

# Apache CarbonData & Spark Meetup

---

12月1日 09:00 - 12:10

深圳市科兴科学园会议中心 2 号会议室

联合主办:



技术社区:



# Apache CarbonData & Spark Meetup



# Apache CarbonData & Spark 历程

- Apache CarbonData Beijing Meetup (2016.10.29)
- Spark Summit East (2017.1.7-9, Boston)
- Apache CarbonData Shanghai Meetup (2017.9.2)
- Realtime BigData Beijing Meetup(2018.9.8)
- OSI conference, bangalore(2018.10.17)
- Tech triveni conference, Delhi(2018.11.18)
- Streaming Analytics Meetup, bangalore (2018.11.24)
- **Apache CarbonData & Spark Shenzhen Meetup(2018.12.1)**

# Apache CarbonData & Spark Meetup

时间	主题	演讲嘉宾
9:00-9:40	What's New in Apache Spark 2.4?	李潇, Databricks Tech Lead Manager, Apache Spark PMC member, 佛罗里达大学博士
9:40-10:20	Apache CarbonData技术介绍与行业实践	李昆, 华为大数据平台架构师, Apache CarbonData PMC成员
10:20-10:35	短休交流	
10:35-11:15	QATCodec: Past, Present, and Future	徐铖, Intel Engineering Manager, Apache Hive、Commons、ORC committer
11:15-11:55	Apache CarbonData的使用实践与性能调优	徐传印, 华为某产品大数据平台系统工程师, Apache CarbonData Committer
11:55-12:10	合影留念	

# Apache CarbonData & Spark Meetup

李潇

*Databricks Tech Lead Manager*  
*Apache Spark PMC member*  
佛罗里达大学博士



## What's New in Apache Spark 2.4?

This talk will provide an overview of the major features and enhancements in Spark 2.4 release and the upcoming releases and will be followed by a Q&A session. The Apache Spark 2.4 comes packed with a lot of new functionalities: new barrier execution mode, flexible streaming sink, the native AVRO data source, PySpark's eager evaluation mode, Kubernetes support, higher-order functions, Scala 2.12 support and a lot of other improvements.

# Apache CarbonData & Spark Meetup

**李昆**

华为大数据平台架构师  
Apache CarbonData PMC成员



---

## Apache CarbonData 技术介绍与行业实践

针对海量明细数据的分析场景，全方位介绍 CarbonData 的架构、技术原理、适用场景以及最佳实践。

## New Apache CarbonData Contributors

Contributors	Name	JIRA	PR
微构科技魏基勇	Fixed the parent version error in MV module	CARBONDATA-3120	<a href="https://github.com/apache/carbondata/pull/2943">https://github.com/apache/carbondata/pull/2943</a>
	Update the document to add spark 2.3.2 and datamap mv compiling method.	CARBONDATA-3133	<a href="https://github.com/apache/carbondata/pull/2955">https://github.com/apache/carbondata/pull/2955</a>
陆金所蒋晓峰	Fix the HiveExample exception	CARBONDATA-3128	<a href="https://github.com/apache/carbondata/pull/2954">https://github.com/apache/carbondata/pull/2954</a>
中科院张源	Fixing the getOrCreateCarbonSession method parameter to an empty string	CARBONDATA-3119	<a href="https://github.com/apache/carbondata/pull/2961">https://github.com/apache/carbondata/pull/2961</a>
孔嘉良	Fix some error links of Readme	CARBONDATA-3111	<a href="https://github.com/apache/carbondata/pull/2934">https://github.com/apache/carbondata/pull/2934</a>
微构科技魏基勇	Carbondata流式入库集成Flink	CARBONDATA-3130	TODO
陆金所蒋晓峰			TODO
信雅达张少华	Add presto/Hive/ doc in parent readme		TODO

# 活动

- 2018.11.15-12.31 参与Apache CarbonData贡献的新贡献者，可以获得华为运动手环/小天鹅音响等奖品。
- Bug/特性/案例等

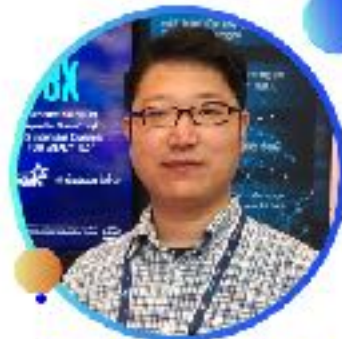




# Apache CarbonData & Spark Meetup

徐铖

*Intel Engineering Manager  
Apache Hive, Commons,  
ORC committer*



## QATCodec: Past, Present, and Future

随着业务的持续增长, 针对计算密集型和高I/O的大数据解决项目利用了 Intel QuickAssist 技术来使用额外的硬件加速器来 offload CPU 压缩计算并进行加速。在这个演讲中, 我们将介绍一些核心的使用 QATCodec 来加速大数据计算的场景。

# Apache CarbonData & Spark Meetup

## 徐传印

华为SmartCare大数据平台系统工程师  
Apache CarbonData Committer



## Apache CarbonData的使用实践与性能调优

2018年以来, CarbonData发布了多个稳定版本 (1.3.x, 1.4.x, 1.5.x), 引入了大量新特性, 同时性能进一步得到提升。如何理解和使用好这些特性, 发挥 CarbonData 的特长是摆在 CarbonData 使用者面前的一道难题。目前 CarbonData 已经在 Huawei SmartCare 产品中作为详单查询引擎处理超过 100TB/天的数据, 并且在将来会进一步作为数据仓库来处理汇总和报表相关的业务。本议题以自己项目组在数据平台中使用 CarbonData 的经历, 讲解如何使用和调优 CarbonData。

# 活动

- 2018.11.15-12.31 参与Apache CarbonData贡献的新贡献者，可以获得华为运动手环/小天鹅音响等奖品。
- Bug/特性/案例等

