


TensorFlow遇上Spark



RayCloud

(/u/49d1f3b7049e)

+关注

2017.02.21 07:07*

字数 1967

阅读 9642

评论 6

喜欢 43

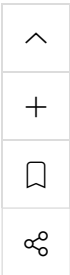
(/u/49d1f3b7049e)

TensorFlowOnSpark 项目是由 Yahoo 开源的一个软件包，实现 TensorFlow 集群服务部署在 Spark 平台之上。

大家好，这次我将分享 TensorFlow On Spark 的解决方案，将 TensorFlow 集群部署在 Spark 平台之上，实现了 TensorFlow 与 Spark 的无缝连接，更好地解决了两者数据传递的问题。



这次分享的主要内容包括 TensorFlowOnSpark 架构设计，探讨其工作原理，通过理解其设计，更好地理解 TensorFlow 集群在 Spark 平台上的运行机制。

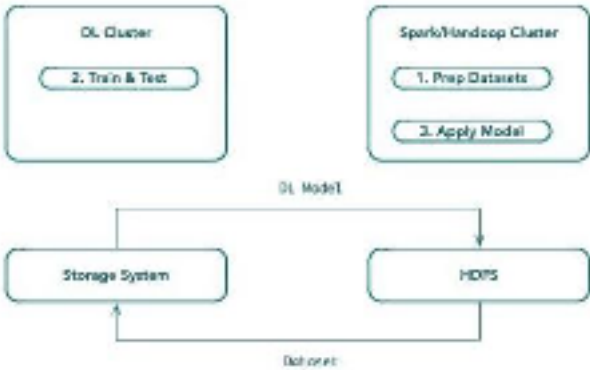




tfos.part3.3_3.pdf.jpg

首先，探讨 TensorFlowOnSpark 的架构与设计。主要包括如下两个基本内容：

- 架构分析
- 生命周期



tfos.part4.4_4.pdf.jpg

在开始之前，先探讨一下 TensorFlowOnSpark 的背景，及其它需要解决的问题。为了实现 Spark 利用 TensorFlow 深度学习，及其 GPU 加速的能力，最常见的解决方案如上图所示。

搭建 TensorFlow 集群，并通过利用既有的 Spark 集群的数据完成模型的训练，最后再将训练好的模型部署在 Spark 集群上，实现数据的预测。

该方案虽然实现了 Spark 集群的深度学习，及其 GPU 加速的能力，但需要 Spark 集群与 TensorFlow 集群之间的数据传递，造成冗余的系统复杂度。

^

+

□

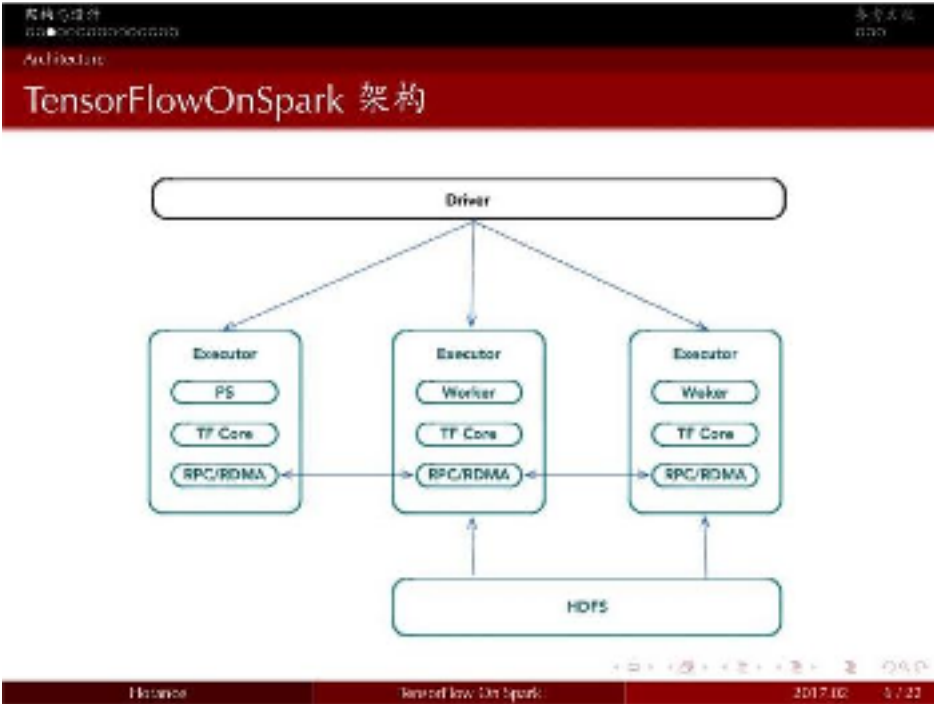
⌂



tfos.part5.5_5.pdf.jpg

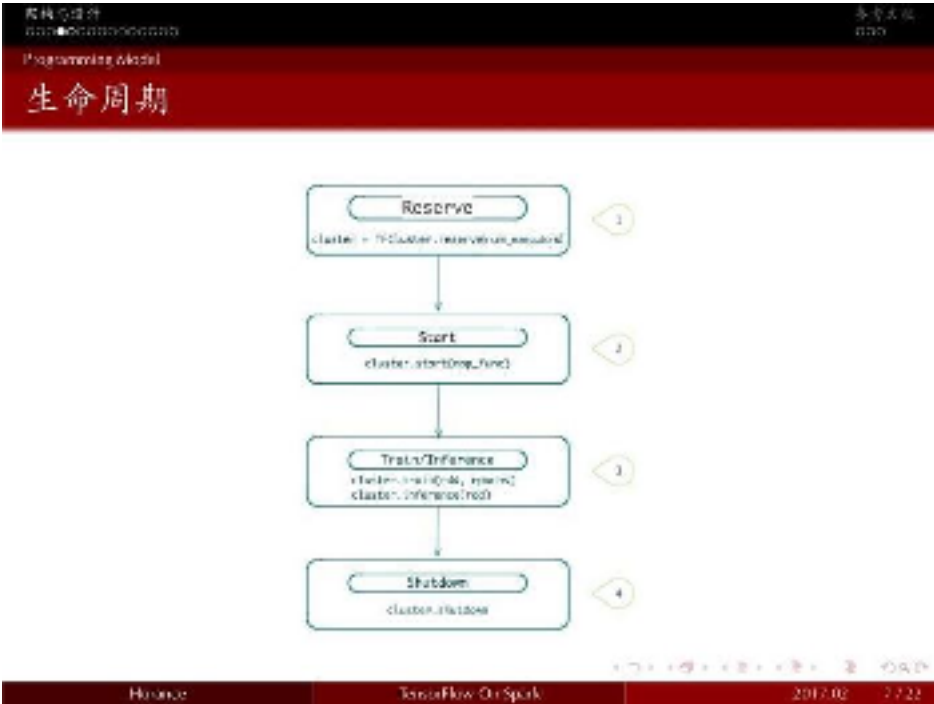
很容易想到，可以将 TensorFlow 集群部署在 Spark 之上，用于解决集群间数据传递的问题。

依次类同，该方案可实现 Caffe 部署在 Spark 集群之上，实现 Spark 集群对多种深度学习框架的支持能力，并兼容既有 Spark 组件的完整性，包括 Spark MLlib, Spark Streaming, Spark SQL 等。



tfos.part6.6_6.pdf.jpg

TensorFlowOnSpark 的架构较为简单，Spark Driver 程序并不会参与 TensorFlow 内部相关的计算和处理。其设计思路像是将一个 TensorFlow 集群运行在了 Spark 上，其在每个 Spark Executor 中启动 TensorFlow 应用程序，然后通过 gRPC 或 RDMA 方式进行数据传递与交互。



tfos.part7.7_7.pdf.jpg

TensorFlowOnSpark 的 Spark 应用程序包括 4 个基本过程。

- Reserve：组建 TensorFlow 集群，并在每个 Executor 进程上预留监听端口，启动“数据/控制”消息的监听程序。
- Start：在每个 Executor 进程上启动 TensorFlow 应用程序；
- Train/Inference：在 TensorFlow 集群上完成模型的训练或推理
- Shutdown：关闭 Executor 进程上的 TensorFlow 应用程序，释放相应的系统资源(消息队列)。

tfos.part8.8_8.pdf.jpg

用户直接通过 spark-submit 的方式提交 Spark 应用程序(mnist_spark.py)。其中通过 --py_files 选项附带 TensorFlowOnSpark 框架(tfspark.zip)，及其 TensorFlow 应用程序(mnist_dist.py)，从而实现 TensorFlow 集群在 Spark 平台上的部署。



tfos.part9.9_9.pdf.jpg

首先看看 TensorFlow 集群的建立过程。首先根据 spark-submit 传递的 num_executor 参数，通过调用 `cluster = sc.parallelize(num_executor)` 建立一个 `ParallelCollectionRDD`，其中分区数为 num_executor。也就是说，此时分区数等于 Executor 数。

然后再调用 `cluster.mapPartitions(TFParkNode.reserve)` 将 `ParallelCollectionRDD` 变换 (transformation) 为 `MapPartitionsRDD`，在每个分区上回调 `TRSparkNode.reserve`。

`TRSparkNode.reserve` 将会在该节点上预留一个端口，并驻留一个 Manager 服务。Manager 持有一个队列，用于完成进程间的同步，实现该节点的“数据/控制”消息的服务。

数据消息启动了两个队列：Input 与 Output，分别用于 RDD 与 Executor 进程之间的数据交换。

控制消息启动了一个队列：Control，用于 Driver 进程控制 PS 任务的生命周期，当模型训练完成之后，通过 Driver 发送 Stop 的控制消息结束 PS 任务。

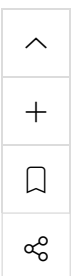
tfos.part10.10_10.pdf.jpg

这是从分区的角度看待 TensorFlow 集群建立的过程，横轴表示 RDD。这里存在两个 RDD，第一个为 `ParallelCollectionRDD`，然后变换为 `MapPartitionsRDD`。

纵轴表示同一个分区(Partition)，并在每个分区上启动一个 Executor 进程。在 Spark 中，分区数等于最终在 `TaskScheduler` 上调度的 Task 数目。

此处，`sc.parallelize(num_executor)` 生成一个分区数为 num_executor 的 `ParallelCollectionRDD`。也就是说，此时分区数等于 num_executor 数目。

在本例中，num_executor 为 3，包括 1 个 PS 任务，2 个 Worker 任务。



tfos.part11.11_11.pdf.jpg

TensorFlow 集群建立后，将生成上图所示的领域模型。其中，一个 TFCluster 将持有 num_executor 个 TFSparkNode 节点；在每个 TFSparkNode 上驻留一个 Manager 服务，并预留一个监听端口，用于监听“数据/控制”消息。

实际上，TFSparkNode 节点承载于 Spark Executor 进程之上。



- 局部修改 TF 程序：少于 10 行
- ClusterSpec：TensorFlow 任务列表
- 启动 Server：每个 Task 启动一个 Server



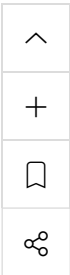
tfos.part12.12_12.pdf.jpg

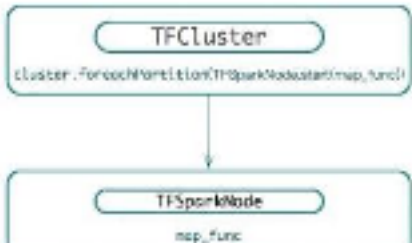
TensorFlow 集群建立后，通过调用 cluster.start 启动集群服务。其结果将在每个 Executor 进程上启动 TensorFlow 应用程序。

此处，需要对原生的 TensorFlow 应用程序进行适配修改，包括 2 个部分：

- Feeding 与 Fetching：数据输入/输出机制修改
- ClusterSpec：TF 集群的构造描述

其余代码都将保留，最小化 TensorFlow 应用程序的修改。





tfos.part13.13_13.pdf.jpg

在 cluster 上调用 foreachPartition(TFSparkNode.start(map_func))，将在每个分区(Executor 进程)上回调 TFSparkNode.start(map_func)。其中，map_func 是对应 TF 应用程序的包装。

通过上述过程，将在 Spark 上拉起了一个 TF 的集群服务。从而使得 Spark 集群拥有了深度学习 和 GPU 加速的能力。



- TensorFlow QueueRunner: FileReaders & QueueRunner
- Spark Feeding: RDD -> Executor -> TensorFlow Graph

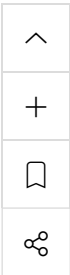


tfos.part14.14_14.pdf.jpg

当 Spark 平台上已经拉起了 TF 集群服务之后，便可以启动模型的训练或推理过程了。在训练或推理过程中，最重要的是解决数据的 Feeding 和 Fetching 问题。

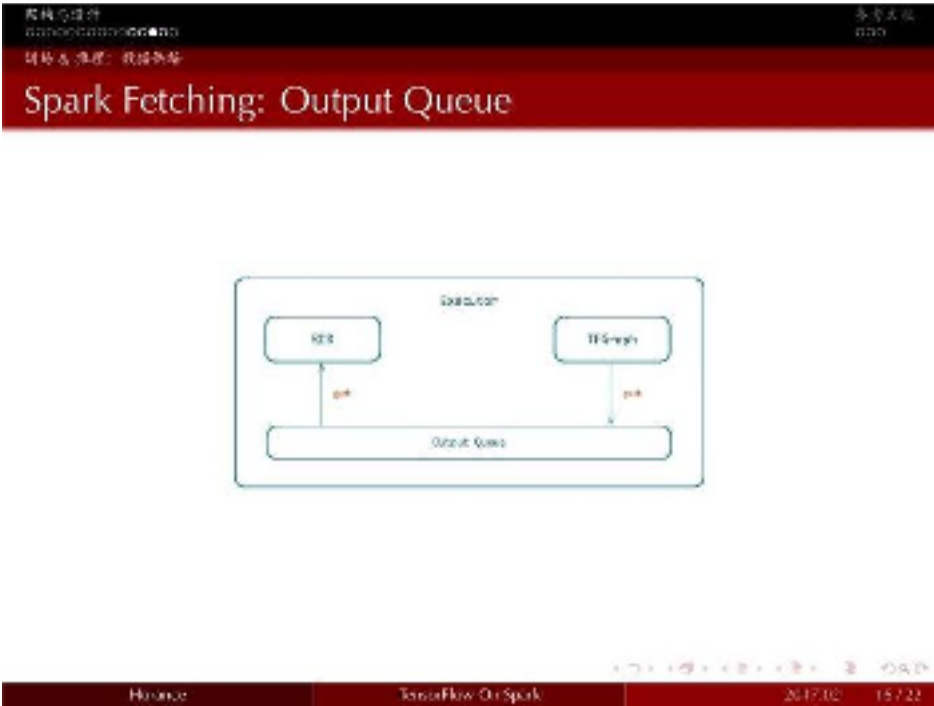
TFoS 上提供了两种方案：

- TensorFlow QueueRunner：利用 TensorFlow 提供的 FileReader 和 QueueRunner 机制。Spark 未参与任何工作，请查阅 TensorFlow 官方相关文档。
- Spark Feeding：首先从 RDD 读取分区数据(通过 HadoopRDD.compute)，然后将其放在 Input 队列中，Executor 进程再从该队列中取出，并进一步通过 feed_dict，调用 session.run 将分区数据供给给 TensorFlow Graph 中。



tfos.part15.15_15.pdf.jpg

Feeding 过程，就是通过 Input Queue 同步实现的。当 RDD 读取分区数据后，阻塞式地将分区数据 put 到 Input 队列中；TFGraph 在 session.run 获取 Next Batch 时，也是阻塞式地等待数据的到来。



tfos.part16.16_16.pdf.jpg

同样的道理， Fetching 过程与 Feeding 过程类同，只是使用 Output Queue，并且数据流方向相反。

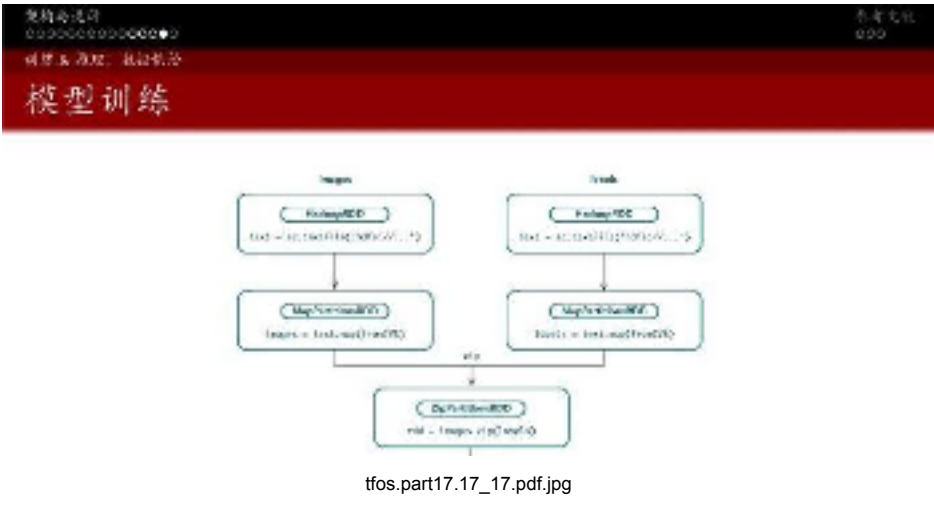
session.run 返回的数据，通过 put 阻塞式地放入 Output Queue，RDD 也是阻塞式地等待数据到来。

^

+

🔖

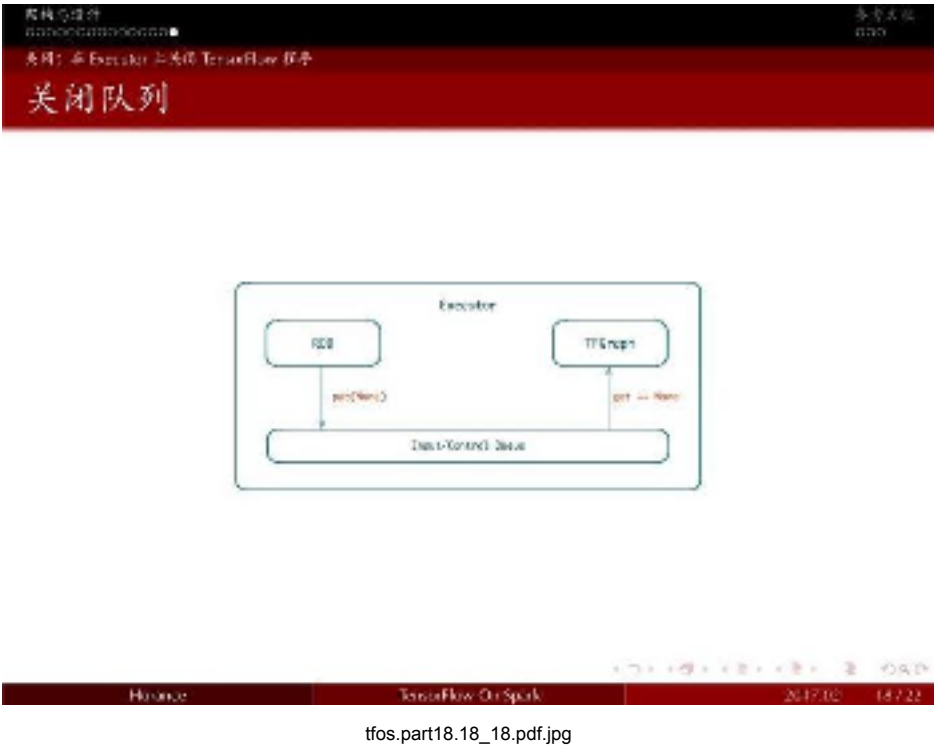
🔗



以模型训练过程为例，讲解 RDD 的变换过程。此处以 Mnist 手写识别为例，左边表示 x，右边表示 y。分别通过 HadoopRDD 读取分区数据，然后通过 MapPartitionsRDD 变换分区的数据格式；然后通过 zip 算子，实现两个 RDD 的折叠，生成 ZipPartitionsRDD。

然后，根据 Epochs 超级参数的配置，将该 RDD 重复执行 Epochs 次，最终将结果汇总，生成 UnionRDD。

在此之前，都是 Transformation 的过程，最终调用 foreachPartition(train) 启动 Action，触发 Spark Job 的提交和任务的运行。



当模型训练或推理完成之后，分别在 Input/Control 队列中投掷 Stop (以传递 None 实现)消息，当 Manager 收到 Stop 消息后，停止队列的运行。

最终，Spark 应用程序退出，Executor 进程退出，整个工作流执行结束。



参考文献

tfos.part19.19_19.pdf.jpg



- tensorflow.org, Google Inc.
- TensorFlowOnSpark, Yahoo Inc.



tfos.part20.20_20.pdf.jpg

推荐资料，强烈推荐直接地源代码阅读；最后欢迎大家关注我的简书。

^

+

🔖

🔗



tfos.part22.22_22.pdf.jpg

小礼物走一走，来简书关注我

赞赏支持

软件匠艺 (/nb/4653592) 举报文章 © 著作权归作者所有

RayCloud (/u/49d1f3b7049e) ♂

写了 70958 字，被 854 人关注，获得了 565 个喜欢

(/u/49d1f3b7049e)

攻城狮

+ 关注

喜欢 | 43

更多分享

(http://cwb.assets.jianshu.io/notes/images/937598f

写下你的评论...

6条评论 只看作者 按喜欢排序 按时间正序 按时间倒序

FishSeeker (/u/d588a2d0e86e)

3楼 · 2017.04.10 19:51

(/u/d588a2d0e86e)

lz写得太好了，可否分享一下ppt学习一下 1817520086@qq.com

1人赞 回复

迷茫在拉格朗日 (/u/5da99902cd82)： 同求：2541224507@qq.com

2017.08.17 19:59 回复

添加新评论

罗成罗群 (/u/ecffae6a9625)

2楼 · 2017.03.06 09:03

(/u/ecffae6a9625)

好

赞 回复

EDWIN (/u/9132ba692cc9)

4楼 · 2017.06.23 16:29

(/u/9132ba692cc9)

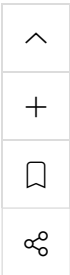
1. 是否是每个tensorflow进程只训练一个分区的数据？

2. 是的话怎么同步训练结果？

赞 回复

RayCloud (/u/49d1f3b7049e)： @EDWIN (/u/9132ba692cc9) 1. 是 2. 取决于TensorFlow的同步方式：异步或同步

2017.07.10 02:21 回复



添加新评论



MessiyQin (/u/efb8204a12b5)
5楼 · 2017.12.05 19:05
(/u/efb8204a12b5)
求分享ppt学习， 541792972@qq.com

赞 回复

被以下专题收入，发现更多相似内容

- + 收入我的专题
- 软件匠艺 (/c/9707a5467cdc?utm_source=desktop&utm_medium=notes-included-collection)
- 程序员 (/c/NEt52a?utm_source=desktop&utm_medium=notes-included-collection)
- 今日看点 (/c/3sT4qY?utm_source=desktop&utm_medium=notes-included-collection)
- 编程 (/c/92d8cf3ffccc?utm_source=desktop&utm_medium=notes-included-collection)
- 机器学习 (/c/dd423eea0f24?utm_source=desktop&utm_medium=notes-included-collection)
- 机器学习 (/c/e20f093d5db7?utm_source=desktop&utm_medium=notes-included-collection)
- 首页投稿 (暂停... (/c/bDHhpK?utm_source=desktop&utm_medium=notes-included-collection)
- 展开更多

推荐阅读 更多精彩内容 > (/)

The Coding Kata: FizzBuzzWhizz in Scala (/p/2b00f4a4651c?utm_cam...

Functional programming leads to deep insights into the nature of computation. -- Martin Odersky 形式化 FizzBuzzWhizz详细描述请自行查阅相关资料。此处以3, 5, 7为例，形式化地描述一下问题。接下来我将...

RayCloud (/u/49d1f3b7049e?
utm_campaign=maleskine&utm_content=user&utm_medium=pc_all_hots&utm_source=recommendation)

代码阅读的姿势 (/p/3e6d4c520719?utm_campaign=... (/p/3e6d4c520719?
utm_campaign=maleskine&utm_content=note&utm
众里寻他千百度，蓦然回首，那人却在，灯火阑珊处。一般地，在一个程序员的
的日常工作中，绝大多数时间都是在「阅读代码」，而不是在「写代码」。...

RayCloud (/u/49d1f3b7049e?
utm_campaign=maleskine&utm_content=user&utm_medium=pc_all_hots&utm_source=recommendation)

我来告诉你，上大学的意义是什么 (/p/b9d760b195cf?... (/p/b9d760b195cf?
utm_campaign=maleskine&utm_content=note&utm
♥有读者在后台问我，说：“他觉得大学上得挺无奈的。刚上大学的他，完全没
了高中的上进努力，平时上课要么睡觉，要么玩手机，老师讲的什么内容也几...

影子影 (/u/9ddf8a34f958?
utm_campaign=maleskine&utm_content=user&utm_medium=pc_all_hots&utm_source=recommendation)


教你写出会有打赏的文章 (/p/05f2c91a3824?utm_cam... (/p/05f2c91a3824?
utm_campaign=maleskine&utm_content=note&utm
写作是这个时代最好的自我投资。在经济腾飞信息爆炸快速迭代的时代，每个
人至少有一次个体崛起的机会。想要在浩瀚的知识海洋里找到一条通往彼岸...
GM小咖 (/u/384c40c1e528?
utm_campaign=maleskine&utm_content=user&utm_medium=pc_all_hots&utm_source=recommendation)

自学画画两个月，我还是那个零基础的小白吗（很多多... (/p/9019b6013c12?
utm_campaign=maleskine&utm_content=note&utm
分享我的自学参考书籍和工具 答案是肯定的！我还是那个小白，零基础的小
白。我想，当我的画纸有一米高的时候，我就算是入门了吧。现在，才十分...
南蛮文子 (/u/94434fd7e3a0?
utm_campaign=maleskine&utm_content=user&utm_medium=pc_all_hots&utm_source=recommendation)

(/p/7a9c4fa298ae?




utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
Apache Spark 2.2.0 中文文档 - Spark Streaming 编程指南 | ApacheCN (/...
Spark Streaming 编程指南 概述 一个入门示例 基础概念 依赖 初始化 StreamingContext Discretized Streams
(DStreams) (离散化流) Input DStreams 和 Receivers (接收器) DStreams...

 Joyyx (/u/5d6219efd1b8?
utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

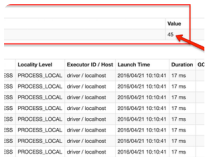
(/p/22c450a71328?




utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
Apache Spark 2.2.0 中文文档 - Spark Streaming 编程指南 | ApacheCN (/...
Spark Streaming 编程指南 概述 一个入门示例 基础概念 依赖 初始化 StreamingContext Discretized Streams
(DStreams) (离散化流) Input DStreams 和 Receivers (接收器) DStreams...

 片刻_ApacheCN (/u/a5d135d71592?
utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

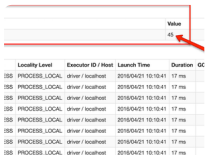
(/p/d43ab8f3b779?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
Apache Spark 2.2.0 中文文档 - Spark 编程指南 | ApacheCN (/p/d43ab8f3...
Spark 编程指南 概述 Spark 依赖 初始化 Spark 使用 Shell 弹性分布式数据集 (RDDs) 并行集合 外部
Datasets (数据集) RDD 操作 基础 传递 Functions (函数) 给 Spark 理解闭包 示例 Local (本地) vs. cl...


 片刻_ApacheCN (/u/a5d135d71592?
utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

(/p/c752c00c9c9f?

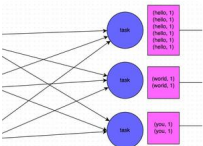


utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
Apache Spark 2.2.0 中文文档 - Spark 编程指南 | ApacheCN (/p/c752c00...
Spark 编程指南 概述 Spark 依赖 初始化 Spark 使用 Shell 弹性分布式数据集 (RDDs) 并行集合 外部

Datasets (数据集) RDD 操作 基础 传递 Functions (函数) 给 Spark 理解闭包 示例 Local (本地) vs. cl...


 Joyyx (/u/5d6219efd1b8?utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

(/p/cae948a08a08?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
[整理] Spark性能优化指南 (/p/cae948a08a08?utm_campaign=maleskine...

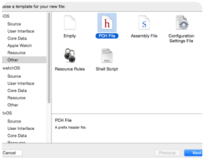
Spark的性能调优实际上是由很多部分组成的，不是调节几个参数就可以立竿见影提升作业性能的。我们需要根据不同的业务场景以及数据情况，对Spark作业进行综合性的分析，然后进行多个方面的调节和优化，才...


 东皇Amrzs (/u/d557aa88529a?utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

(/p/e147329589a7?

utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
Xcode7 添加PCH文件 (/p/e147329589a7?utm_camp...

1.) 打开你的Xcode工程. 在Supporting Files目录下,选择 File > New > File > iOS
> Other > PCH File 然后点击下一步; 2.) 给你的PCH文件起名字TestDemo....



 iOSZHU (/u/a96cd98783e6?


utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

(/p/452788ba9621?




utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
最低调的8种孕期食物：不是大鱼大肉，胜过大鱼大肉 (/p/452788ba9621?u...

在怀孕期间保持健康的饮食是非常重要的。在孕期，孕妇的身体需要额外的营养补充，例如：维生素和矿物质。缺乏关键营养素的饮食可能会对婴儿的发育产生负面影响。不良的饮食习惯和过度的体重增加也可能...

 瞧那一家子 (/u/c311b4527a00?utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

最没能坚持的事和最想坚持的事 (/p/acc52ff9a851?utm_campaign=males...

在我的记忆里，英语总是最让我头疼的学科，从初中开始学习英语烦恼就一直困扰着我，真的很烦唉！虽然我的英语不堪回首，可提起他来却又好多话要说。坎坷的英语求学历程，至今还历历在目。初一初二两年...

 coffee漫 (/u/2d6f6b31478a?


utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

(/p/6ad91903e561?



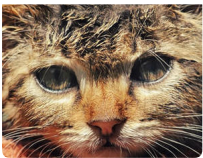
utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
杨幂在《真正男子汉2》秀里朴实抢眼！真人秀外秒变时尚教主！ (/p/6ad91...

上一期包晓贝写了一篇关于宝强离婚的文章。很多读者都在支持宝强，这跟本人一样一如既往的支持宝宝。包晓贝的世界是时尚自媒体，我们立足于时尚立足于包包的世界。无论是娱乐新闻还是专题时尚周刊，我...

 包晓贝的世界 (/u/4e26c8603906?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)


(/p/1be983fd1181?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)

《我和我的小萝莉们》第五十一章 把曦搞哭了 (/p/1be983fd1181?utm_ca...

我战战兢兢地转头朝她望去，却见她也正双目圆睁怒瞪着我！我心想可别惹毛她，我可不能成为勋别第二。
抱定这个想法之后，我尝试着先安抚曦的情绪：“曦，那个.....刚才我.....”曦突然伸手拿着个本子挡住自...

 段雪生 (/u/e9b6ff9e8c77?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

^

+

🔖

🔗