SAMPLE OUTPUT

HARP

12 Threads, 370K Records

```
[guptaapu@j-057 hadoop-2.6.0]$ time hadoop jar contrib-0.1.0.jar edu.iu.rf.RFMapCollective 32 2 12 /data/train_test_files/train_airline_harp_5.csv /data/trai
n_test_files/merged_test_airline_harp_5.csv /out_32_2_12_harp
Get map collective runner
```
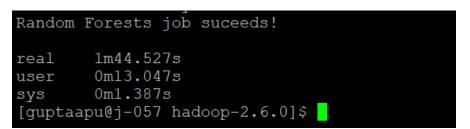
Memory Consumption   ▼ ⑦   INTEL VTUNE AMPLIFIER 2019

Analysis Configuration   **Collection Log**   >

  ⋰⋱   **Collecting Memory Consumption data**   🗑

        Dec 02 2018 16:39:36 The Intel VTune Amplifier 2019 is collecting data. Closing this window will cancel the analysis and terminate the profiled application in
        case the Launch Application target type is selected.

```
Random Forests job suceeds!

real    1m44.527s
user    0m13.047s
sys     0m1.387s
[guptaapu@j-057 hadoop-2.6.0]$ █
```

Memory Consumption Memory Consumption ▼ ⑦   INTEL VTUNE AMPLIFIER 2019

Analysis Configuration   Collection Log   **Summary**   Bottom-up

⊙ **Elapsed Time** ⑦ : **87.646s**

    Allocation Size: 65 MB
    Deallocation Size: 53 MB
    Allocations: 48,447
    Total Thread Count: 63
    Paused Time ⑦ :   0s

⊙ **Top Memory-Consuming Functions**
    This section lists the most memory-consuming functions in your application.

Function	Memory Consumption	Allocation/Deallocation Delta	Allocations	Module
os::malloc	39 MB	8 MB	2,847	libjvm.so
init_tpssmrte_jstring	19 MB	896 KB	10,086	libamplxe_tpssmrte_javaprof_1.0.so
os::Linux::commit_memory_impl	1 MB	1 MB	31	libjvm.so
Interpreter	1 MB	193 KB	704	[Dynamic code]
java::lang::ClassLoader::defineClass1	867 KB	557 KB	2,523	[Compiled Java code]
[Others]	1 MB	152 KB	32,256	

N/A is applied to non-summable metrics.

Collection and Platform Info

This section provides information about this collection, including result set size and collection platform data.

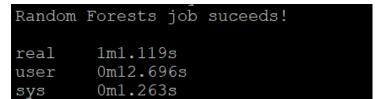|	
Application Command Line:	
Operating System:	3.10.0-862.9.1.el7.x86_64 NAME="Red Hat Enterprise Linux Server" VERSION="7.5 (Maipo)" ID="rhel" ID_LIKE="fedora" VARIANT="Server" VARIANT_ID="server" VERSION_ID="7.5" PRETTY_NAME="Red Hat Enterprise Linux Server 7.5 (Maipo)" ANSI_COLOR="0;31" CPE_NAME="cpe:/o:redhat:enterprise_linux:7.5:GA:server" HOME_URL="https://www.redhat.com/" BUG_REPORT_URL="https://bugzilla.redhat.com/" REDHAT_BUGZILLA_PRODUCT="Red Hat Enterprise Linux 7" REDHAT_BUGZILLA_PRODUCT_VERSION=7.5 REDHAT_SUPPORT_PRODUCT="Red Hat Enterprise Linux" REDHAT_SUPPORT_PRODUCT_VERSION="7.5"
Computer Name:	j-057
Result Size:	65 MB
Collection start time:	21:39:36 02/12/2018 UTC
Collection stop time:	21:41:04 02/12/2018 UTC
Collector Type:	User-mode sampling and tracing

Finalization mode: Fast. If the number of collected samples exceeds the threshold, this mode limits the number of processed samples to speed up post-processing.

CPU

Name:	Intel(R) Xeon(R) E5/E7 v3 Processor code named Haswell
Frequency:	2.3 GHz
Logical CPU Count:	48

24 Threads, 370K Records

```
[guptaapu@j-057 hadoop-2.6.0]$ time hadoop jar contrib-0.1.0.jar edu.iu.rf.RFMapCollective 32 2 24 /data/train_test_files/train_airline_harp_5.csv /data/train_test_files/merged_test_airline_harp_5.csv /out_32_2_24_harp
Get map collective runner
```

```
Random Forests job suceeds!

real    1m1.119s
user    0m12.696s
sys     0m1.263s
```

Memory Consumption Memory Consumption ▾ ⑦   INTEL VTUNE AMPLIFIER

Analysis Configuration   Collection Log   Summary   Bottom-up

Elapsed Time ⑦: 38.964s

Allocation Size:	40 MB
Deallocation Size:	32 MB
Allocations:	23,609
Total Thread Count:	62
Paused Time ⑦:	0s

Top Memory-Consuming Functions

This section lists the most memory-consuming functions in your application.

Function	Memory Consumption	Allocation/Deallocation Delta	Allocations	Module
os::malloc	26 MB	6 MB	1,533	libjvm.so
init_tpssmrte_jstring	9 MB	356 KB	4,766	libamplxe_tpssmrte_javaprof_1.0.so
Interpreter	1 MB	310 KB	554	[Dynamic code]
java::lang::ClassLoader::defineClass1	835 KB	525 KB	2,522	[Compiled Java code]
os::Linux::commit_memory_impl	544 KB	520 KB	12	libjvm.so
[Others]	1002 KB	105 KB	14,222	

*N/A is applied to non-summable metrics.

Collection and Platform Info

This section provides information about this collection, including result set size and collection platform data.

Application Command Line:	
Operating System:	3.10.0-862.9.1.el7.x86_64 NAME="Red Hat Enterprise Linux Server" VERSION="7.5 (Maipo)" ID="rhel" ID_LIKE="fedora" VARIANT="Server" VARIANT_ID="server" VERSION_ID="7.5" PRETTY_NAME="Red Hat Enterprise Linux Server 7.5 (Maipo)" ANSI_COLOR="0;31" CPE_NAME="cpe:/o:redhat:enterprise_linux:7.5:GA:server" HOME_URL="https://www.redhat.com/" BUG_REPORT_URL="https://bugzilla.redhat.com/" REDHAT_BUGZILLA_PRODUCT="Red Hat Enterprise Linux 7" REDHAT_BUGZILLA_PRODUCT_VERSION=7.5 REDHAT_SUPPORT_PRODUCT="Red Hat Enterprise Linux" REDHAT_SUPPORT_PRODUCT_VERSION="7.5"
Computer Name:	j-057
Result Size:	56 MB
Collection start time:	21:43:51 02/12/2018 UTC
Collection stop time:	21:44:30 02/12/2018 UTC
Collector Type:	User-mode sampling and tracing

Finalization mode: Fast. If the number of collected samples exceeds the threshold, this mode limits the number of processed samples to speed up post-processing.
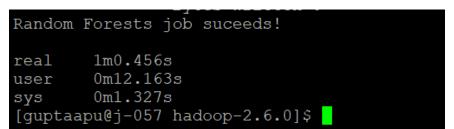
⊙ CPU

Name:	Intel(R) Xeon(R) E5/E7 v3 Processor code named Haswell
Frequency:	2.3 GHz
Logical CPU Count:	48

48 Threads, 370 K Records

```
[guptaapu@j-057 hadoop-2.6.0]$ time hadoop jar contrib-0.1.0.jar edu.iu.rf.RFMapCollective 32 2 48 /data/train_test_files/train_airline_harp_5.csv /data/train_test_files/merged_test_airline_harp_5.csv /out_32_2_48_harp
Get map collective runner
```

```
Random Forests job suceeds!

real    1m0.456s
user    0m12.163s
sys     0m1.327s
[guptaapu@j-057 hadoop-2.6.0]$
```

Ant Memory Consumption Memory Consumption ▾ ⑦ INTEL VTUNE AMPLIFIE

Analysis Configuration   Collection Log   **Summary**   Bottom-up

⊙ Elapsed Time ⑦: 44.688s

Allocation Size:	42 MB
Deallocation Size:	33 MB
Allocations:	24,787
Total Thread Count:	63
Paused Time ⑦:	0s

⊙ Top Memory-Consuming Functions

This section lists the most memory-consuming functions in your application.

Function	Memory Consumption	Allocation/Deallocation Delta	Allocations	Module
os::malloc	28 MB	6 MB	1,617	libjvm.so
init_tpssmrte_jstring	9 MB	384 KB	5,012	libamplxe_tpssmrte_javaprof_1.0.so
Interpreter	1 MB	343 KB	571	[Dynamic code]
os::Linux::commit_memory_impl	864 KB	840 KB	14	libjvm.so
java::lang::ClassLoader::defineClass1	803 KB	493 KB	2,521	[Compiled Java code]
[Others]	1007 KB	169 KB	15,052	

*N/A is applied to non-summable metrics.

Collection and Platform Info

This section provides information about this collection, including result set size and collection platform data.

Application Command Line:
Operating System: 3.10.0-862.9.1.el7.x86_64 NAME="Red Hat Enterprise Linux Server" VERSION="7.5 (Maipo)" ID="rhel" ID_LIKE="fedora" VARIANT="Server" VARIANT_ID="server" VERSION_ID="7.5" PRETTY_NAME="Red Hat Enterprise Linux Server 7.5 (Maipo)" ANSI_COLOR="0;31" CPE_NAME="cpe:/o:redhat:enterprise_linux:7.5:GA:server" HOME_URL="https://www.redhat.com/" BUG_REPORT_URL="https://bugzilla.redhat.com/" REDHAT_BUGZILLA_PRODUCT="Red Hat Enterprise Linux 7" REDHAT_BUGZILLA_PRODUCT_VERSION=7.5 REDHAT_SUPPORT_PRODUCT="Red Hat Enterprise Linux" REDHAT_SUPPORT_PRODUCT_VERSION="7.5"

Computer Name: j-057
Result Size: 56 MB
Collection start time: 21:46:30 02/12/2018 UTC
Collection stop time: 21:47:15 02/12/2018 UTC
Collector Type: User-mode sampling and tracing

Finalization mode: Fast. If the number of collected samples exceeds the threshold, this mode limits the number of processed samples to speed up post-processing.

CPU

Name: Intel(R) Xeon(R) E5/E7 v3 Processor code named Haswell
Frequency: 2.3 GHz
Logical CPU Count: 48

SPARK

12 Threads, 370 K Records

```
[guptaapu@j-057 spark-2.3.2-bin-hadoop2.7]$ time bin/spark-submit rf_spark.py /data/train_test_files/train_airline_spark_5.csv /data/train_test_files/test_ai
rline_spark_5.csv 32
2018-12-02 16:54:41 WARN  NativeCodeLoader:62 - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2018-12-02 16:54:42 INFO  SparkContext:54 - Running Spark version 2.3.2
2018-12-02 16:54:42 INFO  SparkContext:54 - Submitted application: AirlineApp
```

```
Building Random Forest Model..
Measuring test results:
Test Error = 0.0699952
f1 = 0.927965


real    0m25.646s
user    2m42.103s
sys     0m14.374s
```

Memory Consumption   Memory Consumption ▾ ⑦   INTEL VTUNE AMPLIFIER 2019

Analysis Configuration   Collection Log   Summary   Bottom-up

Elapsed Time ⑦: 8.355s

Allocation Size: 2 MB
Deallocation Size: 2 MB
Allocations: 2,864
Total Thread Count: 2
Paused Time ⑦: 0s

Top Memory-Consuming Functions

This section lists the most memory-consuming functions in your application.

Function	Memory Consumption	Allocation/Deallocation Delta	Allocations	Module
sock_recv	2 MB	0 B	645	_socket.so
copy	96 KB	0 B	24	copy.py
__init__	71 KB	9 KB	95	java_gateway.py
__init__	69 KB	3 KB	92	java_gateway.py
IO_StringIO	40 KB	128 B	322	cStringIO.so
[Others]	178 KB	288 B	1,686	

*N/A is applied to non-summable metrics.

48 Threads, 370 K Records

```
[guptaapu@j-057 spark-2.3.2-bin-hadoop2.7]$ time bin/spark-submit rf_spark.py /data/train_test_files/train_airline_spark_5.csv /data/train_test_files/test_ai
rline_spark_5.csv 32
2018-12-02 16:51:02 WARN  NativeCodeLoader:62 - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2018-12-02 16:51:03 INFO  SparkContext:54 - Running Spark version 2.3.2
2018-12-02 16:51:03 INFO  SparkContext:54 - Submitted application: AirlineApp
```

```
Building Random Forest Model..
Measuring test results:
Test Error = 0.0699952
f1 = 0.927965


real    0m27.808s
user    3m0.972s
sys     0m14.865s
```

Memory Consumption Memory Consumption ▾ ⑦   INTEL VTUNE AMPLIFIER 2019

Analysis Configuration   Collection Log   Summary   Bottom-up

⌄ Elapsed Time ⑦: 10.645s

Allocation Size:	2 MB
Deallocation Size:	2 MB
Allocations:	2,869
Total Thread Count:	2
Paused Time ⑦:	0s

⌄ Top Memory-Consuming Functions

This section lists the most memory-consuming functions in your application.

Function	Memory Consumption	Allocation/Deallocation Delta	Allocations	Module
sock_recv	2 MB	0 B	645	_socket.so
copy	96 KB	0 B	24	copy.py
__init__	71 KB	9 KB	95	java_gateway.py
__init__	69 KB	3 KB	92	java_gateway.py
IO_StringIO	40 KB	128 B	322	cStringIO.so
[Others]	190 KB	12 KB	1,691	

N/A is applied to non-summable metrics.

⌄ Collection and Platform Info

This section provides information about this collection, including result set size and collection platform data.

Application Command Line:	
Operating System:	3.10.0-862.9.1.el7.x86_64 NAME="Red Hat Enterprise Linux Server" VERSION="7.5 (Maipo)" ID="rhel" ID_LIKE="fedora" VARIANT="Server" VARIANT_ID="server" VERSION_ID="7.5" PRETTY_NAME="Red Hat Enterprise Linux Server 7.5 (Maipo)" ANSI_COLOR="0;31" CPE_NAME="cpe:/o:redhat:enterprise_linux:7.5:GA:server" HOME_URL="https://www.redhat.com/" BUG_REPORT_URL="https://bugzilla.redhat.com/" REDHAT_BUGZILLA_PRODUCT="Red Hat Enterprise Linux 7" REDHAT_BUGZILLA_PRODUCT_VERSION=7.5 REDHAT_SUPPORT_PRODUCT="Red Hat Enterprise Linux" REDHAT_SUPPORT_PRODUCT_VERSION="7.5"
Computer Name:	j-057
Result Size:	4 MB
Collection start time:	21:51:17 02/12/2018 UTC
Collection stop time:	21:51:28 02/12/2018 UTC
Collector Type:	User-mode sampling and tracing

Finalization mode: Fast. If the number of collected samples exceeds the threshold, this mode limits the number of processed samples to speed up post-processing.

⌄ CPU

Name:	Intel(R) Xeon(R) E5/E7 v3 Processor code named Haswell
Frequency:	2.3 GHz
Logical CPU Count:	48