# TensorFlow估计器和数据集api



当TensorFlow 1.3被释放时。这是将近一年前的事了，TensorFlow已经有了一些更新，其中1.8是撰写本文时的最新版本。是时候重新审视这些api并看看它们是如何进化的了。

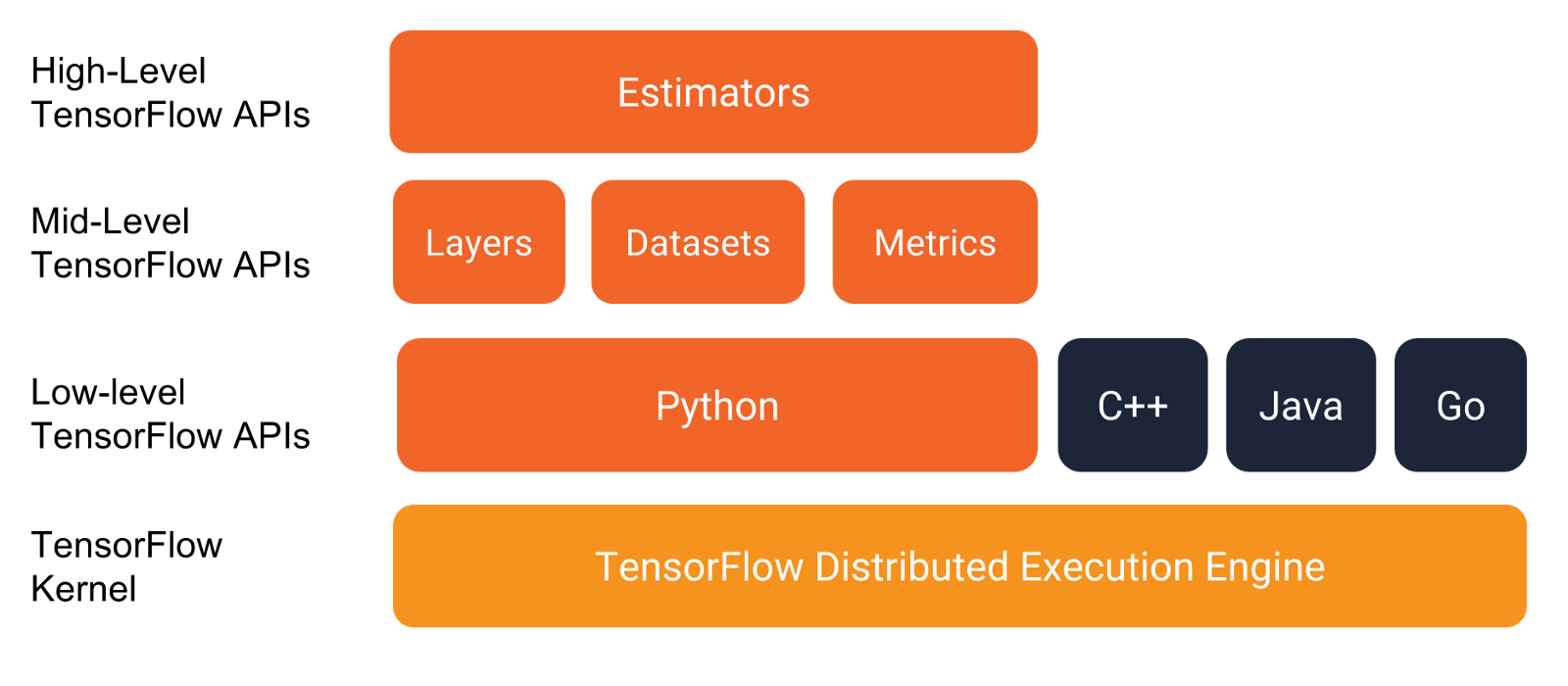
自TF 1.3以来，估计器和数据集api已经变得更加成熟:

我们强烈建议使用以下api编写TensorFlow程序：

估计量，表示一个完整的模型。估计器API提供了训练模型、判断模型精度和生成预测的方法。

数据集，用于构建数据输入管道。Dataset API有一些方法来加载和操作数据，并将其馈送到您的模型中。数据集API与估计器API很好地吻合。

Estimator API提供了一个顶级抽象，并与其他API（如用于构建输入流的Dataset API和用于构建模型架构的Layers API）进行了很好的集成甚至可以用Keras模型构造一个估计量。



在下面，我将概述API。.

### 估计器

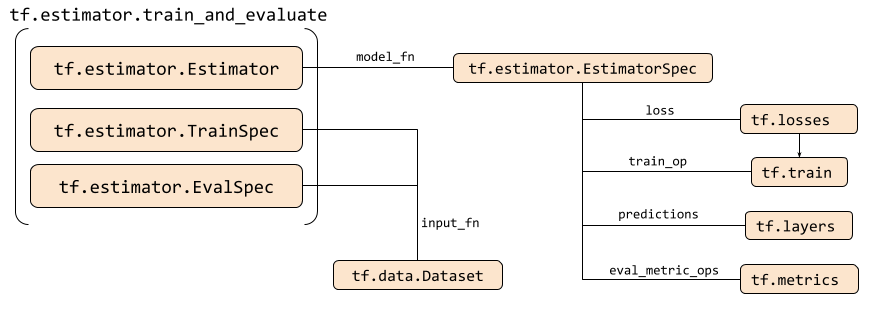
的核心保持稳定，我们仍然可以创建一个估计器，如下所示：

* model是初始化模型的函数。这个函数由一个知道如何生成所有输出、如何训练以及如何评估模型的对象来表示和完全定义。这个模型函数对于如何实现体系结构仍然是不可知的，您可以使用或其他库来实现您的模型体系结构。
* config是一个指定如何运行估计器的对象。它知道保存检查点的位置、保存检查点的次数、何时记录等等。
* params是一个包含模型超参数的对象。在我之前的文章中，这是一个对象。但是在1.8中，不再使用hparam，params可以是字典或其他对象（my中的对象）。

创建估计器后，我们可以使用以下函数对其进行训练：

* model\_估计器是估计器。
* train\_spec是表示训练配置的对象，知道如何构建训练数据馈送器。
* eval\_spec是表示评估配置的对象，知道如何构建评估数据馈送器。

请注意，这与以前的blogpost不同，我们使用TensorFlow类（现在已被弃用）来运行培训。摆脱实验班使一切变得不那么复杂。培训和评估输入函数和钩子现在清楚地分为TrainSpec和EvalSpec，您只需要调用train\_和\_evaluate函数。



### 数据集

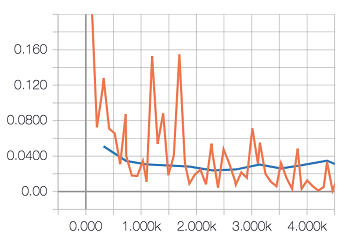
已经完全成熟并从TensorFlow核心库迁移到TensorFlow核心库，现在它允许您构建复杂的输入管道。在中，它用于构建一个输入馈送器，该馈送器对数据进行洗牌，并根据需要以正确的批大小重复数据：

注意，在这个例子中，我只是加载。您可以使用Dataset API构建更多。

### 在本地运行代码

您可以通过以下方式从本地运行代码：

这将启动培训和评估课程。通过使用Estimator API，它还设置了默认日志记录和检查点保存，我们可以通过以下方式可视化：



### 在Google云ML引擎上运行代码

由于对配置的抽象，Estimator API允许它轻松地训练模型：

使用tf.estimator.train\_和\_evaluate，您可以在本地和云中分布的不同设备上使用不同的集群配置运行相同的代码，并且在不更改任何代码的情况下获得一致的结果

I provided a minimal example of [how to run the accompanying code on Google Cloud](https://github.com/peterroelants/tf_estimator_example#training-on-google-cloud). For example, you can train the code on the cloud by running:

In summary, the TensorFlow Estimator API, as well as the Dataset API, have matured a lot. They provide a nice abstraction layer to manage input data streams, models, and training/evaluation configurations.