

上海Spark Meetup第十次聚会

Shanghai Apache Spark Meetup第十次聚会活动将于2016年09月10日12:30 至 17:20在四星级的上海通茂大酒店(浦东新区陆家嘴金融区松林路357号)。距离地铁2、4、6、9号线的世纪大道站1000米，距离地铁4号线浦电路站740米。本次活动分别请到了运营商和高校系统的讲师来分享经验，主题覆盖了从研发到应用的各种不同视角，希望给大家带来耳目一新的感觉。

会议主题

1、中国电信在大数据领域上的创新与探索

演讲嘉宾：丁廷鹤，中国电信股份有限公司信息化研发中心大数据部门数据挖掘组负责人。

主题简介

：中国电信利用流式计算、图计算、机器学习、深度学习等时下热门的分析技术与分布式框架，针对电信集团大数据场景下的DPI、话单、ODS数据，开发出的一系列具有价值的创新应用，包括社交网络、意图识别、用户画像、骚扰电话等等，并探讨我们未来的在深度学习架构体系、业务盈利上面的着力点。

2、函数式编程与RDD

演讲嘉宾

：邢森，上海理想公司大数据团队资深研发工程师，参与构建了电信运营商大规模分布式计算平台，并长期专注于大数据及云计算相关领域。

主题简介

：分析函数式语言的思想和本原理，以及函数式语言在大数据中的应用。涉及mapreduce、spark、Scala。

3、社交网络中的信息传播

演讲嘉宾

：杨洋，浙江大学计算机学院讲师，于2016年毕业于清华大学计算机系PHD，研究兴趣包括社会网络分析、数据挖掘、机器学习等。曾访问美国康奈尔大学（与图灵奖获得者John Hopcroft合作）、比利时鲁汶大学。博士期间共计发表论文12篇，其中以第一作者发表CCF A类论文5篇。获清华大学优秀博士毕业生、北京市优秀博士毕业生、国家奖学金等荣誉。曾担任KDD'15分会主席，并担任ICDM'16、ICWSM'15、ASONAM'15等国际学术会议的程序委员会成员。

主题简介

：在线社交网络的兴起彻底改变了人们沟通交流、分享信息的渠道与方式。理解网络中信息的传播过程对国家安全有着重要的意义，同时也存在着巨大的市场价值。然而，信息传播的机理并不完全清晰，有待深入研究。一方面，不同用户在传播过程中所扮演的角色和发挥的作用难以量化，而另一方面，个体行为与传播的宏观现象之间的关联机理也尚不明确。在这个报告中，我将通

过微观和宏观两个维度，研究社交网络中信息传播的形成机理。主要包括：在微观层面，研究用户的社会角色对信息传播过程的影响力，提出社会角色感知的个体传播行为模型；在宏观层面，对信息传播规模与用户状态的动态演化过程进行联合建模，该模型可以在理论上推导出服从幂定律分布的传播规模；最后，探讨用户个体行为与传播规模长尾效应间之间的关联关系，将微观与宏观层面的信息传播结合在一个统一的框架之中。

4、大数据分析和机器学习

演讲嘉宾：Jason Dai，现任英特尔大数据首席架构师、资深首席工程师，负责领导英特尔全球大数据分析技术和平台的研发工作。在此之前，他曾担任英特尔软件和服务事业部技术总监、首席工程师，负责领导英特尔大数据的技术研发，包括和加州大学伯克利分校合作开发下一代大数据技术（如Apache Spark项目），以及为国内外多个大型互联网公司开发部署下一代大数据平台。他也在微软负责研发大规模云计算和大数据平台，以支撑其多个大型互联网和云计算服务。他在大数据、机器学习、云计算以及并行计算方面是世界公认的专家；他是Apache Spark项目的项目管理委员会（PMC）成员，Strata + Hadoop World北京大会联席主席，在国际学术会议和杂志上发表了超过15 多篇论文，拥有或申请超过20 多项专利，并曾在一些顶级大学教授计算机课程。

5、分布式流式数据处理框架：功能对比以及性能评估

演讲嘉宾

：毛玮，Intel大数据团队资深工程师，开源社区贡献者，主要专注于大数据流式数据处理框架的研究和应用；王华峰：目前任职于英特尔亚太研发中心大数据部门，是英特尔开源流处理系统Gearpump的开发者，长期关注大数据领域和分布式计算，专注于流处理系统的开发和研究。

主题简介

：从流处理的核心概念，到功能的完备性，再到周边的生态环境，全方位对比了目前比较热门的流处理框架：Spark，Flink,Storm和Gearpump。结合不同的框架的设计，为大家进行深入的剖析。与此同时，从吞吐量和延时两个方面，对各个框架进行性能评估。主要技术点：流式数据处理，Spark，Flink,Storm和Gearpump。

日程表

12:30 - 13:00 签到，签署，进入会场 / sign in, entry

13:00 - 13:20 开场白 / Opening

13:20 - 13:30 主题发言 / Keynote

13:30 - 14:10 技术演讲 / session - 《中国电信在大数据领域上的创新与探索》

14:10 - 14:50 技术演讲 / session - 《函数式编程与RDD》

14:50 - 15:30 技术演讲 / session - 《社交网络中的信息传播》

15:30 – 15:50 茶歇 / Tea Break

15:50 – 16:10 英特尔技术演讲开场白/Intel session opening - 《大数据分析和机器学习》

16:10 – 16:50技术演讲 / session - 《分布式流式数据处理框架：功能对比以及性能评估》

16:50 – 17:20 活动收尾及社交/Ending and Social

地图与交通信息

上海浦东新区陆家嘴金融区松林路357号-上海通茂大酒店

报名地址

本活动可以免费报名：[点我免费报名](#)

资料下载

[点击下载PPT](#)

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。
本文链接: **【】**（**）**