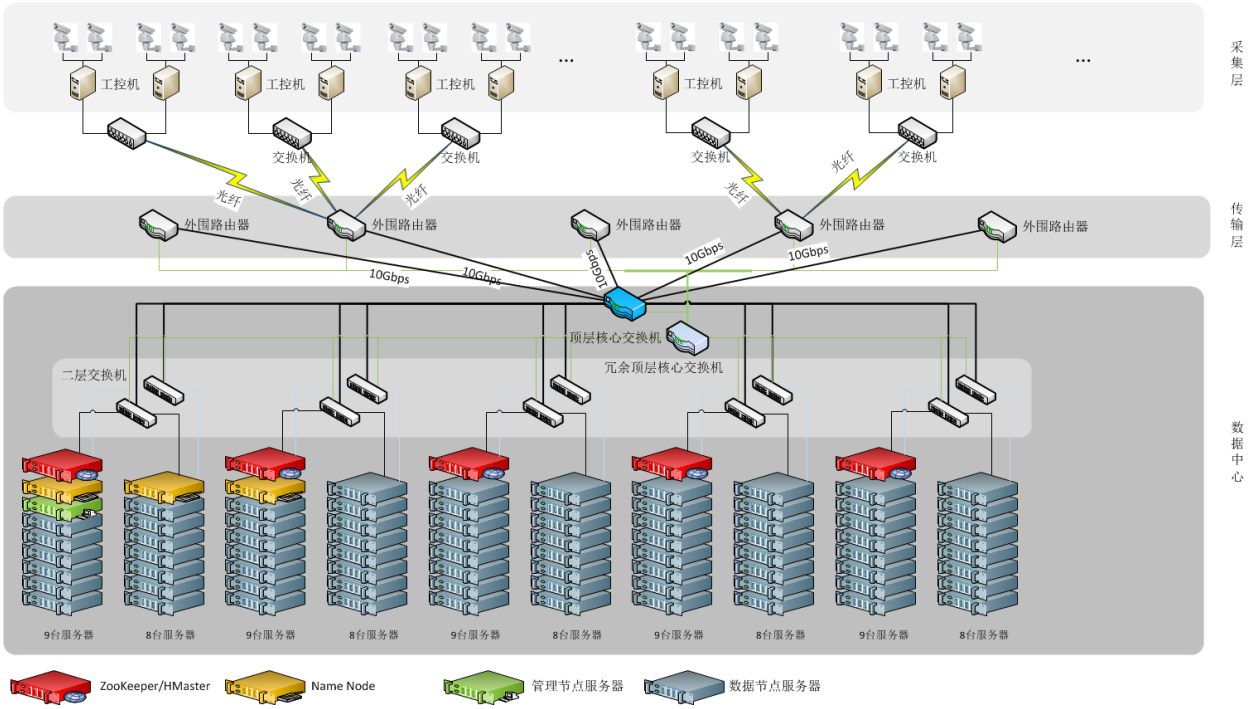
三个主要的途径

- 交通摄像
  - 通常安装在高速和街道
  - 2 百万像素网络摄像头 – 已部署
  - 5 百万像素网络摄像头 – 新安装
  - 每个摄像头1或2个线路
  - 每城市中200~1000 摄像头

- 监控摄像
  - 在街道上安装, 在建筑物周围安装, etc

- **GPS 终端**
  - 在商用车辆上安装
  - 新兴的带有摄像头，GPS和3G网络的平板电脑

智能交通的一个数据中心



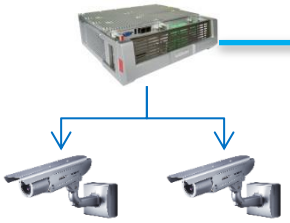
实时路况影像

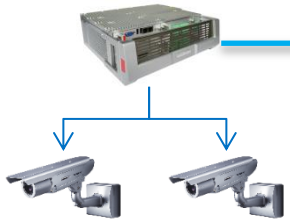
车辆跟踪

实时拥堵状态

基于交通流量的信号灯控制

智能交通的软件架构





数据挖掘 (例如：车辆跟踪)

Hive

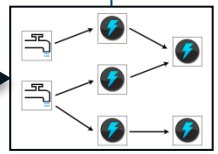
MapReduce

即时查询 (例如：路况信息)

应用程序

HBase

SQL



视频流处理 (例如：实时路况)

当前智能交通的功能

交通管理

- ✓ 向控制中心和监控系统实时报告路况
- ✓ 在定长路段通过平均车速计算超速车辆
- ✓ 检测伪造车牌车辆
- ✓ 通过分析出发地-目的地数据为道路建设提供参考

公共安全

- ✓ 实时跟踪车辆
- ✓ 秒级超速违章检测和模糊查找
- ✓ 黑名单警告和报警，或改变交通模式.
- ✓ 检测某些地点相同车辆反常的高发事件

旅客指引

- ✓ 为驾驶员获取最新的实时路况影像和交通流量状态
- ✓ 在城市中为每一路段进行时间估计

案例分享二
某运营商省公司清账单系统

新详单系统建设的必要性

原清帐单系统建立在小型机及其高端存储设备上。为了实现海量数据存储及快速导入，原系统把明细清单压缩存放到文件系统中，数据库只保留索引信息以满足查询性能的要求。随着时间推移，数据量增长，需找新的解决方案越来越迫切：

- 通过文件存储定长记录的方式，程序难以修改。原有清单中心基于266字节的定长格式，但新融合计费项目上线，清单格式增长至1024字节。
- 文件系统缺乏常规查询语言，如SQL，HIVE等，旧已经不能满足越多越多统计需求。
- 系统需要不断增加新字段，文件系统无法扩展。
- 文件系统不支持数据库常规的更新功能，详单冲销、修正、补信息等功能难以实现。
- 随着新详单格式改变，存储空间及性能相应需要增加5倍。扩容费用高昂。

新清帐单系统关键需求

一、必须能够高效处理海量数据

- ✓ 单月清单数据量约1000亿条×1k/条=100TB，6个月总量高达600TB(6 + 1) ~ 700 T
- ✓ 从600TB清单数据中检索某用户某个月的清单记录，响应时间应小于1秒
- ✓ 支持高峰期每秒2000个并发访问查询
- ✓ 满足现在清帐单业务的查询统计需求(23类)
- ✓ 实时入库，清单文件无积压。（清单文件最大2万条，最小1条记录。实时生产，平均每秒2个20MB的清单文件，高峰期到每秒10个20MB文件）
- ✓ 对联机分析必须提供标准编程接口，支持SQL/JDBC/ODBC等

二、高可扩展和高可用

- ✓ 用户程序查询数据不需要知道底层细节，比如数据分布细节
- ✓ 可以水平扩展
- ✓ 允许多台机器故障的场景下，业务不中断

新清账单系统基于Hbase的测试结果

以下数据是实验情况下对Intel的Hbase基于一个月详单数据的测试结果，随着集群规模扩大，性能还能线性的提高

负载情况	并发线程（并发用户）	性能指标1: 查询个数(用户数)/秒	性能指标2: 清单条数/秒	平均查询延时
高负载	7500	600个查询/秒	400000条清单/秒	0.9秒
中等负载	2500	450个查询/秒	300000条清单/秒	0.5秒
低负载	1000	200个查询/秒	130000条清单/秒	0.3秒

总体性能及成本综合评估:

Hbase在5台查询清单PC的基础上，就达到了最大查询速度476个查询/秒，数据量达到285000条清单/秒，入库速度2.83万条/秒，平均每条记录1KB

在实现同样需求的情况下，相比小型机和高端存储方案，新清账单中心全部选用了**pc server**等设备，我们预计，**成本约降为原来的1 /4，性能提升大于3倍。**

新详单系统实现架构——逻辑架构

平衡各种优缺点后，及对进行大量的性能、功能及稳定性测试后我们最终选择基于Hadoop Hbase作为新详单系统实施方案。下图为新详单系统项目的逻辑架构：

新详单系统项目逻辑架构图

新详单项目实施情况

本项目底层通过PC服务器组构建出一个基于集群，采用INTEL提供的Hadoop产品（分布式文件系统+分布式数据库），上层由合作伙伴开发业务程序，对入库和查询进行业务处理。

- 这种架构有效的屏蔽了底层的功能，对上层来说，只需要调研相关接口即可。数据的分发、复制、任务调度、容错都是由系统软件来控制。大规模的PC具备强大的处理能力和网络带宽，同时具备线性的横向扩展能力。3份冗余的数据保证对硬件的容错和读处理的支持。
- 存储使用72台PC机身硬盘作分布式存储DataNode，每台PC配置6TB磁盘容量，按每份数据存放3份计算，有效容量144TB，保存6+1个月数据，压缩比1:5 。

已经完成的新详单系统，实现功能如下：

- ✓ 实现个人账详单数据存储和实时查询展示
- ✓ 实现集团清单数据存储和批量导出
- ✓ 实现补卡预处理实时查询
- ✓ 实现网站/彩信账单文件、邮寄账单文件、短信账单文件批量导出

新详单项目实施情况（续）

目前新详单系统很好的达到了我们的预期目标，下表是新详单系统上线后的关键性能统计：

序号	名称	效率	说明
1	入库效率	>200 MB/s 或20 万条每秒 ，资源消耗20%	实时入库，清单文件无积压。（清单文件最大2万条，最小1条记录。实时生产，平均每秒2个20MB的清单文件，高峰期到每秒10个20MB文件）
2	并发访问	>2000 笔/秒	支持每秒2000个查询。（旧系统做了每人每天6次查询限制，旧系统高峰期查询达200/秒，预计放开限制后，查询高峰能达2000/秒）
3	平均响应时间	<1秒	每个查询小于1秒

旧详单系统与新详单系统实施效果对比

*2 * P595小型机（48CPU）及DS8300高端存储*

VS

97台 X3650 PC服务器集群

文件系统+关系型数据库

- ❖ 常有文件积压，不能实时入库，系统负荷过重
- ❖ 每秒小于200个并发查询
- ❖ 自定义文件系统，只支持266字，扩展需要重新开发
- ❖ 只有54TB可用空间
- ❖ 灾难恢复需要通过磁带，业务中断时间过长

Before

Now

企业级Hadoop集群

- ✓ 每15分钟加载，不存在积压，平均20%资源消耗
- ✓ 每秒大于2000个并发查询
- ✓ 支持稀疏表，轻松扩展任何字段
- ✓ 138TB可用空间，并提供三份数据冗余
- ✓ 多台服务器同时故障，也不中断业务

案例分享三

某金融客户数据票据详单平台

项目简介：
某金融客户大数据平台是某金融客户处理票据详单，电子保单，某金融客户交易详单平台。
票据数据主要包括消费者在POS机上的消费签购单详和法人单位开具给个人的付费凭证。 票据数据每个月有5000万到6000万条数据。该数据的使用者除了某金融客户的内部用户外，也开放给参与交易的商户。
某金融客户大数据平台也包括电子保险单存储。 每份保单文件为约3MB 大小， 一个城市每年约产出1500万份保单。
某金融客户大数据平台的第三项功能是某金融客户交易详单， 包括某金融客户网存款，取款，转账详单。每月产生40TB到47TB数据。电子保险单存储和某金融客户交易详单平台的用户主要是某金融客户的内部用户。

面临的挑战：采用IDH系统方案之前，某金融客户的方案有困难

1）在数据集达到半年数据规模时，采用XML方案的系统票据查询速度过慢导致无法实际使用该功能；

2）电子保单存储和查询中， 当存储的数据量达到一年的数据量时， 查询速度过慢导致无法实际使用该功能。

3）曾在交易详单项目上尝试采用开源的HBase, 经常发生故障，而且没有管理监控报警功能， 无法投入商业应用。

IDH方案的实现： 新的某金融客户大数据平台项目采用英特尔Hadoop发行版平台，实现了以下功能

1） 将基于XML的票据查询方案转换为基于HBase的方案，外部导入程序将XML文件转换为HBase数据结构，实现了在海量票据数据中一秒之内的查询返回；

2) 电子保单数据和保单文件从关系数据库转换为HBase，使用IDH大对象存储的优化，性能比开源Hadoop成倍数提升；

3) 将开源HBase的交易详单应用部署到英特尔Hadoop发行版，提供更高的性能、稳定性和管理性。

硬件配置：集群规模45个节点， name node采用了ＨＡ， 每个节点配１0块２ＴＢＳＡＴＡ 数据硬盘组成ＪＢＯＤ， 系统盘采用２块５００ＧＢ ＳＡＳ硬盘组成ＲＡＩＤ １， ６４ＧＢ内存，双路６核ＣＰＵ，千兆网络。

效果： 在设计数据量下，票据查询和电子保单查询的速度都在秒级，达到设计标准。

案例分享四

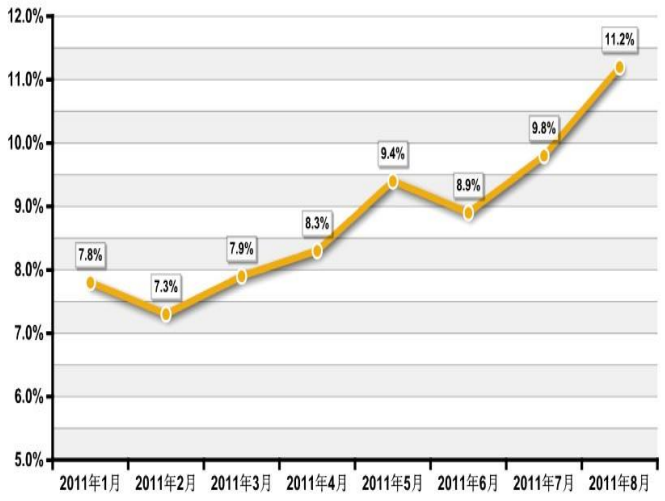
某通信运营商全国用户上网记录

业务背景

随着移动互联网业务的发展，上网记录查询成为用户投诉的焦点

问题来源

- 目前，某省分公司**3G客户数据流量问题争议占3G业务投诉达7-10%**，且近几个月呈上升趋势，个别省分比例高达20%
- 一些用户对3G业务流量产生及计费方式不了解，主观认为自己未使用或使用较少数据流量，**要求运营商提供上网记录**，而现有系统不具备此功能，从而导致投诉升级。



3G流量费争议占总咨询投诉量比率

有限公司北京分公司客户详单
移动业务手机上网流量费详单

客户名称：		业务号码：	
起止日期：	2011年10月1日~2011年10月31日	查询日期：	2011年11月21日
总流量合计：	171289KB	费用合计：	0.000元

每页显示数量: 20 50 100 全部 到 页 GO 第1页 共10页 首页 上页 下页 尾页

序号	网络类型	业务类型	使用起始时间	总流量KB	通信地点	费用
1	3G	计费流量	2011-10-01 07:58:25	41		0.000
2	3G	计费流量	2011-10-01 11:38:02	1308		0.000
3	2G	计费流量	2011-10-01 11:41:28	21		0.000
4	3G	计费流量	2011-10-01 11:41:54	130		0.000
5	3G	计费流量	2011-10-01 11:47:24	2810		0.000
6	3G	计费流量	2011-10-01 12:23:42	1		0.000
7	2G	计费流量	2011-10-01 12:29:14	2		0.000
8	3G	计费流量	2011-10-01 12:30:27	15		0.000
9	2G	计费流量	2011-10-01 13:06:12	17		0.000
10	3G	计费流量	2011-10-01 13:11:41	24		0.000
11	2G	计费流量	2011-10-01 13:27:05	2		0.000
12	3G	计费流量	2011-10-01 13:31:14	1044		0.000

难点分析

上网记录是海量数据

用户每月的上网记录约几万至数十万

- 在Gn（SGSN与GGSN之间）接口上部署采集设备来生成用户上网记录
- 用户手机访问一次网页，约会产生数十条，甚至数百条请求，意味着产生数十条和数百条上网记录
  - 访问手机新浪网首页，约产生20条记录
  - 访问新浪iPad首页，约产生40条记录
  - 在iPad中看一条新浪新闻，产生超过180条记录
  - 访问淘宝触摸屏版，约产生60条记录
  - 大量的DNS查询、推送服务记录（如苹果通知服务）等
- **全国每日新增约10TB数据，每月近万亿条记录，存放6个月，约2PB。**

移动互联网处于快速发展期：每6个月，流量翻一番

- 移动互联网用户快速增加，智能终端迅速普及、户均流量显著增长，上网记录数据将进一步猛增

移动用户上网记录集中查询与分析支撑系统

全国集中的一级架构，国内电信行业首次将Hadoop/HBase引入到商用电信服务系统建设中

关键性能指标

数据存储

- 上网记录入库时间：一般小于30分钟，实际约10分钟
- 具备存储全国移动用户不小于6个月的原始上网记录能力
  - 历史5个月+当前月
- 统计分析的中间报表数据保存不小于5年

数据查询

- 上网记录查询速度：不高于1秒（不含用户访问查询页面的时间）
- 支持并发查询数目：1000请求/秒

解决方案

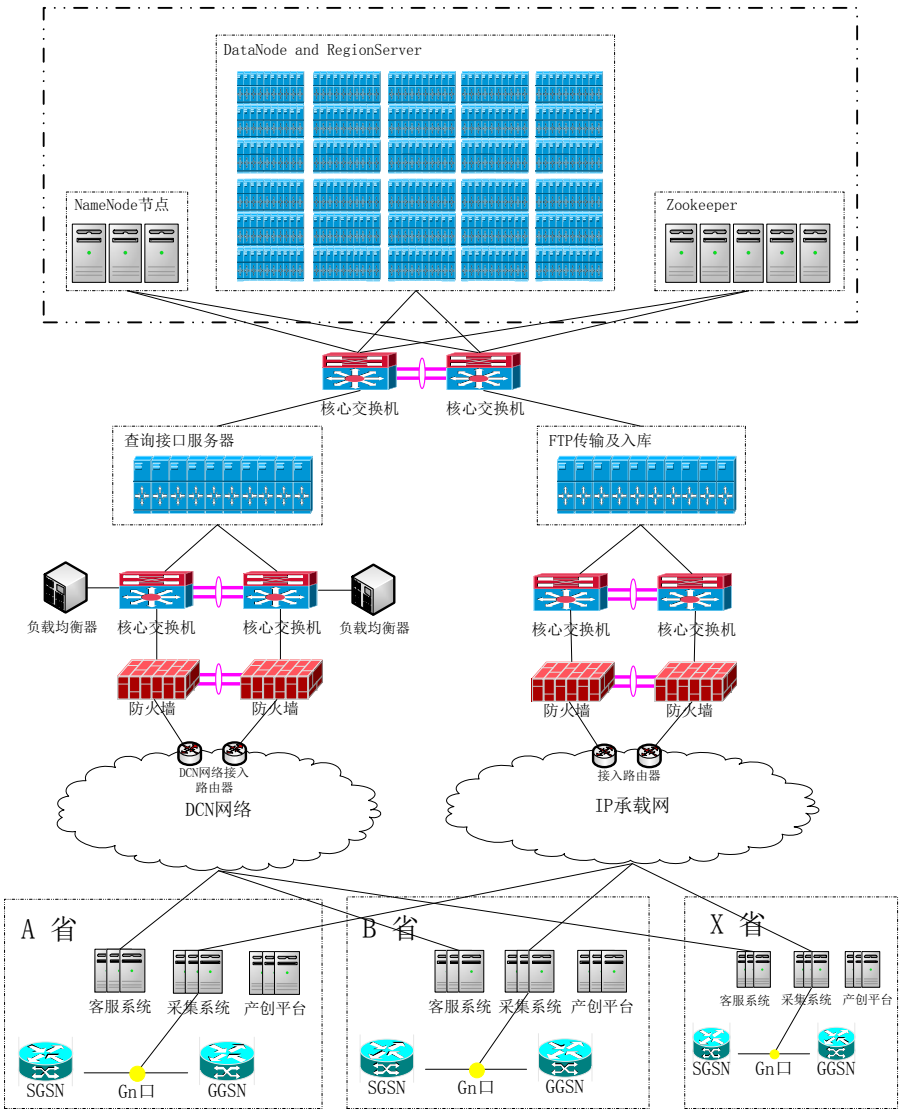
系统构成

- 系统采用全国集中的一级架构方案进行建设，主要包含数据采集子系统、数据入库子系统、数据存储子系统、数据查询与分析子系统

基本技术

- 采用Hadoop/HBase作为上网记录存储方案
- 采用MapReduce/Hive作用统计分析和数据挖掘工具

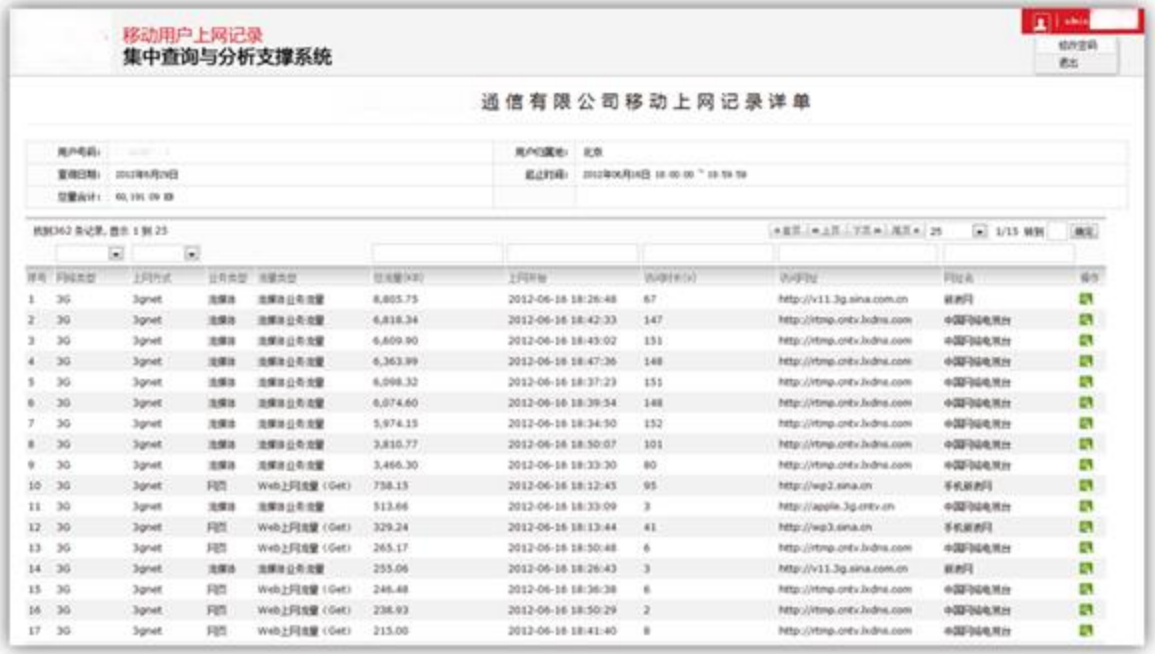
系统部署



系统部署

- 两路x86服务器
  - NameNode节点：3台
  - DataNode（数据存储节点）：~200台
  - Zookeeper节点：7台
  - 集群监控节点：1台
  - 入库服务节点：24台
  - Web查询应用服务节点：20台
- 网络交换设备
  - 机框间通过万兆交换机连接，以完成快速的数据交换
- 英特尔Hadoop发行版
  - 满足高性能的数据导入和快速查询。
  - 稳定、易于部署和管理的企业级方案。

移动用户上网记录集中查询与分析支撑系统



案例分享五

某运营商网络优化

业务背景和需求

业务背景

电信网络优化是指找出手机信号的盲区，调整手机基站，减少信号干扰，简单来说就是测试信号，调整一些无线网络的参数，使无线网络覆盖得更好，信号更稳定。

需求

能够高效处理海量数据

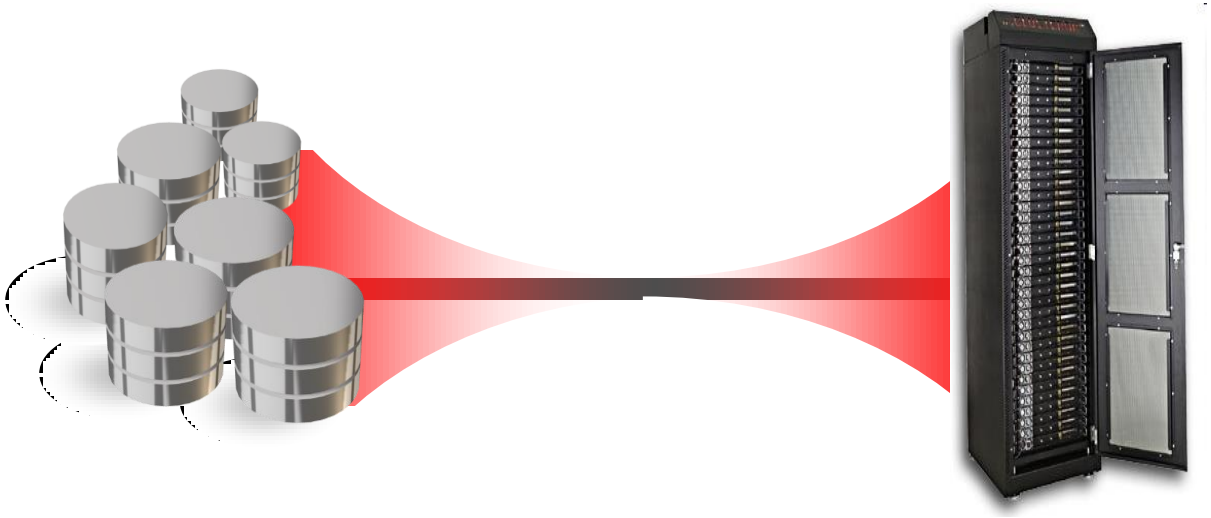
- 网优系统会处理每天PCMD语音通话记录约500GB, A10/A11数据通讯记录约2TB, ，6个月总量高语音记录约为90TB, 互联网使用记录为360TB
- 支持12个月的时间跨度的访问查询
- 实时入库

需求类型 - 从使用者角度的应用场景进行分类，可以基本概括为以下几种类型

- 实时数据处理 -- 针对需要快速给出数据结果处理的需求
- 周期统计数据处理 -- 针对定期的数据事务处理的需求
- 复杂模型数据处理 -- 针对的是临时的一些复杂计算的数据处理需求
- 平滑系统迁移 --针对已有的业务系统进行迁移的需求

传统分析应用数据处理架构面临巨大挑战

传统主机+存储的数据库架构的IO瓶颈问题



传统的解决方案方案

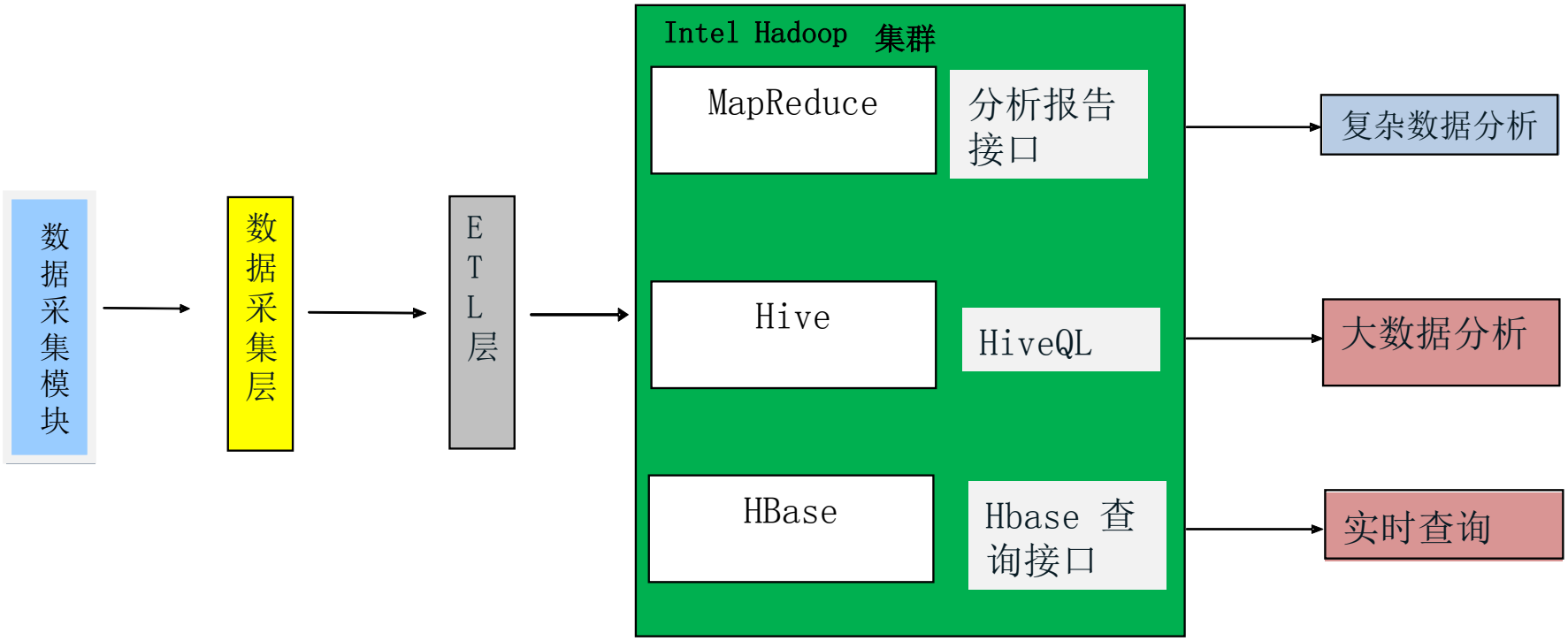
- 利用传统的小型机+磁盘阵列的方案如果要实现上百TB的数据存储，整个系统的投资将非常昂贵
- 大数据量访问存在着明显的I/O瓶颈
- 传统数据仓库Oracle数据库提供单一的访问接口，数据读写速度慢，不能满足海量网优数据的业务处理需求
- 数据难以维护和管理

Hadoop方案优势

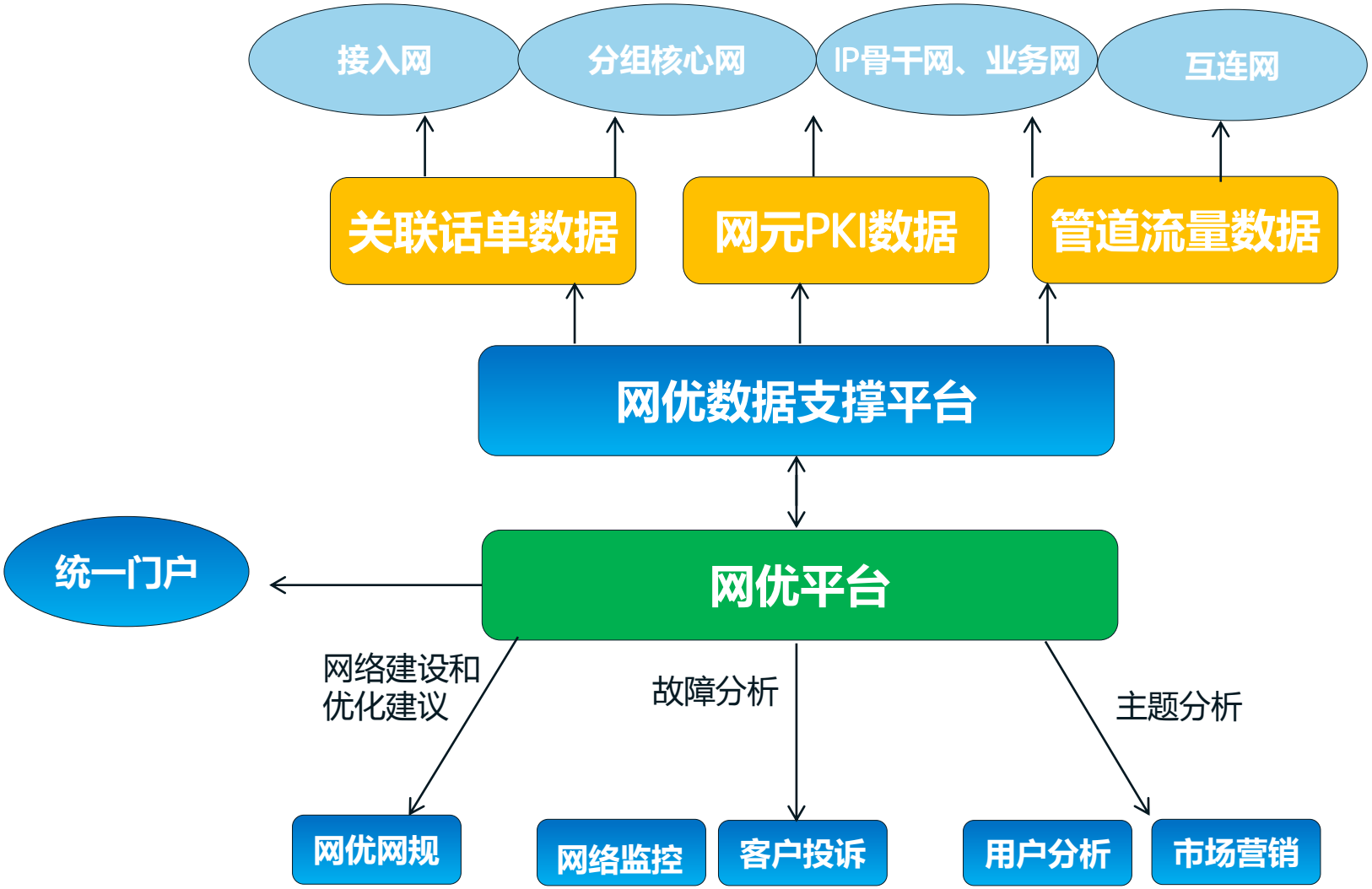
- Hadoop云平台利用相对廉价的服务器提供的HDFS分散存储技术，可以低成本方式实现PB级别的海量数据存储，而且利用集群的水平扩展性，基本消除I/O瓶颈。
- 同时Hadoop云平台还提供了基于MapReduce并行计算技术的的HIVE数据仓库和并行ETL处理架构
- 可以并行将海量数据导入云数据库，节省了数据加载的时间，极大的提高了海量数据的处理效率
- 云平台还可以做列存储实时查询和分析，分析效率高，更有针对性

网优数据支撑平台

网优系统架构采用了基于hadoop的分布式文件系统HDFS,数据存储则采用了分布式数据库hbase,同时结合云计算的其他组件构成.如下图所示



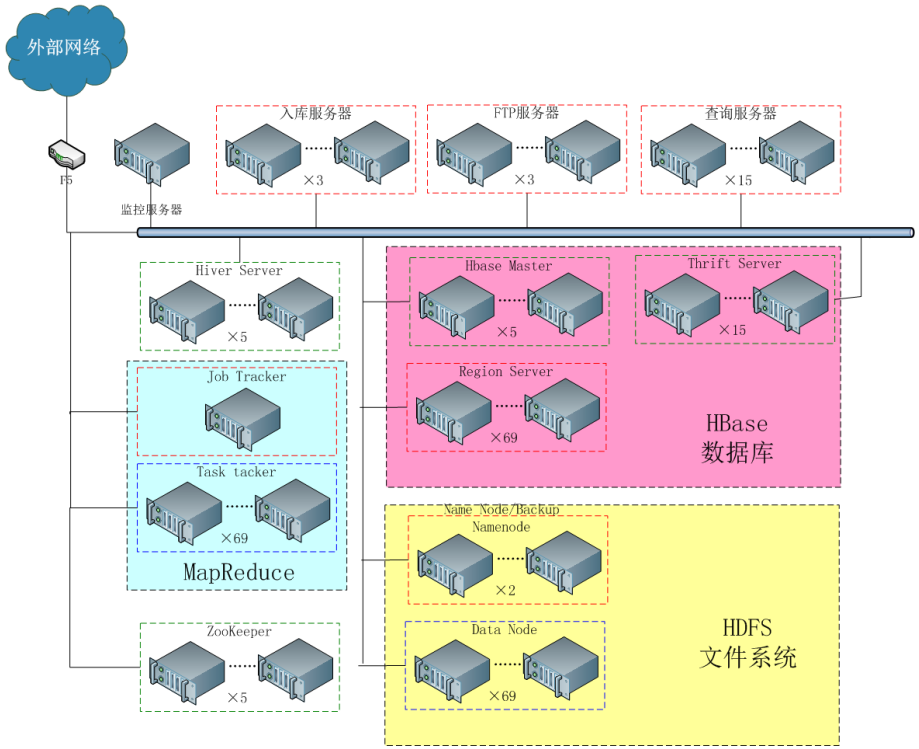
平台集成架构


接入网
分组核心网
IP骨干网、业务网
互连网
关联话单数据
网元PKI数据
管道流量数据
网优数据支撑平台
统一门户
网优平台
网络建设和
优化建议
故障分析
主题分析
网优网规
网络监控
客户投诉
用户分析
市场营销

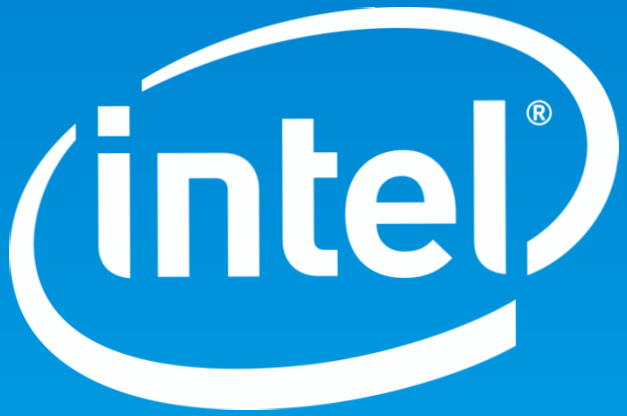
网优大数据平台的部署方案

- 该架构有效的屏蔽了底层的功能，对上层来说，只需要调研相关接口即可。数据的分发、复制、任务调度、容错都是由系统软件来控制。大规模的PC具备强大的处理能力和网络带宽，同时具备线性的横向扩展能力。3份冗余的数据保证对硬件的容错和读处理的支持。

- 存储使用志强服务器本地硬盘作分布式存储DataNode，每台服务器配置12颗1TB磁盘容量，按每份数据存放3份计算。



设备	硬件设备
Hadoop 集群管理节点	Intel X56系列处理器，48GB内存，12*1TB SATA硬盘
Hadoop集群NameNode/JobTracker	Intel X56系列处理器，48GB内存，12*1TB SATA硬盘
NameNode/JobTracker HA备份节点	Intel X56系列处理器，48GB内存，12*1TB SATA硬盘
Secondary NameNode	Intel X56系列处理器，48GB内存，12*1TB SATA硬盘
HBase 集群Master和Zookeeper节点	Intel X56系列处理器，48GB内存，12*1TB SATA硬盘
DataNode/TaskTracker/Region Server	Intel X56系列处理器，48GB内存，12*1TB SATA硬盘
HBase Thrift服务器节点/查询服务器	Intel X56系列处理器，48GB内存，12*1TB SATA硬盘
数据采集， ETL服务器	Intel X56系列处理器，48GB内存，12*1TB SATA硬盘


intel