



Strata Data Conference 北京2017

尊敬的 _____ 先生/女士，您好！

Strata Data Conference 北京2017将于2017年07月在北京召开。

会议介绍

Strata Data Conference: Make data work

Strata Data Conference是关于数据、机器学习及分析如何改变商业和社会本身的领先会议。来自各种规模创新公司的顶尖数据科学家、分析师和管理人员聚集一堂，分享深入、难以获取的知识。

Strata Data Conference 7月12-15日重返北京。我们在寻找讲师来和有才华的技术观众分享有吸引力的数据案例分析、成熟的最佳实践、有效的新分析方法以及不同寻常的技能。我们寻找的主题包括：

会议主题包括：

AI应用

智能应用例如个人助理、聊天机器人、机器人、无人机、无人驾驶汽车正出现在各领域。除了应用和使用案例我们也对一些议题和培训感兴趣，包括深度学习、强化学习、概率机器学习、计算机视觉、自然语言理解、语音和情感技术及相关主题。

数据科学&高级分析

机器学习的最新算法和进展，以及文化演变和团队建设方面的棘手问题。

- 统计、算法及机器学习（包括深度学习）
- 主动学习及其他“人类参与的”机器学习系统
- 数据分析流程、探索、协作、同行评审、记录的再现性和数据来源。
- 通过设计和社交科学技巧来创建更好的试验并提出正确的问题。
- 数据结构和布局（表、图形、网络、时间序列、非结构化文本）
- 欺诈侦测、对抗分析、游戏理论

Hadoop使用案例

Hadoop生态系统真实案例分析，范围包括从初创企业到行业巨头。

Hadoop内核&发展

深入了解这一主流大数据栈，包括实践经验、整合技巧和前瞻。

数据工程和架构

Hadoop之外的工具——例如Cassandra、Storm、Elasticsearch、Kafka还有基于云的服务——以及它们如何融入数据科学工具包。

- 大数据平台及架构 (Hadoop、Elasticsearch、Cassandra、Kafka、Storm等等。)
- 扩展、查询性能、可用性、计算成本、自动化、加密
- 为分析做预处理、清理、整理和增强数据
- 混合内部部署和云数据服务

Spark&更多发展

Apache Spark最佳实践、架构考量、来自于初创公司和大型企业的真实案例研究。

物联网&实时计算

物联网搜集和产生的数据——包括存储、分析和发布这些信息的困难；及从结果洪流中抽取能理解的、有意义的见解。

- 机器、传感器、人群及移动数据搜集
- 分析与物联网
- 开放数据标准与互操作性

安全

数据需要像加密这样的工具来保护安全和隐私；越来越多的数据和算法可以改善我们的安全状况。但是安全团队一直都在和那些试图钻算法漏洞的人展开不断竞赛。该主题系列探讨数据治理和数据在更好的安全性中的角色。

企业应用

企业如何从遗留数据存储向大数据转移以及最佳实践——及障碍——从而成为数据驱动企业

可视化&用户体验

如果数据没产生结果就没任何意义。该主题系列解决增强、用户体验、新界面、交互性及可视化。

- 分析和报告
- 增强和虚拟现实
- 设计、交互性及可视化
- 设计中断和上下文接口
- 用户体验和数据驱动设计

开放数据

在中国有越来越多的开放数据存储库和计划。我们将展示大数据技术、数据科学和开放数据在公共或私有领域的应

用。

关于 Strata Data Conference

Strata Data Conference是最前沿科学与新兴商业基础碰撞和融合的地方。在这里我们会深入探索新兴技术和科技。您将通过深入的辅导课剖析案例研究、发展新技能，分享数据科学中新兴的最佳实践并畅想未来。

该活动之前作为Strata + Hadoop World创建于2012年，O'Reilly和Cloudera将两个成功的大数据会议组合在一起。

议题主席Doug Cutting（Cloudera首席架构师，Apache Hadoop创始人）、Roger Magoulas（O'Reilly研究主管）以及企业家Alistair Croll和议题开发总监Ben Lorica（O'Reilly首席数据科学家）已经安排了一个覆盖整个大数据工具和技术的议题安排。Strata Data Conference涵盖了像人工智能和机器学习等当前热门话题，并且重点放在如何实施数据战略上。

为什么您应该参加

Strata Data Conference 将聚集大数据领域最有影响力的产业决策者、战略专家、架构师、开发人员和分析师，共同打造产业和技术的未来。

- 成为了解如何利用这些巨大变化的最前沿人群，并在所导致的颠覆中存活下来
- 在各个行业和学科找到利用您的数据资产的新方法
- 学习如何从科学项目中提取数据并应用到实际行业中
- 对专业数据人士来讲将发现培训、雇佣和职业机会
- 与其他创新人员和意见领袖面对面交流

体验 Strata Data Conference

3整天的议程包括富于启发的主题演讲、非常实用又有丰富信息的议题，以及很多有趣的社交活动。

- 探索最新的前沿问题、案例研究以及最佳实践
- 与商界领袖、数据专家、设计者和开发者交流的机会
- 为参会者、记者和供应商提供了活跃的“走廊交流会”，使您有机会对重要问题进行探讨和辩论
- 有趣的晚间活动招待会，更重要的是给您更多与参会者和演讲者面对面的时间

您会看到谁

Strata Data Conference 将吸引数据行业最出色的人员：开发者、数据科学家、数据分析师以及其他数据行业的专业人员，包括：

- 商业智能经理和分析师
- 商务经理、战略专家和决策者
- CIO, CTO 以及企业架构师
- 数据驱动设计者，记者以及人类学家
- 数据工程师
- 数据科学家
- 设计者

- 开发者和数据库专业人员
- 创新人士和企业家
- 产品经理
- 研究人员和学术人员
- 风投和投资者
- 副总裁、市场主管或数据仓库主管

培训

Strata Data Conference 培训

*Training courses takes place 9:00am - 17:00pm and are limited in size to maintain a high level of hands-on learning and instructor interaction. **Training passes do not include access to tutorials on Thursday.***

Breaks:

- *Morning break: 10:30 – 11:00*
- *Lunch: 12:30 – 1:30*
- *Afternoon break: 15:00 – 15:30*

July 12 — 13, 2017

Included in Platinum and 2-Day Training passes.

Participants should plan to attend both days of this training course.

09:00–17:00 Wednesday, 2017-07-12

数据科学精髓：互联网金融实例 - 量化线上金融信用与欺诈风险的评估 (Data Science Essentials: Examples from Internet Finance - Quantifying Credit and Fraud Risks Online)

Location: 多功能厅3B (Function Room 3B) 观众水平 (Level): 中级 (Intermediate)



Jike Chong (YiRenDai/CreditEase)

您想了解互联网金融幕后的量化分析流程吗？个人信用是怎样通过大数据被量化的？在实践中，机器学习算法的应用存在着哪些需要关注的方面？怎样通过图谱分析来融合多维数据，为我们区分正常用户和欺诈用户？这套辅导课基于清华大学交叉信息研究院2017年春天新开设的一门“量化金融信用与风控分析”研究生课。其中会用LendingClub的真实借贷数据做为案例，解说一些具体模型的实现。

July 12, 2017
Included in Gold and 1-Day Training passes.

09:00–17:00 Wednesday, 2017-07-12

Apache Spark高级实践和原理解析 (Advanced practice and principle analysis)

Location: 多功能厅5A (Function Room 5A) 观众水平 (Level): 中级 (Intermediate)



Carson Wang (Intel), 俞育才 (Intel), Zhichao Li (Intel), Yiheng Wang (Intel), Daoyuan Wang (Intel)

这几年随着大数据分析和机器学习等等在工业界中越来越广泛的应用，越来越多的人选择在大数据平台比如Apache Spark之上构建大规模数据处理、分析和机器学习，以便利用大量原始数据和扩展架构。如何深入理解大数据关键技术并更好的运用它们？本次课程将结合当前大数据技术的浪潮和趋势，为您介绍Apache Spark的高级实践和原理解析，帮助您加深领会Apache Spark的精华设计思想，以及如何与流式分析、机器学习，深度学习等紧密结合，在数据采集，分析处理，特征提取，机器学习等方面提供一致性和集成性的高级实践。

辅导课

辅导课

请选择8月4日周四的全天或半天辅导课。专家的讲座将带你深入重要议题。请注意：参加辅导课您的注册内容包必须包含周四辅导课；该门票不能参加培训课程。

Thursday, July 13

09:00–12:30 Thursday, 2017-07-13

使用Alluxio(前Tachyon)来加速大数据计算 (Using Alluxio (formerly Tachyon) to speed up big-data analytics)

地点：多功能厅2 (Function Room 2) 观众水平 (Level): 中级 (Intermediate)



Bin Fan (Alluxio), Haoyuan Li (Alluxio)

在这个三个小时的教学课中，我们将向参与者讲授Alluxio基础知识，演示Alluxio如何工作以及如何使用此系统帮助分布式计算引擎（如Spark或MapReduce）以内存速度共享数据。

09:00–12:30 Thursday, 2017-07-13

用TensorFlow进行深度学习 Deep Learning with TensorFlow

地点：报告厅 (Auditorium) 观众水平 (Level): Beginner



TensorFlow是一个流行的开源机器学习库，特别适合进行深度学习。本辅导课会通过实际的例子来介绍机器学习和深度学习。我们会指导参会者自己动手来使用TensorFlow和TensorBoard进行练习。

09:00–12:30 Thursday, 2017-07-13

从简单到复杂：Apache Kafka应用实例详解

地点：多功能厅8A+8B (Function Room 8A+8B) 观众水平 (Level): 中级 (Intermediate)



Jiangjie Qin (linkedin corp)

Apache Kafka作为近年来最流行的消息系统之一，其使用场景已经从最初的集中系统消息队列发展到更为复杂的一系列使用场景，包括流处理，数据库复制，CDC等等。本次演讲将以Kafka在LinkedIn的实践为基础详细介绍Kafka的各种应用场景。

09:00–12:30 Thursday, 2017-07-13

大数据的数据模型Big Data - Data Modeling

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): Beginner



Ted Malaska (Blizzard Entertainment)

The recent advancement in distributed processing engines, from Spark to Impala to Spark Streaming or Storm, has proved exciting. However, if your design only focuses on the processing layer to get speed and power then you may be missing half the story, leaving a significant amount of optimization untapped.

13:30–17:00 Thursday, 2017-07-13

AWS上使用MXNet进行分布式深度学习Distributed Deep Learning on AWS using MXNet

地点：多功能厅2 (Function Room 2) 观众水平 (Level): Intermediate



Damon Deng (AWS)

深度学习正持续地在诸如计算机视觉、自然语言处理和推荐引擎等领域引领最前沿的进步。带来这个进步的一个关键因素就是大量的高度灵活和对开发人员很友好的深度学习框架的出现。在本辅导课里，亚马逊机器学习团队的成员将会就深度学习的背景做一个简短的介绍，主要关注与其相关的应用领域。并会对强大和可扩展的深度学习框架——MXNet——做一个介绍。辅导课的最后，你可以获得上手的机会来获得针对多种应用的经验，包括计算机视觉和推荐引擎等。并可以看到如何使用预先配置好的深度学习AMI和CloudFormation模版来帮助加快开发速度。

13:30–17:00 Thursday, 2017-07-13

使用Apache Spark和BigDL来构建深度学习驱动的大数据分析(Building Deep Learning Powered Big Data Analytics using Apache Spark and BigDL)

地点：报告厅 (Auditorium) 观众水平 (Level): 中级 (Intermediate)



Yiheng Wang (Intel)

深度学习已经在很多的领域（例如计算机视觉、自然语言处理和语音识别等）取得了顶尖水准的表现，对工业界有极大的潜在应用价值。我们应该注意到深度学习和大数据的联系非常得紧密。首先，深度学习的模型需要使用大量的数据来训练，这就是为什么它直到大数据时代才开始蓬勃发展。其次，现在绝大部分的大数据都是视频、音频和文字数据，非常适合使用深度学习算法来处理。为了能释放深度学习的能力，我们就应该把它运用在大数据的环境里。

13:30–17:00 Thursday, 2017-07-13

Modern Streaming Architectures

地点：多功能厅8A+8B (Function Room 8A+8B) 观众水平 (Level): Beginner



Sijie Guo (Twitter), Maosong Fu (Twitter)

The move to streaming architectures from batch processing is a revolution in how companies use data. But what is the state of the art for real-time data stack, including stream computing engine, data storage engine, language and tools. What are the typical challenges in a modern real-time data stack? How will the modern technology impact the streaming architecture and applications in the future?

13:30–17:00 Thursday, 2017-07-13

Hadoop application architectures: Fraud detection

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): Intermediate



Ted Malaska (Blizzard Entertainment)

Ted will walk participants through building a fraud-detection system, using an end-to-end case study to provide a concrete example of how to architect and implement real-time systems via Apache Hadoop components like Kafka, HBase, Impala, and Spark.

议题

11:15–11:55 Friday, 2017-07-14

Driving Southeast Asia Forward With Big Data

地点：紫金大厅B (Grand Hall B) 观众水平 (Level): Non-technical

11:15–11:55 Friday, 2017-07-14

Pluto: A Distributed Heterogeneous Deep Learning Framework

地点：报告厅 (Auditorium) 观众水平 (Level): 中级 (Intermediate)

11:15–11:55 Friday, 2017-07-14

Apache Hadoop 3 Features and Development Update

地点：多功能厅2 (Function Room 2) 观众水平 (Level): Beginner

11:15–11:55 Friday, 2017-07-14

HAP：多流动态实时分析系统

地点：多功能厅6A+B (Function Room 6A+B) 观众水平 (Level): 中级 (Intermediate)

11:15–11:55 Friday, 2017-07-14

Scaling R faster and larger using Apache Spark

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): 中级 (Intermediate)

13:10–13:50 Friday, 2017-07-14

Spinach: 使用Spark SQL进行即席查询

地点：紫金大厅B (Grand Hall B) 观众水平 (Level): 中级 (Intermediate)

13:10–13:50 Friday, 2017-07-14

数据驱动企业增长(Data Drive Enterprise's Growth)

地点：多功能厅6A+B (Function Room 6A+B) 观众水平 (Level): 高级 (Advanced)

13:10–13:50 Friday, 2017-07-14

ING的快速数据——运用流式分析解决方案来创建一个实时、数据驱动的银行(Fast Data at ING - streaming analytics solutions to create a real-time, data-driven bank)

地点：多功能厅2 (Function Room 2) 观众水平 (Level): Intermediate

14:00–14:40 Friday, 2017-07-14

Offheap HBase read-path in production - The Alibaba story

地点：多功能厅6A+B (Function Room 6A+B) 观众水平 (Level): Advanced

14:00–14:40 Friday, 2017-07-14

Distributed Deep Learning at Scale on Apache Spark with BigDL

地点：报告厅 (Auditorium) 观众水平 (Level): 中级 ()

14:00–14:40 Friday, 2017-07-14

Mastering Spark Unit Testing

地点：多功能厅2 (Function Room 2) 观众水平 (Level): Intermediate

14:00–14:40 Friday, 2017-07-14

SDK+FinGraph+Go：用一手行为数据和图谱信息创造商业价值

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): 中级 ()

14:00–14:40 Friday, 2017-07-14

Cost-Based Optimizer Framework for Spark SQL

地点：紫金大厅B (Grand Hall B) 观众水平 (Level): 中级 (Intermediate)

14:50–15:30 Friday, 2017-07-14

机器人的预测性维护实战：解读实时、可扩展的分析管道(Robot Predictive Maintenance in Action: Real-time, Scalable Pipeline Explained)

地点：多功能厅6A+B (Function Room 6A+B) 观众水平 (Level): Intermediate

14:50–15:30 Friday, 2017-07-14

使用Spark/BigDL高级机器学习实现寿险业务再发现(Re-implement Life Insurance Services by using Spark/BigDL Advanced Machine Learning)

地点：报告厅 (Auditorium) 观众水平 (Level): 中级 (Intermediate)

14:50–15:30 Friday, 2017-07-14

Speed Up Big Data Encryption In Apache Hadoop And Spark

地点：多功能厅2 (Function Room 2) 观众水平 (Level): Intermediate

14:50–15:30 Friday, 2017-07-14

大数据时代银行客户社交关系圈研究与应用

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): 中级 (Intermediate)

14:50–15:30 Friday, 2017-07-14

Apache Kudu: 1.0版和未来(Apache Kudu: 1.0 and beyond)

地点：紫金大厅B (Grand Hall B) 观众水平 (Level): Beginner

16:20–17:00 Friday, 2017-07-14

Spark在今日头条的实践(Practices of Spark in JinRi TouTiao)

地点：紫金大厅B (Grand Hall B) 观众水平 (Level): 中级 (Intermediate)

16:20–17:00 Friday, 2017-07-14

从LR到DNN点击率预估系统的进化(The evolution of CTR prediction systems, from LR to DNN)

地点：报告厅 (Auditorium) 观众水平 (Level): 中级 (Intermediate)

16:20–17:00 Friday, 2017-07-14

Spark和TiDB(Spark On TiDB)

地点：多功能厅2 (Function Room 2) 观众水平 (Level): 中级 (Intermediate)

16:20–17:00 Friday, 2017-07-14

ShadowMask: Anonymize your sensitive big data

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): 中级 (Intermediate)

16:20–17:00 Friday, 2017-07-14

Hadoop遇到云上对象存储——实现原理、陷阱和性能优化(When Hadoop meets Object Storage - Implementation Principles, Pitfalls and Performance Optimization)

地点：多功能厅6A+B (Function Room 6A+B) 观众水平 (Level): 中级 (Intermediate)

Saturday, July 15

11:15–11:55 Saturday, 2017-07-15

The Architecture of Decoupling Compute and Storage with Open Source Alluxio

地点：多功能厅6A+B (Function Room 6A+B) 观众水平 (Level): Non-technical

11:15–11:55 Saturday, 2017-07-15

Active Learning in the Real World

地点：报告厅 (Auditorium) 观众水平 (Level): Intermediate

11:15–11:55 Saturday, 2017-07-15

领英大数据平台-超过1万节点，每天15万个作业，智能连接4.7亿职场用户(LinkedIn Big Data Platform - 10,000+ nodes, 150,000+ daily jobs, connecting 470 million members)

地点：多功能厅2 (Function Room 2) 观众水平 (Level):

11:15–11:55 Saturday, 2017-07-15

Angel:面向高纬度的机器学习计算框架(Angel : A Machine Learning Framework for High Dimensionality)

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): 高级 (Advanced)

11:15–11:55 Saturday, 2017-07-15

Apache Kylin 2.0 : 从Hadoop上的OLAP 引擎到实时数据仓库

地点：紫金大厅B (Grand Hall B) 观众水平 (Level): 中级 (Intermediate)

13:10–13:50 Saturday, 2017-07-15

在滴滴出行的最佳实践(Spark best practice in Didi)

地点：紫金大厅B (Grand Hall B) 观众水平 (Level): 中级 (Intermediate)

13:10–13:50 Saturday, 2017-07-15

On-device machine learning: TensorFlow on Android

地点：报告厅 (Auditorium) 观众水平 (Level): Beginner

13:10–13:50 Saturday, 2017-07-15

Transactional Streaming with Apache DistributedLog

地点：多功能厅2 (Function Room 2) 观众水平 (Level): 中级 (Intermediate)

13:10–13:50 Saturday, 2017-07-15

多视图建模与半监督学习：应用于海量用户数据挖掘与行为分析

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): 中级 (Intermediate)

13:10–13:50 Saturday, 2017-07-15

Hyperledger与CDH大数据生态系统的融合以及应用实践(Hyperledger's Integration with CDH Big Data Ecosystem, and Its Application in Real World)

地点：多功能厅6A+B (Function Room 6A+B) 观众水平 (Level): 中级 (Intermediate)

14:00–14:40 Saturday, 2017-07-15

基于深度学习的网络表示

地点：报告厅 (Auditorium) 观众水平 (Level): Intermediate

14:00–14:40 Saturday, 2017-07-15

在领英搭建Hadoop和Kafka之间的桥梁——Hadoop团队的视角(Building the bridge between Hadoop and Kafka at LinkedIn - A Hadoop team's perspective)

地点：多功能厅6A+B (Function Room 6A+B) 观众水平 (Level): 中级 (Intermediate)

14:00–14:40 Saturday, 2017-07-15

欺诈的潜伏性 - 如何利用大数据进行反欺诈检测

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): 中级 (Intermediate)

14:00–14:40 Saturday, 2017-07-15

HBase多数据中心方案及未来的增量备份功能介绍

地点：多功能厅2 (Function Room 2) 观众水平 (Level): 中级 (Intermediate)

14:50–15:30 Saturday, 2017-07-15

Columnar Storage @ Uber

地点：多功能厅6A+B (Function Room 6A+B) 观众水平 (Level): 非技术性 (Non-technical)

14:50–15:30 Saturday, 2017-07-15

Powering Robotics Clouds with Alluxio

地点：报告厅 (Auditorium) 观众水平 (Level): 中级 (Intermediate)

14:50–15:30 Saturday, 2017-07-15

Alluxio缓存策略优化与大规模性能评测

地点：多功能厅2 (Function Room 2) 观众水平 (Level): 中级 (Intermediate)

14:50–15:30 Saturday, 2017-07-15

GraphSQL - 崭新的游戏规则: 一个完整的高效图数据和分析平台(GraphSQL - A Game Changer: A Complete High Performance Graph Data & Analytics Platform)

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): Intermediate

14:50–15:30 Saturday, 2017-07-15

Data service and processing platform for Ads in Ebay

地点：紫金大厅B (Grand Hall B) 观众水平 (Level): 中级 (Intermediate)

16:20–17:00 Saturday, 2017-07-15

人工智能工业应用痛点及解决思路

地点：报告厅 (Auditorium) 观众水平 (Level): Advanced

16:20–17:00 Saturday, 2017-07-15

HDFS纠删码最新探秘

地点：多功能厅2 (Function Room 2) 观众水平 (Level): 中级 (Intermediate)

16:20–17:00 Saturday, 2017-07-15

微软的通用异常检测平台(Common Anomaly Detection Platform at Microsoft)

地点：多功能厅5B + C (Function Room 5B+C) 观众水平 (Level): 非技术性 (Non-technical)

16:20–17:00 Saturday, 2017-07-15

Unified SQL for Big Data on Hadoop

地点：多功能厅6A+B (Function Room 6A+B) 观众水平 (Level): Intermediate

主题演讲



叶杰平 (Ye Jieping)
副总裁, 滴滴出行

叶杰平，滴滴出行研究院副院长，DiDi Fellow，美国密歇根大学终身教授及密歇根大学大数据研究中心的管理委员会成员。2005年美国明尼苏达大学计算机系博士毕业。专业方向为机器学习, 数据挖掘, 以及大数据分析。在机器学习和数据挖掘国际顶级会议及期刊上共发表论文200余篇。曾获KDD和ICML最佳论文奖以及美国国家自然科学基金会生涯奖 (NSF CAREER Award)，并担任多个机器学习和数据挖掘领域顶级会议的主席。现任职机器学习和数据挖掘期刊IEEE TPAMI，DMKD，和 IEEE TKDE的副编委。

大数据在滴滴出行(Big Data at DiDi Chuxing)

Every day, Didi's platform generates over 70TB worth of data, processes more than 9 billion routing requests, and produces over 13 billion location points. In this talk, Ye Jieping will show how AI technologies have been applied

to analyze such big transportation data to improve the travel experience for millions of people in China.



Lukas Biewald
Founder & Chief Data Scientist , CrowdFlower

Lukas Biewald is the founder and Chief Data Scientist of CrowdFlower. Founded in 2009, CrowdFlower is a data enrichment platform that taps into an on-demand workforce to help companies collect training data and do human-in-the-loop machine learning.



Doug Cutting
Chief Architect, Cloudera

Doug Cutting is the chief architect at Cloudera and the founder of numerous successful open source projects, including Lucene, Nutch, Avro, and Hadoop. Doug joined Cloudera from Yahoo, where he was a key member of the team that built and...

周六欢迎致辞

大会日程主席 Jason Dai、Ben Lorica 与 Doug Cutting 致辞开始第二天主题演讲。



Jason (Jinquan) Dai
CTO, Big Data Technologies, Intel

Jason Dai is currently a Senior Principal Engineer and CTO, Big Data Technologies, at Intel. Prior to that, he was a principle architect in Microsoft, responsible for building the large-scale cloud and big data platform that powers some of...

周六欢迎致辞

大会日程主席 Jason Dai、Ben Lorica 与 Doug Cutting 致辞开始第二天主题演讲。



Yuanqing Lin

百度深度学习实验室 (IDL) 主任, Baidu

林元庆，现任百度深度学习实验室 (IDL) 主任，拥有清华大学光学工程硕士学位和宾夕法尼亚大学电气工程博士学位。

林元庆在机器学习和计算机视觉等研究领域拥有多年的研究经验和显著的成果。在加入百度前，曾任NEC美国实验室媒体分析部门主管。在他的带领下NEC研究团队在深度学习、计算机视觉和无人驾驶等领域取得世界领先水平。2005年至今在顶级国际会议和期刊发表论文30余篇，拥有11项美国专利，曾担任NIPS大会领域主席、大规模视觉识别和检索国际研讨会联合主席等。

加入百度后，林元庆致力于带领深度学习实验室研发具有统治级别的人工智能技术，其领导的团队在多个领域实现了技术上重大进展并且应用到百度的多项产品中去，极大地提升了产品的性能以及用户的体验，其带领的团队在多项重要计算机视觉技术在国际测试集上取得世界第一名的好成绩。



Ben Lorica
Chief Data Scientist, O'Reilly Media

Ben Lorica is the chief data scientist at O'Reilly Media. Ben has applied business intelligence, data mining, machine learning, and statistical analysis in a variety of settings, including direct marketing, consumer and market research, targeted advertising, text mining, and financial...

周六欢迎致辞

大会日程主席 Jason Dai、Ben Lorica 与 Doug Cutting 致辞开始第二天主题演讲。



Zhe Zhang
Software Engineer, LinkedIn

Zhe Zhang is an Engineering Manager at LinkedIn where he leads the Core Big Data Services team. The team leverages open source technologies including Hadoop, Spark, TensorFlow, and beyond, to form the storage-compute engine of LinkedIn's big data platform. Zhe...

成长的烦恼-领英大数据平台500倍扩展中应对的挑战 (Growing pains - when your big data platform grows really big)

领英是全球最早应用大数据技术的公司之一。在过去9年的时间里，领英的大数据平台扩展了将近500倍，从20台节点支持10个用户运行MapReduce，到现在超过1万台节点支持几千名工程师和科学家运行从交互式Presto查询到

TensorFlow深度学习的各种大规模数据分析。这个报告会分享领英的大数据平台团队怎样解决大规模和高速增长带来的各种挑战。

会议嘉宾



黄文宇
广发银行股份有限公司
总行数据中心总经理



马洪宾
Kyligence Inc.
高级软件架构师



马晓宇
PingCAP
技术主管



顾佳盛
中国人寿
数据科学家



陈雨强
第四范式
首席研究科学家



费辉
阿里云
资深开发工程师



王玮
中国人寿
大数据项目负责人



王振华
Huawei Technologies
Research Engineer



杨帆
Lenovo
Senior Manager



杨军
阿里巴巴
高级算法专家/算法架构师



李浒
今日头条
软件工程师



李呈祥
万达网络科技集团
大数据技术专家



李元健
百度
软件工程师



张铭
北京大学
教授



吴炜
万达网络研究院
资深研究员



俞育才
intel
大数据架构师



丛宏雷
万达网络科技集团有限公司
资深研究员



Xiaoyong
Microsoft
Program Manager II



Zhe Zhang
LinkedIn
Software Engineer



Xuefu Zhang
Uber
Software Engineer



ximeng zhang
GrowingIO
CEO



张李晔
新智新氨科技
大数据架构师



Pengfei Yue
intel
Senior Engineering Man...



姚舜扬
intel
软件工程师



Yu Xu
GraphSQL
CEO



Qinyan
中国人寿
大数据机器学习项目经理



Tony Xing
Microsoft
Senior Product Manager



吴中
DataVisor
Director of Engineering



Mingxi Wu
GraphSQL Inc.
Vice President of Engine...



Binggang Wo
Cloudera
Solutions Consultant



Yiheng Wang
intel
Software Engineer



王海华
滴滴出行
研发工程师



Daoyuan Wang
intel
软件工程师



Carson Wang
intel
Software Engineer



Andrew Wang
Cloudera
Software Engineer



Daniel Templeton
Cloudera
Software Engineer



Chen Sammi
Intel
Senior Software Engineer



Jiangjie Qin
Linkedin Corp
Staff Software Engineer



Gene Pang
Alluxio
Software Engineer



Ted Malaska
Blizzard Entertainment
Group Technical Architect



Zhenxiao Luo
Uber
Senior Software Engineer



Ben Lorica
O'Reilly Media
Chief Data Scientist



刘轶
eBay
Architect



Shaoshan Liu
PerceptIn
Co-Founder



Zhichao Li
intel
Senior Software Engineer



Haoyuan Li
Alluxio
CEO



Fangshi Li
Linkedin Corp
Software Engineer



Shengsheng Huang
Intel
Big Data Software Archit...



ming huang
腾讯
技术专家



Ron Hu
Huawei Technologies
Database System Architect



Hao Hao
Cloudera
Software Engineer



Yufeng Guo
Google
Developer Advocate



Sijie Guo
Twitter
Staff Software Engineer



Bas Geerdink
ING
IT Manager



Maosong Fu
Twitter
Software Engineer



Darren Fu
eBay
Data Engineer



Bin Fan
Alluxio
Software Engineer



Mateusz Dymczyk
H2O.ai
Software Engineer



Mathieu Dumoulin
MapR Technologies
Data Scientist



Damon Deng
AWS
解决方案架构师



Jason (Jinquan) Dai
intel
CTO, Big Data Technolo...



Doug Cutting
Cloudera
Chief Architect



Jike Chong
YiRenDai/CreditEase
Chief Data Scientist



Feng Cheng
Grab
Data Engineer



Haifeng Chen
intel
Senior Software Architect



Biao Chen
Cloudera
资深架构师



Lukas Biewald
CrowdFlower
Founder & Chief Data S...



蒋守壮
万达网络科技集团有限公司
资深大数据开发工程师



顾荣
南京大学
助理研究员



叶杰平
滴滴研究院
副院长



林元庆
百度深度学习实验室
主任



莫云
宜人贷
数据工程师



Yu Li
Alibaba
Senior Technical Expert



Edwin Law
Grab Taxi
Lead Data Engineer

团购优惠政策（请联系客服）：

如果一个公司：

注册3-5人则享受八折优惠。

6-9人：七五折优惠

10人或10人以上：七折优惠

温馨提示：白金门票已售罄

会议门票如下：

	白金门票 2017年7月12-15日	黄金门票 2017年7月12-15日	白银门票 2017年7月13-15日	青铜门票 2017年7月14-15日
	¥9295 现在注册	¥7303 现在注册	¥5471 现在注册	¥4451 现在注册
2天培训 (周三-周四)	●			
1天培训 (周三)		●		
全部教学辅导课 (2017年7月13日)		●	●	
全部主题演讲&议题演讲 (2017年7月14-15日)	●	●	●	●
展示厅&全部现场社交活动 (2017年7月14-15日)	●	●	●	●
90天Safari Books Online 会员资格	●	●	●	●

请注意：白金门票和黄金门票不包含周四的辅导课；标准折扣不适用。白银门票和青铜门票不包括周三和周四的培训课程。

培训门票

	2天培训门票 2017年7月12-13日	1天培训门票 2017年7月12-13日
	¥8261 现在注册	¥6029 现在注册
2天培训 (周三-周四)	●	
1天培训 (周四)		●
展示厅&全部现场社交活动 (2017年7月14-15日)	●	●
90天Safari Books Online会员资格	●	●

请注意：这些培训门票不包括周四的辅导课。标准折扣不适用。

取消和转让政策

如果您必须取消一定要在June 14, 2017之前书面通知我们。请联系我们。会议开始前30天之内取消是不退款的。June 28, 2017之前您可以把注册转让给其他人，一定要发送授权。确认并完成支付后取消的参会者、或者超过截止日期取消的参会者要承担全部会议费用。遇到极端情况该会议取消O'Reilly Media, Inc.的责任仅限于退回支付的注册款项。

行为守则

所有参会者必须遵守我们的行为守则，其核心想法是：O'Reilly会议对每个人都应该是一个安全、富有成效的环境。

母婴室

会议现场会提供一个附近私密空间方便母亲和孩子哺乳等。

摄影 & 视频

我们的目的是捕捉会议中激动人心的时刻，您可能看到一些摄影师，包括我们请来的摄影师，来记录本次活动。我们拍摄的照片和视频可能会在网站上发布，也可能在未来的市场宣传中使用。

请注意会议中是不允许拍摄视频的。

隐私政策

根据各自的隐私政策注册人联系信息将会被会议主办方分享和使用——O'Reilly/Strata Conference 以及 Cloudera/Hadoop World。

