

# Real Time Stream Processing in Astronomy

**Veljko Vujčić**  
*Astronomical Observatory Belgrade*

VELJKO@AOB.RS

**Darko Jevremović**  
*Astronomical Observatory Belgrade*

DARKO@AOB.RS

## Abstract

One of problems with “Big Data” is when large quantities of data have to be analyzed fast - (almost) in real time. So far in astronomy those situations were not very common. In the next few years a real time processing will become more relevant and the field should prepare for the advent of new instruments and techniques.

**Keywords:** astroinformatics, complex event processing, stream processing

## 1. Introduction

A nature of most astronomical objects is that they change slow or very slow and are hard to detect on timescales available to astronomers. Of course there are some objects/phenomena which are changing very rapidly. Timescales of changes of physical properties of astronomical objects range from milliseconds to millions of years.

Historically there were long term observational programs where specific, more interesting objects, which showed unusual changes, were followed for many months or years. Also, there were short term observational campaigns which were focused on specific objects, trying to understand their nature.

The Zwicky Transient Facility (ZTF) is a new time-domain survey that started operations at Palomar Observatory in the beginning of 2018. ZTF searches and catalogs rare and exotic transients and its data will drive novel studies of supernovae, variable stars, binaries, AGN and asteroids ZTF started to disseminate part of their detected events in mid-2018 (?). This is precursor of Large Synoptic Survey (LSST, which will come online 2022/23, ?) and the cadence of observations is similar (though depth is significantly shallower). Number of alerts about detected events per ZTF visit is in range of few hundreds to thousand while with LSST we can expect few tens of thousands.

This paper deals with technologies and services which will possibly allow astronomers / observatories to easily handle vast quantities of data coming from these new facilities.

## 2. Event processing concepts

Event processing is an umbrella term for technologies that are conceptually centered around ingestion, manipulation and dissemination of events. An event could be any discrete occurrence, depending on the system ontology and granularity of its concepts. Event processing enables real-time pattern application, suitable for large input streams where automatized reaction is crucial.

Main difference between event processing applications and common transactional systems is their query paradigm. While traditional applications make queries and serve previ-

ously stored data per request, event processing applications run continuous queries through streams of new data and deliver discoveries if and when they happen.

Some of the use cases for event processing are:

- Real time analytics
- Monitoring, fault detection
- Diagnostics made in conjunction with historical data
- Data is incoming continuously, organized in streams
- Huge loads of data which are a challenge to process with traditional applications

Concept of *data windows* is as autochthonous and inherent Complex Event Processing mechanism. Windows are basis for aggregation and most of the CEP patterns. Pattern matching is another key characteristic of CEP systems. Patterns shape multiple simple events into a single complex event. Pattern formulation is commonly given through definition of presence or absence of events, in a certain temporal interdependence.

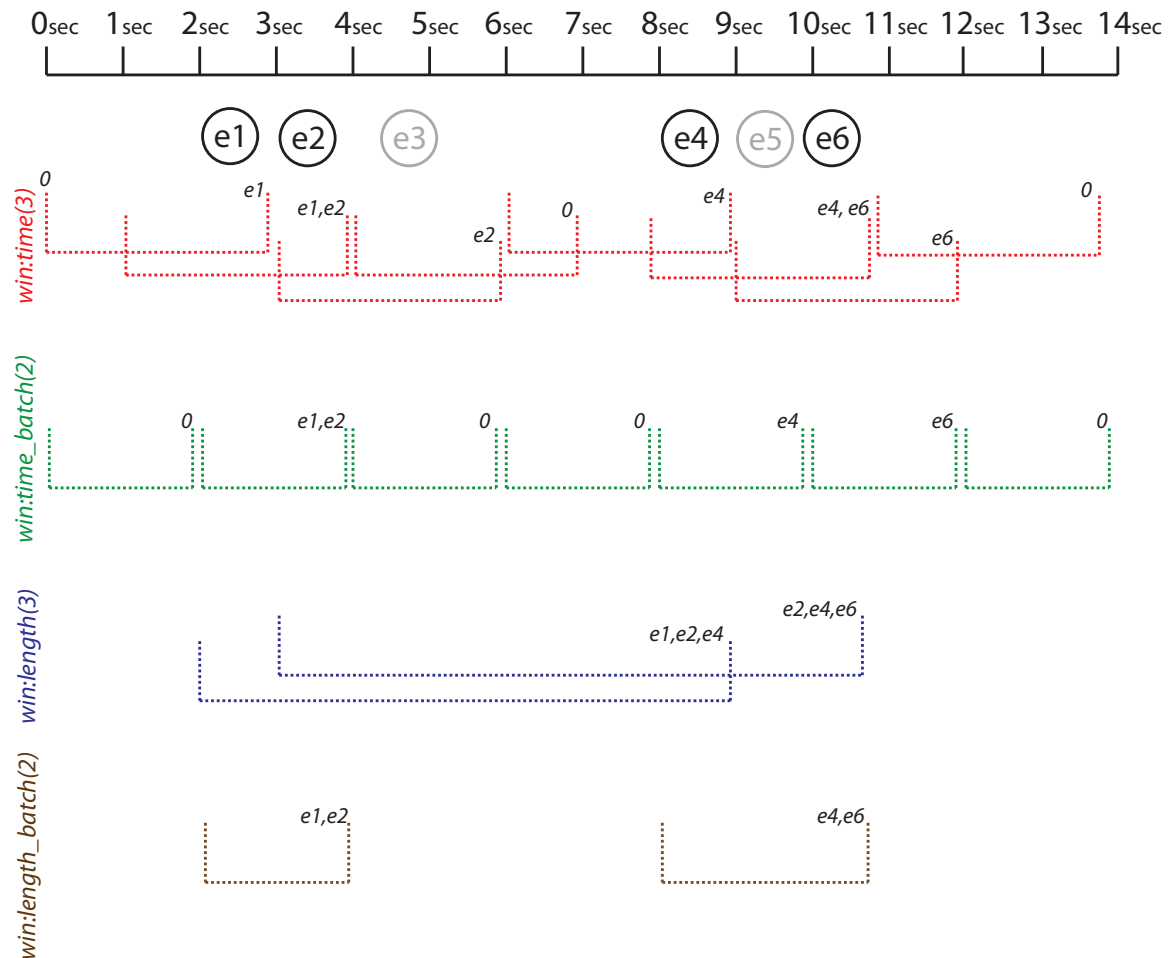


Figure 1: Four basic flavors of event processing "windows"

### 3. Application of Event Processing to Astronomy

Esper EPL is a specialized language for event processing based on SQL query language. We chose to present CEP concepts with examples coded in EPL.

Given here is the mathematical model of a “mag per hr” pattern, given in pseudo code. It detects change in brightness of an object and triggers an alert if it increases at a rate of 1 magnitude per hour or quicker (with a constraint that object was previously undetected). The pattern may be used for GRB detection, or with modified constraints for detection of some other phenomena.

```
select * from LsstStream
  match_recognize (
    partition by ra, decl
    measures A.ra as ra, A.decl as decl, A.detectionTime as firstDetection,
    B.detectionTime as secondDetection, A.mag as firstMag, B.mag
    as secondMag, Math.abs(B.mag - A.mag)/( (B.detectionTime -
    A.detectionTime)/3600000 ) as ratePerHour
    pattern (A B)
    define
      A as A.diaObject=false,
      B as Math.abs(B.mag - A.mag)/( (B.detectionTime -
      A.detectionTime)/3600000 ) >=1
  )
```

### 4. Conclusion

In the new data rich environment which will start at early 2020’s one can expect that real time stream analytics need tools both scalable and expressive. We described one possible direction in achieving this using CEP engines.

### References

- J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, C. Affeldt, and et al. First low-latency LIGO+Virgo search for binary inspirals and their electromagnetic counterparts. , 541:A155, May 2012. doi: 10.1051/0004-6361/201218860.
- B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, and et al. Multi-messenger Observations of a Binary Neutron Star Merger. , 848:L12, October 2017. doi: 10.3847/2041-8213/aa91c9.
- James F Allen. Maintaining knowledge about temporal intervals. In *Readings in qualitative reasoning about physical systems*, pages 361–372. Elsevier, 1990.
- C. Catala and B. H. Foing, editors. *MUSICOS: Multi-Site Continuous Spectroscopy : 1 : 1988*, 1988.
- C. Catala and B. H. Foing, editors. *MUSICOS: Multi-Site Continuous Spectroscopy : 2 : 1990*, 1990.

- Thomas Dudziak. Storm & esper, 2011. URL <https://tomdzk.wordpress.com/2011/09/28/storm-esper/>. [Online; accessed 05-2014].
- Opher Etzion, Peter Niblett, and David C Luckham. *Event processing in action*. Manning Greenwich, 2011.
- FF Gardner and DK Milne. The supernova of ad 1006. *The Astronomical Journal*, 70:754, 1965.
- D. A. Green. Historical supernovae and their remnants. *Highlights of Astronomy*, 12: 350–353, 2002.
- Zeljko Ivezić, JA Tyson, B Abel, E Acosta, R Allsman, Y AlSayyad, SF Anderson, J Andrew, R Angel, G Angeli, et al. Lsst: from science drivers to reference design and anticipated data products. *arXiv preprint arXiv:0805.2366*, 2008.
- David C Luckham. *Event processing for business: organizing the real-time enterprise*. John Wiley & Sons, 2011.
- F. Masci, S. R. Kulkarni, M. Graham, T. Prince, and G. Helou. The Zwicky Transient Facility Public Alert Stream. *The Astronomer’s Telegram*, 11685, June 2018.
- Brenda M Michelson. Event-driven architecture overview. *Patricia Seybold Group*, 2(12): 10–1571, 2006.
- Roy Schulte. An overview of event processing software, 2014. URL <http://www.complexevents.com/2014/08/25/an-overview-of-event-processing-software/>. [Offline; accessed 08-2015].
- M. Tsimelzon. Complex event processing: Ten design patterns, 2006. URL <http://complexevents.com/wp-content/uploads/2007/04/Coral8DesignPatterns.pdf>. [Online; accessed 03-2018].
- Petr Zapletal. Introduction into distributed real-time stream processing, 2015. URL <https://www.cakesolutions.net/teamblogs/introduction-into-distributed-real-time-stream>. [Online; accessed 03-2018].
- Petr Zapletal. Comparison of apache stream processing frameworks: Part 1, 2016. URL <https://www.cakesolutions.net/teamblogs/comparison-of-apache-stream-processing-framework>. [Online; accessed 03-2018].