

上海Spark Meetup第六次聚会

上海Spark Meetup第六次聚会将于2015年08月08日下午1:30 PM to 5:00 PM在上海市场浦云计算创新基地发展有限公司举办，详细地址上海市杨浦区伟德路6号云海大厦13楼。本次聚会由Intel举办。

大会主题

主讲题目：Tachyon: 内存为中心可容错的分布式存储系统

摘要：在越来越多的大数据应用场景诸如机器学习,数据分析等,内存成为保证和提升性能的关键. Tachyon正是以内存为中心而构建的可容错的分布式存储系统, Tachyon全局管理和调度服务器集群的内存资源并通过lineage来实现内存数据丢失的容错. 通过Tachyon, 不同的计算平台如Hadoop MapReduce, Spark可以充分利用内存的速度优势来存取数据, 并无缝实现数据的共享.

Tachyon项目诞生于UC Berkeley AMPLab的研究项目并迅速开源. 在仅仅两年多的时间里. Tachyon项目取得了飞速的成长. 系统被超过50家公司部署并应用. 目前社区里有100名以上的开发者, 来自超过30家不同的学校, 机构和公司(包括Baidu, IBM, Intel, Yahoo等).

嘉宾简介：范斌博士就职于Tachyon Nexus, 是目前Tachyon开源项目的核心开发者之一. 加入Tachyon Nexus之前范斌曾于谷歌工作, 从事下一代大规模分布式存储系统的研究与开发, 并因此获得谷歌Technical Infrastructure Award. 范斌曾在卡耐基梅陇大学(Carnegie Mellon University)攻读并获得计算机博士学位, 有多篇学术论文发表在SIGCOMM, SOSP, NSDI等网络和分布式系统方向的顶级计算机学术会议.

主讲题目：Spark与SequoiaDB联手打造互联网金融大数据平台

摘要：分享我们使用Spark在金融行业的一些应用场景，以及我们和Spark对接的关键技术包括NoSQL数据库和Spark结合的方式与意义，并且结合我们在一家金融企业中历史数据平台项目的场景来展开讨论。

嘉宾简介：王涛，SequoiaDB巨杉数据库联合创始人，他曾就职于IBM多伦多实验室IBM TorontoLab (DB2 UDB Development Lab)，曾经是DB2领域的专家，作为IBM DB2全球最高技术专家小组的成员，参与IBM下一代大数据平台的架构规划，精通数据库内核及体系结构。在IBM多伦多实验室工作八年后，王涛选择回国创业。目前担任巨杉数据库CTO及总架构师，成功研发了国内唯一一款开源新一代分布式数据库——SequoiaDB（巨杉数据库）。作为公司创始人之一，王涛参与公司发展战略规划，负责整个数据库研发项目的管理，包括数据库系统的构建，管理公司的整体核心技术，组织制定和实施重大技术决策和技术方案等。目前，公司产品已经开源，客户遍布互联网、金融、电信等各大行业。

主讲题目：Spark DataFrames: 更高效的结构化数据分析

摘要：DataFrame是Spark自1.3.0起为结构化数据分析加入的新API。DataFrame API简单易用，而且让用户可以充分利用Catalyst模块为执行计划进行优化，高效地进行数据分析。同时，对于常规操作，通过DataFrame API用户可以利用Spark核心引擎的一些专有数据结构进一步提高性能表现。本报告将讨论DataFrame API背后的技术细节与常见的使用方式。

嘉宾简介：王道远，英特尔亚太研发有限公司大数据团队软件工程师，Hive/Spark开源项目贡献者，目前专注于Spark SQL模块开发。在进入Spark社区前，他主要负责为IDH中的Hive开发SQL92支持。同时还是《Spark快速大数据分析》一书译者。

日程表

1. 13:00 -13:30 checkin
2. 13:30 - 13:40 Opening
3. 13:40 - 14:30 Tachyon: 内存为中心可容错的分布式存储系统 -- 范斌 @Tachyon Nexus
4. 14:30 - 15:20 Spark与SequoiaDB联手打造互联网金融大数据平台 -- 王涛 @ SequoiaDB巨杉数据库
5. 15:20 - 15:40 茶歇
6. 15:40 - 16:30 Spark DataFrames: 更高效的结构化数据分析 -- 王道远 @ Intel
7. 17:00 end

地图与交通信息

上海市杨浦区伟德路6号云海大厦13楼（上海市杨浦云计算创新基地发展有限公司）

报名地址

<http://www.meetup.com/Shanghai-Apache-Spark-Meetup/events/224628824/>

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：[【】（）](#)