

## 上海(Shanghai) Apache Spark Meetup第十四次聚会

第十四次Shanghai Apache Spark Meetup聚会，由中国平安银行大力支持。活动将于2017年12月23日12:30~17:00在上海浦东新区上海海神诺富特酒店三楼麦哲伦厅举行。举办地点交通方便，靠近地铁4号线浦东大道站。座位有限，先到先得。

### 大会主题

- 《Spark在金融领域的算法实践》(13:20 - 14:05)  
演讲嘉宾：  
：潘鹏举,平安银行大数据平台架构师，曾供职于携程、尼尔森等公司担任BI经理、算法专家、数据分析专家等职位，专注于机器学习算法、深度学习、AI在业务的实践，在数据应用领域有丰富的实践经验。  
主题简介：  
：基于数据的应用是公司关注的命题。面对海量数据的建模，如何快速迭代建模，同时部署应用也是大家关注的重点。在金融领域，算法有哪些切入点来帮助业务提升。此次分享会对这些问题进行探讨，并分享在应用实践中碰到的一些问题和踩过的坑。
- 《Data layout in Spark》(14:05 - 14:50)  
演讲嘉宾：范文臣, Databricks 软件工程师，Apache Spark PMC Member，主要研究方向为Spark SQL 的研发。  
主题简介：当使用 Spark 来分析数据的时候，如何快速地访问数据是一个关键问题。本次演讲中文臣会给大家讲解 data layout 是如何影响 Spark 程序的性能，以及如何根据使用的 workload 来选择合适的 data layout。
- 《Spark中的压缩编解码最佳实践》(15:05 - 15:50)  
演讲嘉宾：孙晓霞(Sophia) 英特尔大数据部门软件工程师，专注于大数据平台spark上 workload的性能分析和调优，对于大数据基准测试TPC-DS, TPCx-BB, HiBench等workload在大规模集群上的有丰富的调优经验。  
主题简介：  
：面对海量数据的当代，数据的有效压缩，对于存储、I/O和网络数据传输意义重大。然而，解压数据所带来的额外CPU开销是大数据分析不可忽视的要素，如何在数据压缩率和编解码效率之间找到平衡，在工程方面是一个值得探讨的话题。Sophia将分享Intel大数据团队在为Spark实现 ISA-L(iGZIP), LZ4-IPP, ZLIB-IPP 和 ZSTD等针对IA硬件平台优化的算法支持，并使用基准测试集(TPC-DS/HiBench)对这些压缩编解码在Spark上的性能表现进行的详细分析和对比。
- 《基于Spark和Cloudera数据科学家工作平台的分布式机器学习实践》(15:50 - 16:35)  
演讲嘉宾：苏昌锋 (Kevin)，Cloudera资深售前顾问，曾在Oracle、Teradata、HP等公司从事研发、架构和咨询工作，主要专注与BI、大数据领域的应用落地。  
主题简介：本次演讲介绍基于Spark以及Cloudera Data Science Workbench的自动化数据科学工作的过程，并介绍通过BigDL深度学习算法包进行文本数据的分词和NLP算法处理来通过CDSW来简化各部门的协同来加速“从算法到结果”的周期。

### 日程表

- 12:30 – 13:00 签到，签署，进入会场 / sign in, entry
- 13:00 – 13:20 主题发言 / Keynote
- 13:20 – 14:05 技术演讲 / session - 《Spark在金融领域的算法实践》
- 14:05 – 14:50 技术演讲 / session - 《Data layout in Spark》
- 14:50 – 15:05 茶歇 / Tea Break
- 15:05 – 15:50 技术演讲 / session - 《Spark中的压缩编解码最佳实践》
- 15:50 – 16:35 技术演讲 / session - 《基于Spark和Cloudera数据科学家工作平台的分布式机器学习实践》
- 16:35 – 17:00 活动收尾及社交/Ending and Social

## 地图与交通

上海浦东新区上海海神诺富特酒店三楼麦哲伦厅

## 报名地址

本活动免费报名，报名截止时间：2017-12-22

10:00:00，报名地址：[http://huiyi.csdn.net/activity/product/goods\\_list?project\\_id=3730](http://huiyi.csdn.net/activity/product/goods_list?project_id=3730)

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：[【】（）](#)