

# Threading Performance Measurements with VTune



Christopher Jones *FNAL*



# *What is VTune?*

VTune is an Intel performance measurement system

Can measure

Concurrency efficiency

Cross threads locks and waits

CPU hardware counters

GPU performance

Can do measurements from the GUI or command line tool

Data generated by command line tool can be read into the GUI for analysis

OpenLab has a cross CERN license

[https://twiki.cern.ch/twiki/bin/view/Openlab/IntelTools#Intel\\_Parallel\\_Studio\\_XE\\_201\\_ANI](https://twiki.cern.ch/twiki/bin/view/Openlab/IntelTools#Intel_Parallel_Studio_XE_201_ANI)

# *Using VTune with cmsRun*



VTune did not work 'out of the box' on cmsRun

Steps needed to work

**Had to use most recent VTune beta**

older releases all had assertion failures

`/afs/cern.ch/sw/IntelSoftware/linux/x86_64/xe2016/vtune_amplifier_xe_2016.1.0.424694`

**Tell VTune to ignore processes started by ROOT**

`-strategy=ld-linux.so.2:nt:nt,ld-linux-x86-64.so.2:nt:nt`

VTune is unable to instrument these processes

**Had to enable signal 38**

Problem seen

**GUI sometimes only shows measurements from some of the threads**

**Do not know if failure was during data collection or in viewing**



# Measurement

---

Replicated Tier 0 job

cmsRun configuration changes

Max events of 992

Number of threads 8

VTune configuration

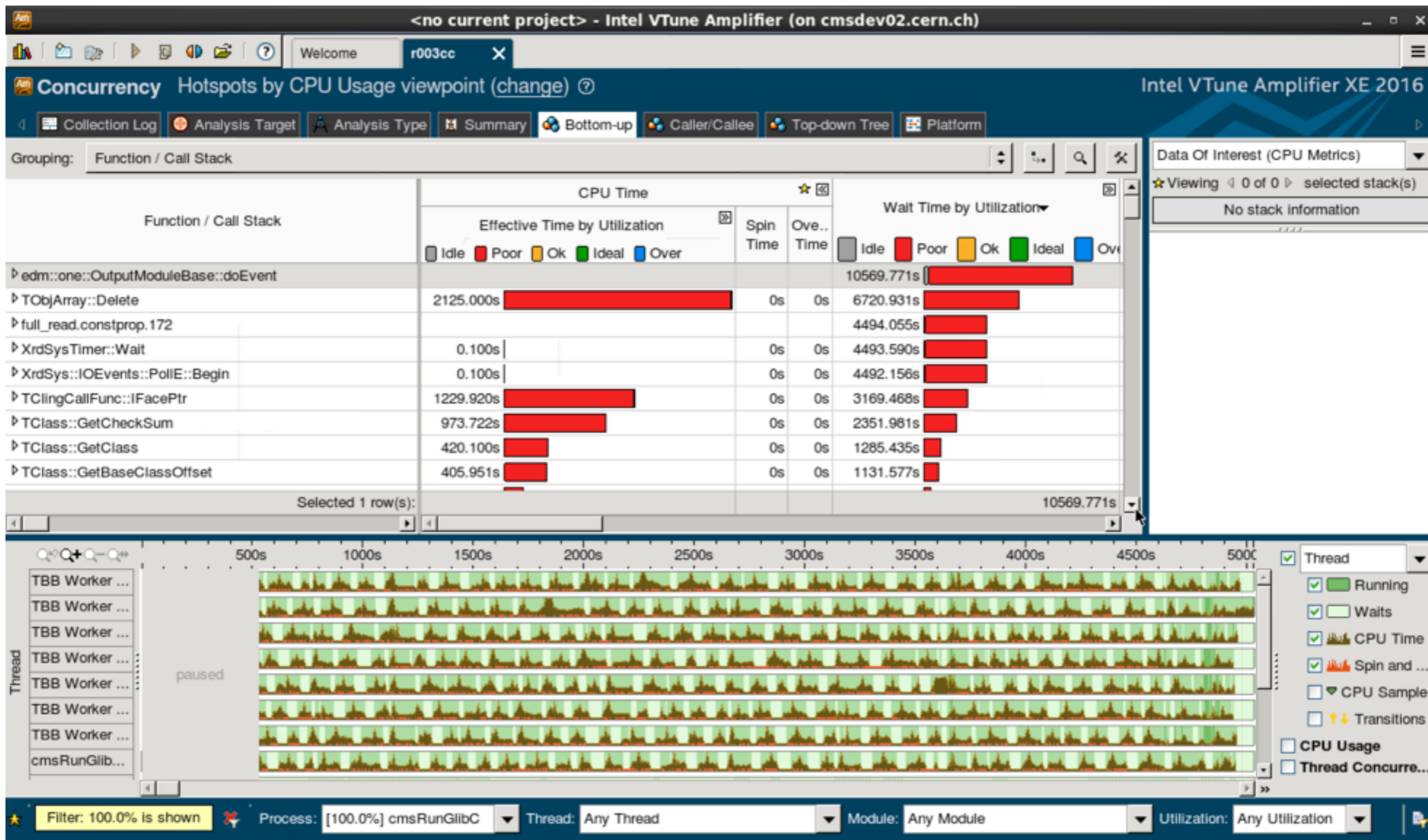
concurrency measurement

skip first 5 minutes of job

sample every 100us

unlimited data collection size

# VTune GUI



# VTune GUI

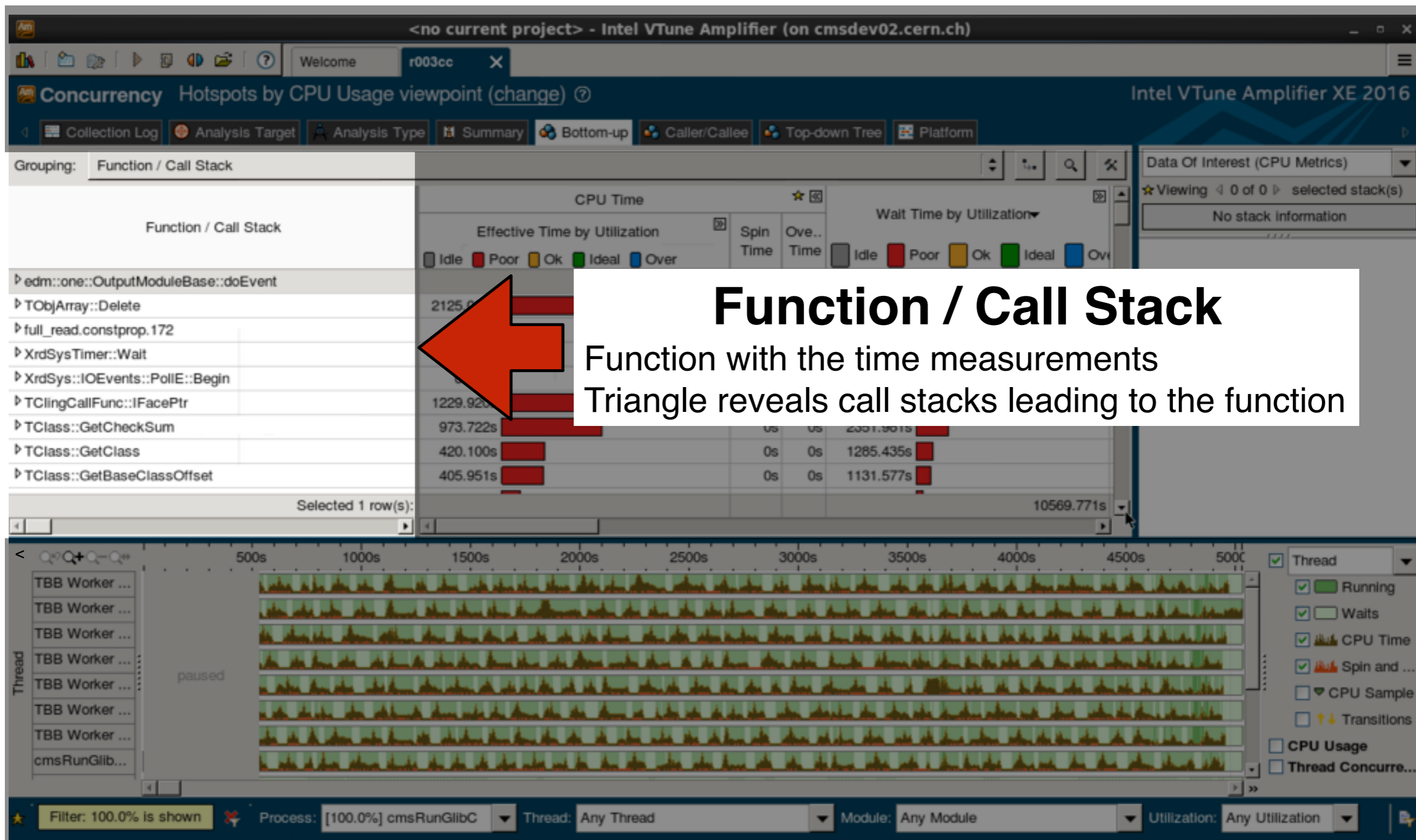


The screenshot shows the Intel VTune Amplifier XE 2016 interface. At the top, the title bar reads "<no current project> - Intel VTune Amplifier (on cmsdev02.cern.ch)". Below the title bar, there's a navigation bar with tabs for "Collection Log", "Analysis Target", "Analysis Type", "Summary", "Bottom-up", "Caller/Callee", "Top-down Tree", and "Platform". The "Bottom-up" mode is currently selected. The main area displays a call stack with columns for "Function / Call Stack", "CPU Time", "Effective Time by Utilization", and "Wait Time by Utilization". A red arrow points to the "Bottom-up" tab. Below the call stack, there's a "Thread" view showing a timeline of CPU usage for multiple threads, with a legend on the right indicating "Running", "Waits", "CPU Time", "Spin and ...", "CPU Sample", "Transitions", "CPU Usage", and "Thread Concurr...".

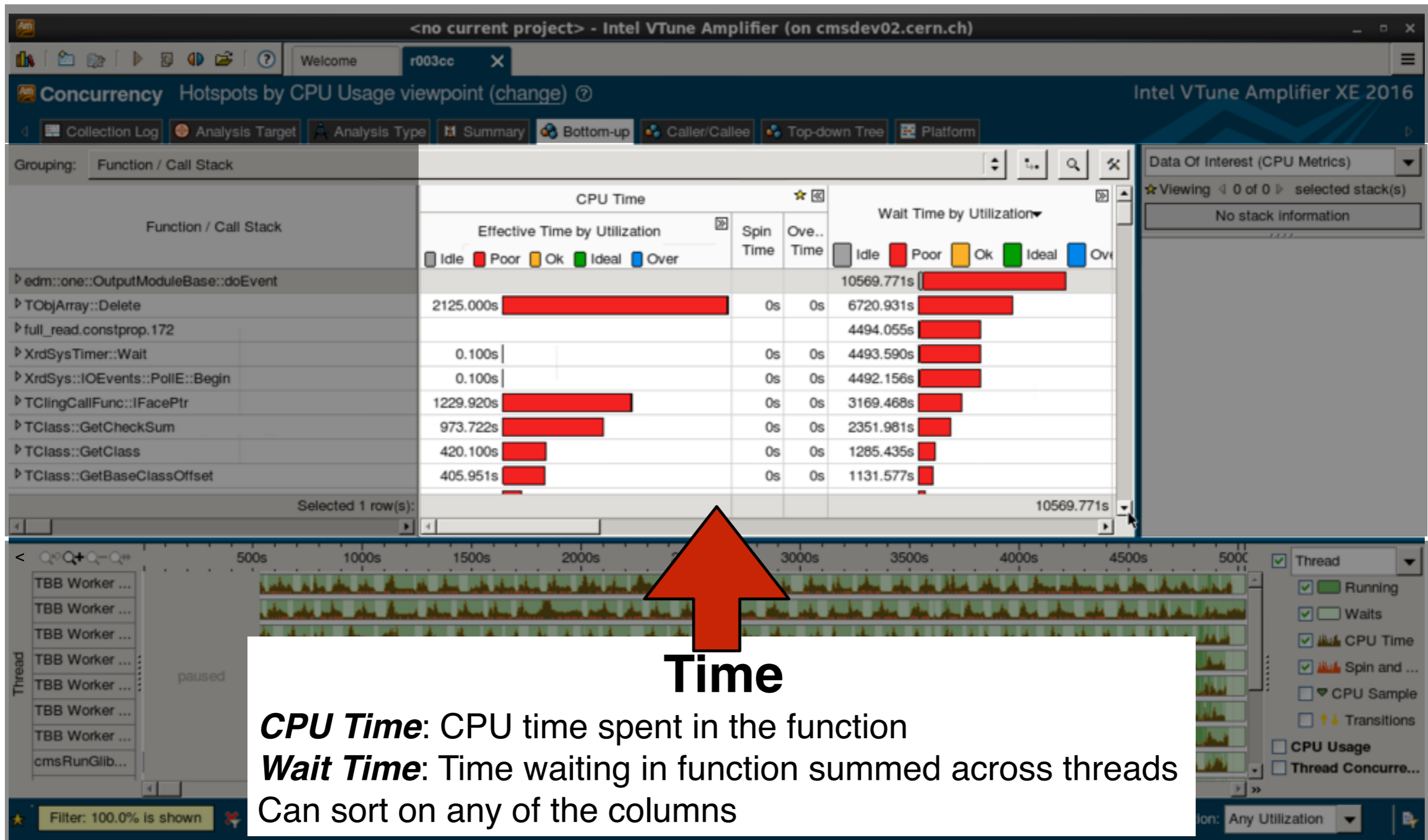
**Modes**

- Summary:** high level view of results
- Bottom-up:** accumulated time in function for all stack traces
- Caller/Callee:** show time in call stack for selected function
- Top-down:** tree view of time spent in stack traces

# VTune GUI



# VTune GUI





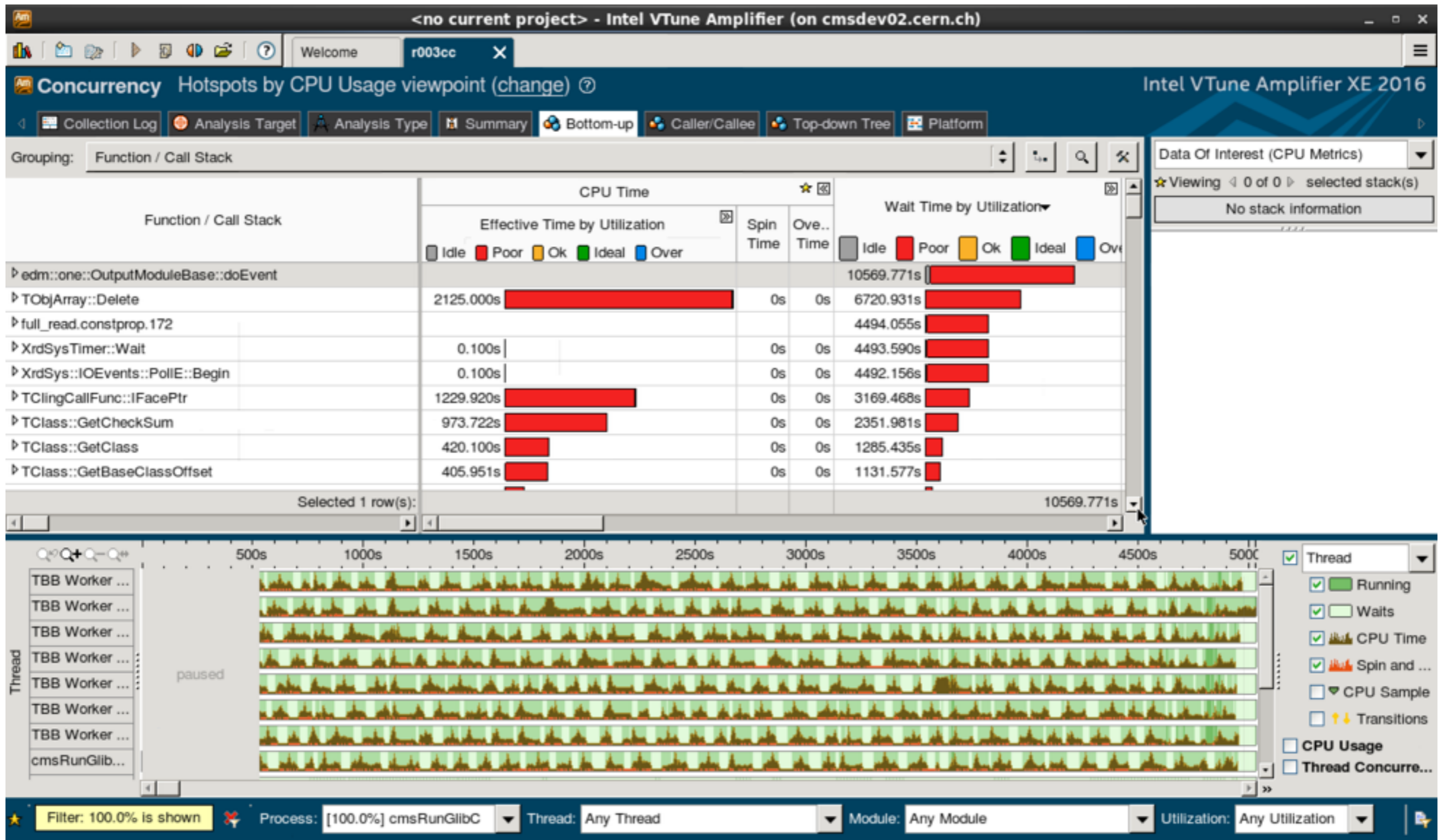
# VTune GUI



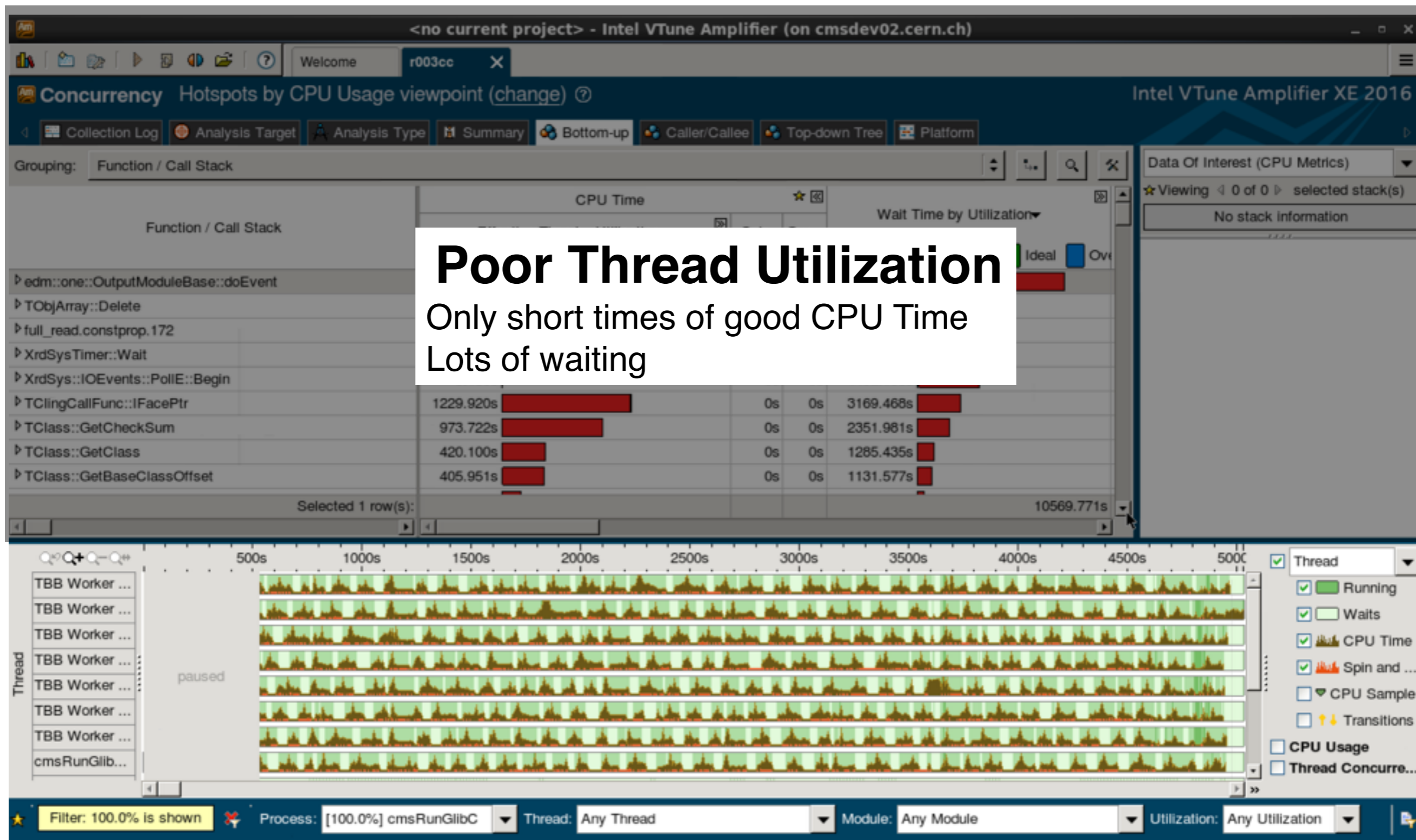
**Thread View**  
Utilization of each thread over time  
Can filter based on time  
Can select waits and see function selected above

The screenshot shows the Intel VTune Amplifier XE 2016 interface. The top window displays 'Concurrency Hotspots by CPU Usage viewpoint'. Below this, there are navigation tabs for 'Collection Log', 'Analysis Target', 'Analysis Type', 'Summary', 'Bottom-up', 'Caller/Callee', 'Top-down Tree', and 'Platform'. The main area shows a table of CPU Time for various functions. A red arrow points from the text box to the 'Thread View' section at the bottom of the screenshot, which displays a timeline of thread utilization for several 'TBB Worker' threads and 'cmsRunGlib...' threads. The timeline shows CPU Time (green) and Waits (yellow) for each thread. The bottom status bar includes filters for 'Filter: 100.0% is shown', 'Process: [100.0%] cmsRunGlibC', 'Thread: Any Thread', 'Module: Any Module', and 'Utilization: Any Utilization'.

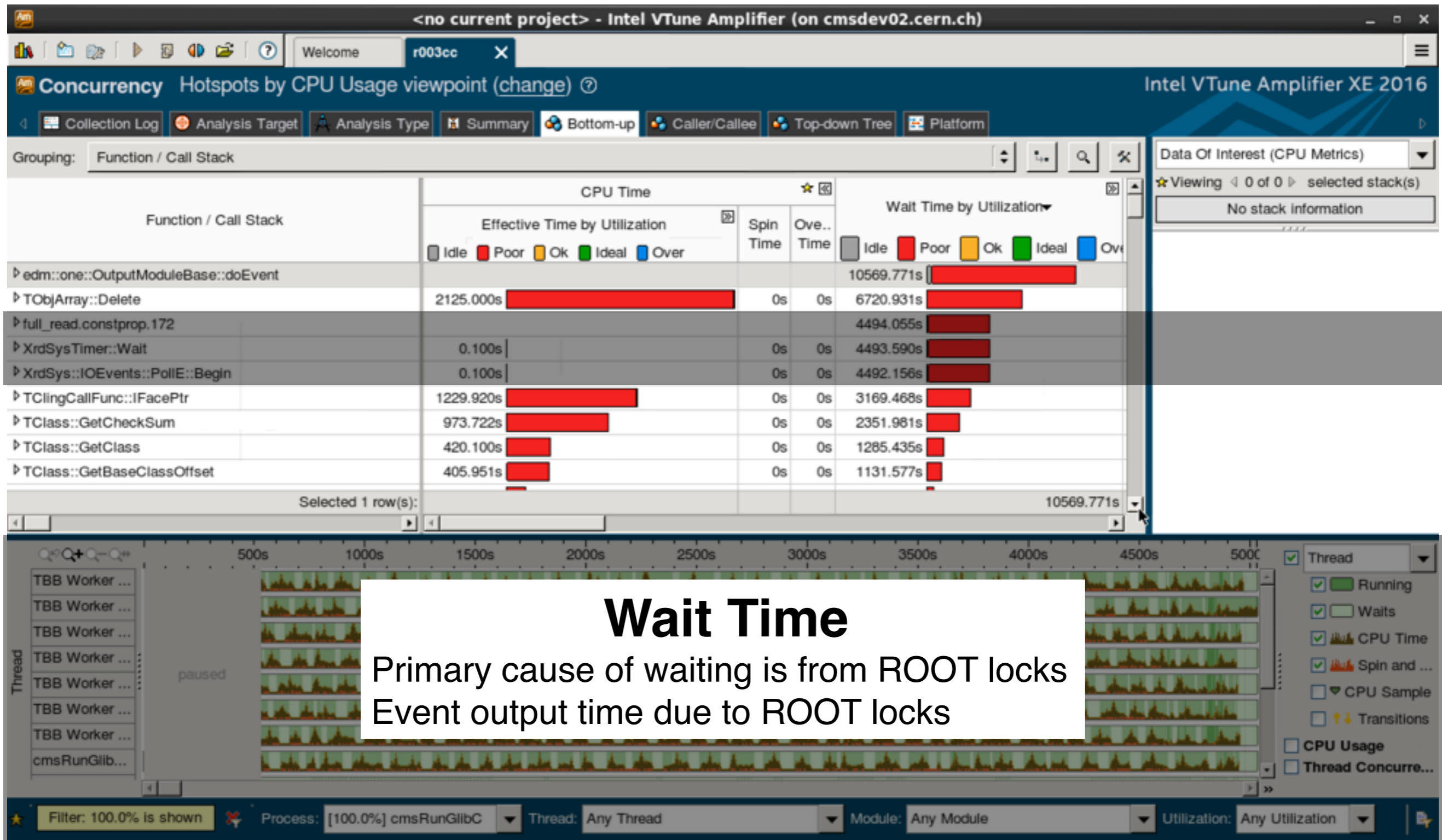
# Initial Measurement



# Initial Measurement



# Initial Measurement



# *Initial Findings*

Waiting on TFormula

A new TFormula made each time a jet correction was called

Waiting on ROOT calls from cut parser

Calling functions via the ROOT interface used always took a lock

Waiting on output

1/2 time waiting to talk to a particular instance of output module

1/2 minutes waiting on ROOT lock during TTree::Fill

Vast majority from TClass::GetChecksum

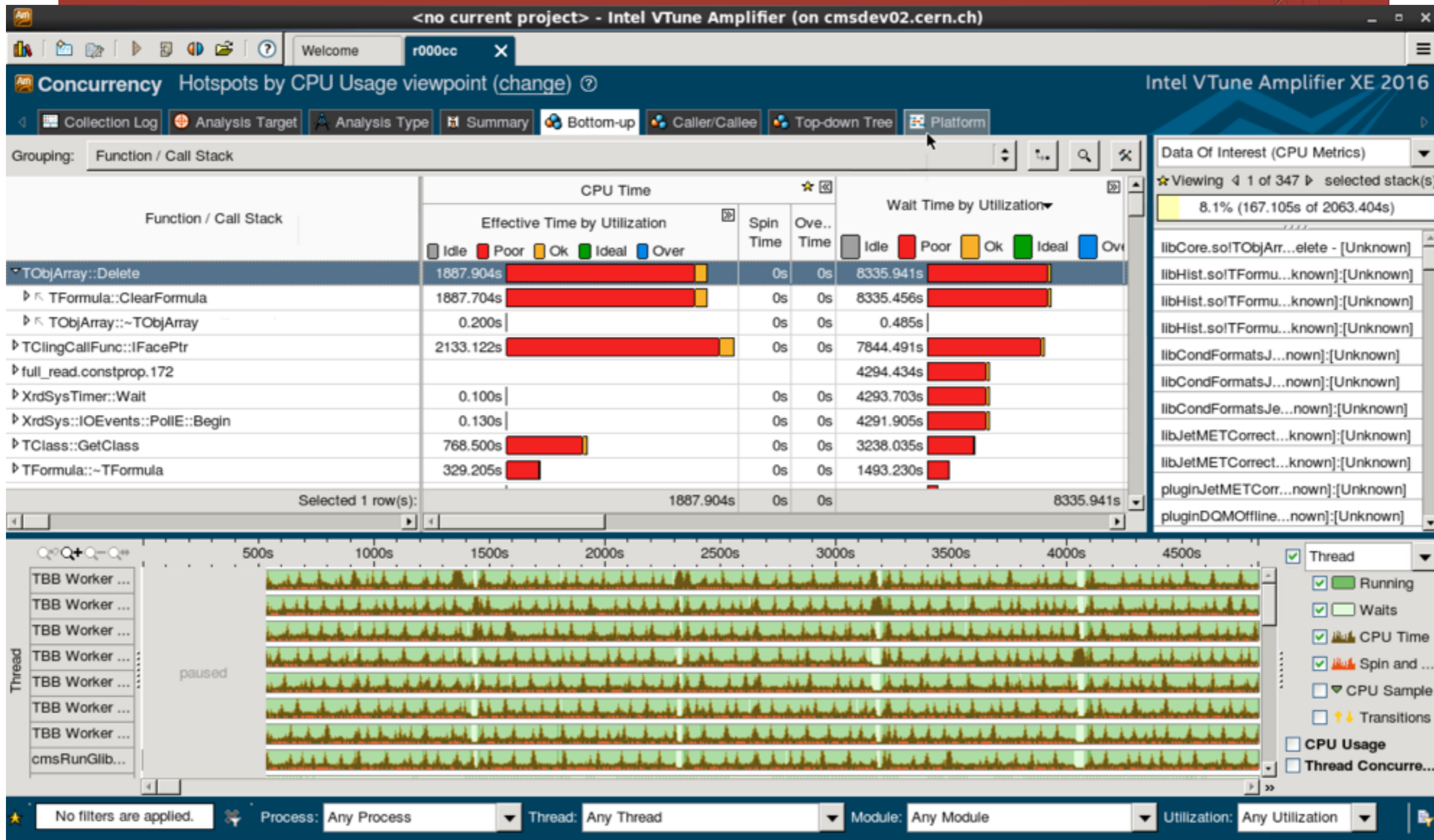
Philippe Canal made changes to ROOT to reduce lock use

TFormula copying should avoid taking locks

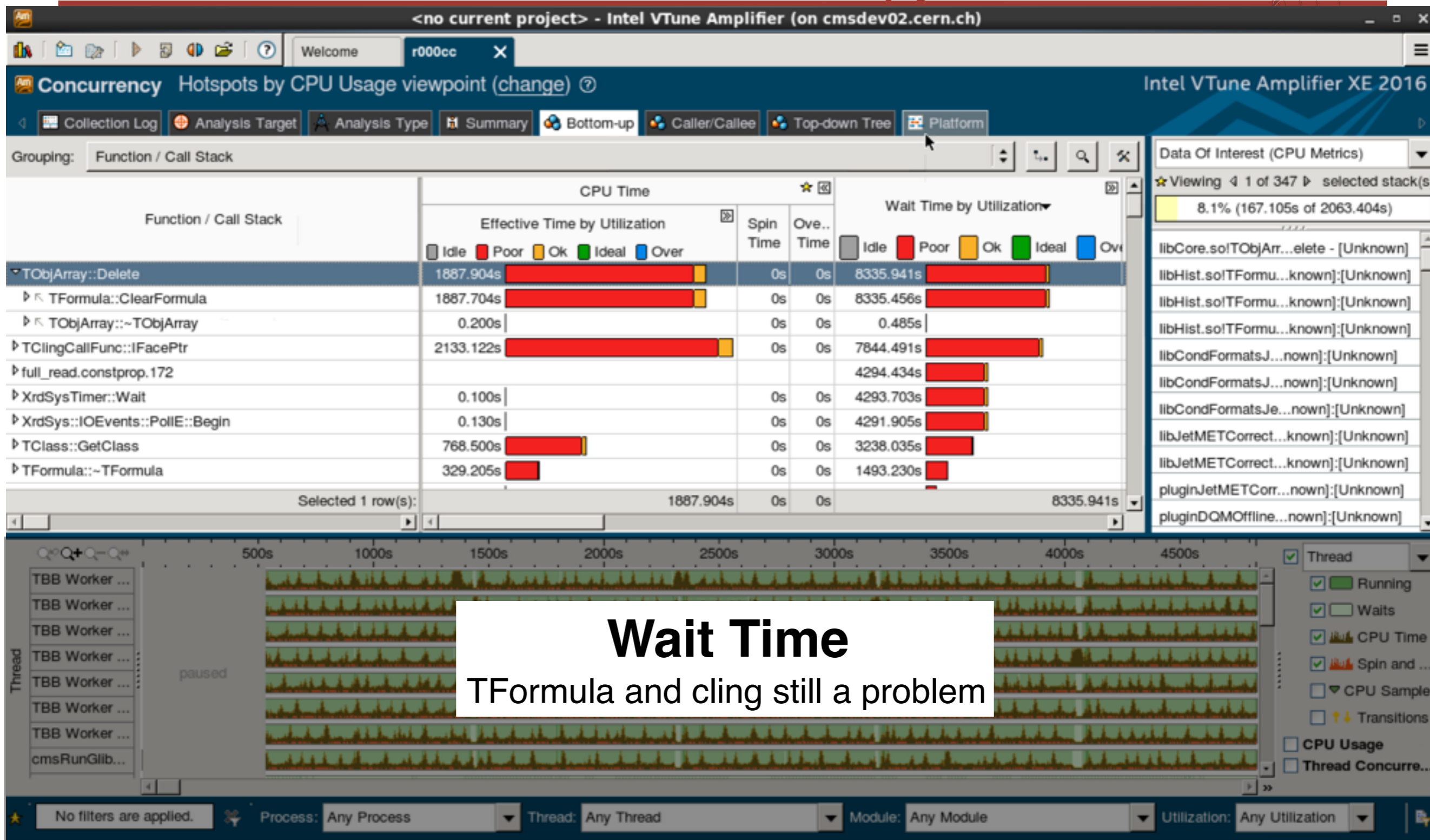
Calling TTree::Fill will avoid most locks

Decreasing the cost of using ROOT to find base class offsets

# Post ROOT Changes



# Post ROOT Changes





# *Post ROOT Findings*

---

Cut parser changes needed

Philippe Canal added API to TMethod to allow caching of function pointer from cling

TFormula changes were insufficient

Problem mitigated in ROOT 6.04 with new TFormula implementation

Replaced with a hand made parser and executor

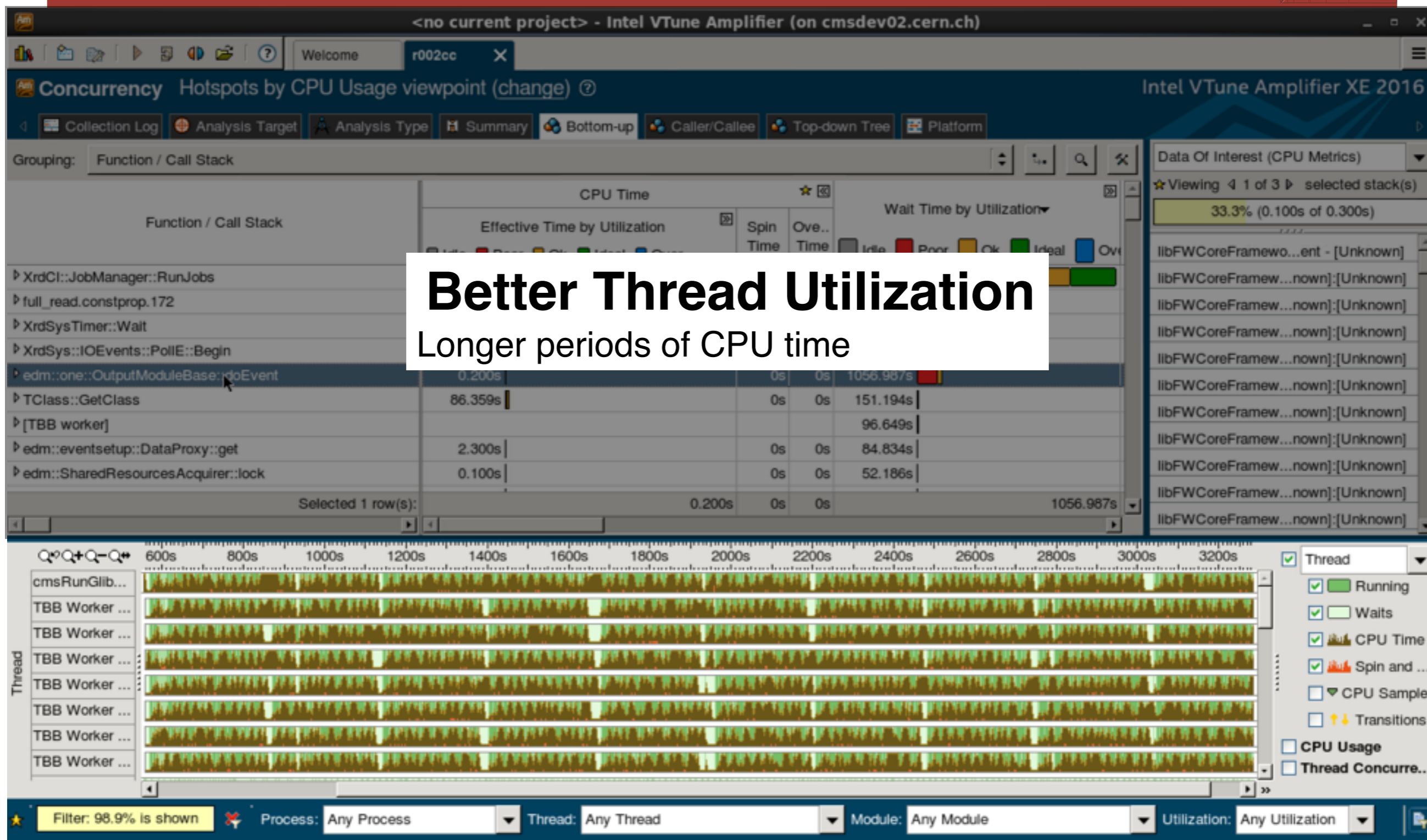
- handles a subset of TFormula expressions

- fully stateless parser and executor so extremely thread efficient

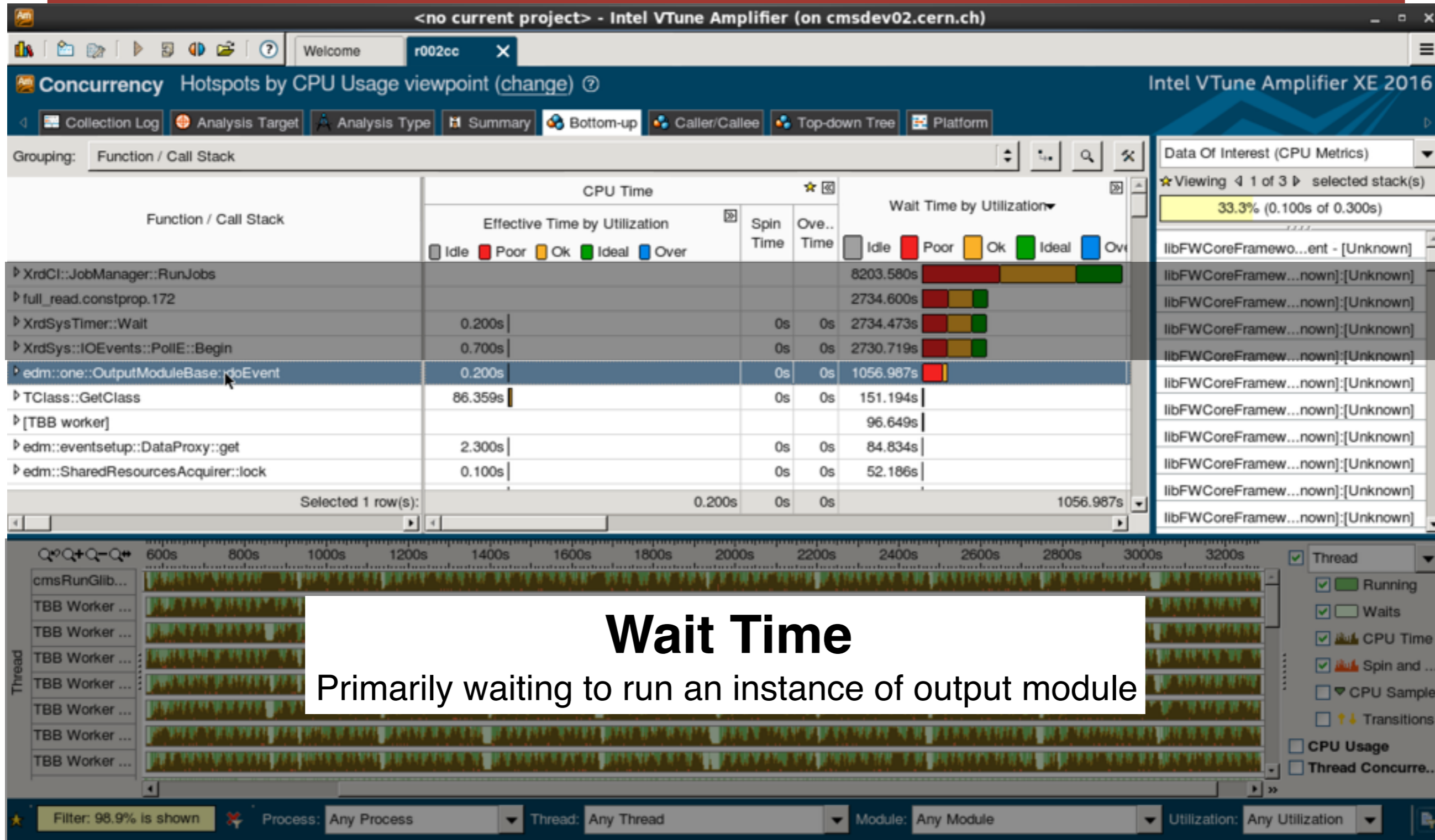
- small memory footprint



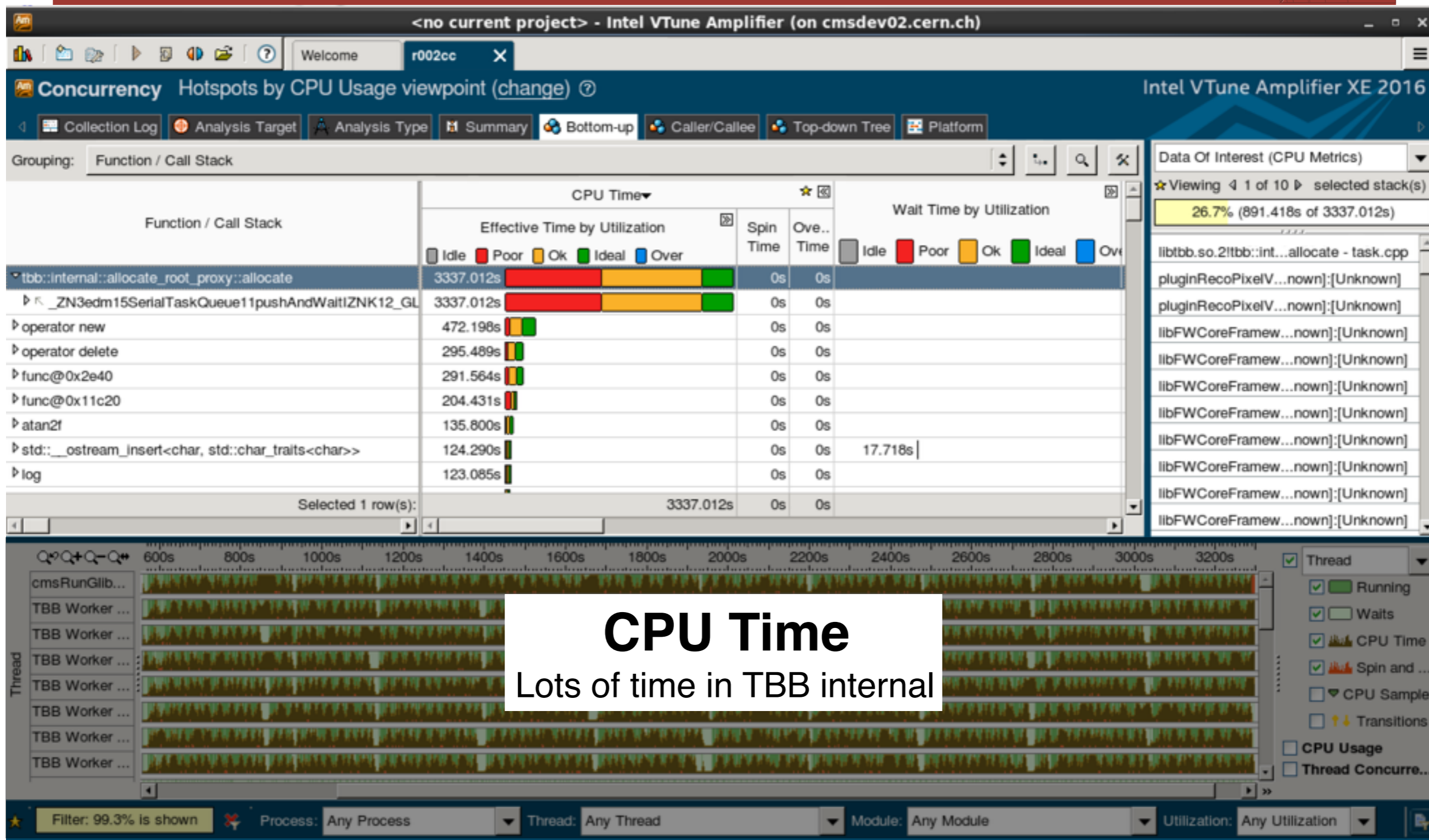
# First CMSSW Changes



# First CMSSW Changes



# First CMSSW Changes



# Findings

TBB time from calls to SerialTaskQueue

CMS class meant to protect a resource from simultaneous access

SerialTaskQueue inappropriately used in an ‘event’ data class

Class was doing a lazy evaluation of elements in its container

First request for each item put task into the SerialTaskQueue

Code wasn’t actually thread safe anyway

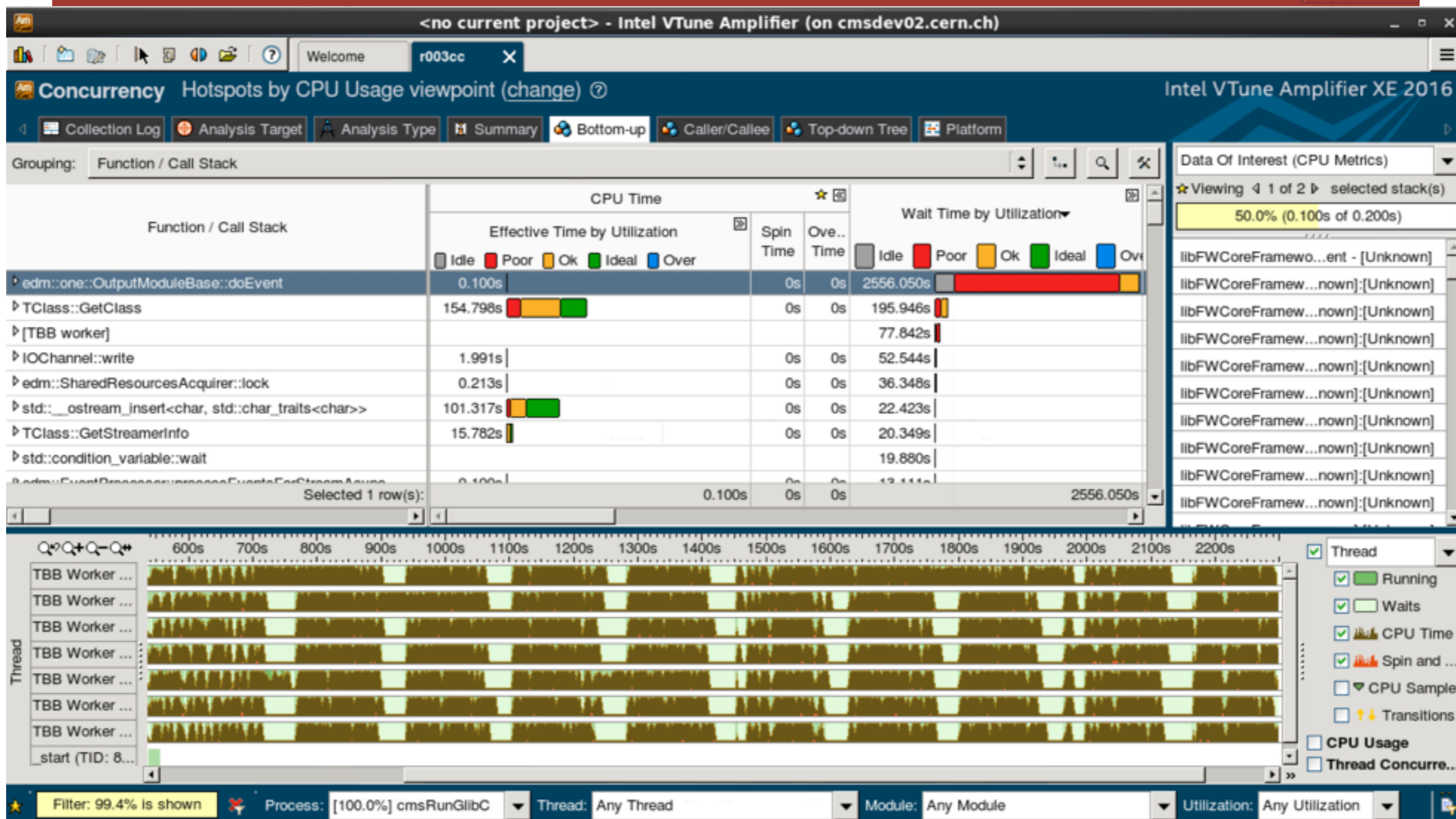
other pieces of code were accessing the container outside of the SerialTaskQueue

Change: remove the use of SerialTaskQueue

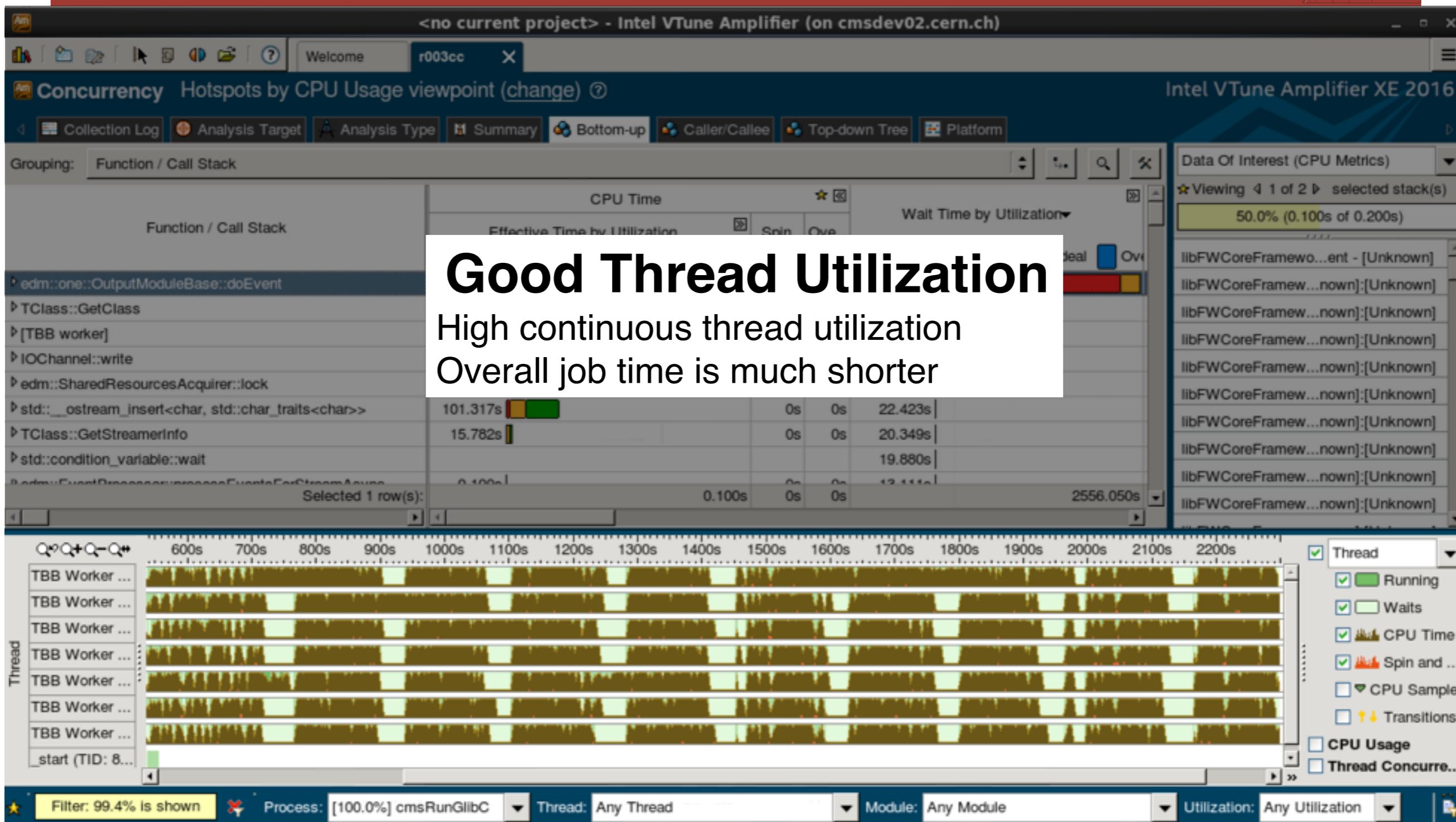
Removing the lazy evaluation had no impact on performance

the code had been prematurely optimized

# Final Results

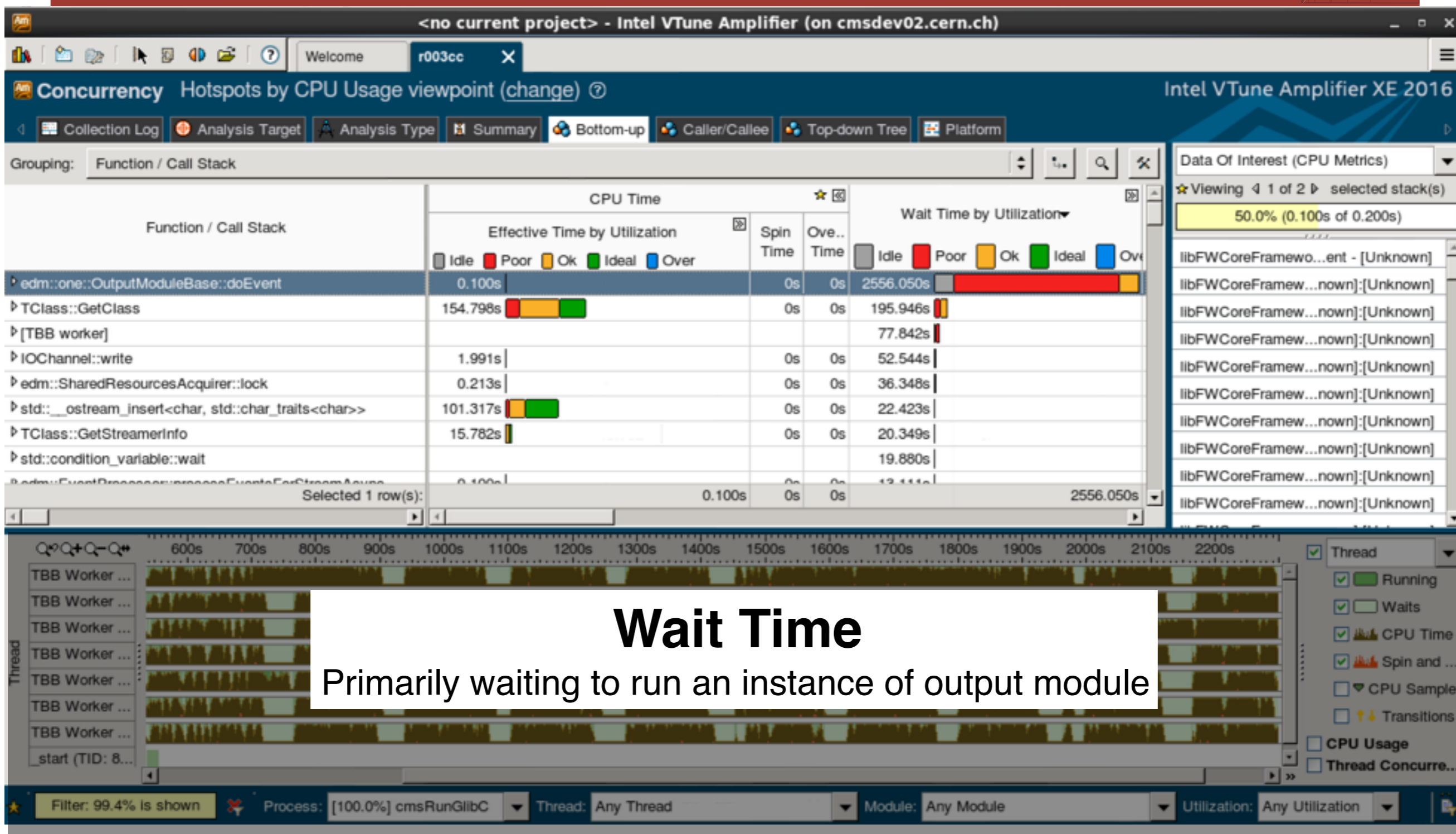


# Final Results

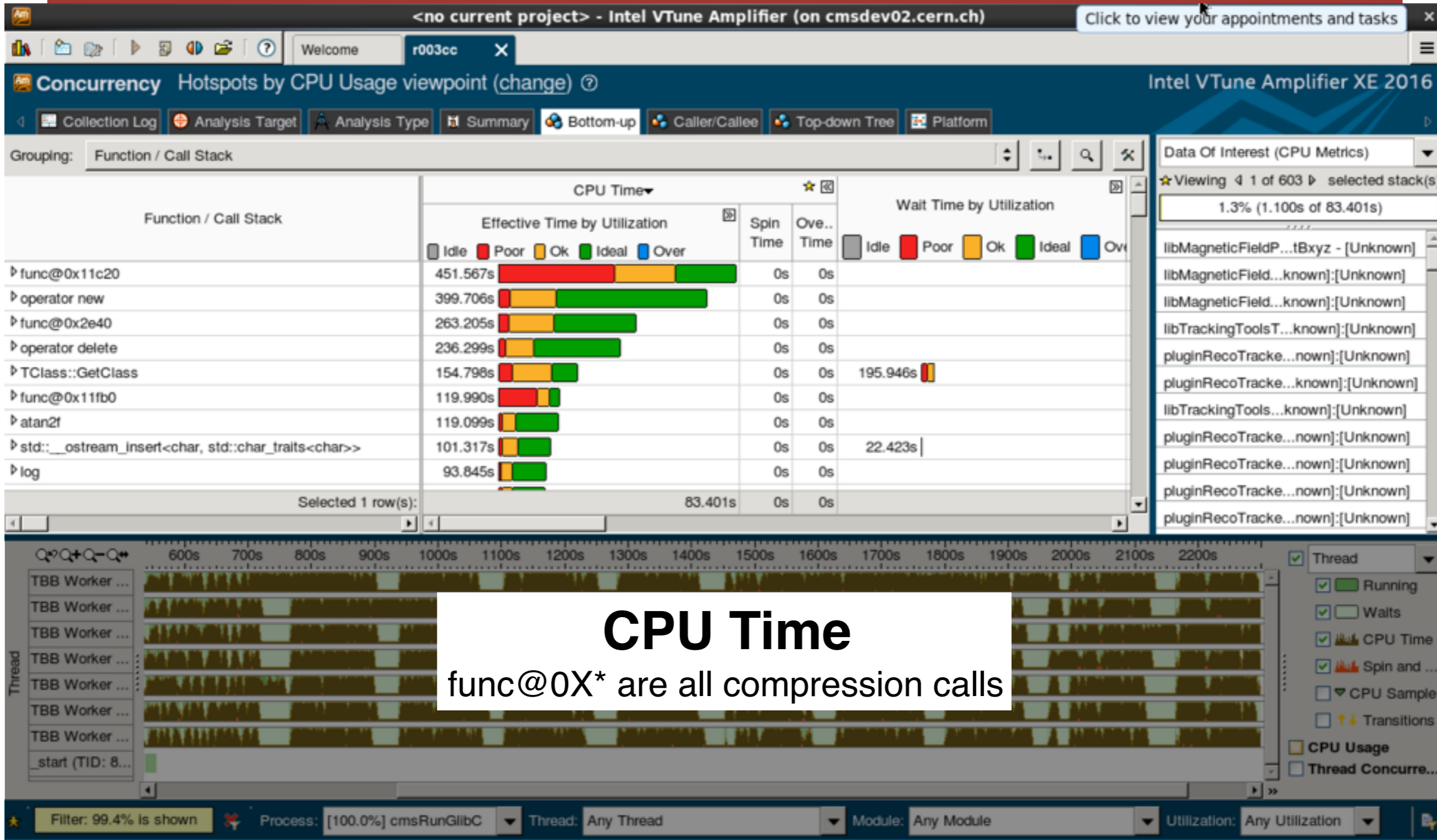


**Good Thread Utilization**  
High continuous thread utilization  
Overall job time is much shorter

# Final Results



# Final Results





# Next Step

Full CMSSW threading framework implementation

Switching to TBB task per module will work around output module

If event 1 is running the output module event 2 could run a different module

Work on output module to make more thread efficient

Decrease time in a single thread

Look into utilizing multiple threads when compressing buffers

# VTune Calibration



The value VTune gives for ‘concurrency’ does not correspond to timing measurements of program

VTune much more pessimistic

Increased VTune concurrency does correspond to real world speedups

Program Changes	Speedup	Regular Job Time (s)	VTune Concurrency	VTune Job Time (s)
Initial	5.5	1826	2.1	16440
ROOT	5.7	1748	2.6	9884
CMSSW I	6.1	1641	7.2	3363
Final	6.2	1605	5.0*	2429

\* fraction of the job was in the end of job single threaded part. GUI histogram of thread utilizations showed most time spent at 7 cores active with second most at 1.

# Conclusion

VTune is a useful tool

It is capable of running on HEP frameworks

It does identify code that hampers threading performance

Not without its problems

Sometimes fails to report work done on some threads

Data gathering appears to strongly affect job running time and distort findings

The problems are not sufficient to stop me from using VTune

**TO BE  
CONTINUED...** 