PROCEEDINGS OF SCIENCE

The Fluid Database Paradigm: A Prototype

A. Weinstein[*a], **L. Fortson**[b], **T. Brantseg**[a], **C. Rulten**[b], **R. Lutz**[c], **J. Haupt**[d], **M. Kadkhodaie Elyaderani**[d]

E-mail: amandajw@iastate.edu

[a] *Department of Physics and Astronomy, Iowa State University, USA;* [b] *Department of Physics & Astronomy, University of Minnesota, USA;* [c] *Department of Computer Science, Iowa State University, USA;*
[d] *Department of Electrical and Computer Engineering, University of Minnesota, USA;*

The process of event-building—i.e gathering and associating data from multiple sensors or sub-detectors that arises from a common physical event—is used in many fields, including high-energy physics and gamma-ray astronomy. The problem of fault tolerance in event-building is a difficult one, and one that becomes increasingly difficult with higher data throughput rates and increasing numbers of sub-detectors. We draw on biological self-assembly models in the development of a novel event-building paradigm that treats each packet of data from an individual sensor or sub-detector as if it were a molecule in solution. Bonds (analogous to chemical bonds) are defined between data packets using metadata-based discriminants. A database, which plays the role of a beaker of solution, quasi-randomly and continually selects pairs of assemblies to test for bonds, allowing single tiles and small assemblies to aggregate into larger assemblies. During this process higher-quality associations supersede spurious ones. The database thereby becomes fluid, dynamic, and self-annealing rather than static. We will describe lessons learned from early prototypes of the fluid database as well as future directions.

38th International Conference on High Energy Physics
3-10 August 2016
Chicago, USA

*Speaker.

PoS(ICHEP2016)930

1. Introduction

In high-energy physics and astrophysics, event-building (the process of associating data from multiple sensors or detector components in a coherent description of a physical event) is a key part of online data processing. There is often a tension between needing to event-build for real-time analysis, which provides immediate feedback, and the need to find and correct faults or gaps in data (caused by data corruption or data transmission lags) before those faults become frozen in an archive. We consider here a prototype *fluid database* designed to address these issues, implemented for a test case based on the gamma-ray observatory VERITAS.

2. The fluid database concept

In a branch of nanoscale engineering known as *self-assembly*, regular structures (*assemblies*) are produced via the random association of molecules in solution. Algorithmic self-assembly models these processes by a pool of abstract objects with simple bonding rules, coupled with repeated random draws of pairs of assemblies from that pool [1]. This methodology can be used to study mechanisms for achieving *fault tolerance* in the self-assembly process.

The fluid database draws to some degree on these concepts. It is a random access database coupled to a continuously iterated process. Data packets are treated as if they were molecules in a beaker of solution. At each iteration, pairs of data packets are drawn from a pool of data packets and associated by means of specified bonding rules based on identifying metadata (e.g. event numbers, timestamps). Fault tolerance is implemented through the existence of weak bonds, which allow data packets to associate even when some metadata is corrupt. Higher quality bonds are allowed to supersede and destroy existing lower-quality bonds. The assembly process ends when the concentration of incomplete assemblies reaches an equilibrium state.

The fluid database schema described here is not an exact cognate to algorithmic self-assembly. The goal of such self-assembly processes is typically the production of many copies of a single machine from a few different types of building block. In this scenario it is irrelevant which copy of a particular building block is chosen for a particular assembly. Only the block's type and its place in the assembly matters. For event-building, on the other hand, there is a unique choice of each type of building block that corresponds to a valid event.

3. Fluid database implementation

The prototype fluid database structure remains similar to that described in [2]. An initial process takes sets of data packets from the individual sensors, associates them based on a primary metadatum, and then inserts the results into the fluid database as preliminary assemblies. We refer to this process as *pre-seeding*. The full constituent bond strengths of these preliminary assemblies are calculated before beginning the self-assembly process. It should be noted that the pre-seeded assemblies, while imperfect, can be used as-is for a real-time analysis of the data. The self-assembly process, which anneals the pre-seeded assemblies into their final form, can proceed on a longer timescale.

The self-assembly function of the fluid database is controlled by a master process, which monitors the number of incomplete assemblies and selects a pair of assemblies to be bond-tested. These

pairs are dispatched to a programmable number of slave processes, which perform the actual bond calculations and update the database if it determines that the original assemblies should be modified. We obtained a roughly factor of 8 speed increase in the assembly process by replacing the original random draw process[2] with a a semi-random selection method. Asssembly one is randomy chosen, but is then paired with a likely interaction candidate by searching for the incomplete assembly located nearest along a chosen metadata axis. By repeatedly cycling through the various metadata axes for choosing interaction matches, the assembler is able to more quickly converge on a solution, if one exists, for each incomplete assembly remaining after preseeding. We are still studying how this semi-random method scales to a large number of metadata criteria and whether the speed gain depends on using sortable metadata axes. To further improve convergence, assemblies that are *complete* (i.e. have a full complement of components from all expected sensors) and whose bond strengths all exceed a user-defined threshold *precipitate* out, i.e. they are removed from consideration by the assembler.

4. Performance tests

VERITAS consists of four imaging atmospheric Cherenkov telescopes (IACTs)[3]. The packets of data from each telescope are stamped with two quantities that permit data packets from a common event to be associated: a centrally-assigned 32-bit event number and a GPS timestamp from a clock local to the telescope[3]. Our bond criteria are defined based on these metadata. Under ideal conditions, event-building could be done by simply matching event numbers. However, the event number transmitted to a given telescope can experience bitwise corruption in transmission, and the GPS clock timestamps can experience variations due to thermal drift. These effects are complicated by a small amount of actual data loss from individual telescopes. These failure modes and their interactions make VERITAS an excellent small-scale test case for the fluid database paradigm.

Three roughly equally-sized samples of VERITAS raw data were used for the fluid database tests. Since the time to convergence is expected to improve with greater parallelization of the assembly process (which itself depends on the number of slave processes running on independent cores), the assembly process was tested using a master and 2-6 slave processes (3-7 cores respectively). The number of slaves could not be increased further due to hardware limitations of the prototype system. Figure 1 summarizes the convergence process, as a function of number of cores, for two of the three runs. We can see that in each case the assembly proceeds in a similar manner: an initial, nearly linear phase where the semi-random algorithm finds almost entirely successful interaction pairs, an inflection point when these "obvious" matches are used up, and the algorithm begins to find unsuccessful interaction pairs, and then convergence to final equilibrium. When number of incomplete assemblies is at the percent level (samples 1 and 2) the convergence timescale is relatively short (under an hour at maximum parallelization) and scales roughly linearly overall with the number of cores used. The slope of the linear regime also scales noticeably with the number of cores used. For the third sample, an extreme case where the corruption was at the 20% level, convergence is slower ($\sim$ 1 day) and the slope of the linear regime is insensitive to the number of cores used. The cause of this phenomenon has not yet been identified. In each case a residue

of between 28 and 421 incomplete assemblies remains, which appears to arise not from metadata corruption but hard data loss at one or more of the telescopes.

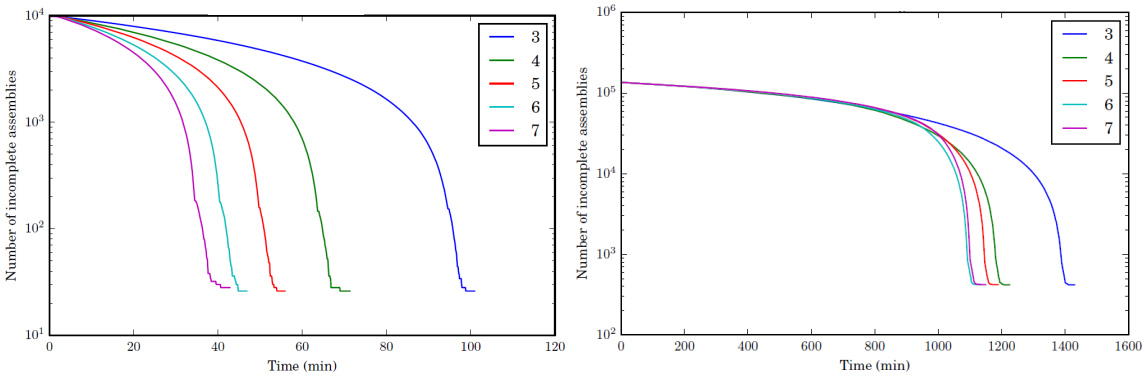

Figure 1: Number of incomplete assemblies vs. time for two of the three sample runs, color-coded by the total number of parallel processes used. The fraction of initial incomplete assemblies is less than 2% for the run on the left, $\sim 20\%$ for that on the right.

5. Conclusions

We present here an in-development prototype of the *fluid database*, designed to recover links between data from different sensors that would otherwise be lost due to data corruption. Initial performance benchmarks for a simple test case (4 sensors) suggest that such recovery is possible on a reasonable timescale as long as the fraction of data affected is at the 20% level or less. More detailed investigation of these benchmarks and their implications for scaling to the generic n-sensor case are still in progress.

6. Acknowledgements

This research is supported by Award #NSF/PHY-1419259 and Award #NSF/PHY-1419250 from the US National Science Foundation. We gratefully acknowledge the use of test data from VERITAS, which is supported by grants from the US Department of Energy Office of Science, the National Science Foundation, and the Smithsonian Institution, and by NSERC in Canada. We acknowledge the excellent work of the technical support staff at the Fred Lawrence Whipple Observatory and at the collaborating institutions in the construction and operation of this instrument. The VERITAS Collaboration is grateful to Trevor Weekes for his seminal contributions and leadership in the field of VHE gamma-ray astrophysics.

References

[1] E. Winfree. PhD thesis, California Institute of Technology, June, 1998.

[2] A. Weinstein, L. Fortson, T. Brantseg, C. Rulten, R. Lutz, J. Haupt, M. Kakhodaie Elyaderani, and J. Quinn *ArXiv e-prints* (Sept., 2015) [arXiv:1509.0220].

[3] J. Holder, R. W. Atkins, H. M. Badran, et al. *Astroparticle Physics* **25** (July, 2006) 391–401, [astro-ph/0604119].