WITHIN THE EYE OF THE STORM

INTRODUCTION TO STORM FRAMEWORK

Sjoerd Mulder / @frontendless
Science Rockstars / @sciencerockstar

SCIENCE ROCKSTARS

- Good science leads to great business
- Productize research work
- Featured on Techcrunch, Wired and Forbes

WHAT IS STORM?

- Distributed realtime computation system
- Since 2010
- Created by Backtype (aquired by Twitter)
- Most starred JVM project on Github

USE CASES

- Realtime analytics
- Online machine learning
- Continuous computation
- Distributed RPC

KEY CONCEPTS

TUPLE
AN ORDERED LIST OF OBJECTS

```
[ 198735697, "foobar", { "ip" : "10.0.0.1" } ]
```

STREAM
AN UNBOUNDED SEQUENCE OF TUPLES

SPOUT
GENERATES TUPLES FROM OTHER SOURCES
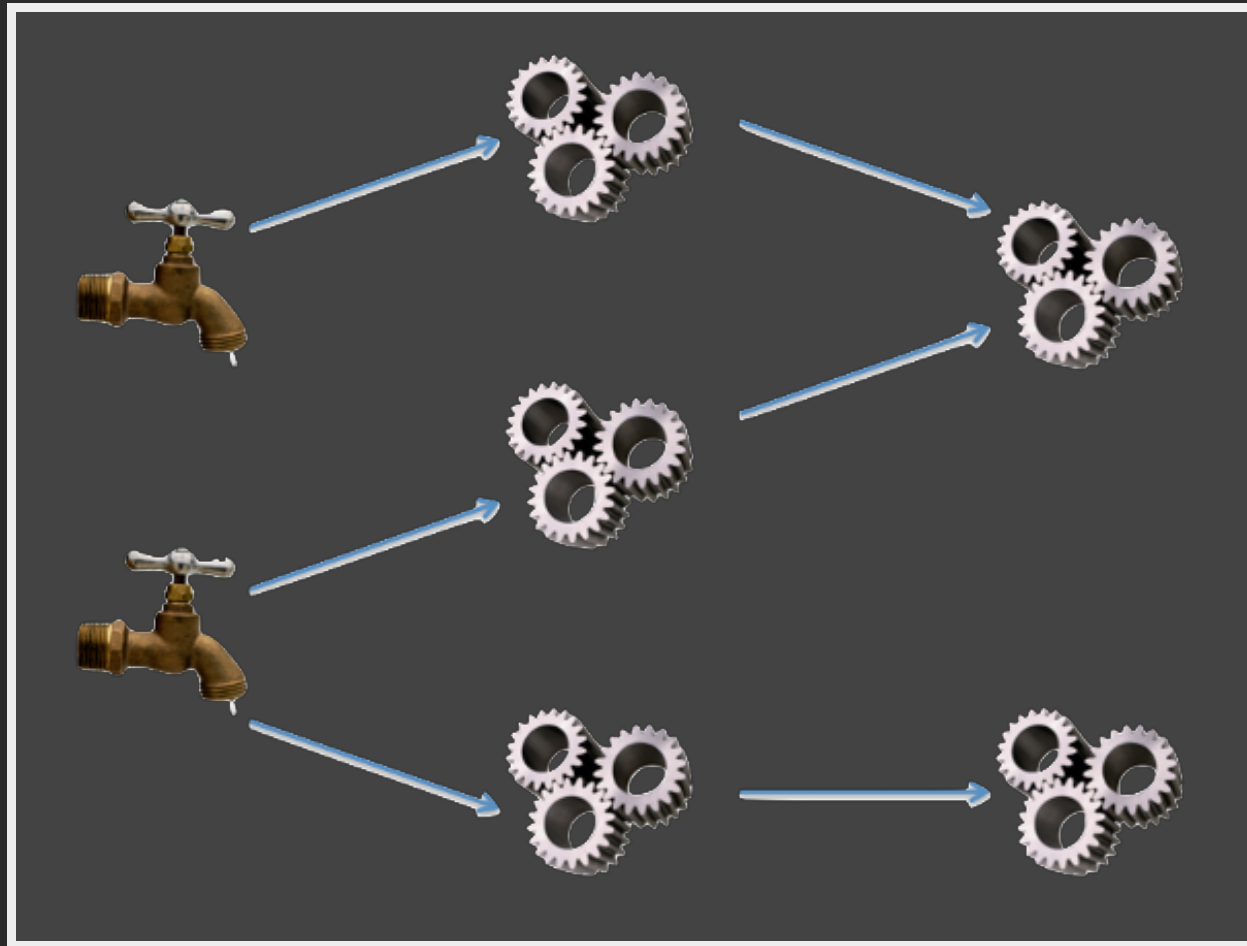
- event data
- log files
- queues

BOLT
PROCESSES A SINGLE TUPLE

- Can be any functionality
- Filter, computation, aggregation, join
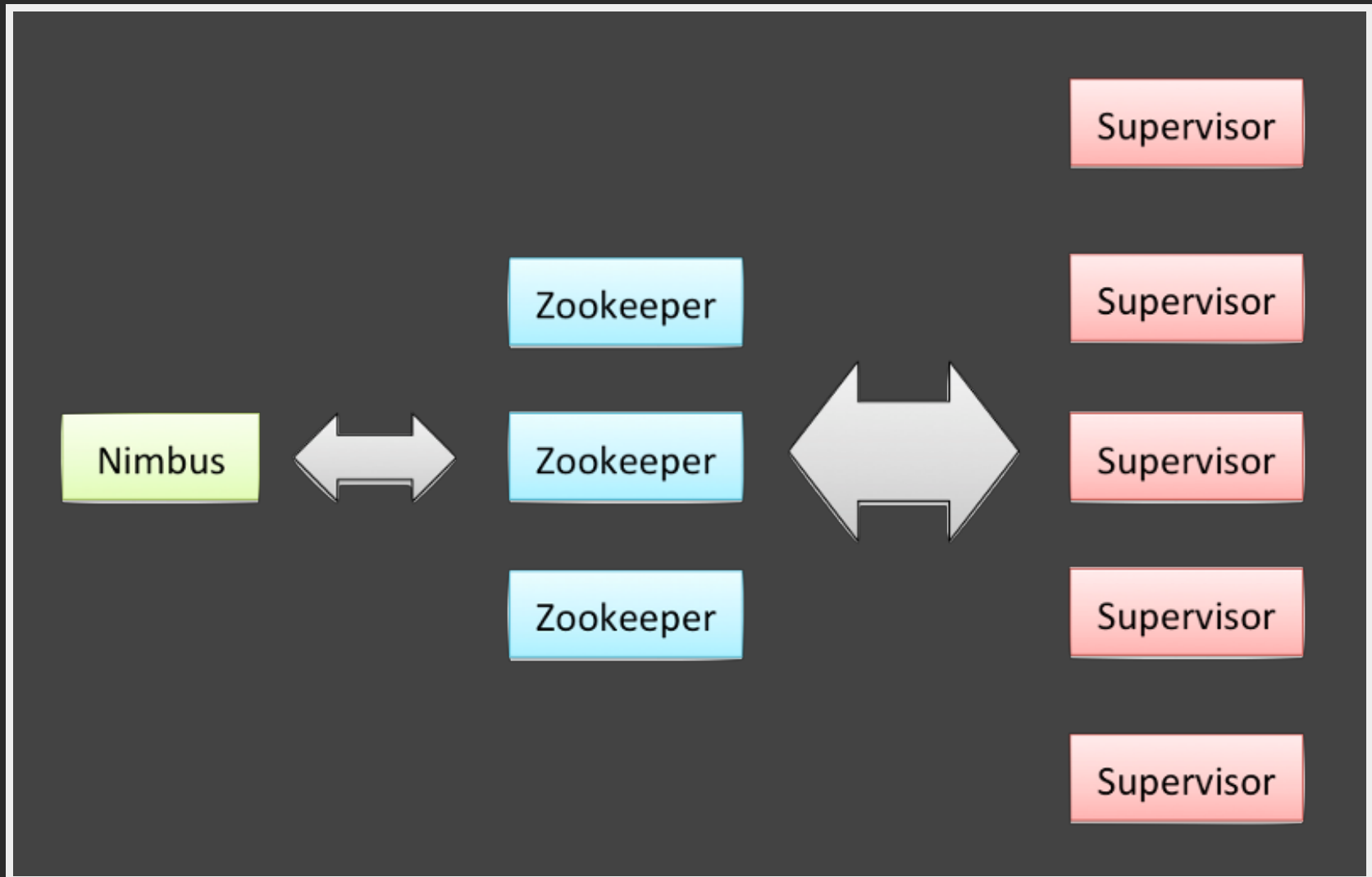- Can create a new stream

TOPOLOGY
CONFIGURED GRAPH OF SPOUTS AND BOLTS
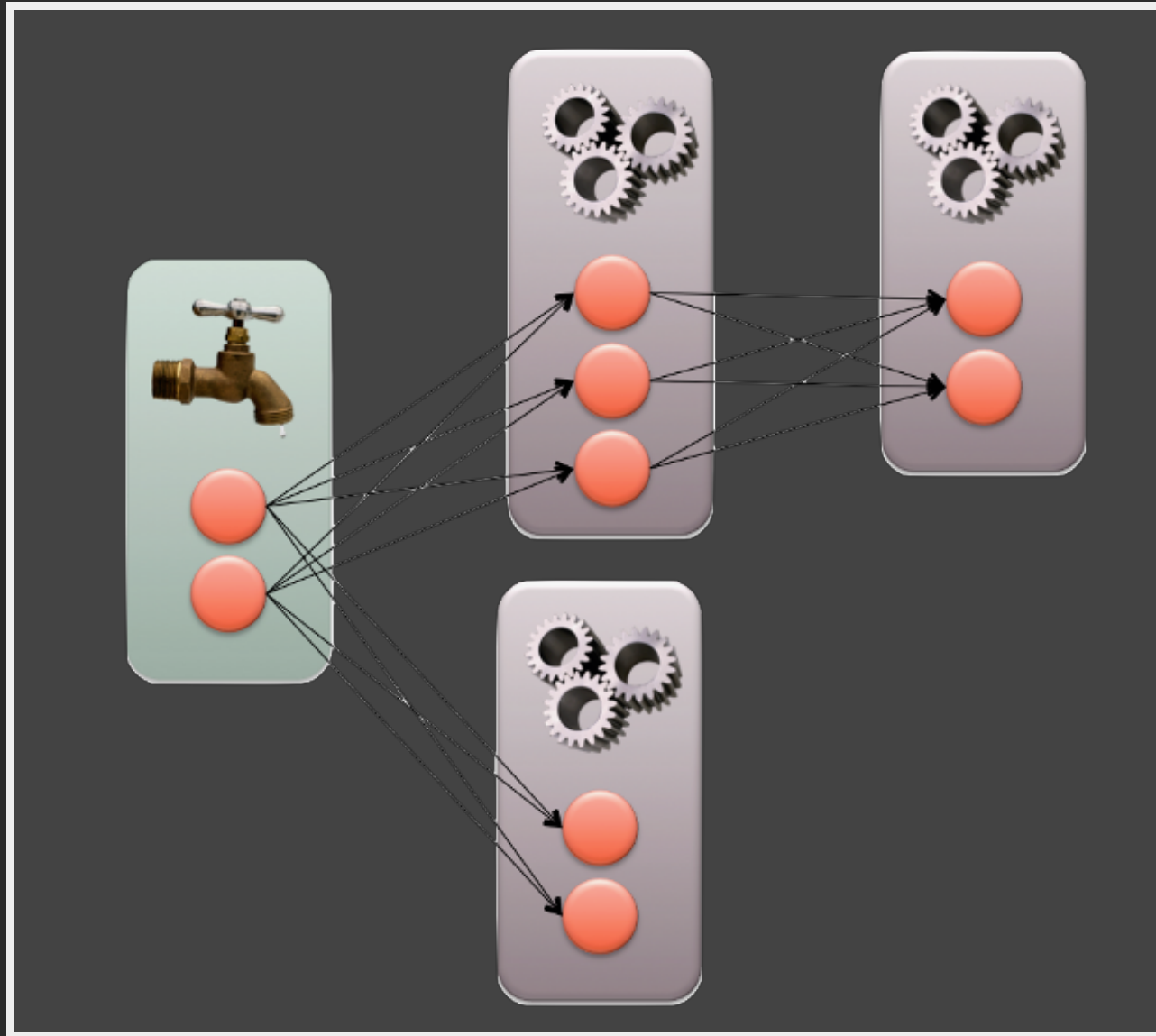
CLUSTER

ZOOKEEPER

NIMBUS

- Master node
- Manages the topologies
- Job Tracker in Hadoop

```
$ storm jar mycode.jar org.example.MyTopology demo
```

SUPERVISOR

- Manages workers
- Worker is a thread spawn by supervisor to do work
- Task Tracker in Hadoop

WORKER

STREAM GROUPING

WHEN A TUPLE IS EMITTED, WHICH TASK RECEIVES IT?

- **Shuffle** grouping does random task distribution
- **Field** grouping groups by specific field (like Hadoop partitioning behaviour)
- **All** grouping replicates to all tasks (carefull)
- **Global** grouping sends an entire stream to one task

WORD COUNT EXAMPLE

```java
TopologyBuilder builder = new TopologyBuilder();
builder.setSpout("spout", new RandomSentenceSpout(), 5);

builder.setBolt("split", new SplitSentence(), 8)
        .shuffleGrouping("spout");
builder.setBolt("count", new WordCount(), 12)
        .fieldsGrouping("split", new Fields("word"));
```

RANDOM SENTENCE SPOUT

```java
public class RandomSentenceSpout extends BaseRichSpout {
    SpoutOutputCollector collector;
    Random random;
    String[] sentences = new String[] {
            "the cow jumped over the moon",
            ...
            "i am at two with nature"};

    @Override
    public void nextTuple() {
        Utils.sleep(100);
        String sentence =
            sentences[random.nextInt(sentences.length)];
        collector.emit(new Values(sentence));
    }
}
```

SPLIT SENTENCE BOLT

```java
public class SplitSentence extends BaseBasicBolt {

    @Override
    public void execute(Tuple tuple, OutputCollector collector) {
        String sentence = tuple.getString(0);
        for(String word : sentence.split(" ")) {
            collector.emit(new Values(word));
        }
    }

    @Override
    public void declareOutputFields(OutputFieldsDeclarer ofd) {
        ofd.declare(new Fields("word"));
    }

}
```

WORD COUNT BOLT

```java
public class WordCount extends BaseBasicBolt {
    Map<String, Integer> counts = new HashMap<>();

    @Override
    public void execute(Tuple tuple, OutputCollector collector) {
        String word = tuple.getString(0);
        Integer count = counts.get(word);
        count = (count == null) ? 1 : count + 1;
        counts.put(word, count);
        collector.emit(new Values(word, count));
    }

    @Override
    public void declareOutputFields(OutputFieldsDeclarer ofd) {
        ofd.declare(new Fields("word", "count"));
    }
}
```

LOCAL CLUSTER
DURING DEVELOPMENT

```java
Config conf = new Config();
conf.setDebug(true);
conf.setMaxTaskParallelism(3);

LocalCluster cluster = new LocalCluster();

cluster.submitTopology("word-count", conf,
                builder.createTopology());

Thread.sleep(10000);

cluster.shutdown();
```

HOW DO WE USE STORM?

- Amazon AWS
- Automated using Chef
- Storm UI
- Graylog2 for log monitoring

PERSUASIONAPI

- Based on behavioral science PhD research
- An individual 'how' recommender
- SaaS-based Machine Learning
- Already served 200.000.000+ advices
- Running in production on Storm

AND WE ARE HIRING

- Looking for rockstars!
- q@sciencerockstars.com
- Contact me

WANT TO GET STARTED OR QUESTIONS?

- http://storm-project.net/
- https://github.com/nathanmarz/storm
- https://github.com/nathanmarz/storm-starter
- https://github.com/nathanmarz/storm/wiki/Understanding the-parallelism-of-a-Storm-topology