



# 皮皮访谈录:

# 大数据时代 名家高端访谈

皮丽华 编著



汇聚百度、阿里、腾讯、京东、滴滴等中国互联网名企及微软、  
Dell等国际软硬件巨头的高端技术访谈

首度揭晓IT金字塔尖高人气、强实力牛人的职场故事

你值得拥有!



皮皮访谈录:

# 大数据时代 名家高端访谈

皮丽华 编著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

## 内容简介

这是一个数据爆发的时代，尤其是在互联网+的风口上，大数据与云计算红得发紫，IT架构也随之演变，各种前瞻技术如雨后春笋般铺天盖地。大数据会给这个世界带来哪些影响？互联网+的时代下，如何重塑IT架构？如果你渴望答案，那么我相信你可以从这本书里找到启发。

风风火火的大数据与云计算到底是什么关系呢？它们就像一枚硬币的两面，彼此相互依存，大数据离不开云计算的分布式处理、分布式数据库和云存储等技术。本书以对话的方式，邀请IT业界大咖坐镇访谈，以风趣幽默、通俗易懂的语言、形象具体的案例、励志感人的故事娓娓道来，为读者讲述了海量数据分析与处理、云计算架构与优化、数据库性能优化与实践、大型分布式软件架构、IDC与数据中心运维、职业生涯与创业六大主要内容。相信读完此书，你会受益匪浅，从一名大数据世界的菜鸟华丽蜕变成行家。

在皮皮访谈录中，我邀请到了来自百度、阿里、腾讯、京东、网易、奇虎、去哪儿、携程，58同城、滴滴、易车等互联网企业，IBM、微软、戴尔、SAP、华为、Hortonworks等软硬件巨头，以及听云、百分点、星环科技、永洪科技等云计算与大数据厂商的大咖坐镇，通过与嘉宾互动采访的形式来讲述IT精英的职场经历，希望读者能直观感受到不同企业的文化，能够体验诸如微信等优质产品以及电商等高可用架构背后的秘诀，洞察大数据、云计算、分布式架构、IDC运维等前沿热点话题的精髓，深入了解与借鉴CTO、技术总监、架构师、DBA等技术大咖的成功经验和人生感悟。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。  
版权所有，侵权必究。

### 图书在版编目(CIP)数据

皮皮访谈录：大数据时代名家高端访谈 / 皮丽华编著. —北京：电子工业出版社，2016.1  
ISBN 978-7-121-27474-9

I. ①皮… II. ①皮… III. ①数据处理 IV. ①TP374

中国版本图书馆CIP数据核字(2015)第253566号

策划编辑：张月萍

责任编辑：徐津平

特约编辑：顾慧芳

印刷：北京京师印务有限公司

装订：北京京师印务有限公司

出版发行：电子工业出版社

北京市海淀区万寿路173信箱 邮编：100036

开本：787×980 1/16 印张：21.25 字数：408千字

版次：2016年1月第1版

印次：2016年1月第1次印刷

定 价：69.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888。

质量投诉请发邮件至zltts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

服务热线：(010) 88258888。

这本书独辟蹊径，实用性很强，主持人与嘉宾的对话互动真实地展现了很多 IT 精英的职场历程，他们的成功有运气的成分，但更多的是靠后天的努力。一路走来实属不易。皮皮访谈录如同一面镜子，让我们有幸目睹了 IT 行业里的真人真事，也让我们对自己看得更清楚，期待这种正能量能够鼓励所有的 IT 爱好者为梦想而努力。

——金官丁，热璞科技 CTO

赢家从不满足于现状，他们会在自己的专业领域中，不断地努力做好每件事，让事情从平凡上升到卓越的境界。他们的成功历程值得我们借鉴，更值得我们传播。《皮皮访谈录》汇聚了互联网名企的架构师、DBA 和运维经理等 IT 成功人士，深度解读了大数据、云计算、分布式架构等 IT 前沿技术。他们为何能站在 IT 金字塔塔尖？他们背后有哪些励志的职场故事？为何双 11 不宕机？QQ 和微信架构有何不同？带着一串串的好奇心，相信这本书可以帮你找到答案。本书紧扣最前沿的 IT 技术，再现了一幕幕真实的实战经历。本书富有内涵、兼智慧与趣味为一体，值得 IT 人士细细品读。

——皮光明，前家度联合创始人

《皮皮访谈录》以行云流水的文风，站在互联网+的风口上，与资深架构师、DBA、IT 高管进行深度对话，庖解大型电商的网站架构、海量数据存储以及云计算等前瞻技术，带领大家领略不一样的 IT 人生。

——孙灵飞，1 号店技术部高级总监

大云时代，很多人都在盲目地崇拜和追求，看别人用什么我也用什么，度娘怎么说我就怎么做，但其实大数据只是一种数据处理的指导思想，并非体现在某种技术的实现层面上。云计算作为一种理念，是为了帮助我们更好地利用和整合资源。你的产品并不意味着一定要用大数据，你的系统也不是放在云上就一定好。如果你真的需要大云，那该怎么用？用什么？如何架构？用过之后又会遇到随之而来的各种坑，各种麻烦。且看各位大神是如何经历这个从无到有，一路掉坑、爬坑、越野过来的历程和体会，一定会对大家有巨大的帮助！

感谢皮皮不辞辛劳地为我们带来诸位大神如此宝贵的经验财富，看过此书的人可以少走弯路，避免踩坑，从此就算开着超跑也能直奔大云了！

——张建华，云贝国际控股集团 CTO，历任搜狐、捉鱼、2345 等公司技术高管

大数据驱动的信息化变革影响深远，让越来越多的公司认识到数据的价值，开始洞察、利用数据，并变革决策及运营体系。《皮皮访谈录》以访谈的形式，讲述了一些知名企业的大咖在大数据及云计算领域方面的经验、历程及技术见解，是大数据领域难得的佳作，对希望在大数据领域有所突破的公司和从业者非常有借鉴价值。

——张炜，美上门技术总监

# 序言

---

IT 圈子里流行着两大社区——ITPub 和 ChinaUnix，国内数一数二的架构师、DBA、运维师、技术专家大抵集结于此，好比江湖圈子流行一个光明顶，集结了各路英雄好汉、武侠高手。金庸用他的小说体记录了光明顶上群雄逐鹿和飞檐走壁。而我，作为 ITPUB 和 ChinaUnix 两大社区运营负责人，利用“IT 名人堂”这个专门针对架构师、DBA、运维专家等 IT 牛人的访谈类栏目，采访了来自百度、阿里、腾讯、京东、网易、奇虎、去哪儿、携程、58 同城、滴滴、易车、IBM、微软、戴尔、SAP、华为、Hortonworks，以及听云、百分点、星环科技、永洪科技等 IT 豪门的 CTO、技术总监、架构师、DBA、运维专家等。

很多人问我，为什么策划“IT 名人堂”？和很多的北漂族一样，我也曾呐喊，也曾努力，在寻找怒放的生命，在刷新自己的存在感。千里之行，始于足下。我放飞梦想，也带着传递正能量的信念，开始走上了皮皮访谈录的生涯，花了一年多的时间来策划 IT 名人堂。

有些话，年轻的时候羞于启齿，等到张开嘴时，已经人近中年，且远隔万重山水。同样，有些人，有些事，经历了如果没有留下痕迹，以后再回忆就淡忘了。每一个值得传播的观点都有一个值得讲述的故事，每一个神秘的故事背后其实是一幕幕真实的励志人生。我试图从一个女性和 IT 媒体人的角度，去深入了解这些被贴上“技术控”、“宅男”、“码农”、“极客”标签的 IT 成功人士。

本书六脉精通，涵盖了“海量数据分析与处理篇”，“云计算架构与优化篇”，“数据库性能优化与实践篇”，“大型分布式软件架构篇”，“IDC 与数据中心运维篇”，“职场生涯与创业篇”。

如果你是一名 DBA，如果你是一名架构师，如果你是一名运维人员，又或者只是一名 IT 从业者，那么无论你是职场老兵，还是新人菜鸟，当你遇到挫折时，当你需要安慰时，当你乘公交坐地铁时，当你饭后困顿时，在这本书里，总有那么一瞬间，你会被某个嘉宾的采访所感染；也总会有那么一瞬间，因为他们的故事，你会爱上某个企业，对某款产品、某个技术背后的架构而心生好奇，或微信，或电商。

最后，我要特别感谢我的父母和我的哥哥皮光明给予我的各种支持与帮助，感谢洪钊峰、王蓓、黄围围、石松、周荣茂、高可、王莹欣、张辉、张梦雨、张莹、刘刚等对我的帮助，感谢 ITPub 和 ChinaUnix 社区的所有网友，感谢所有帮助过我的人！

# 目录

---

<b>第 1 章 海量数据分析与处理篇</b> .....	1
1.1 百度地图大数据科学家吴海山：从百度地图到百度迁徙看时空大数据演变... 1	
1.2 Intel特邀大数据技术顾问谭磊：百度、阿里、腾讯、360、乐视、京东，其实都需要我.....	6
1.3 Hortonworks亚太技术总监Jeff Markharm：他们贡献了80%的Hadoop源码....	11
1.4 星环科技CTO孙元浩：上海外滩踩踏惨案背后的大数据反思.....	19
1.5 百分点研发总监刘国栋：不止于平台，大数据操作系统重磅来袭.....	24
1.6 华为中间件首席架构师彭渊：他发明了Hadoop的双胞胎.....	32
1.7 商业智能开拓者永洪科技CEO何春涛：大数据小数据，一天实现可视化分析....	45
<b>第 2 章 云计算架构与优化篇</b> .....	52
2.1 阿里巴巴资深平台架构师九峰：他们部门是“去IOE”的始作俑者.....	52
2.2 腾讯云资深技术专家于江：腾讯云的两大杀手锏.....	57
2.3 网易云技术总监陈谔：网易云如何支撑云音乐+云课堂+云笔记+云阅读.....	62
2.4 网易云存储架构师来东敏：他设计的网易云存储支撑了网易30多个产品.....	67
2.5 戴尔任意云专家NewMan Wong/Lijun Zhang：他为什么说，云计算不是双刃剑.....	71
2.6 戴尔任意云专家NewMan Wong/Lijun Zhang：戴尔任意云如何做到“任意”.....	79
2.7 SAP中国顾问协会创始人伍昊献：SAP未来是否更名为SAP HANA Cloud.....	86
2.8 柴静“穹顶之下”的网友献计献策：云计算如何治理雾霾.....	91

<b>第3章 数据库性能优化与实践篇</b> .....	96
3.1 阿里巴巴RDS首席产品架构师何云飞：阿里云数据库RDS架构的演进之路....	96
3.2 阿里巴巴架构师杨传辉：双11支付宝核心数据库OceanBase的前世今生.....	103
3.3 阿里巴巴分布式数据库负责人沈询：6年间他只做一件事——淘宝分布式数据库TDDL/DRD.....	108
3.4 微软技术中心架构师韩凯：SQL Server 2014与Windows Azure云的整合.....	116
3.5 微软技术中心架构师韩凯：看完SQL Server 2014的ITPUB网友互动，想不升级都难.....	125
3.6 携程网高级DBA任赞婷：携程DBA女神如何管理300台MySQL和Mongo.....	133
3.7 58同城架构师沈剑：一切抛开业务的数据设计都是耍流氓.....	137
3.8 易车网高级DBA高继伟：MySQL比SQL Server差在哪儿.....	140
3.9 达梦董事长冯裕才：坚持30年，只为数据库中国梦.....	146
3.10 达梦国产数据库DBA黄海明：选择达梦，国家情感优先还是客户价值优先....	154
3.11 全球母婴第一社区宝宝树运维总监刘秋歧：2亿张宝宝照片背后的DBA之道.....	157
3.12 Oracle ACE杨志洪：12年DBA心得——数据库故障三分天注定，七分是人为.....	166
3.13 蓝色曙光数据库总监史跃东：房子、车子、妻子、孩子，做DBA一切都有.....	171
3.14 前东软高级DBA敬勇：为什么这7年他一直钟情于“社会保障系统”.....	180
3.15 ITPUB版主DBA怀晓明：是什么让他对DBA坚守十年.....	185
3.16 80后摇滚DBA高强：做音乐的人如何做IT.....	190
3.17 ITPUB DBA网友热议国产数据库：国产数据的雄起，靠去IOE还是自身实力.....	196
3.18 ITPUB DBA网友热议首家银行系统“去IOE”：Oracle渐去，阿里巴巴来袭.....	205
<b>第4章 大型分布式软件架构篇</b> .....	213
4.1 阿里巴巴高级架构师强琦：淘宝天猫双11秒杀背后的高并发核心架构.....	213
4.2 腾讯社交运维专家周小军：2亿QQ峰值在线并发背后的分布式存储架构.....	218
4.3 京东高级架构师袁航：双11秒杀背后的NoSQL数据库与分布式存储.....	226
4.4 京东开放云高级总监：京东618零宕机背后的京东云核心架构.....	230
4.5 京东云平台架构师桂创华：他主导的京东分布式存储平台，支持京东数百个业务.....	235



4.6	1号店技术总监黄哲铿：大型电商网站的IT架构.....	238
4.7	奇虎360高级专家王超：360如何做到任意一款产品开发周期不超过一个月....	244
4.8	美团网业务核心系统架构师陈义宏：美团“狂拜访，狂上单”背后的高并发核心架构.....	249
4.9	滴滴打车技术总监杨振麟：管理靠盯、技术靠练、办法靠想、潜力靠逼....	254
4.10	超级课程表CTO茹云峰：他如何做到23岁成为网易最年轻系统架构师.....	259
4.11	PMC存储架构师张冬：大话存储.....	264
4.12	ITPUB网友热议淘宝数据吹牛：TPS最大是工行不是淘宝，QPS最大是移动不是淘宝.....	269
<b>第5章</b>	<b>IDC与数据中心运维篇.....</b>	<b>274</b>
5.1	听云总裁张涛：抛开性能问题谈体验都是耍流氓.....	274
5.2	沃趣科技创始人陈栋：数据库运维高枕无忧之秘诀.....	279
5.3	ITPUB网友热议IBM开放POWER：是好事还是坏事.....	284
5.4	ITPUB网友热议IT运维：运维就是消防员，只要宕机夜里3点也得“救火”....	291
5.5	ITPUB网友热议去哪儿网：“真假保单”系统故障Or逐利玄机.....	296
5.6	ITPUB网友热议密码泄露风波：密码不安全，换脸靠谱吗.....	300
<b>第6章</b>	<b>职场生涯与创业篇.....</b>	<b>304</b>
6.1	Linux创业达人陈荣友：青春岁月，身不由己.....	304
6.2	SaaS创业达人陈磊：创业方向最重要.....	310
6.3	机器人创业达人皮光明：我眼中的机器人.....	315
6.4	智能家居创业达人周成虎：创业是一种人生信仰.....	319
6.5	让我们一起追忆IT男女的那些苦与乐，酸与甜.....	325

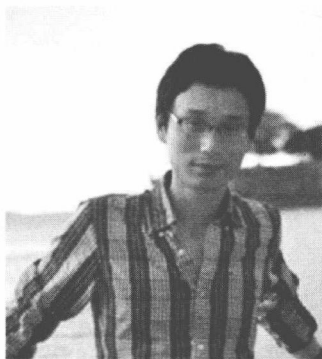
# 第1章 海量数据分析与处理篇

---



## 1.1 百度地图大数据科学家吴海山：从百度地图到百度迁徙看时空大数据演变

吴海山简介：于 2011 年从复旦大学计算机学院博士毕业，博士的研究方向主要是基于计算机视觉的群体行为量化研究。毕业后加入 IBM 中国研究院，进行商业数据分析与优化的相关研究。2014 年 9 月加入百度研究院大数据实验室，负责进行百度时空数据的研究。



在互联网+热潮下，谈的最火的话题之一是大数据，而在大数据领域做得风生水起的当属百度和阿里了。从世界杯预测、高考预测，到奥斯卡奖项预测，百度大

数据吊足了大家的胃口，百度地图每天响应来自手机用户的上百亿次定位请求，数亿的地图位置检索请求，可以收集到移动互联网用户海量的出行数据，百度迁徙更是在春节期间一炮走红，百度热力图、百度天眼等产品也相继问世。天地之间，无论何时何地，我们的世界都漂浮着海量的时空大数据，站在互联网+的风口上，这些时空大数据能飞起来吗？它们究竟有哪些用武之地呢？本期名人堂皮皮邀请到了百度研究院大数据实验室数据科学家吴海山坐镇，为我们揭晓百度大数据的神秘面纱。

**皮皮（Q1）：**你好！很高兴有机会采访到你，我了解到你此前在 IBM 研究院从事商业数据的研究与优化，现在负责百度时空大数据的挖掘研究，很多人可能对时空大数据感到困惑，究竟什么样的数据是时空大数据呢？你觉得 IBM 的大数据和百度的大数据来源有什么区别？

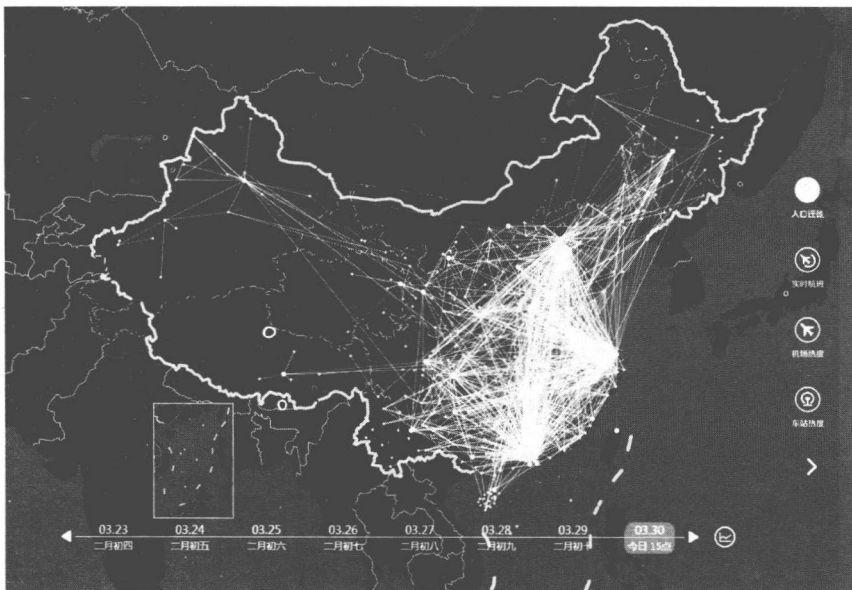
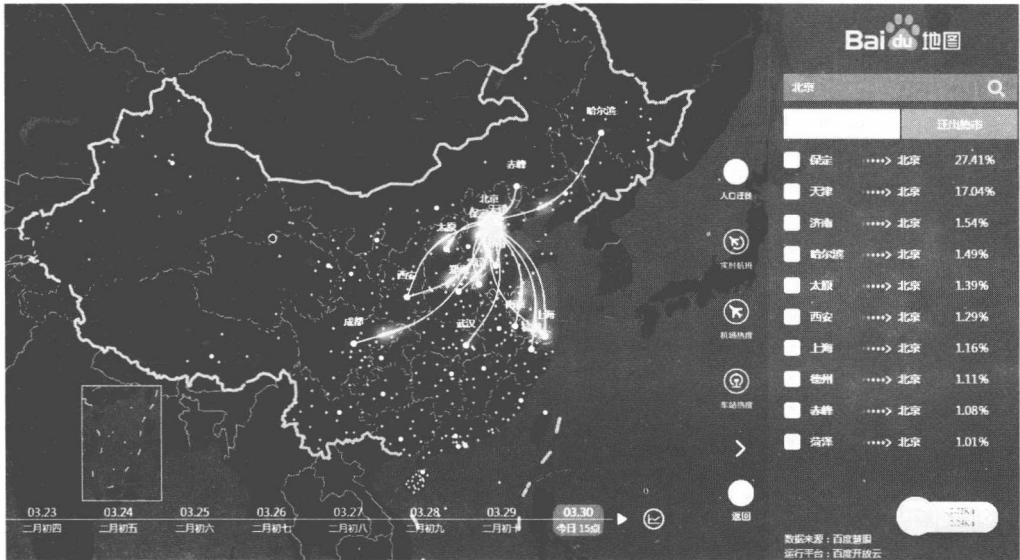
**吴海山（A1）：**所谓的时空大数据（spatial-temporal big data），指的是随着时间变化的空间数据，最为典型的就人出行时产生的历史轨迹，根据轨迹我们就可以知道你什么时候（时）出现在什么地方（空）。近年来，随着移动互联网的普及，有定位传感器的智能手机覆盖面也越来越广，每天会不断产生海量的人的位置数据，也就是时空大数据。

大数据可以说是 IBM 炒红的一个概念。IBM 主要以提供商业解决方案为主，主要的客户来自交通、电力和银行等行业。从某种意义上来说，IBM 本身不产出数据，主要提供对数据的存储硬件设备和相应的数据分析软件平台。和 IBM 不同的是，百度是一个互联网公司。放眼望去，全国网民每天可产生近百亿的搜索数据，像百度地图、百度搜索等都产生了海量的数据。

**皮皮（Q2）：**今年春节前夕，业界首个以“人群迁移”为主题的大数据可视化项目——百度迁徙精彩亮相了，有人把它比喻成从上帝的视角来看待大数据的魅力，大家可以实时查看全国人口的流动情况、查询实时航班信息，还能嗅探到全国火车站和飞机场的热度等，这些以烟花般美丽炫酷的信息图背后到底用到了百度的哪些技术呢？

**吴海山（A2）：**百度迁徙是一个非常棒的项目，尤其是今年，百度迁徙还加入了百度天眼的航班信息、火车站信息等，效果非常炫。我本人并没有参与这个项目，所以无法对具体的技术细节做讲解。不过总体来讲包括这样几个部分：其一，后端实时流数据的传输和计算，每天由百度地图产生的各种数据量非常大，百度迁徙需

要尽量缩短数据的清洗、处理等带来的时间延迟，但百度拥有自己强大的流计算平台；其二，前端的数据可视化技术，这个用到了百度自己开发的 Echarts, 可以高效地呈现可视化的时空数据，如下图所示。

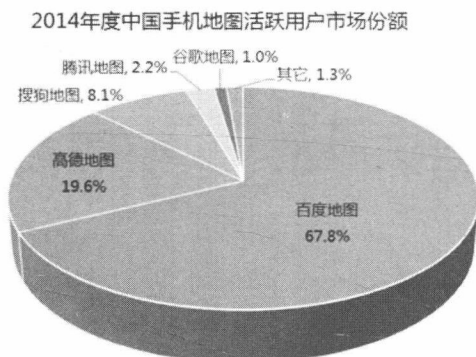


皮皮 (Q3): 百度地图覆盖 500 万公里路网和全国 95% 的高速公路, 支持全国

337 个地级城市高清卫星图和热力图服务，百度地图每天响应来自手机用户的上百亿次定位请求，数亿的地图位置检索请求，它是如何做到数据的实时定位的呢？我们知道百度推出的很多 O2O 服务都以百度地图作为第一流量入口，比如天气应用、同城类应用等第三方应用，能不能和我们分享一下百度地图发力 O2O 的案例？

吴海山（A3）：百度拥有自己研发的定位 SDK，可以通过多种方式，如通信基站、GPS 卫星、Wi-Fi 等方式在不同情景下实现精确的实时定位，当然不同的定位方式，精度有所不同。

作为百度研究院的大数据实验室（big data lab, BDL），我们主要结合公司的长期战略，进行相对前沿的研究，孵化出相关有潜力的产品 demo。比如我们最近有一个关于智能化、个性化的地图搜索技术，当大家在百度地图上搜索“某某饭店”时，我们会根据用户的历史位置和检索偏好等信息，准确预测用户到访该饭店的时间，然后饭店可以提前为用户排号、检查停车位等服务，并根据到访的用户量优化资源配置，从而实现打通线上、线下服务。我们还推出了算法，可以预测用户下一个要搜索位置的类型，超前一步，将用户最想找的位置放在最前面，通过预测性位置搜索（predictive location search）提高用户的体验。又如我们通过大数据分析，得出了 2014 年度中国手机地图活跃用户的市场份额，如下图所示。



数据说明：手机地图活跃用户是指在2014年度月均使用该手机地图4次以上的用户即为该手机地图的活跃用户，一个用户若同为多款手机地图的活跃用户按多个活跃用户计算。

数据来源：比选咨询（BigData）数据中心

皮皮（Q4）：百度在大数据上动作频频，比如世界杯预测、高考预测、奥斯卡奖预测吊足了大家的胃口，春节期间百度迁徙更是一炮走红，百度热力图、百度天眼等产品也相继出世，而支撑这背后的团队无疑归功于像你一样的，一群百度研究

院大数据实验室的工程师们，能否和我们介绍一下百度研究院大数据实验室？在研发诸如上述的百度大数据产品时有哪些创意和趣闻轶事呢？

吴海山（A4）：百度研究院（Baidu Research）有三个实验室，除了我们的大数据实验室（BDL），还有深度学习实验室（IDL）和在美国硅谷的人工智能实验室（AI Lab）。研究院的院长是 Andrew Ng 博士，我们实验室的主任是著名的机器学习科学家张潼教授。除了时空数据的研究，我们实验室的研究方向还有智能预测引擎（如百度高考预测、世界杯预测等）、大规模机器学习算法研究、智能医疗等方向，实验室的主要成员都来自国内外的知名研究院所。作为公司的研究部门，我们本身并不负责一个完成产品的开发，而更侧重根据研究成果开发原型（prototype），然后会有相关的产品部门负责产品化。

皮皮（Q5）：我国人口众多，重大文体活动、节假日集会等活动中，容易出现因人群过度拥挤而引发的危险乃至事故。那么百度大数据是否可以提前预测，做到事前预警呢？请结合上海踩踏事件谈谈你的想法。

吴海山（A5）：说起人群拥挤的研究，还要从我在美国普林斯顿大学从事的博士后研究说起。我一直对自然界的群体行为十分感兴趣，当时主要研究的就是鱼的群体行为，就是想探索为什么鱼可以成群地游泳、背后有哪些玄机？我有个非常不错的研究成果，最近刚刚发表在美国科学院院刊（PNAS）上，我们通过计算机视觉、机器学习的方法探索了信息在鱼群中的传播机理。对于人群来讲，行为其实更加复杂。尽管国内外也有不少学者对人群的异常行为做了一些研究，但是要说预测，还是非常困难的事情。我个人觉得，我们最多能做的是可以提前发现异常，进行预警还是相对可行的。

上海外滩发生踩踏悲剧后，今年1月2日，我们实验室就曾用百度的定位数据进行了可视化分析，然后负责开展百度大数据的外滩踩踏事件分析工作。通过和其他几位同事一起尝试融合百度的多个数据源，试图还原当时的场景，看是否可以在事后寻找一些可以对将来有帮助的线索。根据我个人的研究经验，发生踩踏事情是难以预测的，国内外都发生过多多次这样的悲剧，比如当时很多媒体报道认为，由于人多导致了悲剧，我们的分析结果表明，上海南京东路附近同样达到过类似的人群密度，但却没有发生悲剧。我们当时的研究也主要集中在预警而不是预测上。根据我们的研究，我们发现通过融合多个数据源，是有可能进行预警的：我们发现用户在去目的地之前，会提前使用百度地图进行搜索，进行路径规划后到达现场，这个

时间的提前量在半个小时到两个小时不等，我们还用了鸟巢、首体等地点大型活动的数据进行了验证，也发现了同样的现象。而且地图位置的搜索量和人口定位量也高度相关，这意味着我们可以提前预测人口流量。如果人口流量超过一定阈值，相关部门则可以进行提前布置警力等来预防。此外，我们还研究了预测算法，来更准确地预测人流量。目前我们正在开发该产品，希望将来百度大数据能为公共安全贡献一份力量。

**皮皮 (Q6)：**被誉为互联网下一个入口的是 LBS，百度、阿里、腾讯三大互联网巨头纷纷在地理位置服务领域加大了布局，互联网地图的“入口”之争将会愈演愈烈，高德地图对百度地图步步紧逼，百度 LBS 开放平台横跨了 Web 端、移动端、服务端，能否和我们谈谈百度 LBS 有哪些优势？

**吴海山 (A6)：**时空大数据挖掘是我们实验室的一个重要研究方向，产出的研究成果，将会为百度 LBS 提供强大的“百度时空大脑”：比如我们重点研究的个性化和预测性的位置搜索（personalized and predictive location search）将可以服务百度地图搜索；我们对基于百度大数据的公共安全领域的研究，将可以帮助百度 LBS 在未来拓展该领域的市场；我们还有一些更加前沿的项目研究，这些研究结果不但可以更好地理解百度用户的行为，实现更加智能的人与服务的链接，还可以让公众看到，这些研究未来将受益所有的中国人。通过数据洞察当今的中国，是我们的使命。

## 1.2 Intel特邀大数据技术顾问谭磊：百度、阿里、腾讯、360、乐视、京东，其实都需要我

**谭磊简介：**浙江省企业信息化促进会互联网分会会长，英特尔公司特邀大数据顾问，资深技术管理、数据分析挖掘专家，在美国微软总部服务 13 年，也是《New Internet——大数据挖掘》、《数据掘金——电商数据运营》、《盛宴背后——互联网金融揭秘》三本书的作者。



我是主持人皮皮。有人把数据比作新的石油，认为大数据将彻底改变人类文明的发展脉络，重塑我们对于世界、对于生活的认知。互联网的对决不再止于速度，深度的较量将越演越烈，而如何挖掘数据这座巨大而未知的宝藏，让数据会说话，已成为时下企业扭转乾坤的关键突破口。

谭磊曾经说过，“数据是最底层的東西，它是 0 和 1，是对事物最原始的记录，数据之上才有信息。”如果说内容为王是媒体行业经久不衰的秘诀，那么数据为王则是互联网企业决战大数据时代的枪支弹药。本期我们很荣幸地邀请到了数据分析挖掘专家谭磊（社区 ID：SeattleRaymond）做客名人堂，为我们分享大数据的真知灼见。

皮皮（Q1）：谭老师，你好！我了解到你在美国微软总部服务了 13 年，后回国创业，可见你的人生阅历很丰富，能否和我们介绍一下自己？在人生的道路上，你有哪些心得体会？能否和我们分享一下你的故事？

谭磊（A1）：我人生的前 28 年是非常顺利的，除了考初中和 TOEFL、GRE 之外，高中和大学是直升的，就连研究生班也是直升的。我在复旦读研一的时候，拿到了美国杜克大学的全额奖学金去读博士，在 1995 年我也是学校唯一一个通过微软面试去做实习的博士生。进了微软之后，我的职业生涯一帆风顺，每年升一个级别，很快就开始带团队做项目了。

不过正因为一开始就特别顺利，在过去的二十年职业生涯中，我犯过非常多的错误。等到哪天我功成名就了，我会写一本《Raymond 的 99 个错误》，希望大家能从我过往的失败中吸取教训。

在这里，我给大家分享一个小故事。小时候盖茨是我的偶像和榜样，当我刚进微软的时候，我就和大家说过，我进微软就是因为 Bill Gates。哪一天他离开，也就是我离开的时候。结果在 13 年之后的 2008 年，在盖茨向全公司宣布他要离开微软之前的几个小时，我群发邮件给我微软所有的朋友们，宣布那天是我在微软工作的最后一天，虽然当时我也并没有完全想好打算要做什么。

我一直都喜欢数字的游戏，在美国的时候经常去拉斯维加斯，一边在玩，一边在计算每种不同玩法对应的赌场胜率（odds）以及每笔赌注的投资回报率。和别的去赌场的人心态不同，在去之前我就定好一个数额，比如 \$5000，这笔钱会花在拉斯维加斯。“What happens in Vegas, Stay in Vegas.” 如果输了，那就最多 \$5000。如



果赢了，也在当地花掉，不带回来。这样的话，去赌场就变成了娱乐。你可以这么想，每小时花 \$30，在这么好的场地，请这么多人陪你一起玩，太合算了。注释一下，如果玩每把 \$50 的 21 点，平均每小时庄家会发 60 轮牌，按照正常的打法，赌场的胜率在 51%，也就是平均每小时赌场会从我身上赚 \$30。

在美国让我最难以接受的地方就在于节奏太慢了。而反过来看中国，大家的想法很多，机会非常多，而有执行力的人也不少。我有时候在想，应该和唐骏或者熊明华他们一块儿回来的。不过换一种心态来想，一切都很释然了。我们今生在什么时间什么环境会遇上什么样的人，发生了什么事，是冥冥之中注定的，过去的一切都是积累。如果没有这些年在微软的沉淀和努力，我对于数据和技术的理解恐怕难以达到这么深的高度。如果之前没有在青岛和北京做过一段时间，我不会下决心回到国内。如果不是在杭州结识了一群兄弟，我也不会网络营销上有实际操作的经验。

皮皮 (Q2)：很多人认为数据是死的，人是活的，数据分析到底是一种艺术还是一门科学？

谭磊 (A2)：准确地说，数据分析不是艺术，也不是科学，而是用来帮我们做决策的一种手段和方法。在现有数据上做数据分析，只有统计专业毕业的同学才能胜任，而随着今天各种工具的出现，具有商业运营背景的同学也能做数据分析了。不过，真正要把数据分析做好，其终极目标是要做基于数据的运营。

谈到数据分析，我想举一个司空见惯的例子。一旦你曾经在搜索引擎上搜索过某一件东西，你的互联网世界里都将离不开这个东西的影子，它的广告图片会出现在你浏览的网站上。只因为你之前输入了搜索关键字，这些图片的针对性是很强的，而出现得太多就让人厌烦了。图片的出现是数据分析，而图片出现的频率可能就接近艺术了。

每个公司都有很多自己的数据，但极少数公司能够充分用好这些数据，有些公司能够通过这些数据做一些分析，而绝大多数公司的数据都是沉睡在那里的。2014 年在大连的 iTechClub 峰会上，我和大家说过这样一个事情：每家公司，包括在座的百度、阿里、腾讯、360、乐视和京东等，其实都需要一个 mini-Raymond，把数据真正使用起来。这个意思并不是说有多强，而是说当你想要使用数据的时候，你需要把技术、商业和运营都考虑在内。数据是基础，商业是土壤，而运营是在这