tf.variable_axis_size_partitioner

```
variable_axis_size_partitioner(
    max_shard_bytes,
    axis=0,
    bytes_per_string_element=16,
    max_shards=None
)
```

Defined in `tensorflow/python/ops/partitioned_variables.py`.

See the guide: Variables > Variable Partitioners for Sharding

Get a partitioner for VariableScope to keep shards below `max_shard_bytes`.

This partitioner will shard a Variable along one axis, attempting to keep the maximum shard size below `max_shard_bytes`. In practice, this is not always possible when sharding along only one axis. When this happens, this axis is sharded as much as possible (i.e., every dimension becomes a separate shard).

If the partitioner hits the `max_shards` limit, then each shard may end up larger than `max_shard_bytes`. By default `max_shards` equals `None` and no limit on the number of shards is enforced.

One reasonable value for `max_shard_bytes` is `(64 << 20) - 1`, or almost `64MB`, to keep below the protobuf byte limit.

Args:

- `max_shard_bytes`: The maximum size any given shard is allowed to be.
- `axis`: The axis to partition along. Default: outermost axis.
- `bytes_per_string_element`: If the `Variable` is of type string, this provides an estimate of how large each scalar in the `Variable` is.
- `max_shards`: The maximum number of shards in int created taking precedence over `max_shard_bytes`.

Returns:

A partition function usable as the `partitioner` argument to `variable_scope`, `get_variable`, and `get_partitioned_variable_list`.

Raises:

- `ValueError`: If any of the byte counts are non-positive.

Last updated November 2, 2017.

Blog

GitHub

Twitter

Support

Issue Tracker

Release Notes

Stack Overflow

English

Terms | **Privacy**