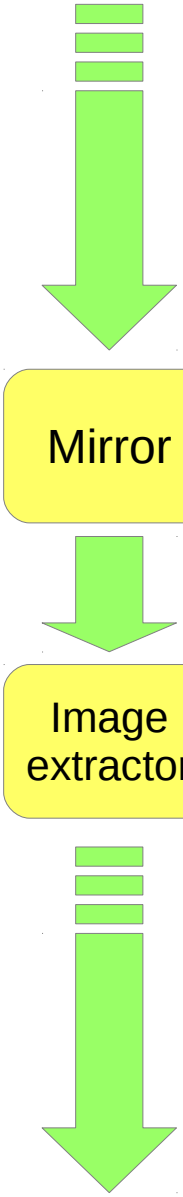
External receiver



Kafka consumer and producer
in the same process.

Based on the Kafka-mirror concept
from Apache Kafka.

Kafka archive

Kafka logs, split into
topic per day.

Long term backup
storage system.

Image archive

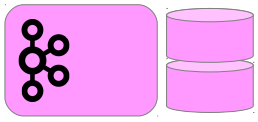
Add more space as
the data grows.

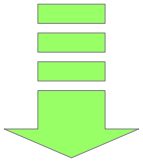
Space per
day
calculation ...

Inline images
replaced by internal
URLs.

- Backup before/after image extraction ?
- Mirror and image extractor in the same process ?
- Storage technology ?

Internal processing

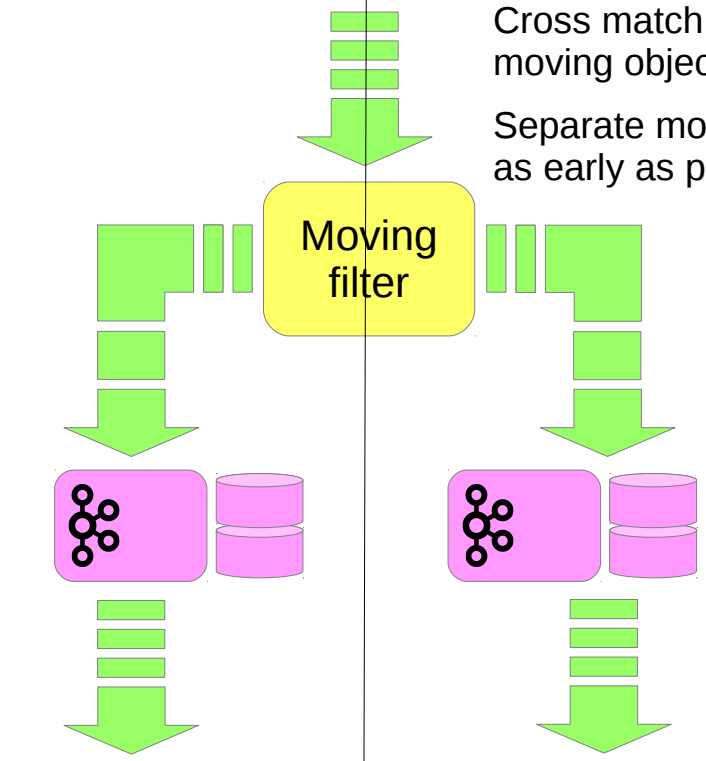




Internal Kafka acting as
an inter-process buffer.

D.Morris
Institute for Astronomy,
Edinburgh University
May 2018

Moving objects



Cross match with known moving objects.

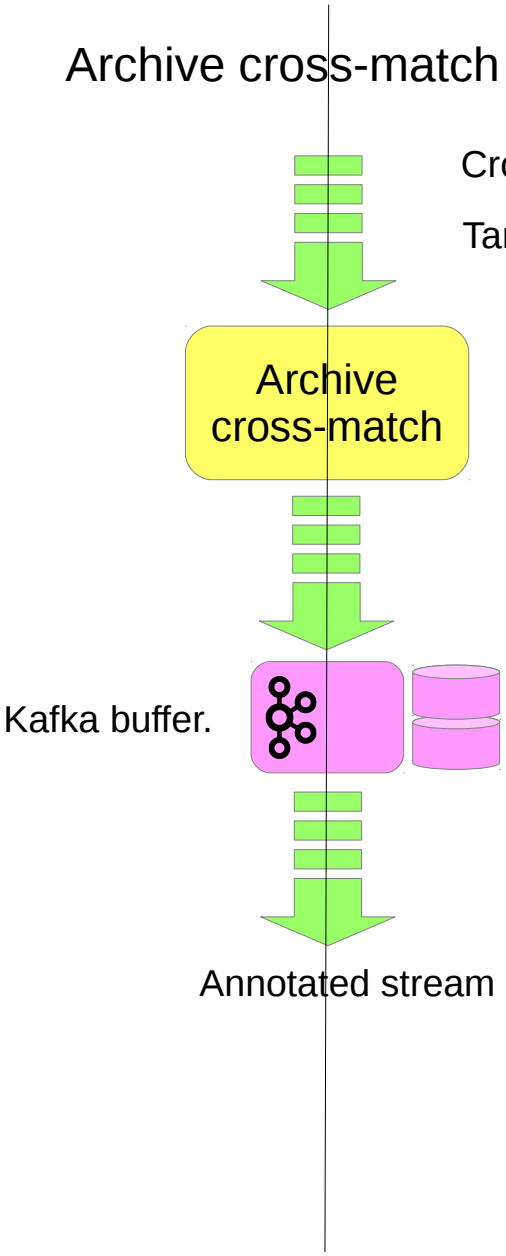
Separate moving/static as early as possible ?

Internal Kafka acting as an inter-process buffer.

- Can we do this at the required data rate ?
- Are moving and non-moving objects handled separately ?
- Does it make sense to skip cross-match with static archives for moving objects ?

D.Morris
Institute for Astronomy,
Edinburgh University
May 2018

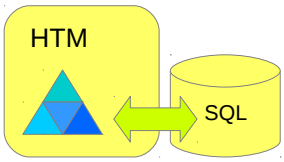
Archive cross-match



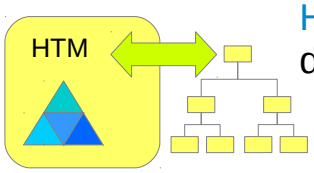
Cross match with main archives.

Target rate < 1ms per match

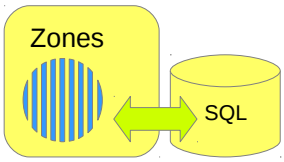
- Plug-and-play component API.
  - Adding new algorithms.
  - Adding new archives.



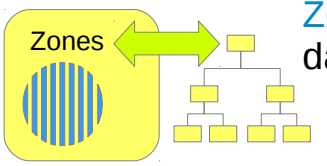
HTM match with data in database.



HTM match with data in memory.



Zone and vector with data in database.



Zone and vector with data in memory.

Multiple interfaces

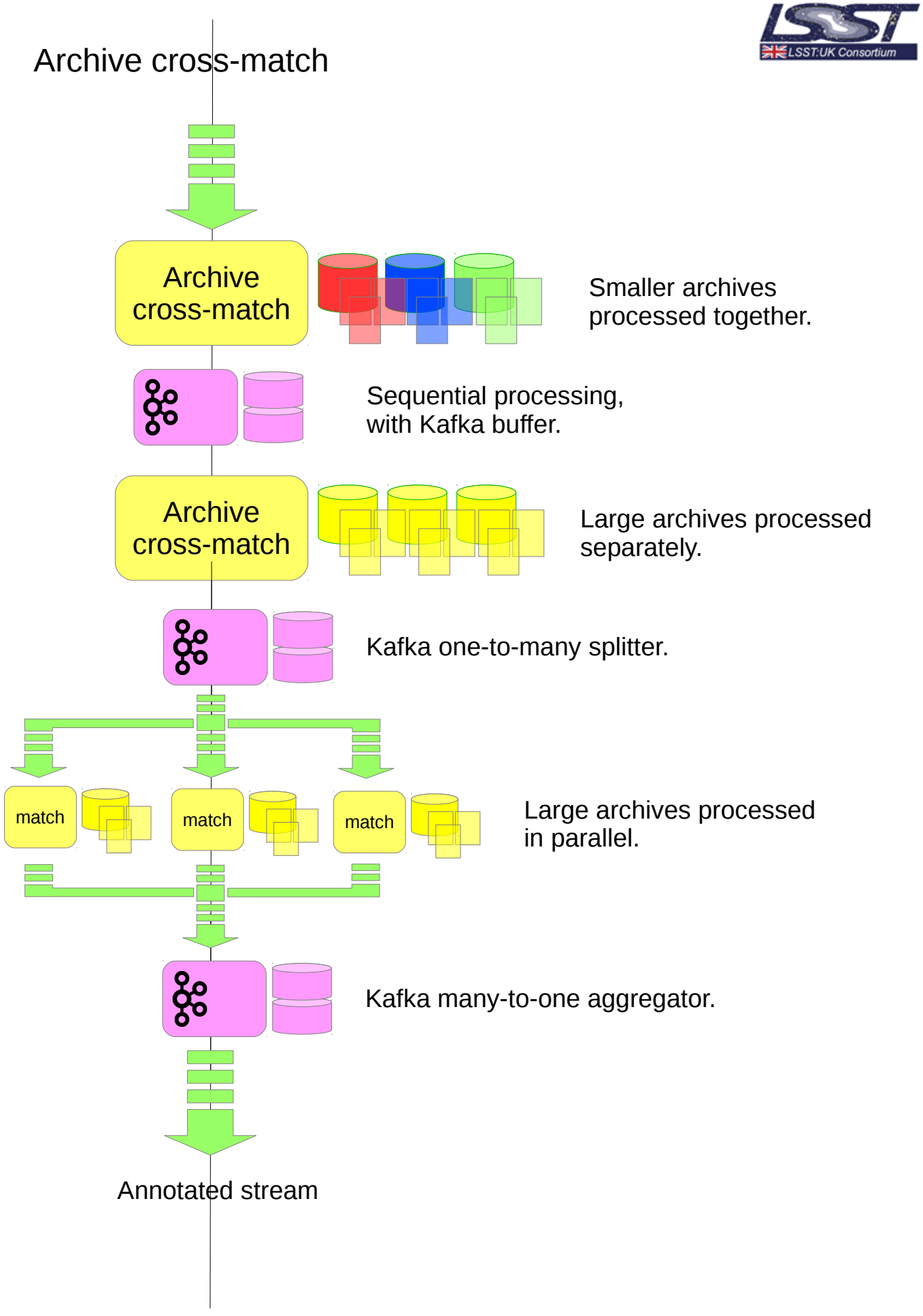
JSON/REST webservice
- Control and configuration
- Cone search
- Cross match

Kafka stream component
- Consumer/Producer API ?
- Kafka Connect component ?

IVOA conesearch
- Publish as IVOA services ?

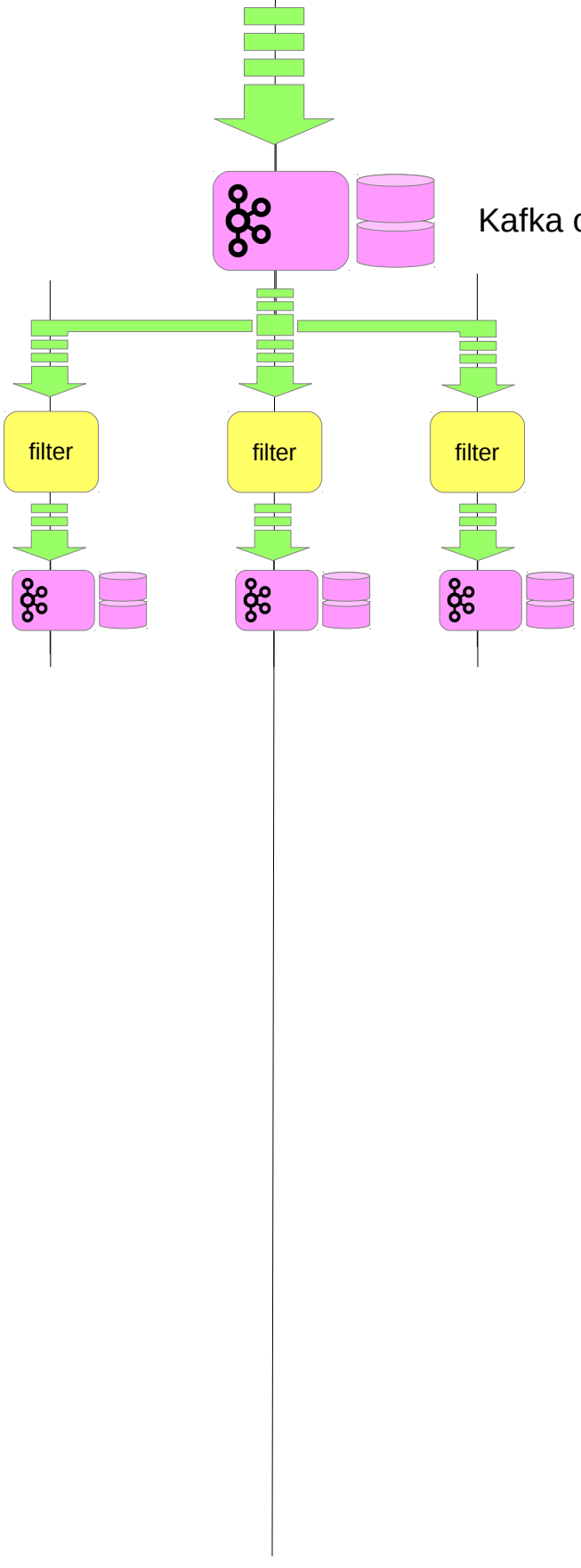
D.Morris
Institute for Astronomy,
Edinburgh University
May 2018

Archive cross-match



D.Morris
Institute for Astronomy,
Edinburgh University
May 2018

Selective filtering

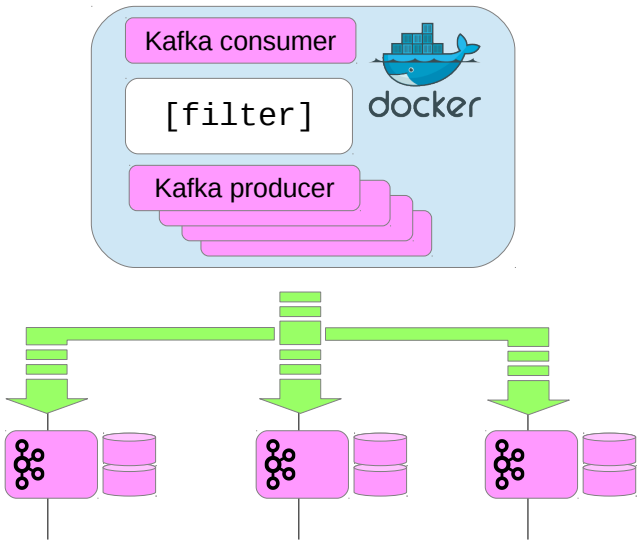


Kafka one-to-many splitter.

Filters create custom streams based on annotations.

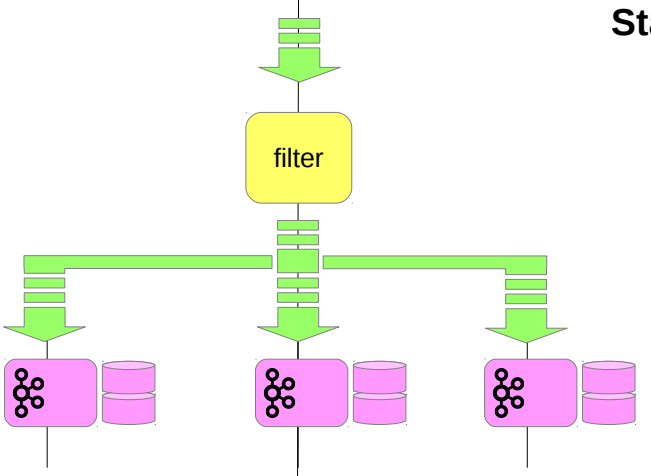
- User defined code (Python).
- User defined code (Java).
- System filters (user = system).

- Multiple output filters



D.Morris
Institute for Astronomy,
Edinburgh University
May 2018

Selective filtering



Single output yes/no filter

```
if (test)
    {
    output ..
    }
```

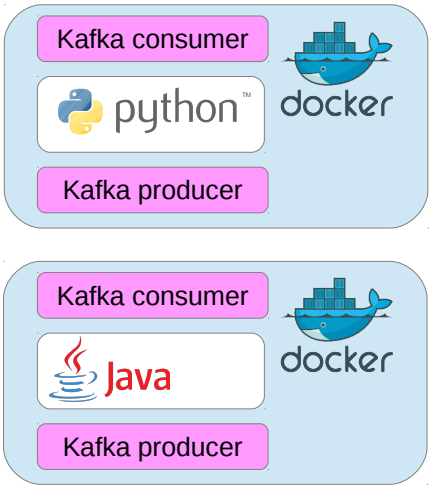
Dual output left/right filter

```
if (test)
    {
    output ..
    }
else {
    output ..
    }
```

Multiple output filter

```
switch (test)
    {
    case xx :
        output ..
    case yy :
        output ..
    case zz :
        output ..
    default :
        output ..
    }
```

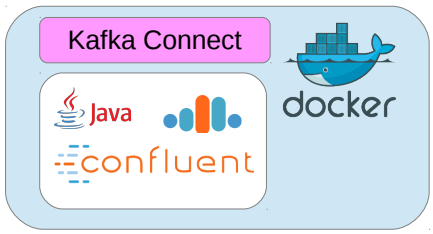
Standard Docker containers



Standard input and output interfaces.

Similar to the OGSA-DAI Activity class.

Confluent Kafka Connect



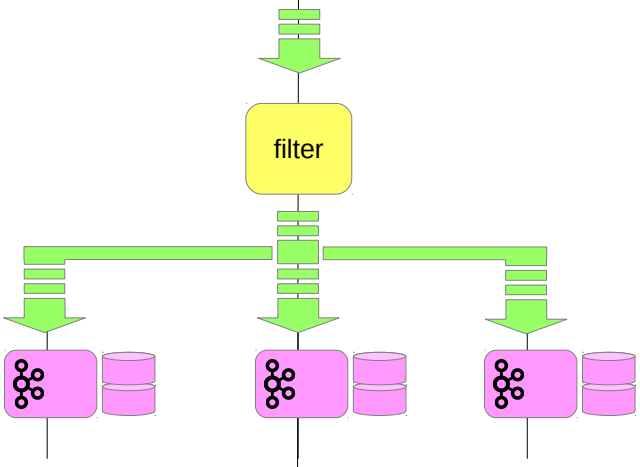
How much of theKafka Connect stack do we use ?

How much do we create ourselves ?

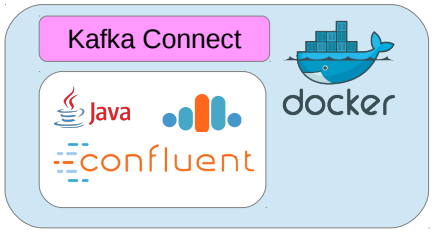
Kafka Connect provides a range of tools for data import and export .. but it adds yet another way of handling message schema.

D.Morris
Institute for Astronomy,
Edinburgh University
May 2018

Selective filtering



Kafka Connect
Import/export and schema

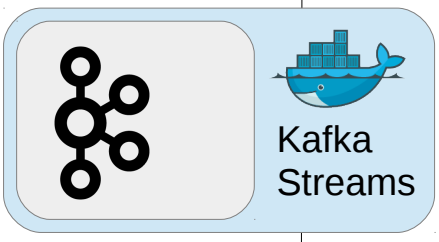


KSQL
Streaming SQL for Kafka

SQL query language

```
CREATE TABLE
    error_counts
AS SELECT
    error_code,
    count(*)
FROM
    monitoring_stream
WINDOW TUMBLING
    (SIZE 1 MINUTE)
WHERE
    type = 'ERROR
```

Kafka Streams
Data streaming with Kafka



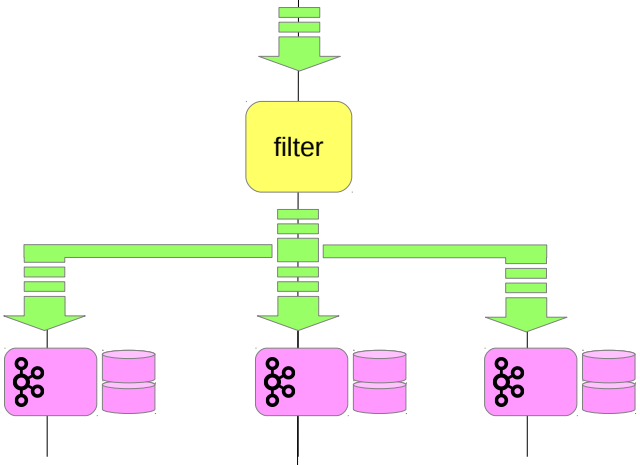
Stream processing API

```
KStream<String, Long> stream = ...;

stream.foreach(
    new ForeachAction<String, Long>()
        {
        @Override
        public void apply(
            String key,
            Long value
            ){
            ....
            }
        }
    );
```

D.Morris
Institute for Astronomy,
Edinburgh University
May 2018

Selective filtering



Internal event schema, critical set of attributes needed for filtering.

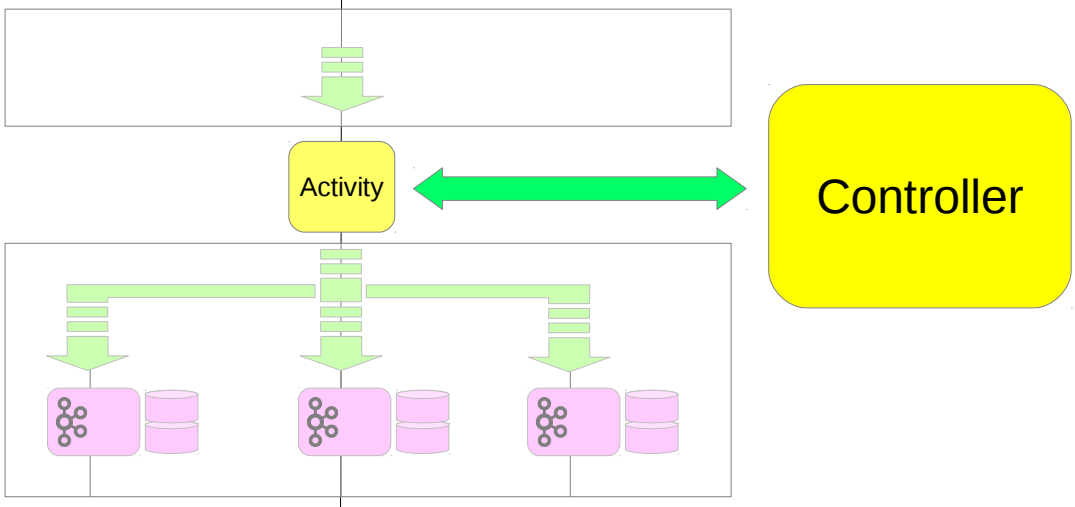
Event-expander, correlates mini-even with original event and creates a new event with additional params.

HTTP webservice with in memory cache ?

memcache ?

D.Morris
Institute for Astronomy,
Edinburgh University
May 2018

REST webservice control API



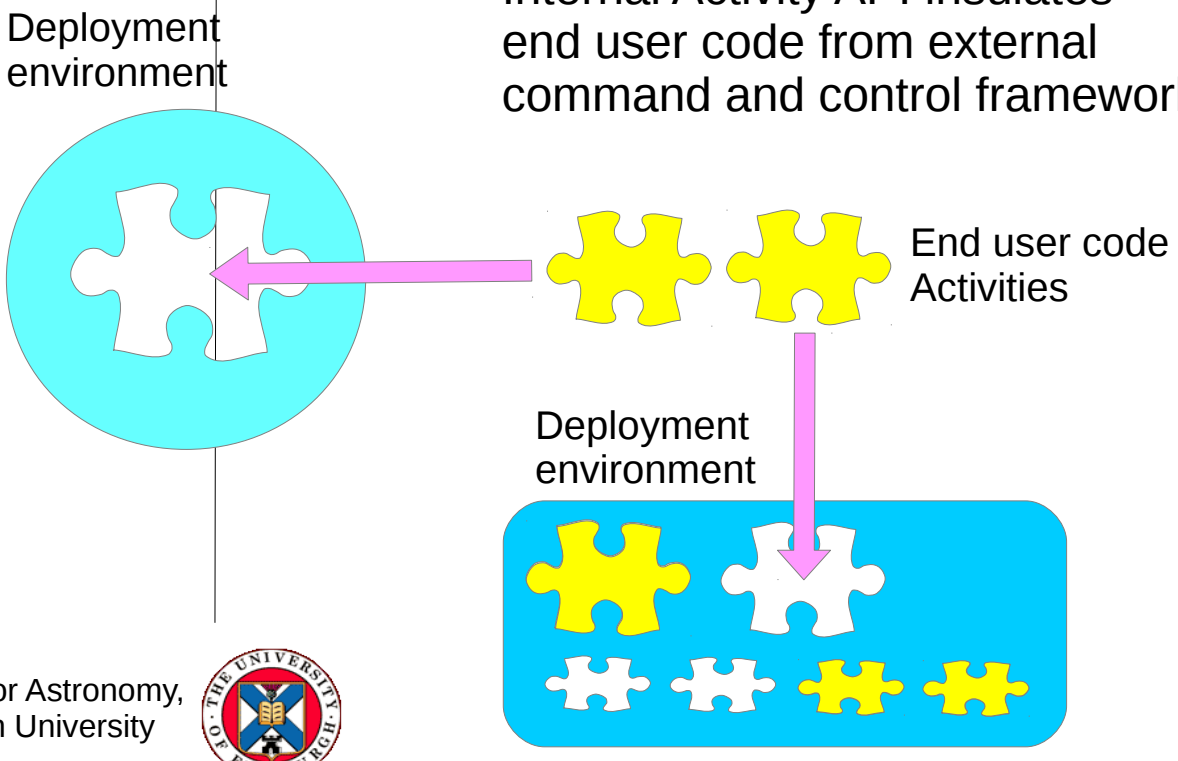
Spring-cloud … yet another framework

Very easy to start .. harder to control the direction.

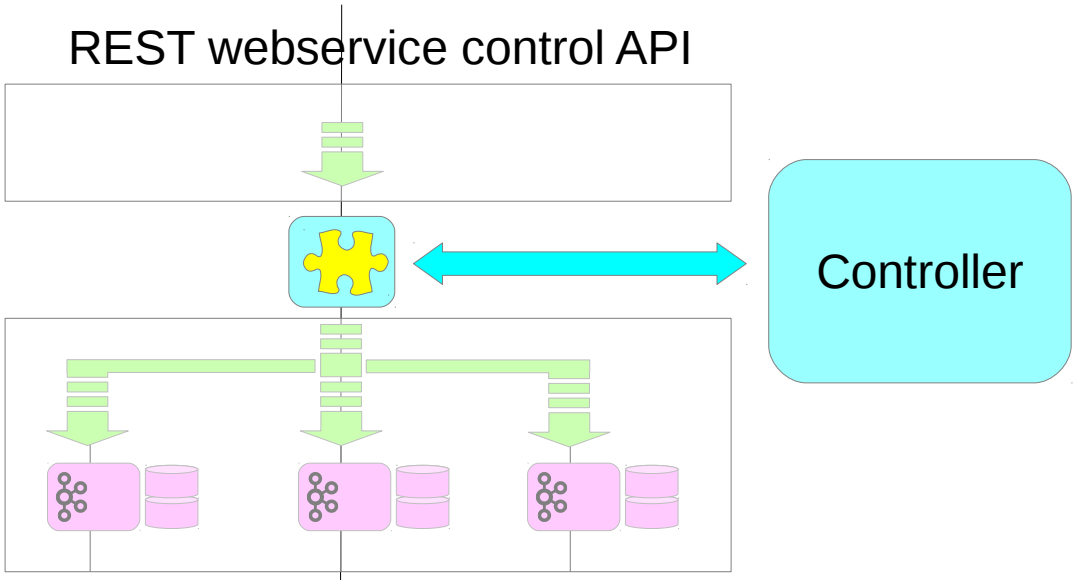
Our own REST API for deploying and controlling Kaffa nodes

Separation of concerns
- Processor Activity interface
- Command and control API

Internal Activity API insulates end user code from external command and control framework.



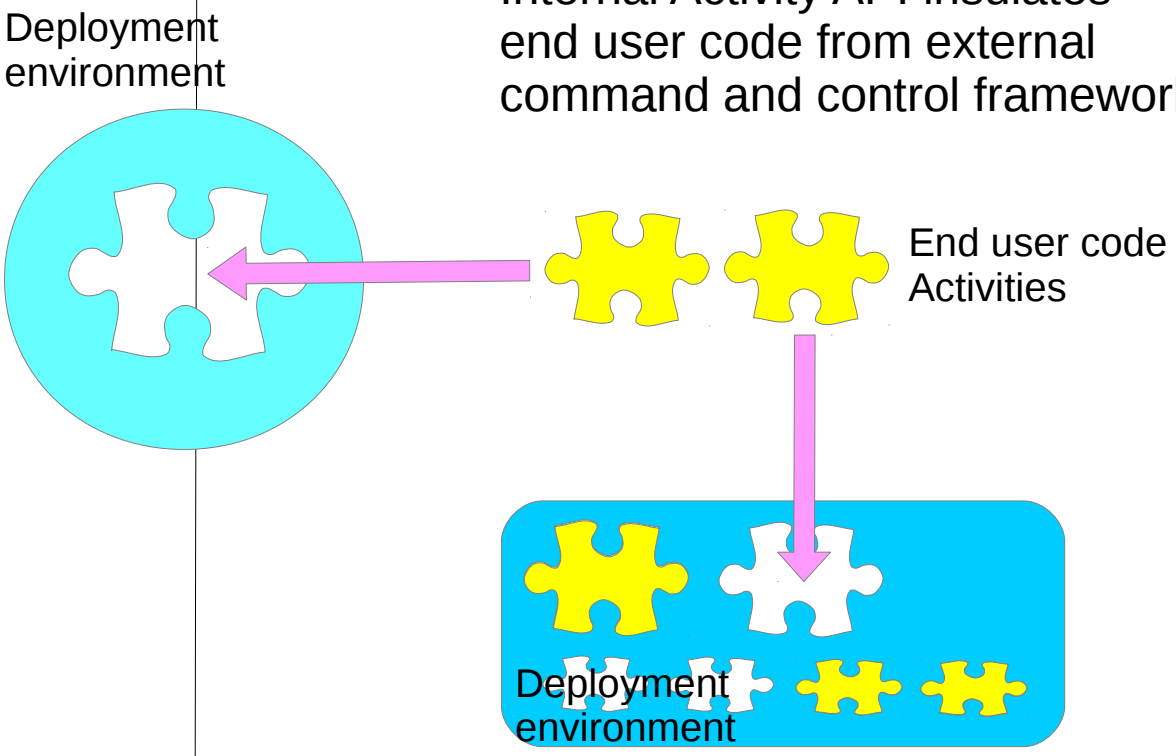
D.Morris
Institute for Astronomy,
Edinburgh University
May 2018



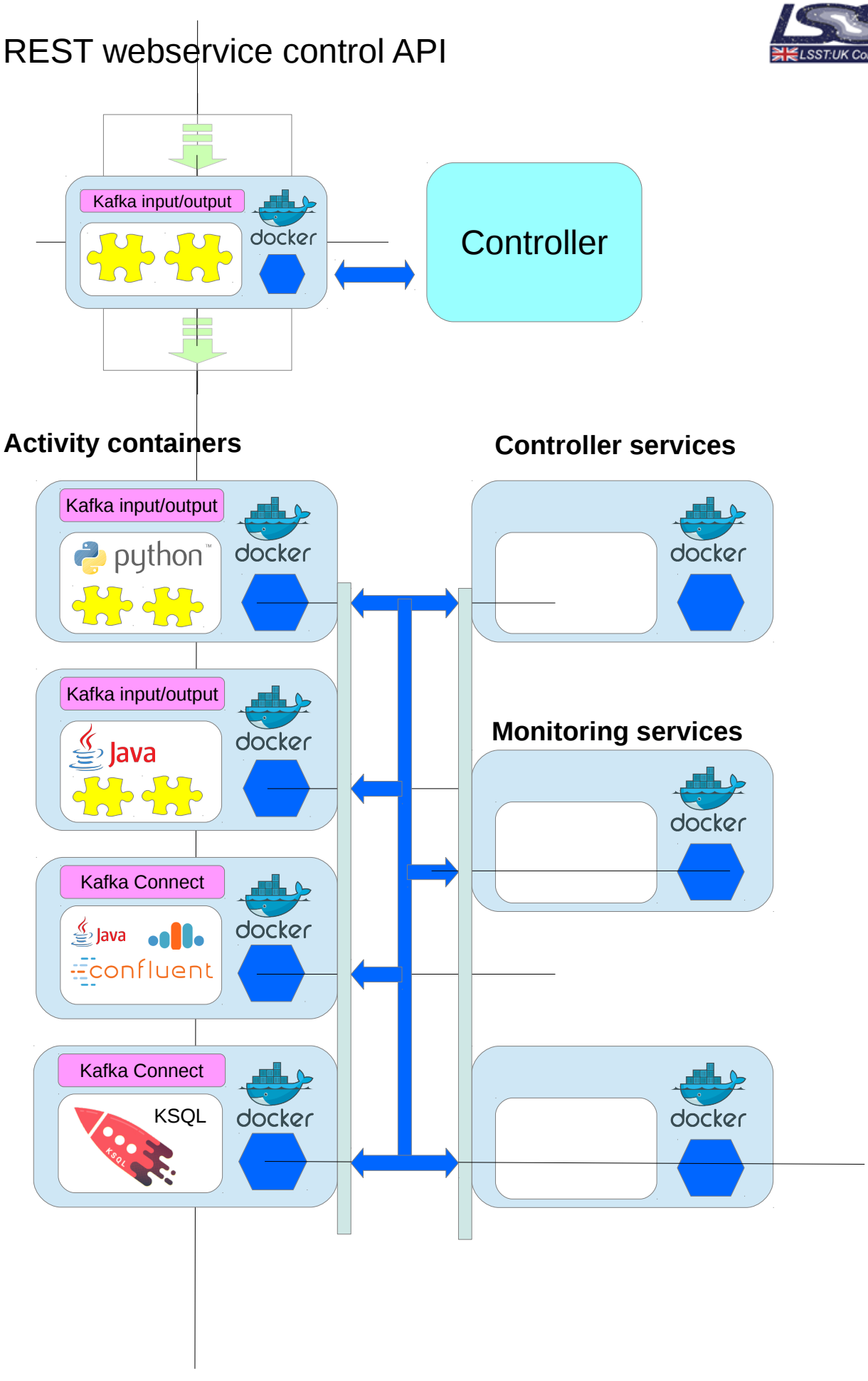
REST API for deploying and controlling Kaffa nodes

Separation of concerns
- Processor Activity interface
- Command and control API

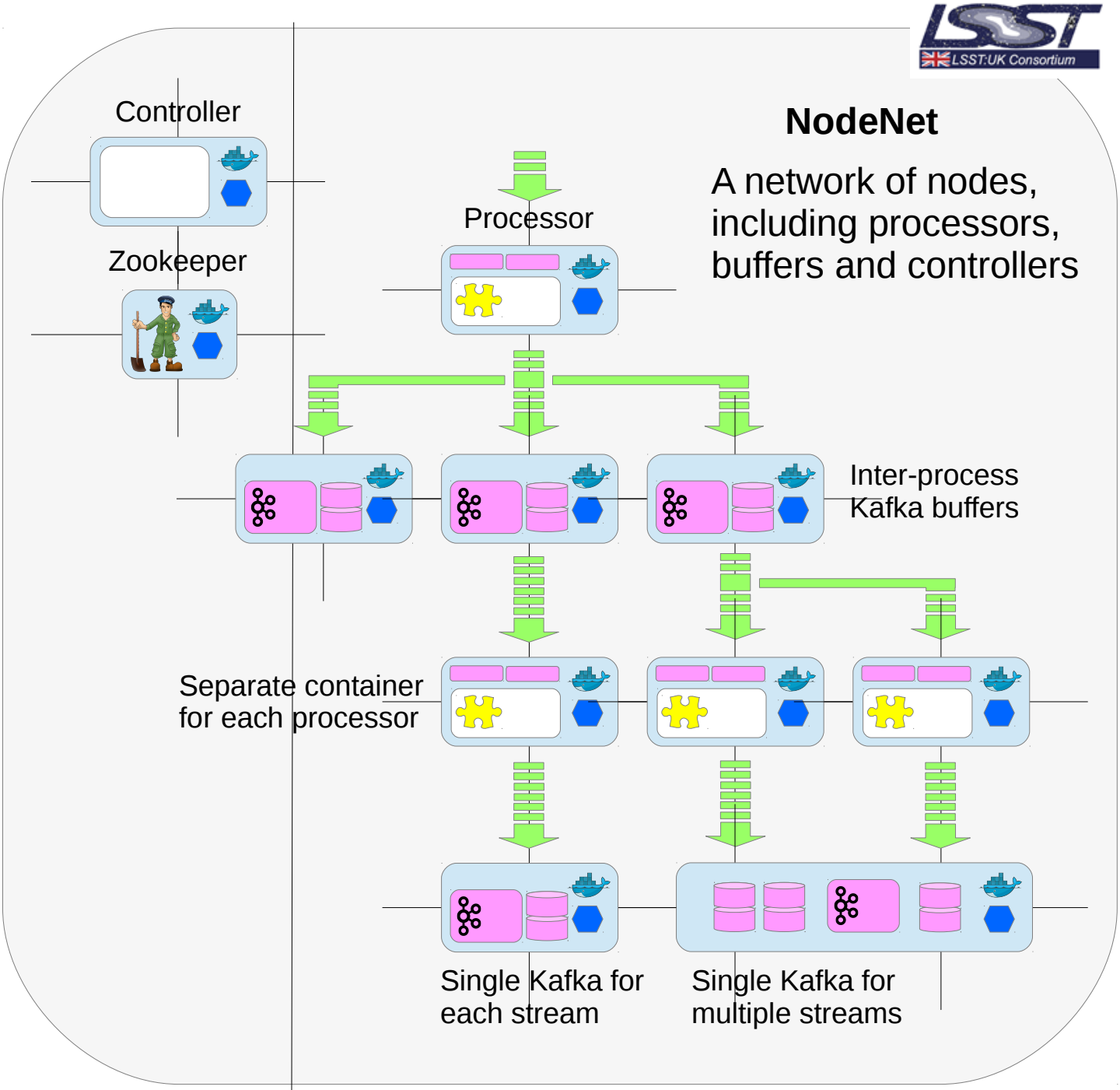
Internal Activity API insulates end user code from external command and control framework.



D.Morris
Institute for Astronomy,
Edinburgh University
May 2018



D.Morris
Institute for Astronomy,
Edinburgh University
May 2018



We need to be able to create multiple NodeNet networks.

Development, testing and live services.

Single VM local instances.

Multiple physical machine deployments.

Start with a controller 'seed' and a set of configuration files.

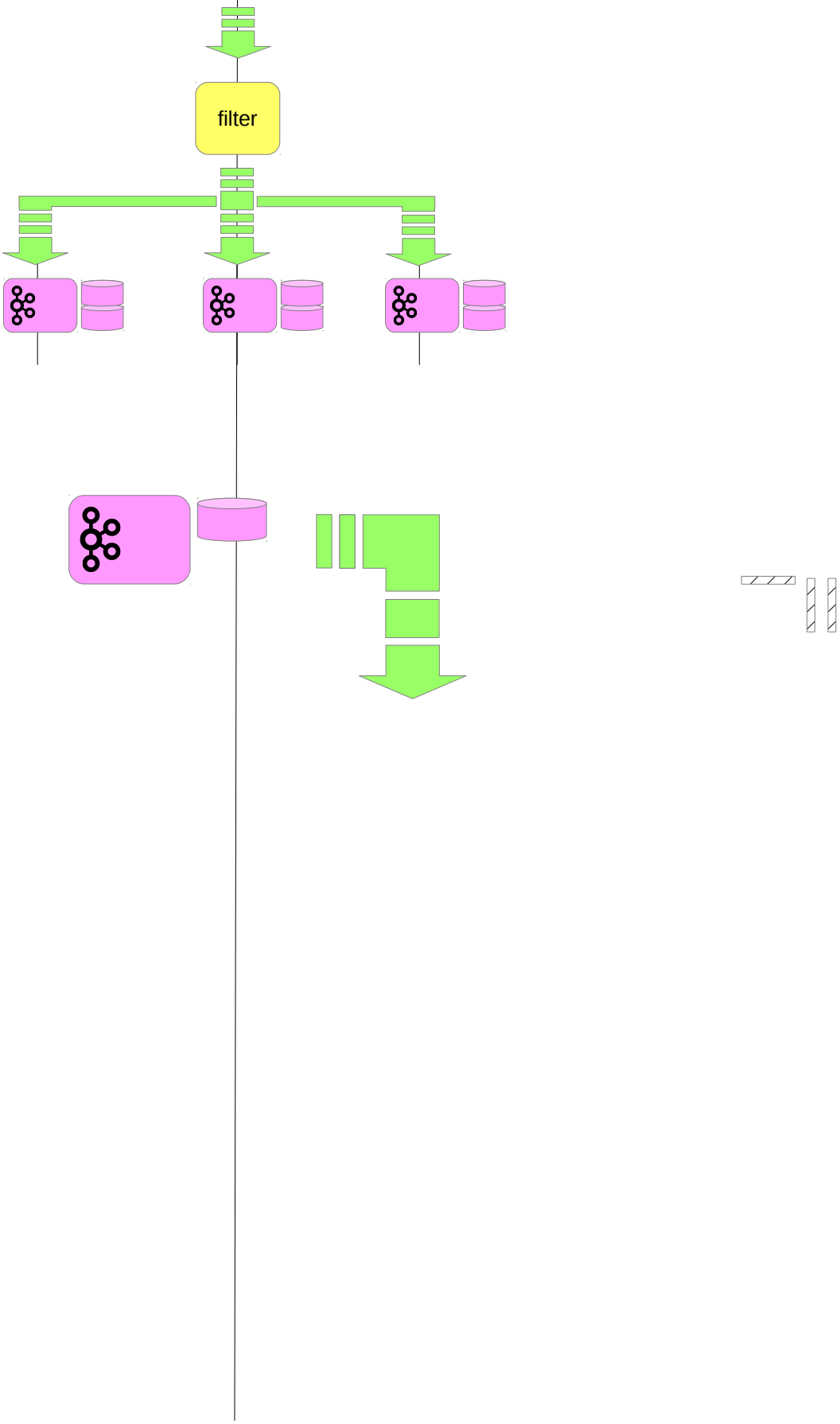
D.Morris
Institute for Astronomy,
Edinburgh University
May 2018

Jupyter visualization

Elasticsearch

D.Morris
Institute for Astronomy,
Edinburgh University
May 2018

Selective filtering



D.Morris
Institute for Astronomy,
Edinburgh University
May 2018