TensorFlow与Apache Flink的结合

陈戊超 · 阿里巴巴 / 技术专家

Apache Flink Community China

CONTENT

01

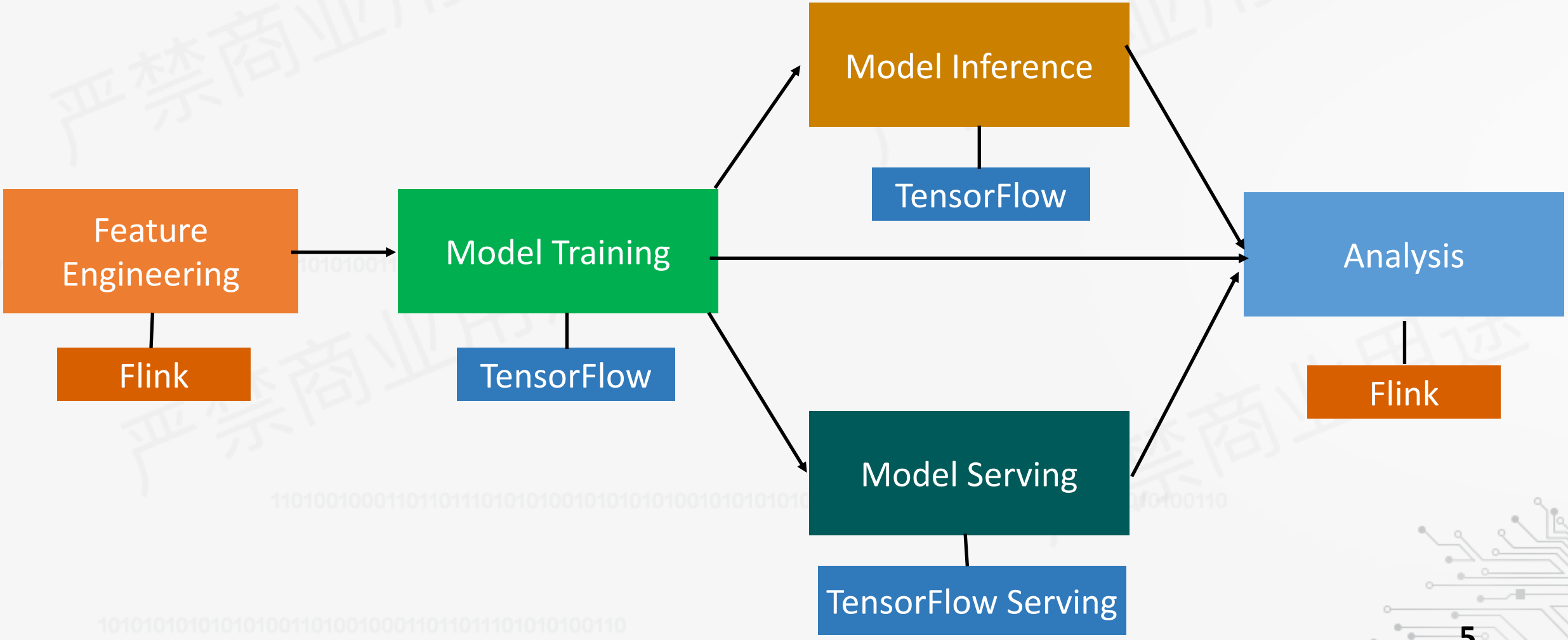
Background

Background

TensorFlow is an open source software library for numerical computation using data flow graphs and is the most popular **AI computing framework**.

Flink is a framework and distributed processing engine for stateful computations over unbounded and bounded data streams。 **Flink is widely used in data processing and Feature Engineering**

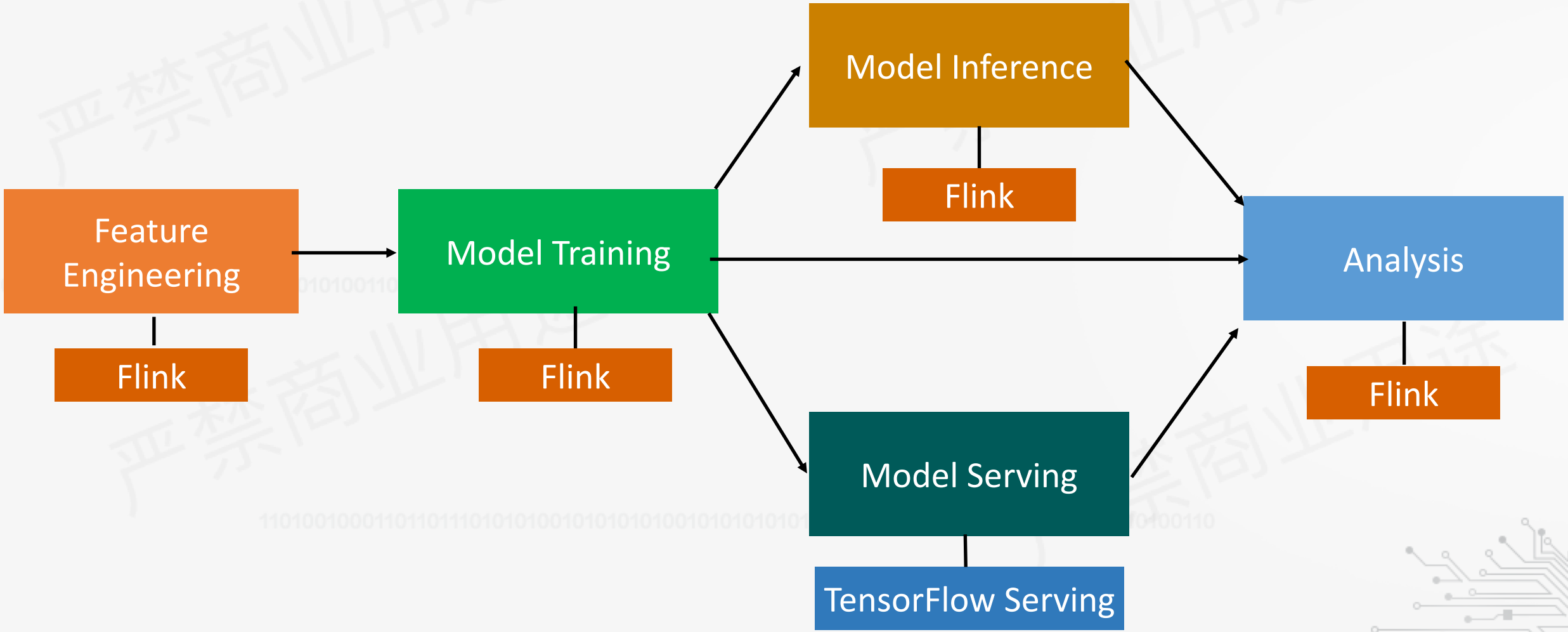
Machine learning workflow



Problems

- *Users do feature engineering, model training and model prediction with **two framework**.*
- *Distributed programs often run in clusters but **it's not friendly** to use TensorFlow for distributed training to determine IP and port first.*
- *TensorFlow Distributed Running Can't **Failover Automatically**。*

Goal


Model Inference
Flink
Feature Engineering
Model Training
Analysis
Flink
Flink
Flink
Model Serving
TensorFlow Serving

Apache Flink

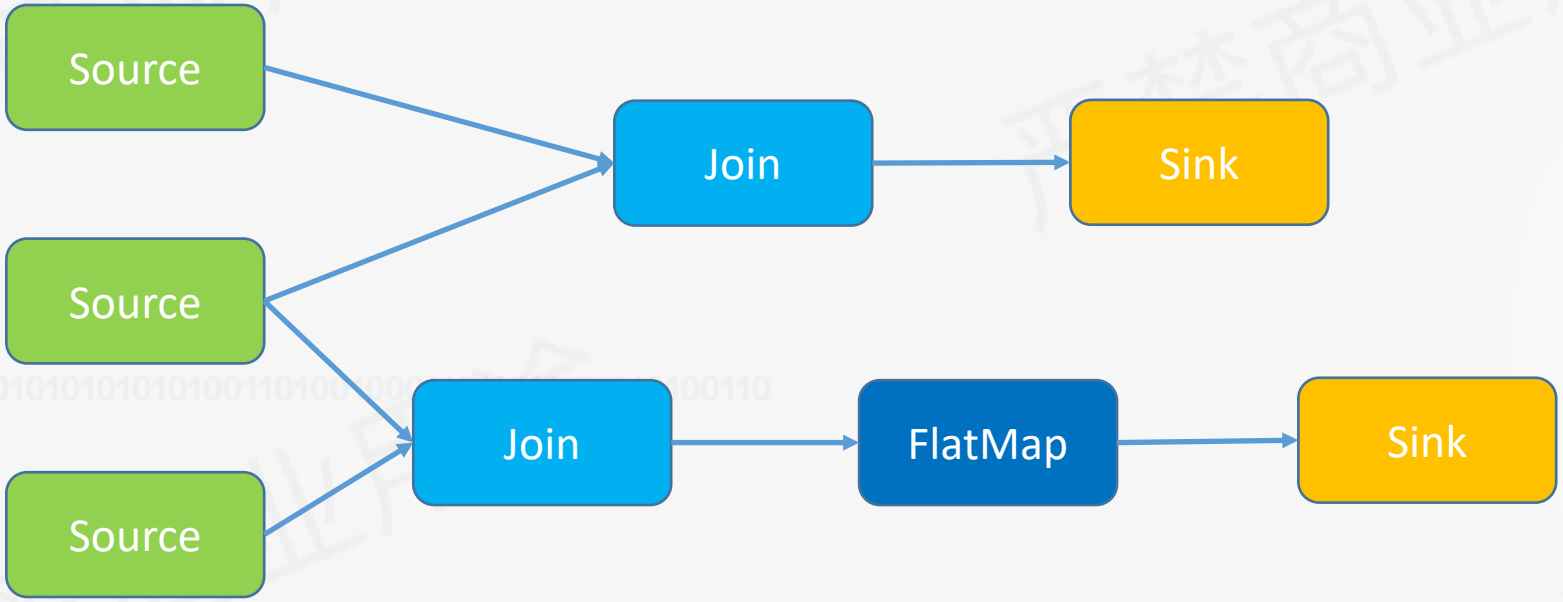
flink-ai-extended

Github: https://github.com/alibaba/flink-ai-extended

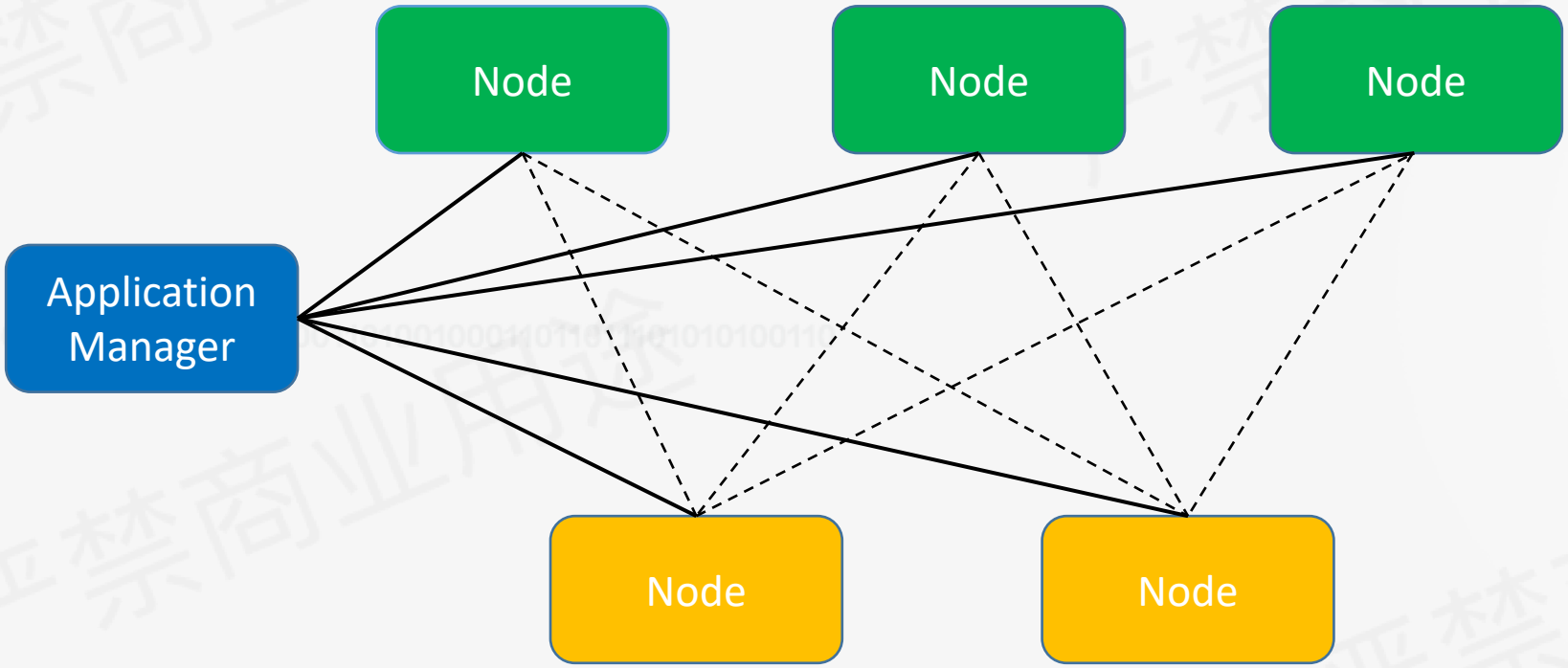
02

Machine Learning On Flink

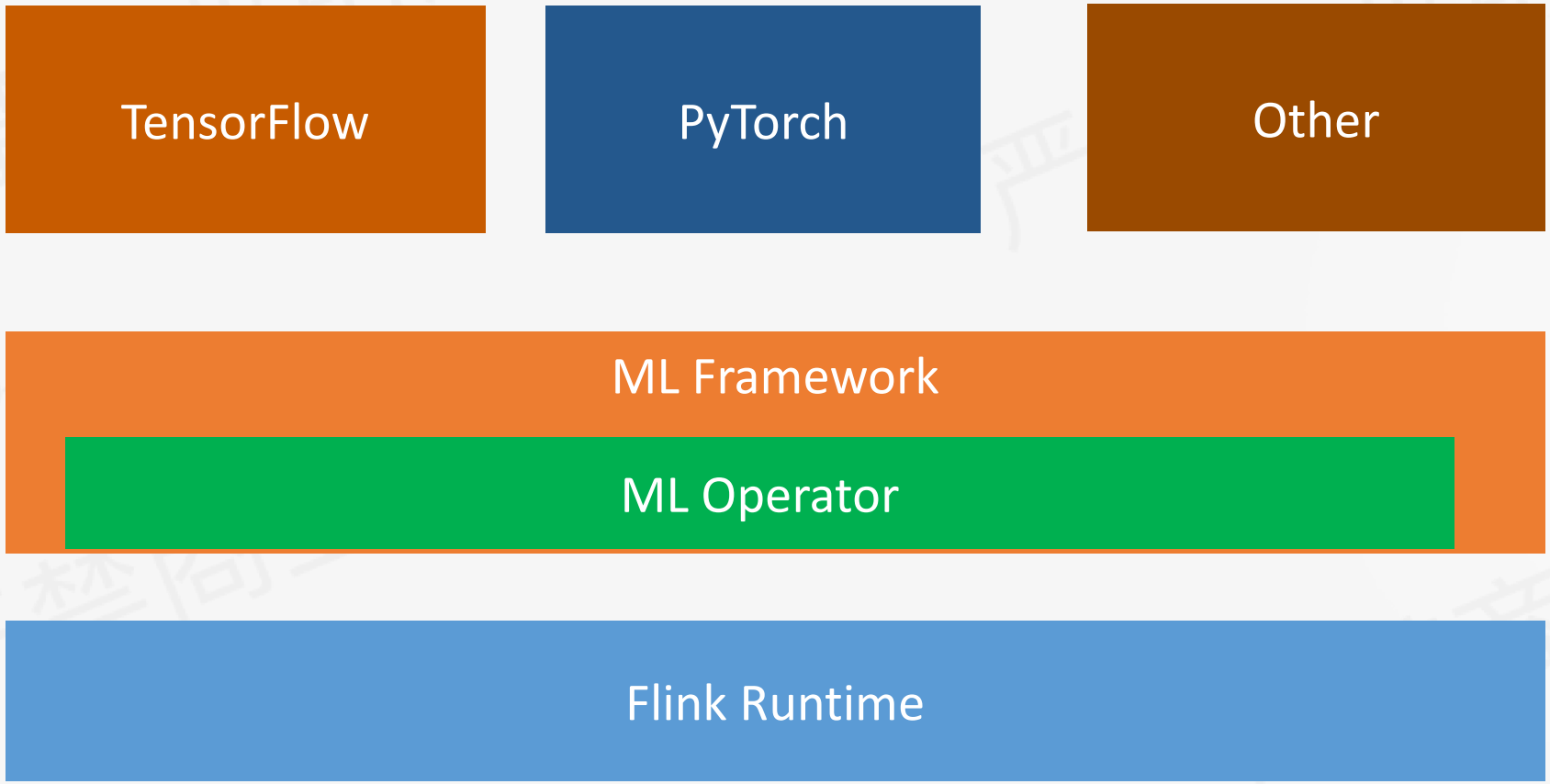
Flink



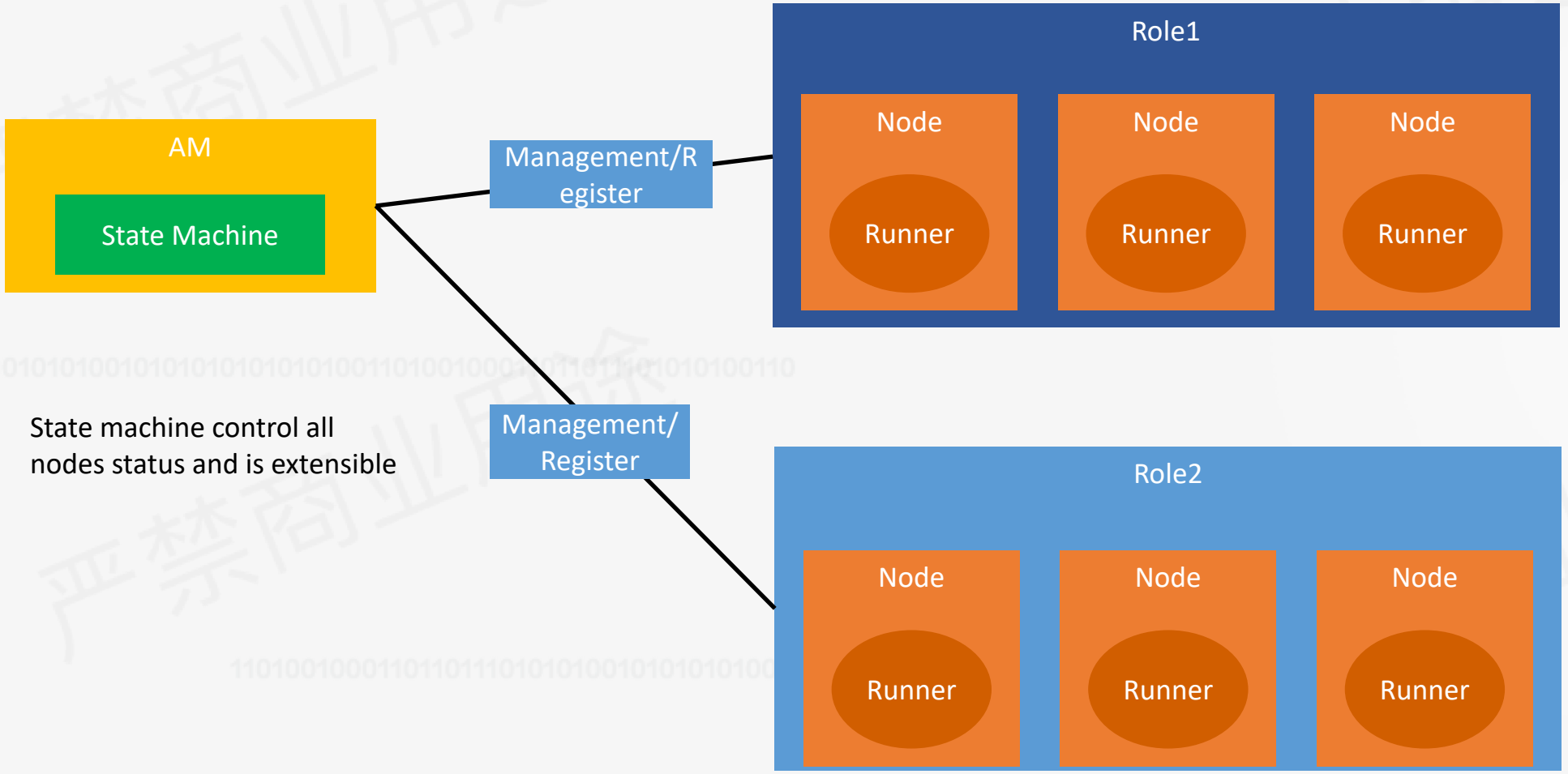
Machine learning cluster



Machine Learning On Flink



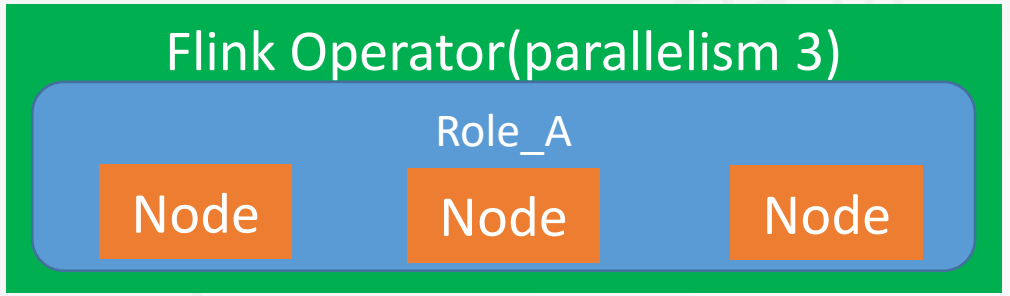
ML Framework

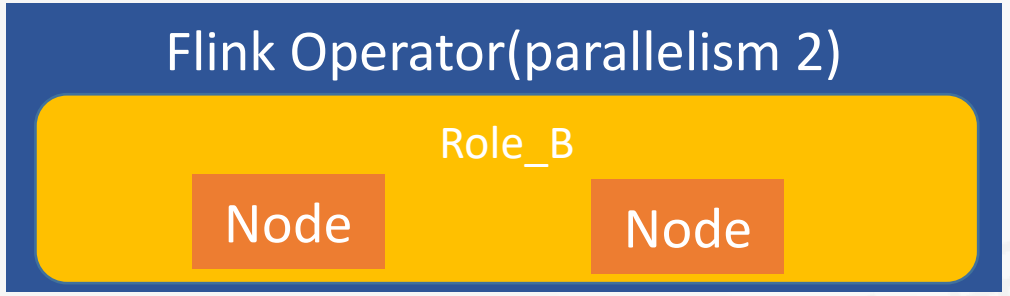


State machine control all nodes status and is extensible

Runner define how to start ML process and is extensible

ML Operator



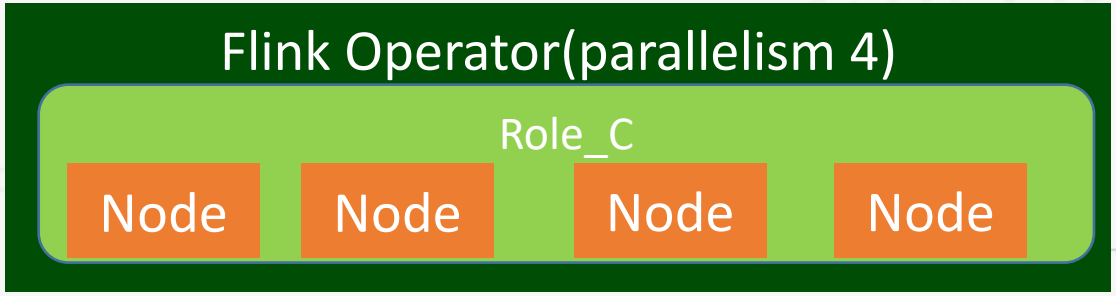


ML Operator provide：

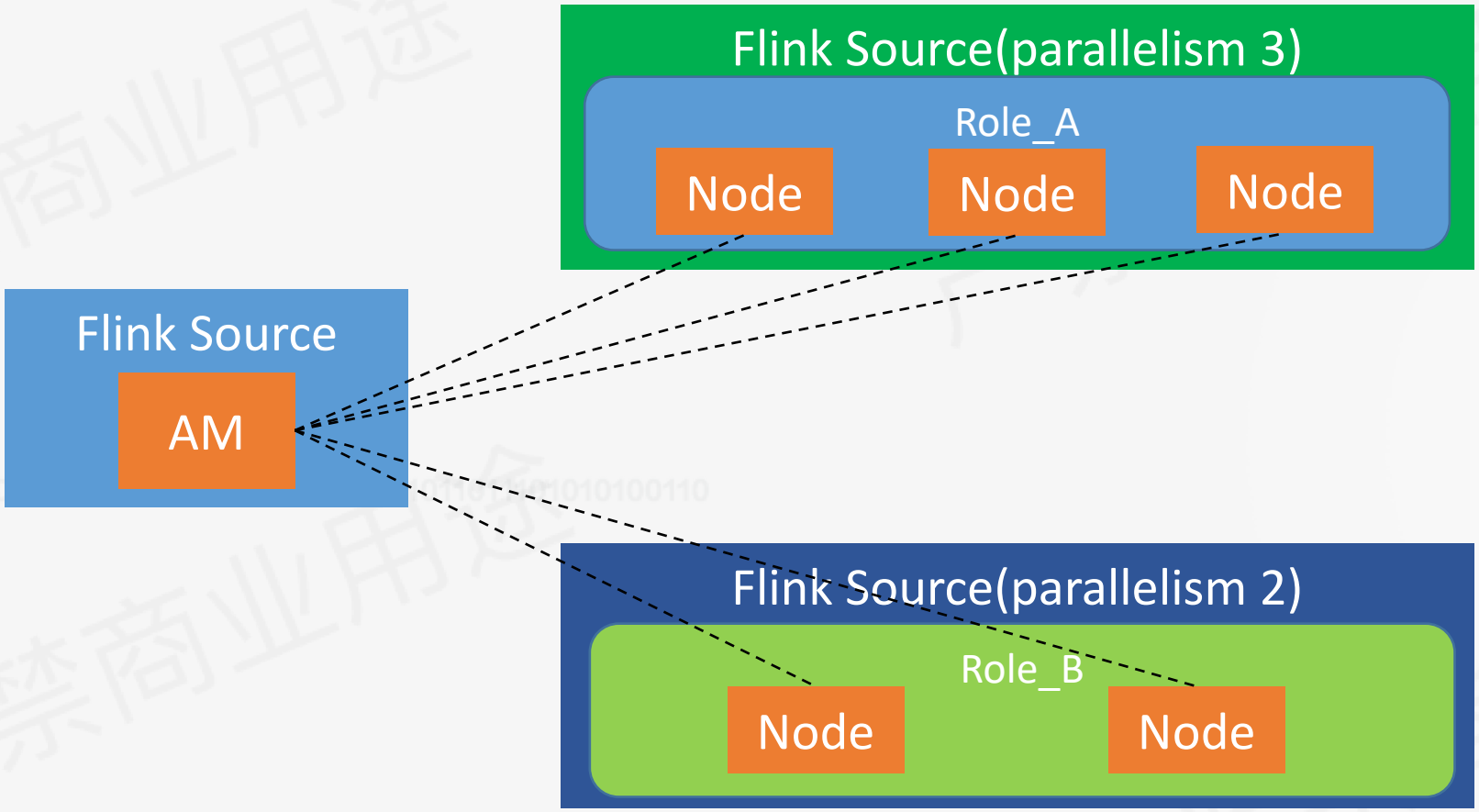
1. addAMRole(Config)
2. addRole(RoleName, Config)

Create ML Job Example:

1. addAMRole(Config)
2. addRole(Role_A, Config)
3. addRole(Role_B, Config)
4. addRole(Role_C, Config)

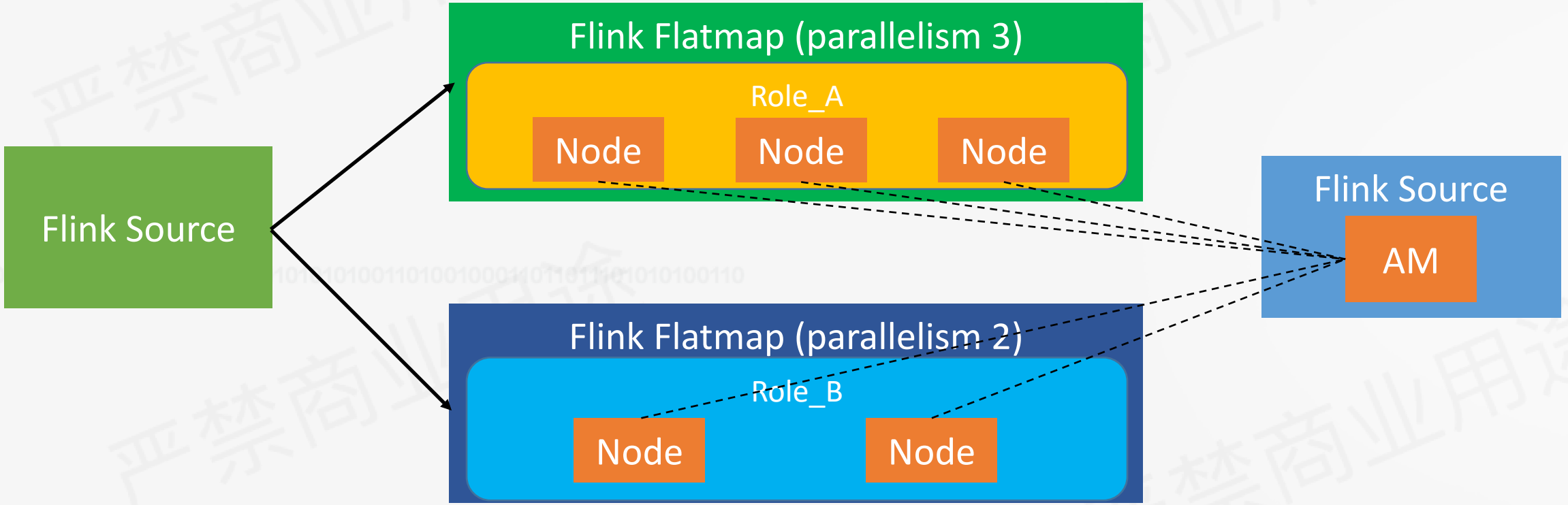


Batch Mode



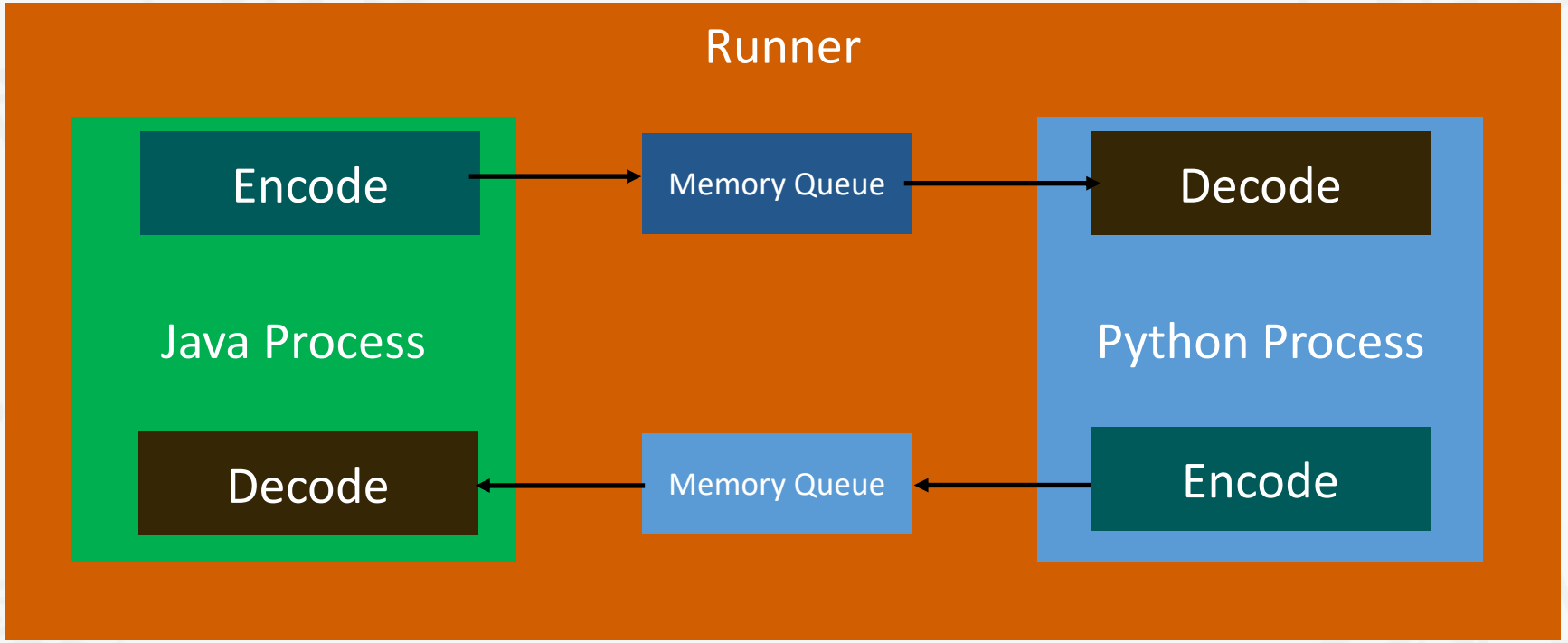
Do not read data from flink ，plan role as flink source operator

Stream Mode



Read data from flink ， plan role as flink flatmap operator

Data Exchange



1. Encode transfer user define object to byte[]
2. Decode transfer byte[] to user define object

Encode and Decode is extensible

Summary

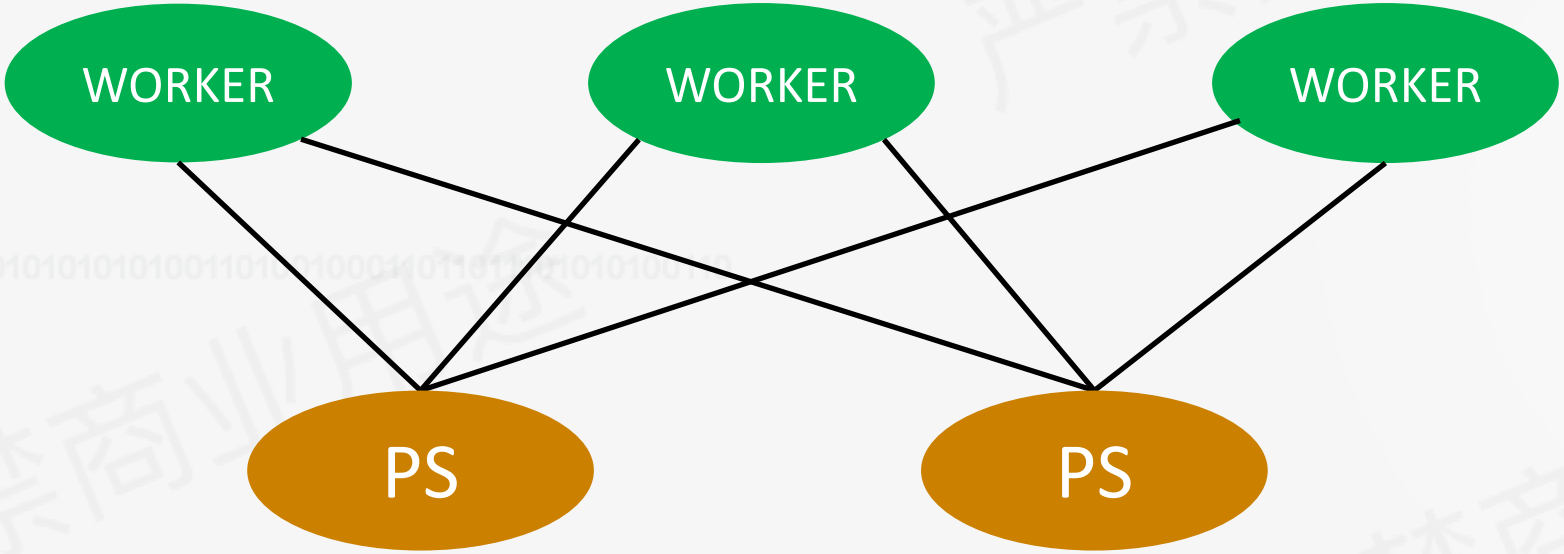
1. Cluster creation and state management

2. Java and python data exchange

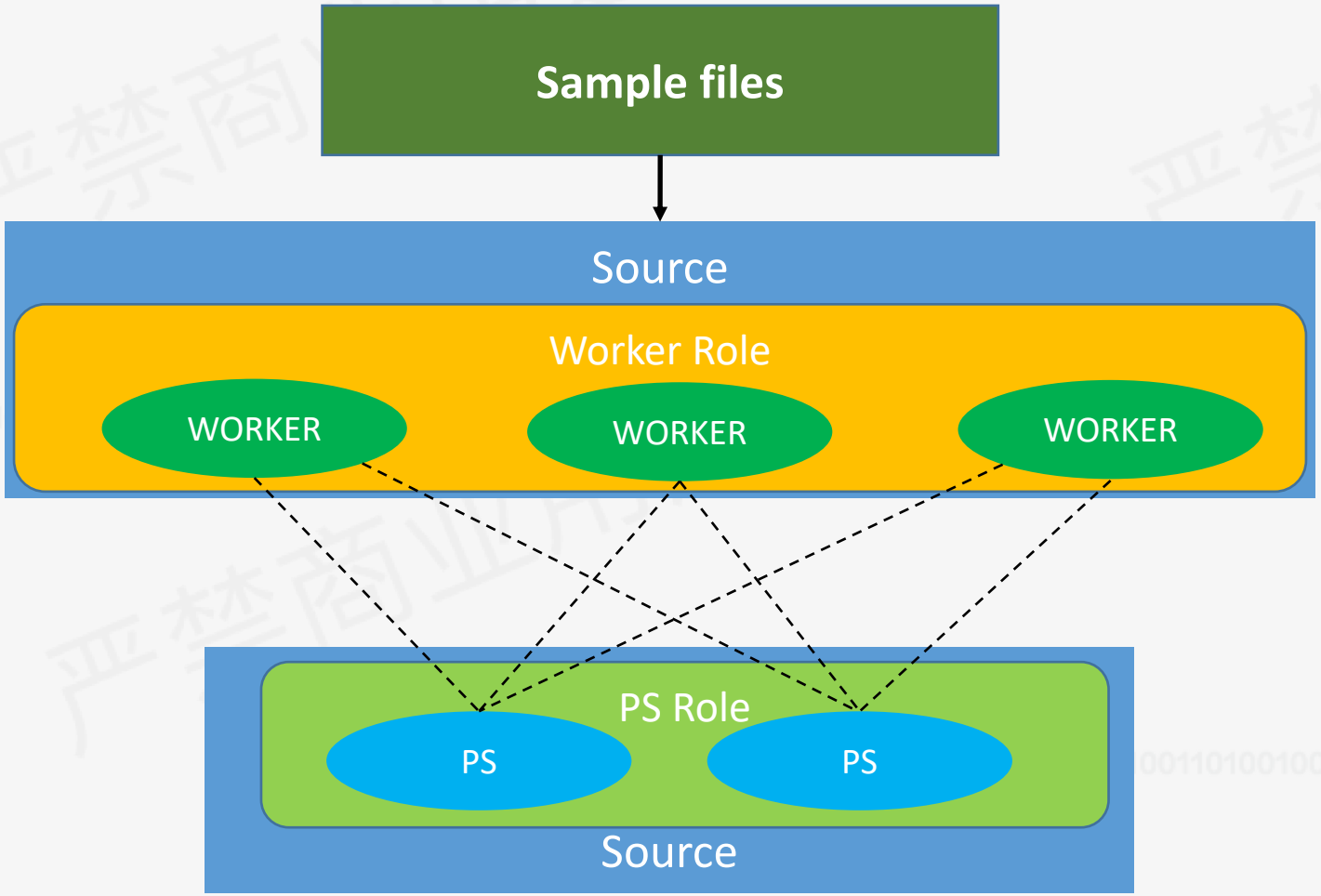
03

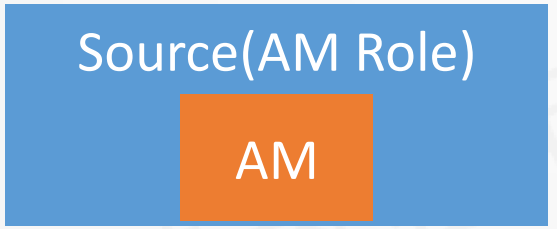
TensorFlow

TensorFlow Training

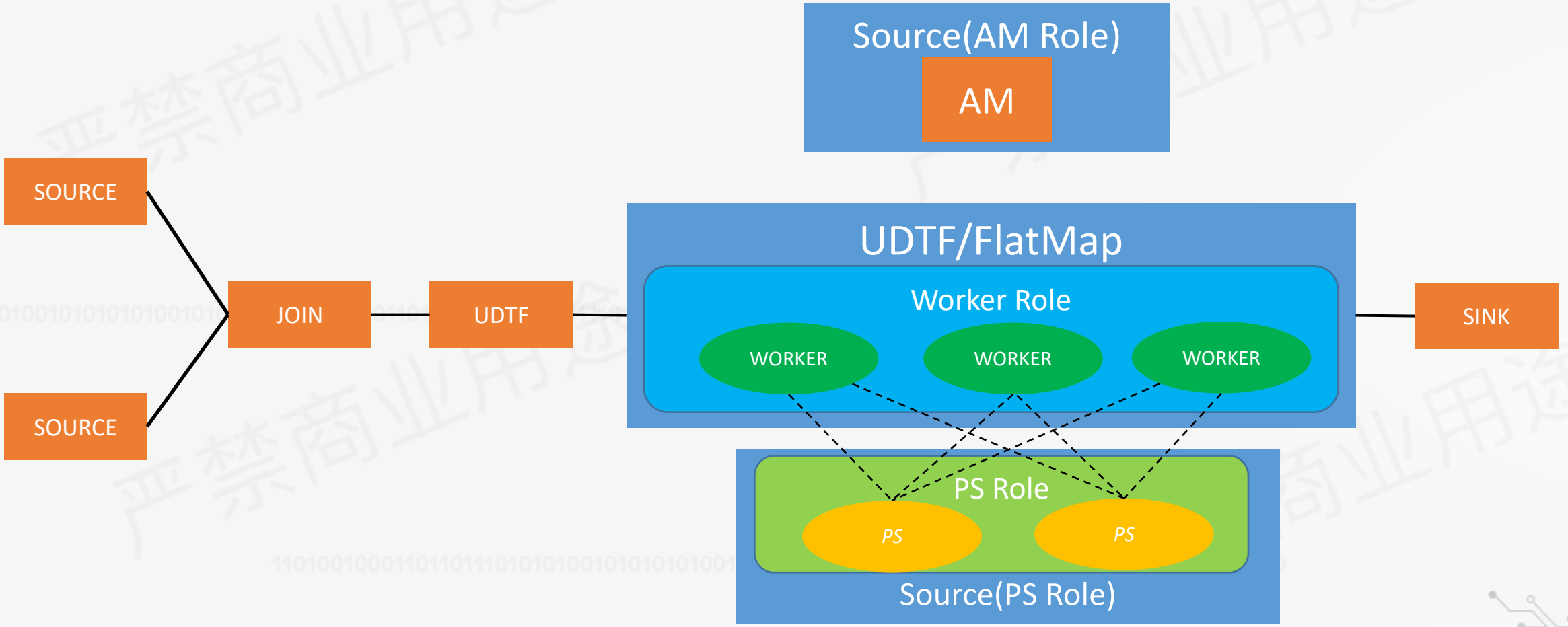


TensorFlow Batch Training

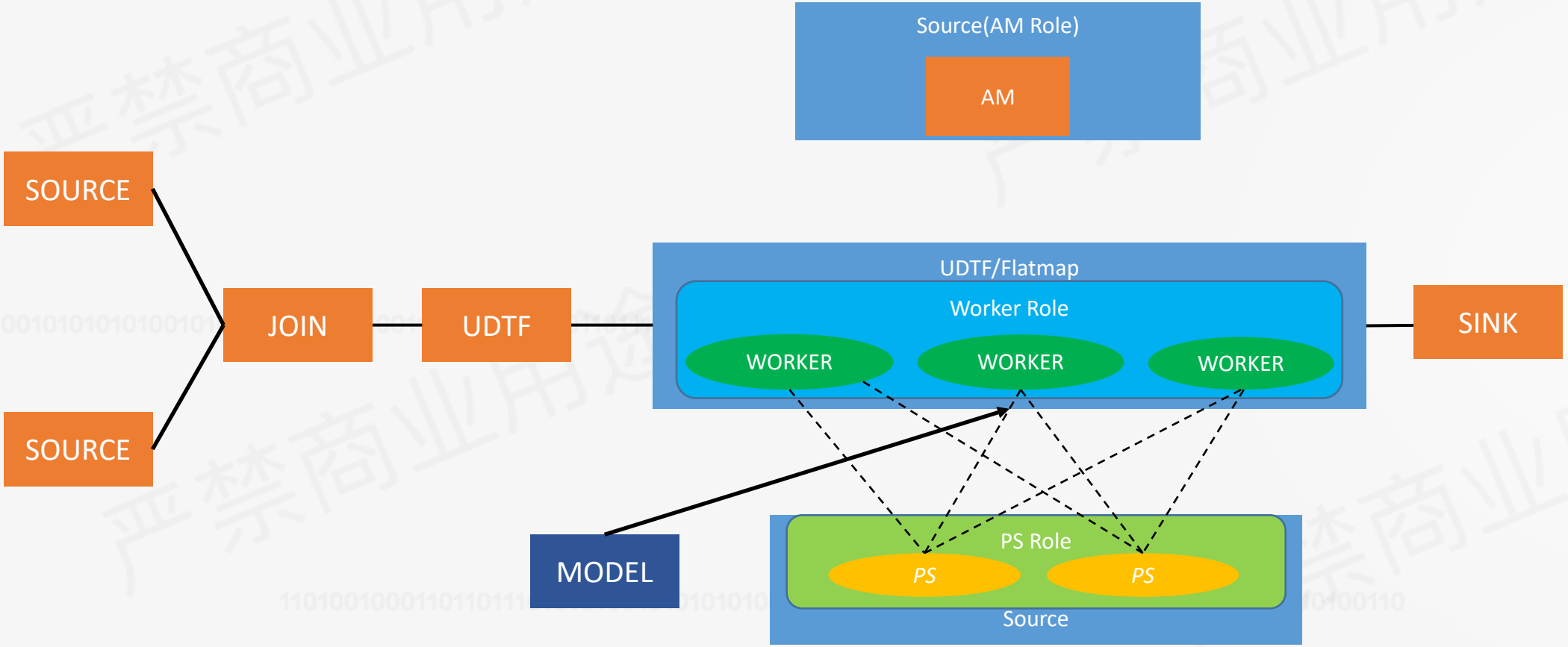




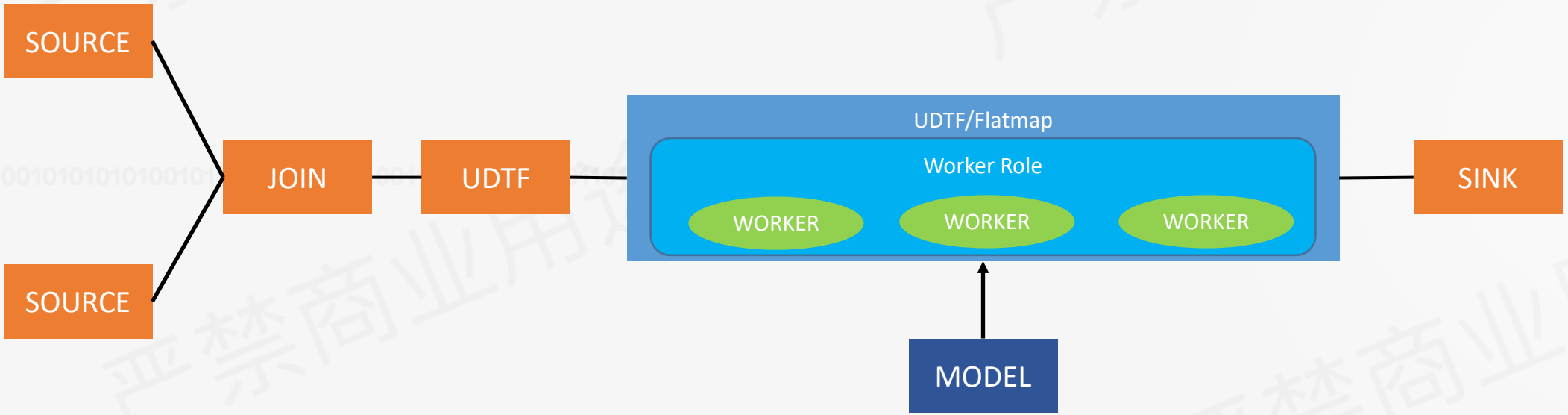
TensorFlow Stream Training



TensorFlow Stream Inference



TensorFlow Stream Inference



TensorFlow Example

```
import tensorflow as tf
cluster = tf.train.ClusterSpec({
   "worker": [
      "A_IP:2222",
      "B_IP:1234",
      "C_IP:2222"
   ],
   "ps": [
      "D_IP:2222",
   ]})
isps = False
if isps:
   server = tf.train.Server(cluster, job_name='ps', task_index=0)
   server.join()
else:
   server = tf.train.Server(cluster, job_name='worker', task_index=0)
   with tf.device(tf.train.replica_device_setter(worker_device='/job:worker/task:0', cluster=cluster)):
      w = tf.get_variable('w', (2, 2), tf.float32, initializer=tf.constant_initializer(2))
      b = tf.get_variable('b', (2, 2), tf.float32, initializer=tf.constant_initializer(5))
      addwb = w + b
      mutwb = w * b
      divwb = w / b
```

Apache Flink

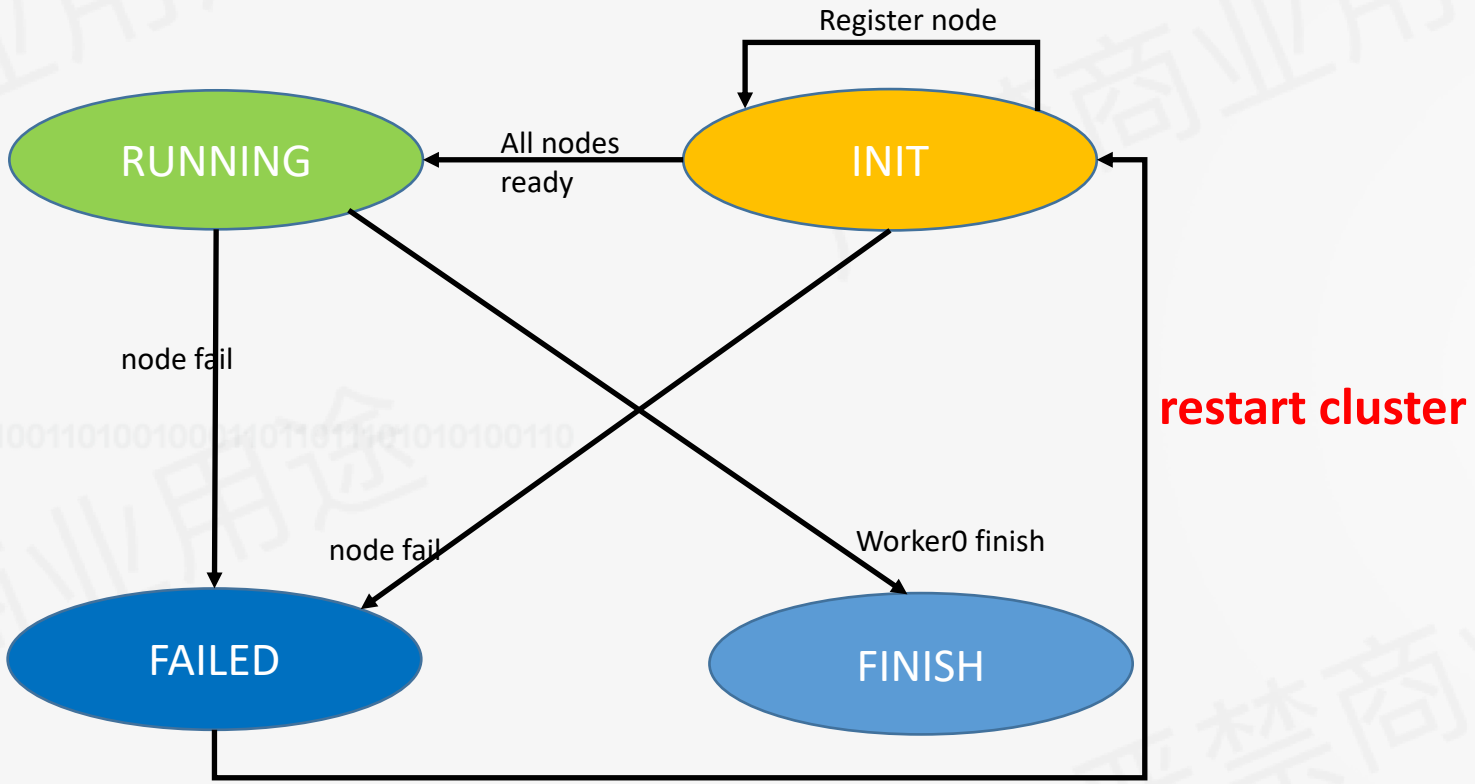
TensorFlow Example

```
import tensorflow as tf

def run_main(context):
    cluster = context.get_cluster()
    job_name = context.get_job_name()
    task_index = context.get_task_index()
    if 'ps' == job_name:
        server = tf.train.Server(cluster, job_name=job_name, task_index=task_index)
        server.join()
    else:
        server = tf.train.Server(cluster, job_name=job_name, task_index=task_index)
        with tf.device(tf.train.replica_device_setter(worker_device='/job:worker/task:0', cluster=cluster)):
            w = tf.get_variable('w', (2, 2), tf.float32, initializer=tf.constant_initializer(2))
            b = tf.get_variable('b', (2, 2), tf.float32, initializer=tf.constant_initializer(5))
            addwb = w + b
            mutwb = w * b
            divwb = w / b

if __name__ == "__main__":
    stream_env = StreamExecutionEnvironment.get_execution_environment()
    train(3, 1, run_main, properties=None, stream_env=stream_env, input_ds=None, output_row_type=None)
```

TensorFlow Failover



Apache Flink

TensorFlow Applications

1. Search Ranking
2. Recommender System

THANKS

flink-ai-extended交流群

3人

扫一扫群二维码，立刻加入该群。