CLARA

Software LEGO System

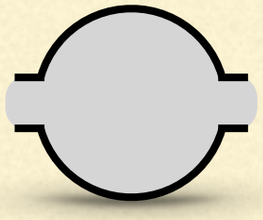
Reactive data-stream processing framework

Vardan Gyurjyan, JLAB November 2018

OUTLINE

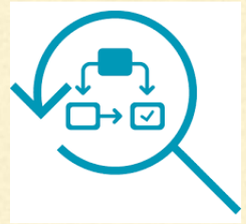
- CLARA in nutshell
- Version 4.3.4 release
- CLAS12 reconstruction application benchmarks

CLARA BASIC COMPONENTS



Data Processing Station
(*service*)

Data-Stream Pipe
(*supports multiple protocols*
pub-sub, p2p, inproc, etc)



Workflow Management System
(*Orchestrator*)

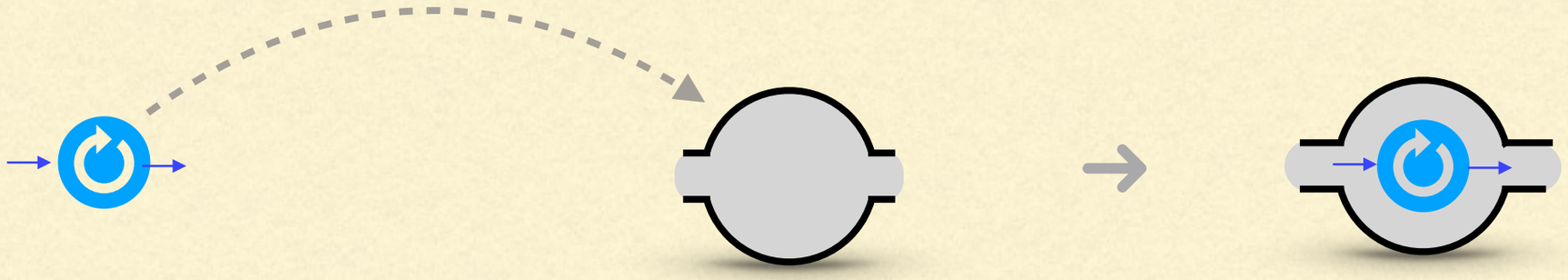
HOW IT WORKS?

Build your own software LEGO brick

User Data Processing Engine

Data Processing Station

Clara Service

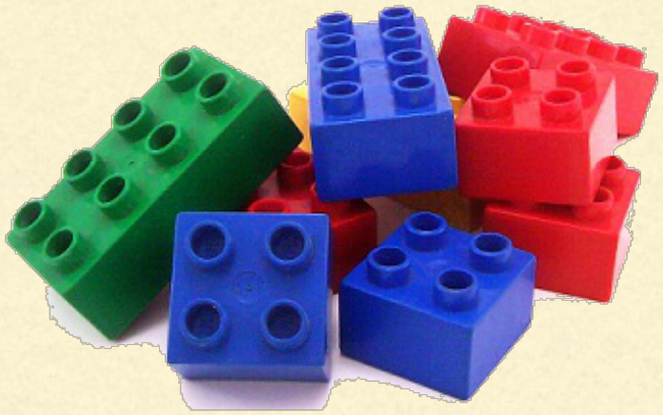


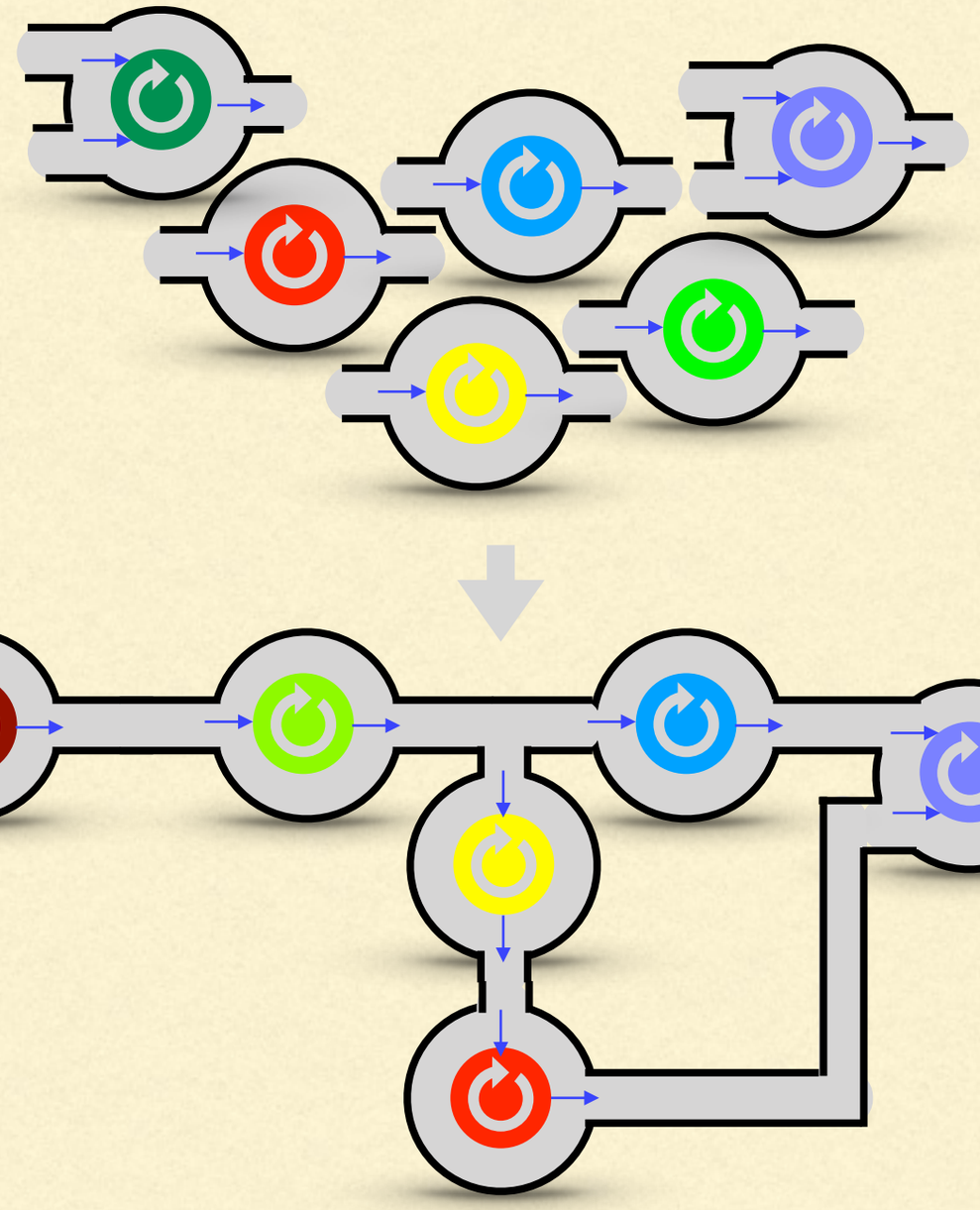
Developed and tested independently

(no knowledge of CLARA except of a single interface/pure virtual class.
No worries about where from data is coming and where it should go
Can be written in C++, Java or Python.Based on a load and/or request CLARA service will scale user engine, i.e.auto threading.)

APPLICATION DESIGN

(no programming is required)



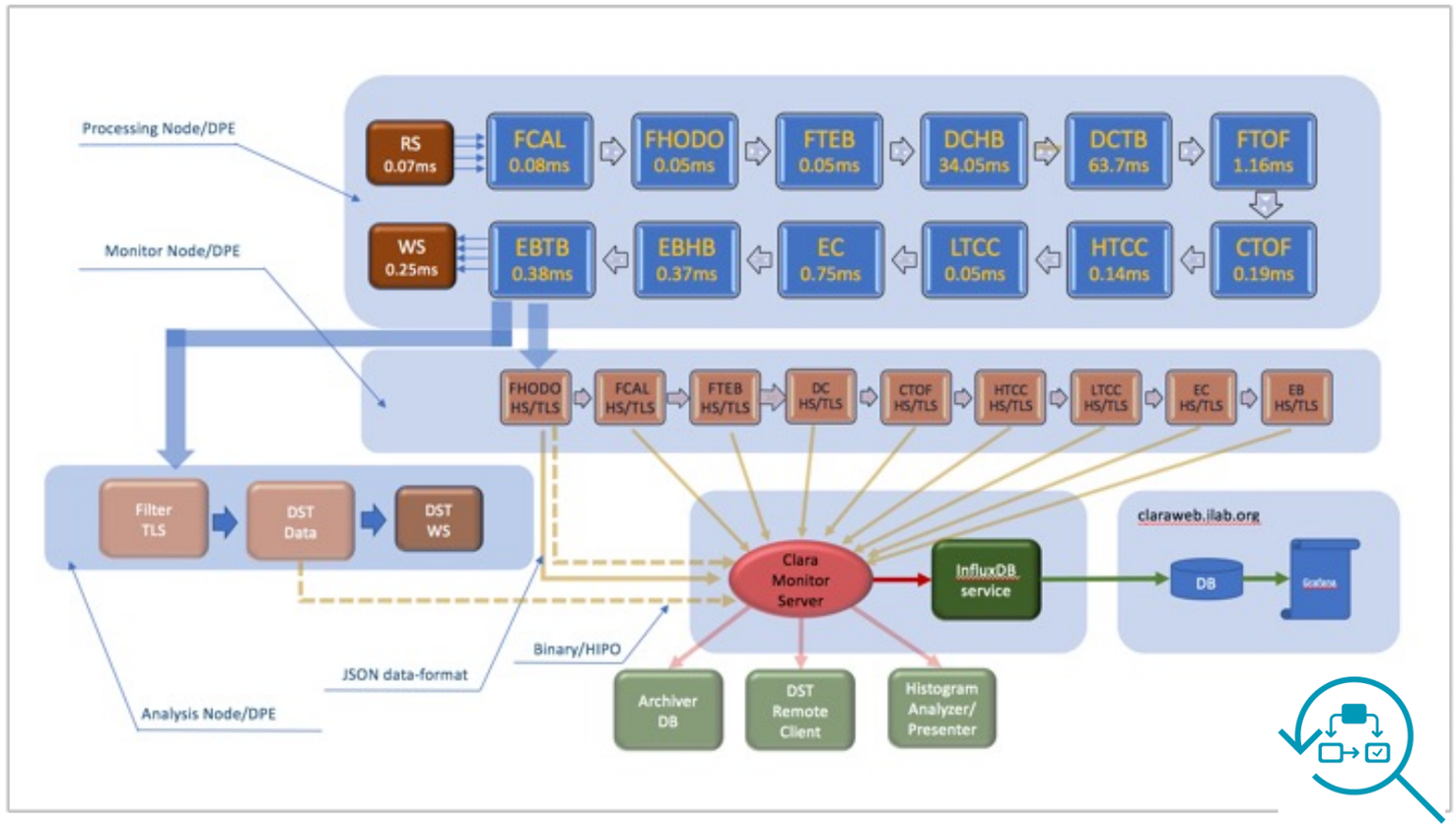


Application is build simply by graphically connecting services together with Clara data pipes

CLAS12 RECONSTRUCTION

Application Composition

Graphical Design



Descriptive Design (YAML)

```
io-services:
  reader:
    class: org.jlab.clas.std.services.convertors.EtRingToHipoReader
    name: EtRingToHipoReader
  writer:
    class: org.jlab.clas.std.services.convertors.HipoToHipoWriter
    name: HipoToHipoWriter
services:
  - class: org.jlab.rec.ft.cal.FTCALEngine
    name: FTCAL
……..
configuration:
  global:
    magnet:
      torus: -1
      solenoid: -1
    ccdb:
      run: 101
      variation: custom
    runtype: mc
    runmode: calibration
  io-services:
    reader:
      system: /tmp/clara-et-system
      host: localhost
      port: 11111
    writer:
      compression: 2
  services:
    EC:
      variation: cosmic
      timestamp: 333
mime-types:
  - binary/data-hipo
```

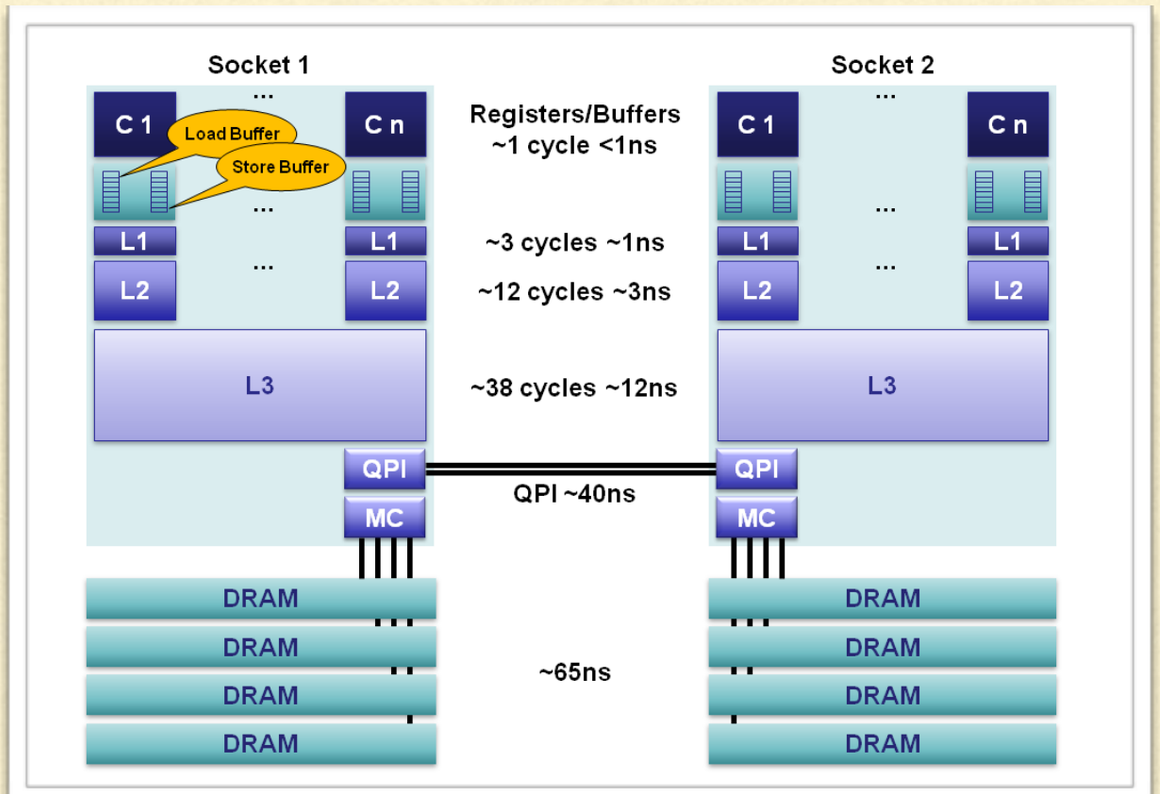
TO SUM UP

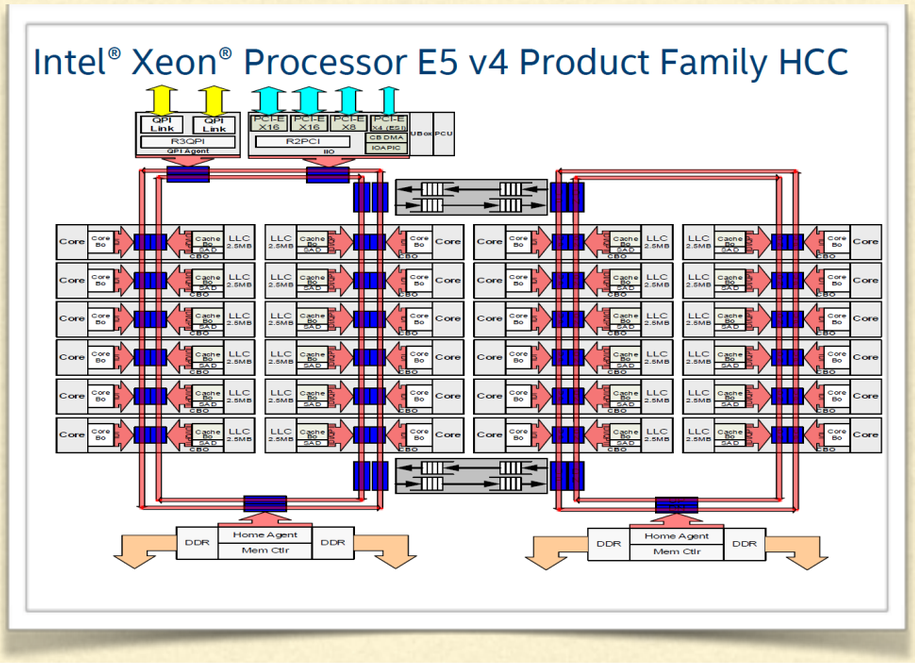
CLARA is a framework for design heterogeneous, distributed, data-stream processing applications

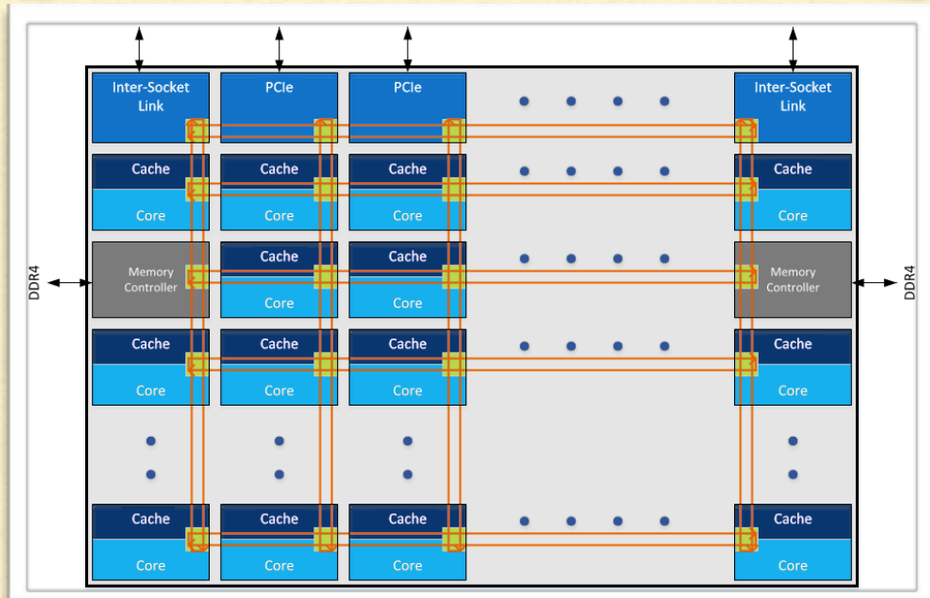
- **User algorithm container** (service, application building block)
  - Services are small
    - Reduced develop-deploy-debug cycles
    - Easy to contribute
  - Services are independent
    - Improved fault isolation
    - Independent scaling and optimizations
    - Easy to embrace new technologies
  - Services are reactive and stateless
- **Data abstraction**
    - Data format agnostic
- **Data transport** (data pipes)
    - Defined outside of the user engine
- **Workflow management system**
    - Data-flow and service monitoring
    - Dynamic elasticity
    - Fault tolerance , profiling , data-set provisioning, etc.

HARDWARE

- Linux Kernel schedulers are becoming more complex (e.g. CFS) in multi-core architectures.
- Complex governor algorithms, such as cpuidl, cpufreq, cpupower, etc.







CLARA V4.3.3/4

Reconstruction performance sudden degradation

- Introducing farm.exclusive mode
- JLAB farm hardware aware
- Sub-node horizontal scaling
- Thread affinity

CLAS12 Reconstruction. Farm18, threads = 64, clas_004013.hipo. Affinity

```
2018-09-26 16:03:23.805: Benchmark results:
2018-09-26 16:03:23.806:   READER          5000 events   total time =    0.27 s   average event time =    0.05 ms
2018-09-26 16:03:23.807:   MAGFIELDS       5000 events   total time =    0.04 s   average event time =    0.01 ms
2018-09-26 16:03:23.808:   FTCAL           5000 events   total time =    0.53 s   average event time =    0.11 ms
2018-09-26 16:03:23.808:   FTHODO          5000 events   total time =    0.89 s   average event time =    0.18 ms
2018-09-26 16:03:23.809:   FTEB            5000 events   total time =    0.25 s   average event time =    0.05 ms
2018-09-26 16:03:23.809:   DCHB            5000 events   total time = 1970.02 s   average event time =  394.00 ms
2018-09-26 16:03:23.809:   FTOFHB          5000 events   total time =    7.00 s   average event time =    1.40 ms
2018-09-26 16:03:23.810:   EC              5000 events   total time =    4.18 s   average event time =    0.84 ms
2018-09-26 16:03:23.810:   CVT             5000 events   total time =  208.42 s   average event time =   41.68 ms
2018-09-26 16:03:23.811:   CTOF            5000 events   total time =    9.34 s   average event time =    1.87 ms
2018-09-26 16:03:23.811:   CND             5000 events   total time =    6.72 s   average event time =    1.34 ms
2018-09-26 16:03:23.812:   HTCC            5000 events   total time =    0.57 s   average event time =    0.11 ms
2018-09-26 16:03:23.812:   LTCC            5000 events   total time =    0.62 s   average event time =    0.12 ms
2018-09-26 16:03:23.812:   RICH            5000 events   total time =    1.32 s   average event time =    0.26 ms
2018-09-26 16:03:23.812:   EBHB            5000 events   total time =    5.97 s   average event time =    1.19 ms
2018-09-26 16:03:23.813:   DCTB            5000 events   total time =  580.20 s   average event time =  116.04 ms
2018-09-26 16:03:23.813:   FTOFTB          5000 events   total time =  110.40 s   average event time =   22.08 ms
2018-09-26 16:03:23.813:   EBTB            5000 events   total time =    7.99 s   average event time =    1.60 ms
2018-09-26 16:03:23.813:   WRITER          5000 events   total time =   13.25 s   average event time =    2.65 ms
2018-09-26 16:03:23.814:   TOTAL           5000 events   total time = 2927.96 s   average event time =  585.59 ms
2018-09-26 16:03:23.814: Average processing time  =  30.82 ms
2018-09-26 16:03:23.814: Total processing time    =  154.12 s
2018-09-26 16:03:23.815: Total orchestrator time  =  160.10 s
```

CLAS12 Reconstruction. Farm18, threads = 64, clas_004013.hipo

```
2018-09-26 16:14:46.893: Benchmark results:
2018-09-26 16:14:46.894:   READER          5000 events   total time =    0.45 s   average event time =    0.09 ms
2018-09-26 16:14:46.895:   MAGFIELDS       5000 events   total time =    0.09 s   average event time =    0.02 ms
2018-09-26 16:14:46.896:   FTCAL           5000 events   total time =    0.99 s   average event time =    0.20 ms
2018-09-26 16:14:46.896:   FTHODO          5000 events   total time =    1.55 s   average event time =    0.31 ms
2018-09-26 16:14:46.897:   FTEB            5000 events   total time =    0.46 s   average event time =    0.09 ms
2018-09-26 16:14:46.897:   DCHB            5000 events   total time = 1438.51 s   average event time =  287.70 ms
2018-09-26 16:14:46.897:   FTOFHB          5000 events   total time =   11.68 s   average event time =    2.34 ms
2018-09-26 16:14:46.898:   EC              5000 events   total time =    7.35 s   average event time =    1.47 ms
2018-09-26 16:14:46.898:   CVT             5000 events   total time = 14639.97 s   average event time = 2927.99 ms
2018-09-26 16:14:46.899:   CTOF            5000 events   total time =   13.90 s   average event time =    2.78 ms
2018-09-26 16:14:46.899:   CND             5000 events   total time =    7.33 s   average event time =    1.47 ms
2018-09-26 16:14:46.899:   HTCC            5000 events   total time =    1.26 s   average event time =    0.25 ms
2018-09-26 16:14:46.900:   LTCC            5000 events   total time =    1.21 s   average event time =    0.24 ms
2018-09-26 16:14:46.900:   RICH            5000 events   total time =    2.44 s   average event time =    0.49 ms
2018-09-26 16:14:46.901:   EBHB            5000 events   total time =   11.25 s   average event time =    2.25 ms
2018-09-26 16:14:46.901:   DCTB            5000 events   total time =  439.07 s   average event time =   87.81 ms
2018-09-26 16:14:46.901:   FTOFTB          5000 events   total time =  122.08 s   average event time =   24.42 ms
2018-09-26 16:14:46.902:   EBTB            5000 events   total time =   15.70 s   average event time =    3.14 ms
2018-09-26 16:14:46.902:   WRITER          5000 events   total time =   11.41 s   average event time =    2.28 ms
2018-09-26 16:14:46.903:   TOTAL           5000 events   total time = 16726.70 s   average event time = 3345.34 ms
2018-09-26 16:14:46.903: Average processing time  =  53.41 ms
2018-09-26 16:14:46.903: Total processing time    =  267.07 s
```

BENCHMARKS

CLAS12 Reconstruction Performance on JLAB farm nodes

	Farm18	Farm16	Farm14	Farm13	Qcd12s
CPU	6148 CPU @ 2.40GHz	E5-2697 v4 @ 2.30GHz	E5-2670 v3 @ 2.30GHz	E5-2650 v2 @ 2.60GHz	E5-2650 0 @ 2.00GHz
N Nodes in the farm	90	50	100	22	170
N Cores	80 (40/40)	72 (36/36)	48 (24/24)	32 (16/16)	32 (16/16)
N Cores/Socket	20 (10/10)	36 (18/18)	24 (12/12)	16 (8/8)	16 (8/8)
N Sockets	4	2	2	2	2
1 Thread [ms]/[Hz]	251.06 / 3.98	289.9 / 3.44	369.89 / 2.7	348.19 / 2.87	467.61 / 2.1
1 Socket [ms]/[Hz]	27.22 / 36.7	20.85 / 47.9	33.34 / 29.9	47.03 / 21.3	65.64 / 15.2
1 Node [ms]/[Hz]	8.36 / 119.6	10.62 / 94.2	16.26 / 61.5	23.81 / 41.9	32.8 / 30.5

Thank you

IMPROVEMENTS

Veronique Ziegler, David Heddle, Bruno Bankel

CLAS12 Reconstruction DCHB and DCTB Services Improvements

(Measurements done on farm16 node)

Release/Branch	5b.6.2	vg-optimized	vg-optimized	vg-optimize	5c.7.0
RK4 implementation	-	V	V	V	V
KF iterations, and adoptive step size	-	-	V	V	V
Code clean-up	-	-	V	V	V
Fast Math libraries	-	-	-	V	V
JRE warmup latency	-	-	-	-	V
Thread affinity	-	-	-	-	V
Object pool implementation	-	-	-	-	-
Code vectorization	-	-	-	-	-
KF service on GPU	-	-	-	-	-
1 Thread [ms]	1855+ (unstable)	1104+ (unstable)	787+ (unstable)	734+ (unstable)	289.9 (stable)
Rate/Node [Hz]	19.0	33.0	45.6	49.1	94.2