// This file defines protos that store the results of autotuning various

// operations.

//

// They are in proto format because we want to log them structured. They offer

// tremendous statistical, testing, and debugging value.

syntax = "proto3";

package tensorflow;

import "google/protobuf/any.proto";

import "google/protobuf/duration.proto";

import "tensorflow/stream\_executor/dnn.proto";

option go\_package = "github.com/tensorflow/tensorflow/tensorflow/go/core/protobuf/for\_core\_protos\_go\_proto";

message CudnnVersion {

int32 major = 1;

int32 minor = 2;

int32 patch = 3;

}

message ComputeCapability {

int32 major = 1;

int32 minor = 2;

}

message AutotuneResult {

enum FailureKind {

UNKNOWN = 0;

// Algorithm wrote memory outside its output buffers.

REDZONE\_MODIFIED = 1;

// Algorithm gave a different result from a reference algorithm.

WRONG\_RESULT = 2;

// Algorithm was rejected for failing to run or for known bugs.

DISQUALIFIED = 3;

}

message FailureResult {

FailureKind kind = 1;

string msg = 2;

// For failure\_kind == WRONG\_RESULT, this field indicates the reference

// configuration that we compared against.

//

// Note that the reference algorithm isn't always correct. However,

// empirically it's more correct, as it's "algo 0", less fancy than the

// compared one.

oneof key {

ConvKey reference\_conv = 11;

GemmKey reference\_gemm = 12;

CudaConvPlanKey reference\_cuda\_conv\_plan = 14;

stream\_executor.dnn.AlgorithmProto reference\_algorithm = 15;

}

int64 buffer\_address = 13;

}

// Legacy and unused in new data; superseded by AlgorithmProto.

message ConvKey {

int64 algorithm = 1;

bool tensor\_ops\_enabled = 2;

}

message GemmKey {

int64 algorithm = 1;

}

// Legacy and unused in new data; superseded by AlgorithmProto.

message CudaConvPlanKey {

string exec\_plan\_id = 1;

}

int64 scratch\_bytes = 8;

google.protobuf.Duration run\_time = 9;

FailureResult failure = 7;

oneof key {

ConvKey conv = 5;

GemmKey gemm = 6;

CudaConvPlanKey cuda\_conv\_plan = 15;

stream\_executor.dnn.AlgorithmProto algorithm = 16;

}

// Next ID: 17

}

message AutotuningLog {

google.protobuf.Any instr = 1;

// Records all auto-tuning results per algorithm.

repeated AutotuneResult results = 2;

CudnnVersion cudnn\_version = 3;

ComputeCapability compute\_capability = 4;

// stream\_executor::DeviceDescription::pci\_bus\_id.

string device\_pci\_bus\_id = 5;

string blas\_version = 6;

// Next ID: 7

}