SageMaker Debugger Profiling Report

SageMaker Debugger auto generated this report. You can generate similar reports on all supported training jobs. The report provides summary of training job, system resource usage statistics, framework metrics, rules summary, and detailed analysis from each rule. The graphs and tables are interactive.

Legal disclaimer: This report and any recommendations are provided for informational purposes only and are not definitive. You are responsible for making your own independent assessment of the information.

In [4]:

```
# Parameters
processing_job_arn = "arn:aws:sagemaker:us-east-1:909528142112:processing-job/pytorch-training-2022-08-1-profilerreport-046dd0e2"
```

Training job summary

The following table gives a summary about the training job. The table includes information about whe
how much time initialization, training loop and finalization took. Your training job started on 08/12/202
seconds.

#		Job Statistics
0	Start time	15:10:09 08/12/2022
1	End time	16:09:59 08/12/2022
2	Job duration	3590 seconds
3	Training loop start	15:12:23 08/12/2022
4	Training loop end	16:09:52 08/12/2022
5	Training loop duration	3448 seconds
6	Initialization time	134 seconds
7	Finalization time	7 seconds
8	Initialization	3 %
9	Training loop	96 %
10	Finalization	0 %

System usage statistics

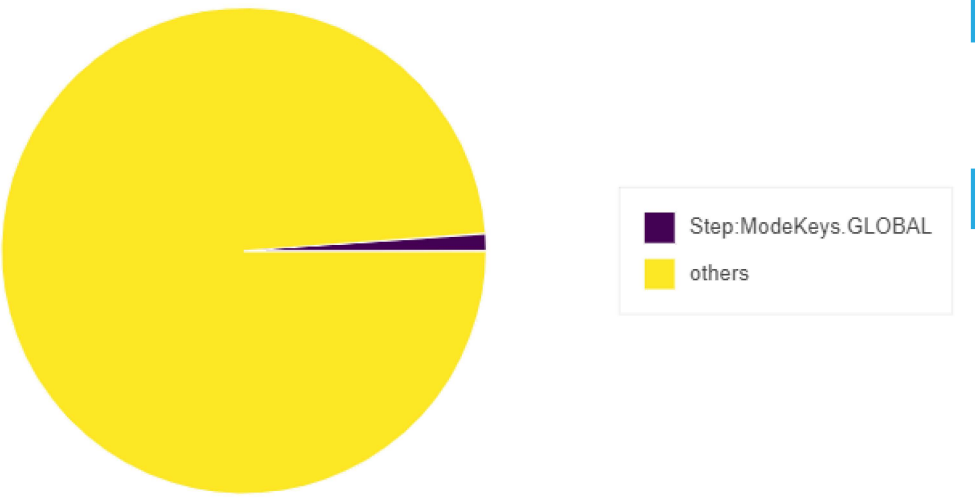
The following table shows statistics of resource utilization per worker (node), such as the total CPU a
and GPU. The table also includes the total I/O wait time and the total amount of data sent or receivec
as p99, p90 and p50 percentiles.

#	node	metric	unit	max	p99	p95
0	algo-1	Network	bytes	60280428.75	0	0
1	algo-1	CPU	percentage	100	100	100
2	algo-1	CPU memory	percentage	28.52	27.54	26.46
3	algo-1	I/O	percentage	39.7	2.08	0.82

Framework metrics summary

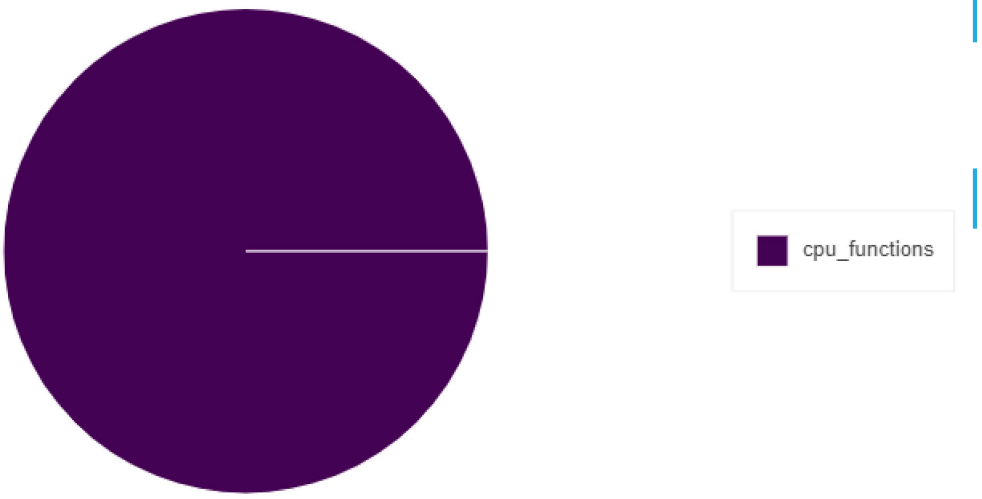
The following two pie charts show the time spent on the TRAIN phase, the EVAL phase, and others. the next step has started). Ideally, most of the training time should be spent on the TRAIN and EVAL GLOBAL.

The ratio between the time spent on the TRAIN/EVAL phase and others



The following piechart shows a breakdown of the CPU/GPU operators. It shows that 100% of training

The ratio between the time spent on CPU/GPU operators



Overview: CPU operators

The following table shows a list of operators that ran on the CPUs. The most expensive operator on t

#	Percentage	Cumulative time in microsecon	CPU operator
0	18.47	60325219	aten::conv2d
1	18.47	60323126	aten::convolution
2	18.46	60321389	aten::_convolution
3	18.46	60315720	aten::mkldnn_convolution
4	5.57	18202961	aten::batch_norm
5	5.57	18199247	aten::_batch_norm_impl_index
6	5.57	18195781	aten::native_batch_norm
7	4.21	13758444	enumerate(DataLoader)#_Sing
8	2.62	8564752	aten::max_pool2d
9	2.6	8480015	aten::max_pool2d_with_indices

Rules summary

The following table shows a profiling summary of the Debugger built-in rules. The table is sorted by the rules that triggered the most frequently. During your training job, the StepOutlier rule was the most frequently triggered. It processed 422 datapoints and was triggered 36 times.

	Description	Recommendation	Number of times rule triggered	Number of datapoints	R
StepOutlier	Detects outliers in step duration. The step duration for forward and backward pass should be roughly the same throughout the training. If there are significant outliers, it may indicate a system stall or bottleneck issues.	Check if there are any bottlenecks (CPU, I/O) correlated to the step outliers.	36	422	
CPUBottleneck	Checks if the CPU utilization is high and the GPU utilization is low. It might indicate CPU bottlenecks, where the GPUs are waiting for data to arrive from the CPUs. The rule evaluates the CPU and GPU utilization rates, and triggers the issue if the time spent on the CPU bottlenecks exceeds a threshold percent of the total training time. The default threshold is 50 percent.	Consider increasing the number of data loaders or applying data pre-fetching.	0	10952	cp gp
GPUMemoryIncrease	Measures the average GPU memory footprint and triggers if there is a large increase.	Choose a larger instance type with more memory if footprint is close to maximum available memory.	0	0	

	Description	Recommendation	Number of times rule triggered	Number of datapoints	R
IOBottleneck	Checks if the data I/O wait time is high and the GPU utilization is low. It might indicate IO bottlenecks where GPU is waiting for data to arrive from storage. The rule evaluates the I/O and GPU utilization rates and triggers the issue if the time spent on the IO bottlenecks exceeds a threshold percent of the total training time. The default threshold is 50 percent.	Pre-fetch data or choose different file formats, such as binary formats that improve I/O performance.	0	10952	i gp
LowGPUUtilization	Checks if the GPU utilization is low or fluctuating. This can happen due to bottlenecks, blocking calls for synchronizations, or a small batch size.	Check if there are bottlenecks, minimize blocking calls, change distributed training strategy, or increase the batch size.	0	0	thr tl
MaxInitializationTime	Checks if the time spent on initialization exceeds a threshold percent of the total training time. The rule waits until the first step of training loop starts. The initialization can take longer if downloading the entire dataset from Amazon S3 in File mode. The default threshold is 20 minutes.	Initialization takes too long. If using File mode, consider switching to Pipe mode in case you are using TensorFlow framework.	0	422	

	Description	Recommendation	Number of times rule triggered	Number of datapoints	R
Dataloader	Checks how many data loaders are running in parallel and whether the total number is equal the number of available CPU cores. The rule triggers if number is much smaller or larger than the number of available cores. If too small, it might lead to low GPU utilization. If too large, it might impact other compute intensive operations on CPU.	Change the number of data loader processes.	0	10	mi ma
BatchSize	Checks if GPUs are underutilized because the batch size is too small. To detect this problem, the rule analyzes the average GPU memory footprint, the CPU and the GPU utilization.	The batch size is too small, and GPUs are underutilized. Consider running on a smaller instance type or increasing the batch size.	0	7060	cpu_thr gpu_thr gpu_memory_th
LoadBalancing	Detects workload balancing issues across GPUs. Workload imbalance can occur in training jobs with data parallelism. The gradients are accumulated on a primary GPU, and this GPU might be overused with regard to other GPUs, resulting in reducing the efficiency of data parallelization.	Choose a different distributed training strategy or a different distributed training framework.	0	0	

Analyzing the training loop

Step duration analysis

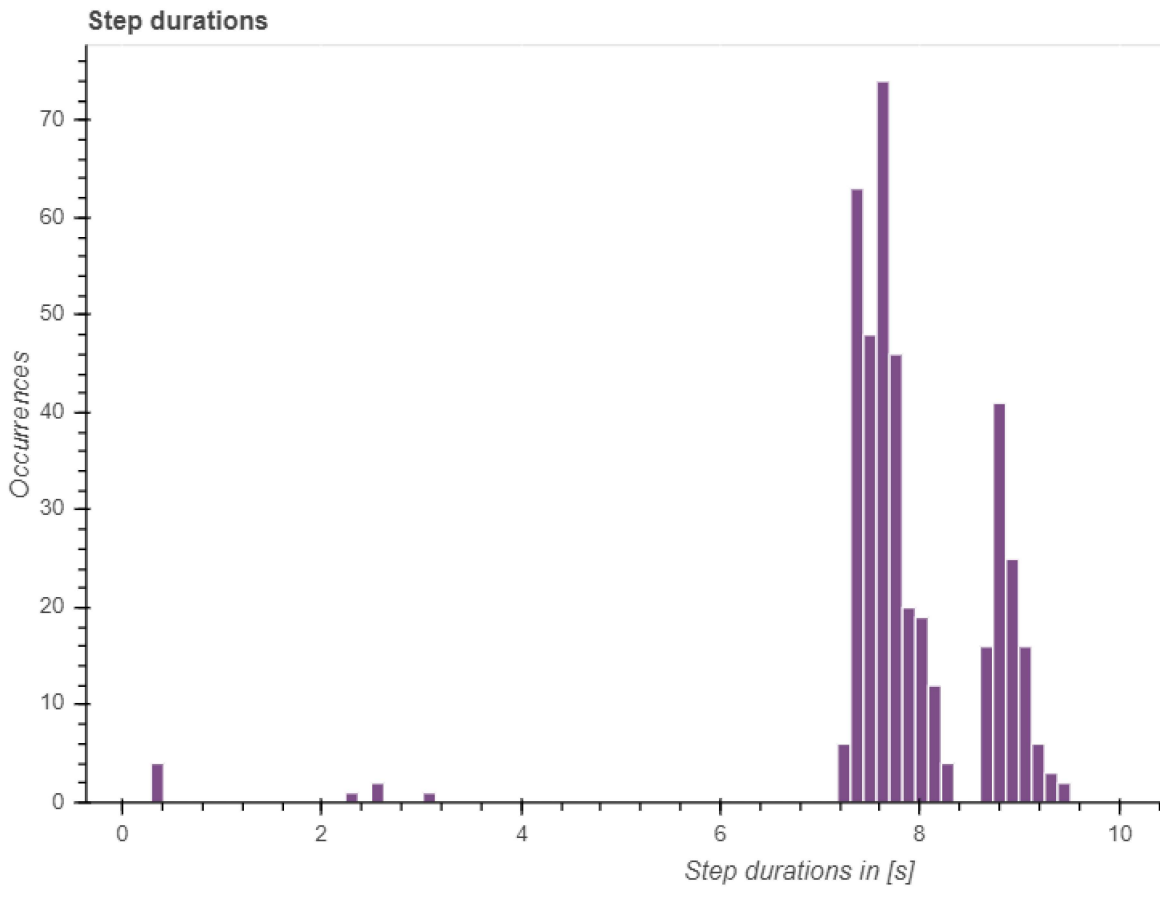
The StepOutlier rule measures step durations and checks for outliers. The rule returns True if duratic
also takes the parameter mode, that specifies whether steps from training or validation phase should
as None. Typically the first step is taking significantly more time and to avoid the rule triggering imme
outliers to ignore. n_outliers was set to 10. The rule analysed 422 datapoints and triggered 36 times.

Step durations on node algo-1-26:

The following table is a summary of the statistics of step durations measured on node algo-1-26. The rule has analyzed the step duration from Step:ModeKeys.GLOBAL phase. The average step duration on node algo-1-26 was 8.13s. The rule detected 12 outliers, where step duration was larger than 3 times the standard deviation of 1.89s

	mean	max	p99	p95	p50	min
Step Durations in [s]	8.13	18.28	16.97	9.14	7.71	0.29

The following histogram shows the step durations measured on the different nodes. You can turn on or turn off the visualization of histograms by selecting or unselecting the labels in the legend.



GPU utilization analysis

Usage per GPU

The LowGPUUtilization rule checks for a low and fluctuating GPU usage. If the GPU usage is consis
bottlenecks or a small batch size. If usage is heavily fluctuating, it can be due to bottlenecks or block
and 5th percentile of GPU utilization on 500 continuous datapoints and found 0 cases where p95 wa
If p95 is high and p5 is low, it might indicate that the GPU usage is highly fluctuating. If both values a
machine is underutilized. During initialization, the GPU usage is likely zero, so the rule skipped the fi
0 datapoints and triggered 0 times.

Workload balancing

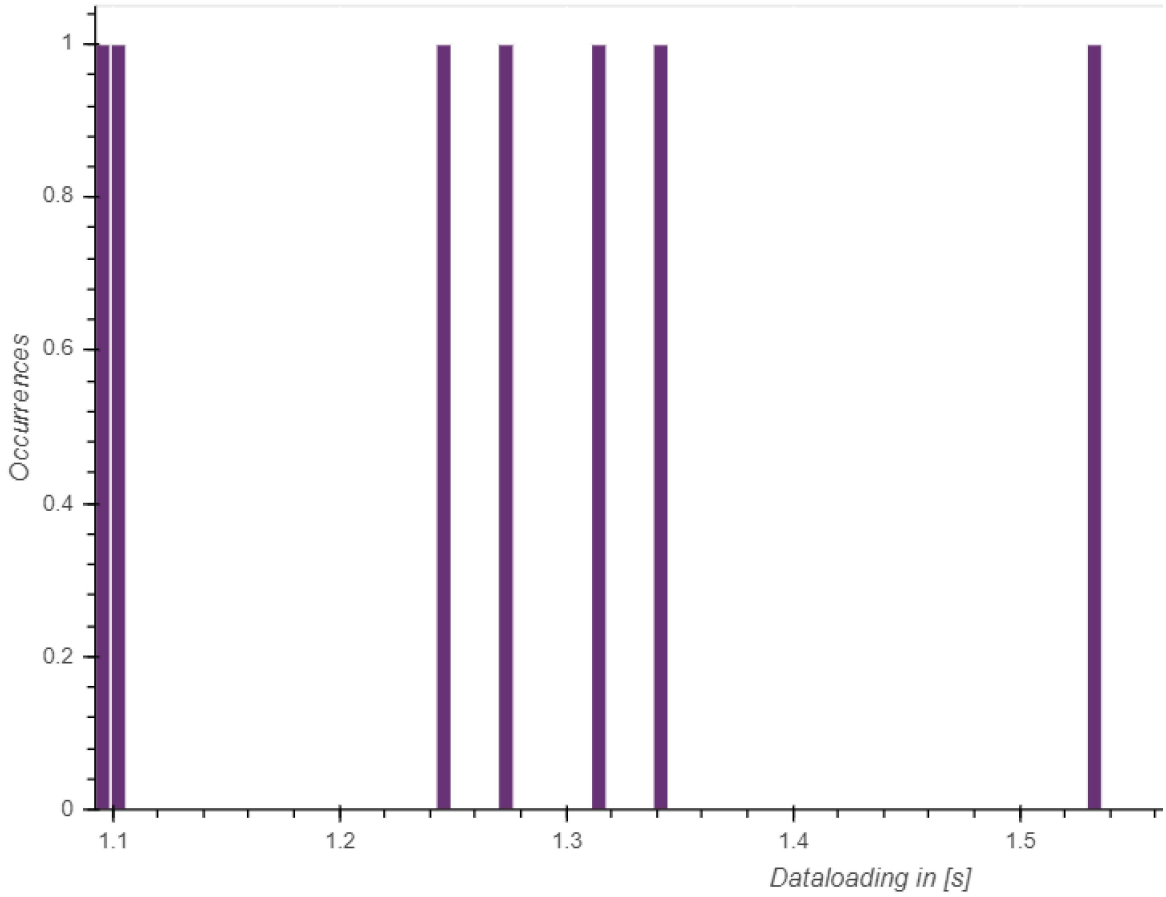
The LoadBalancing rule helps to detect issues in workload balancing between multiple GPUs. It com
GPU and compares then the similarity between histograms. The rule checked if the distance of histo
initialization utilization is likely zero, so the rule skipped the first 1000 data points.

Dataloading analysis

The number of dataloader workers can greatly affect the overall performance of your training job. The
that have been running in parallel on the training instance and compares it against the total number o
smaller than 70% or larger than 200% the total number of cores. Having too few dataloader workers
underutilization. Having too many dataloader workers may hurt the overall performance if you are rur
rule analysed 10 datapoints and triggered 0 times.

Your training instance provided 2 CPU cores, however your training job only ran on average 1 datalo
increase the number of dataloader workers. Using pinned memory also improves performance becau
GPUs. The rule detected that your training job was not using pinned memory. In case of using PyTor
pin_memory=True.

The following histogram shows the distribution of dataloading times that have been measured throug
was 1.3278s. The 95th percentile was 1.7509s and the 25th percentile was 1.2537s



Batch size

The BatchSize rule helps to detect if GPU is underutilized because of the batch size being too small. GPU memory footprint, CPU and GPU utilization. The rule checked if the 95th percentile of CPU utili 70%, the 95th percentile of GPU utilization is below gpu_threshold_p95 of 70% and the 95th percent gpu_memory_threshold_p95 of 70%. In your training job this happened 0 times. The rule skipped the computed the percentiles over window size of 500 continuous datapoints. The rule analysed 7060 da

CPU bottlenecks

The CPUBottleneck rule checked when the CPU utilization was above cpu_threshold of 90% and GF initialization utilization is likely to be zero, so the rule skipped the first 1000 datapoints. With this conf of the total time. This is below the threshold of 50% The rule analysed 10952 data points and triggere

I/O bottlenecks

The IOBottleneck rule checked when I/O wait time was above io_threshold of 50% and GPU utilizatic utilization is likely to be zero, so the rule skipped the first 1000 datapoints. With this configuration the time. This is below the threshold of 50%. The rule analysed 10952 datapoints and triggered 0 times.

GPU memory

The GPUMemoryIncrease rule helps to detect large increase in memory usage on GPUs. The rule cl more than 5.0%. So if the moving average increased for instance from 10% to 16.0%, the rule would so the rule skipped the first 1000 datapoints. The moving average was computed on a window size o violations where the moving average between previous and current time window increased by more t 0 times.