

上海大数据流处理(Big Data Streaming)资料分享

第二期上海大数据流处理(Shanghai Big Data Streaming 2nd Meetup)于2015年12月6日下午12:45在上海世贸大厦22层英特尔(中国)有限公司延安西路2299号进行，分享的主题如下：

一、演讲者1/Speaker 1: 张天伦 英特尔大数据组软件工程师

个人介绍/BIO: 英特尔开源流处理系统Gearpump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。

分享主题/Topic: Storm over Gearpump

概要/Abstract: Storm是目前业界广泛使用的流处理方案，有大量的流处理应用运行在Storm上。为了能让现有Storm用户零成本地迁移至Gearpump，我们实现了对Storm应用的透明兼容，用户无需修改代码重新编译，即可将已有的应用运行在Gearpump上。本此演讲将重点介绍Gearpump兼容Storm应用的实现原理，功能支持及性能数据。

二、演讲者2/Speaker 2: 王新春 大众点评数据中心架构师

个人介绍

/BIO：2013年1月加入大众点评，在数据中心团队主要负责实时计算平台、数据平台工具链相关系统（开发者平台、数据传输、在线服务等），推动包括Storm等实时计算框架在点评的发展。著有《Storm技术内幕与大数据实践》一书。

分享主题/Topic: Storm计算平台在点评的实践

概要/Abstract: 目前Storm在点评的业务场景包括流量实时化、广告、推荐和安全等领域，每天处理的数据超过百亿条。分享的主要内容包括：

- 1) Storm计算平台的构建
- 2) Storm使用的业务场景
- 3) 使用经验和教训

三、演讲者3/Speaker 3: 程浩 Intel大数据技术团队工程师

个人介绍/BIO: Spark开源社区活跃开发者，专注于Spark SQL组件开发和性能优化，合著有《Spark大数据处理技术》一书。

分享主题/Topic: StreamingSQL on Spark

概要/Abstract: SQL和Streaming是Spark生态中使用最多的组件，但是两个组件本身几乎没

有交集，如何使用SQL的方式来无缝集成Spark流式和批处理数据操作，是很多用户迫切需要的功能，英特尔大数据计算团队的开源项目StreamingSQL正是为这个目的而生，本次分享将介绍该项目的使用方式，揭秘它的设计原理以及对比其它Streaming SQL产品。

四、演讲者4/Speaker 4: Todd Lipcon, Cloudera明星工程师

个人介绍/Bio: Todd是社区里面神一样的人物, 他是Apache基金会的成员, Hadoop PMC和Committer, HBase PMC和Committer.

他主导设计和开发了Hadoop的几个重量级功能，包括highly-available metadata journaling (QJM) and automatic failover for HDFS. 自2012年起，开始在Cloudera领导Kudu项目。

分享主题/Topic : Fast Analytics on fast data

概要/Abstract:

This session will talk about how storage layer impact real time computing. We will investigate the trade-offs between real-time transactional access and fast analytic performance in Hadoop from the perspective of storage engine internals. We will also discuss recent advances, evaluate benchmark results from current generation Hadoop technologies, and propose potential ways ahead for the Hadoop ecosystem to conquer its newest set of challenges.

资料下载

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: [【】（）](#)