ИТМО

Apache Storm, a brief overview

Grigoryev Mikhail, J4133c

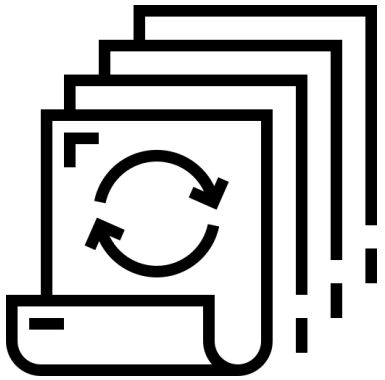
Apache Storm — What Is It?

APACHE STORM™

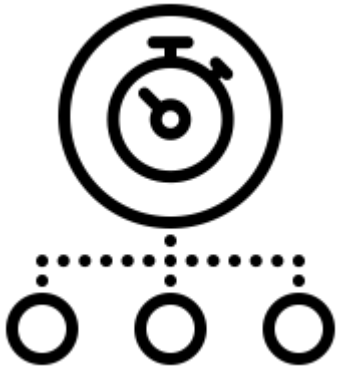
Big Data framework by Apache specialized in parallel and real-time process

Batch vs Real Time Processing

Batch processing – efficient processing of very high volumes of data (not immediate). Requires separate programs for input, processing and output.



Real time processing – continuous input, processing and output of data (immediate, responsive).



Batch Processing

Benefits:
- Efficiency
- Convenient time

Drawbacks:
- Time lag

Real Time Processing

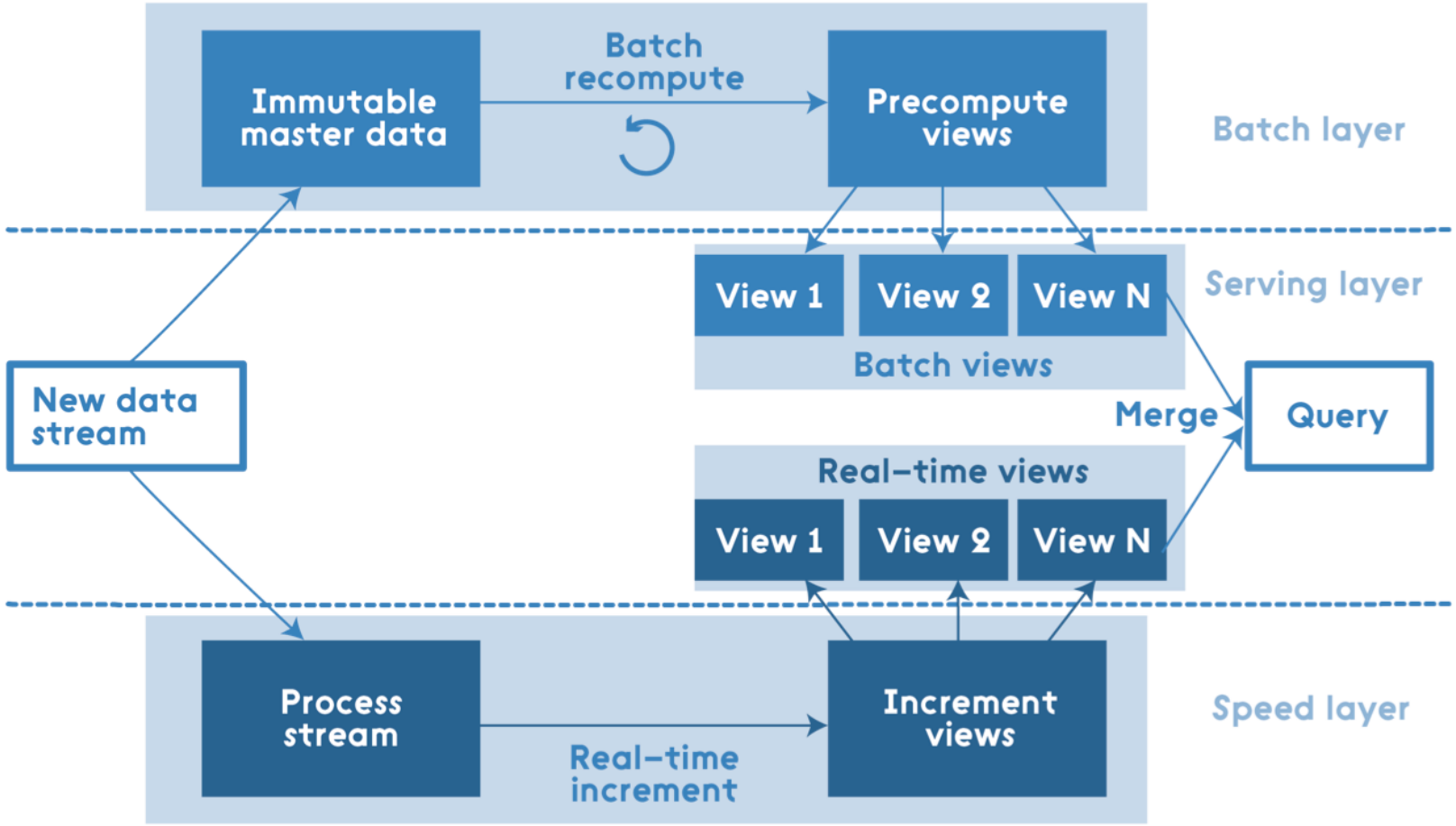
ITMO

Benefits:
- Immediacy
- Ability to take quick action
- Responsiveness

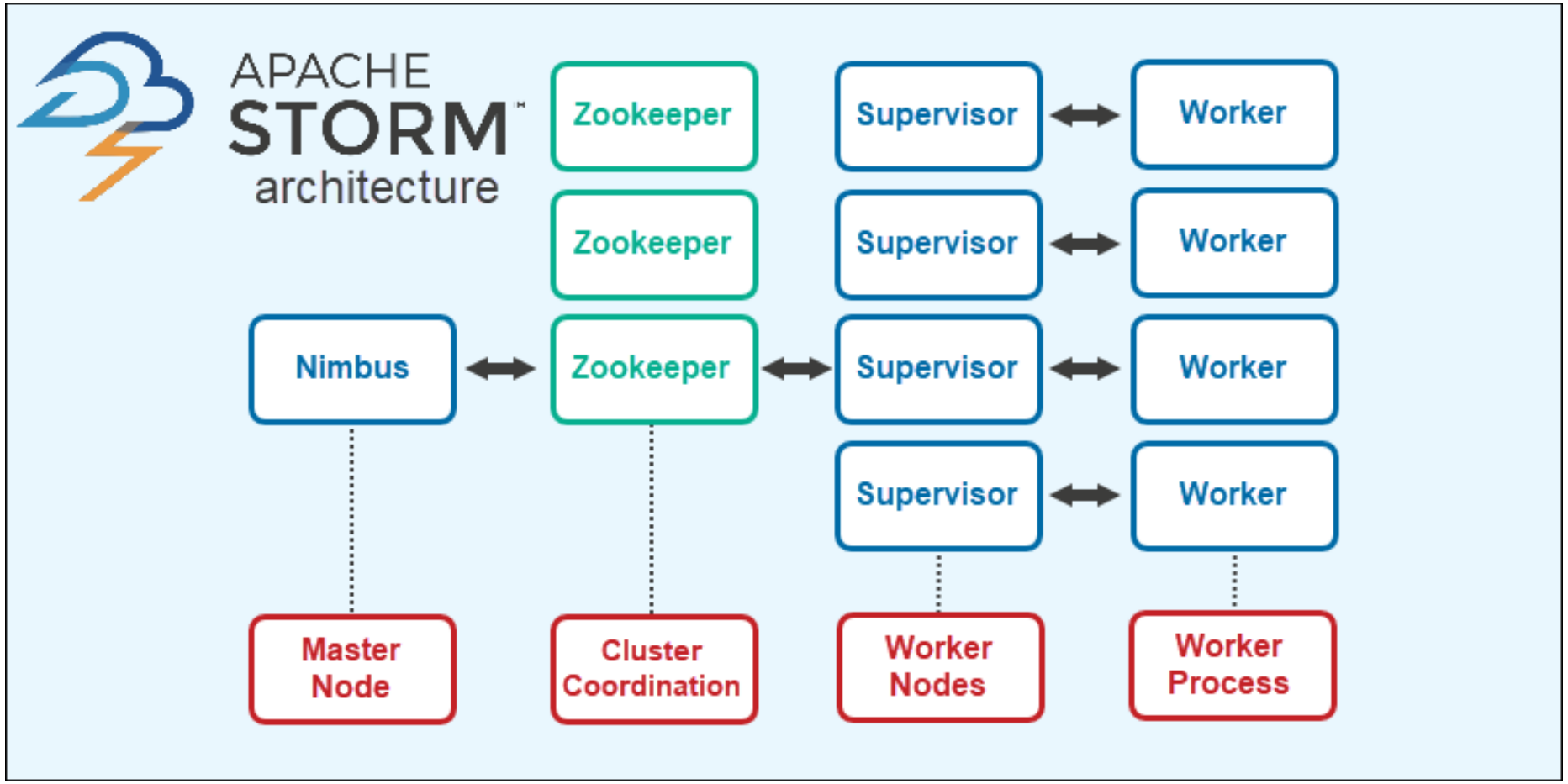
Drawbacks:
- Expensive computations
- Cumbersome setup

APACHE STORM™

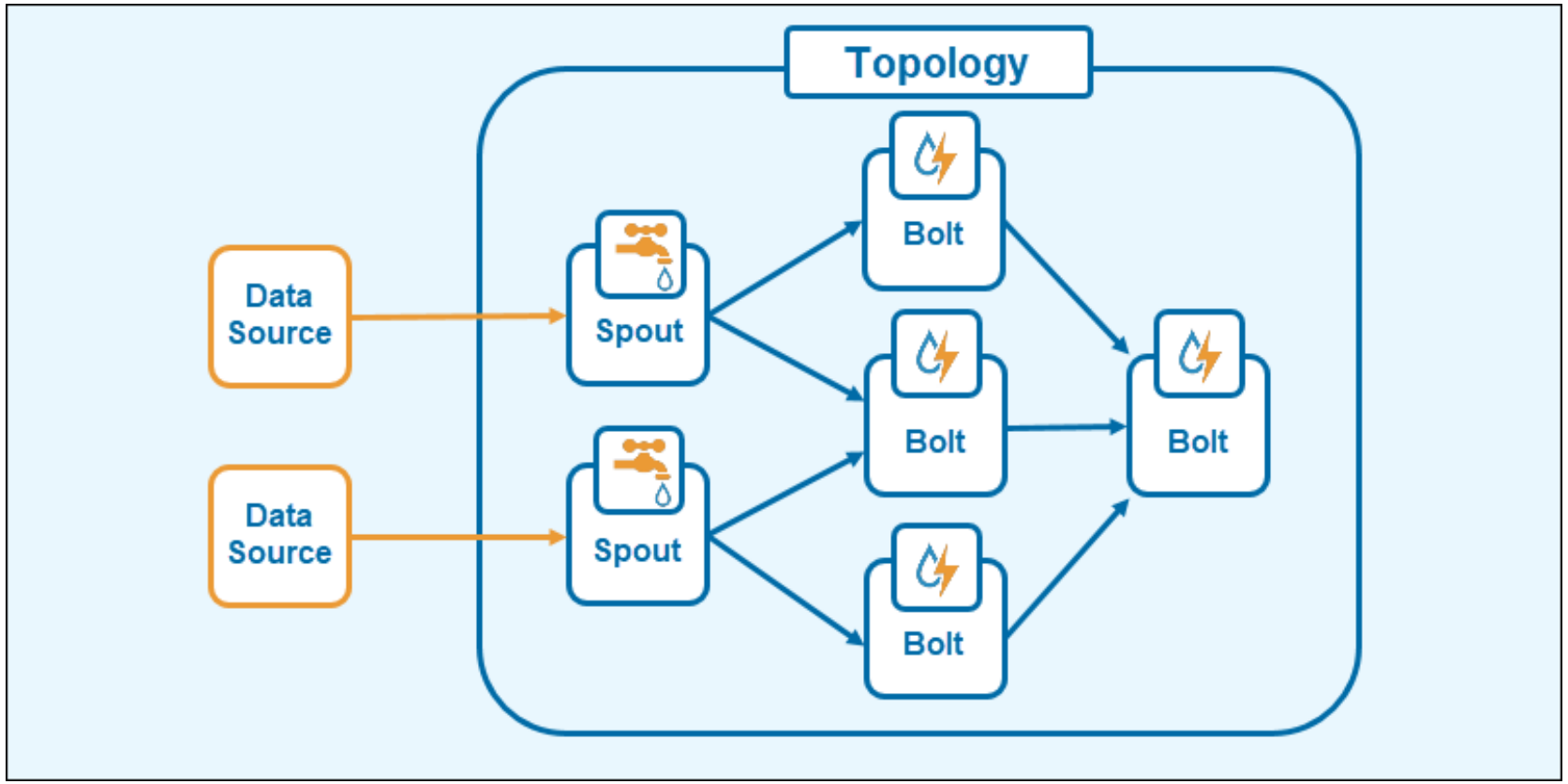
Lambda Architecture



Storm Components (Architecture)



Storm Topology Components



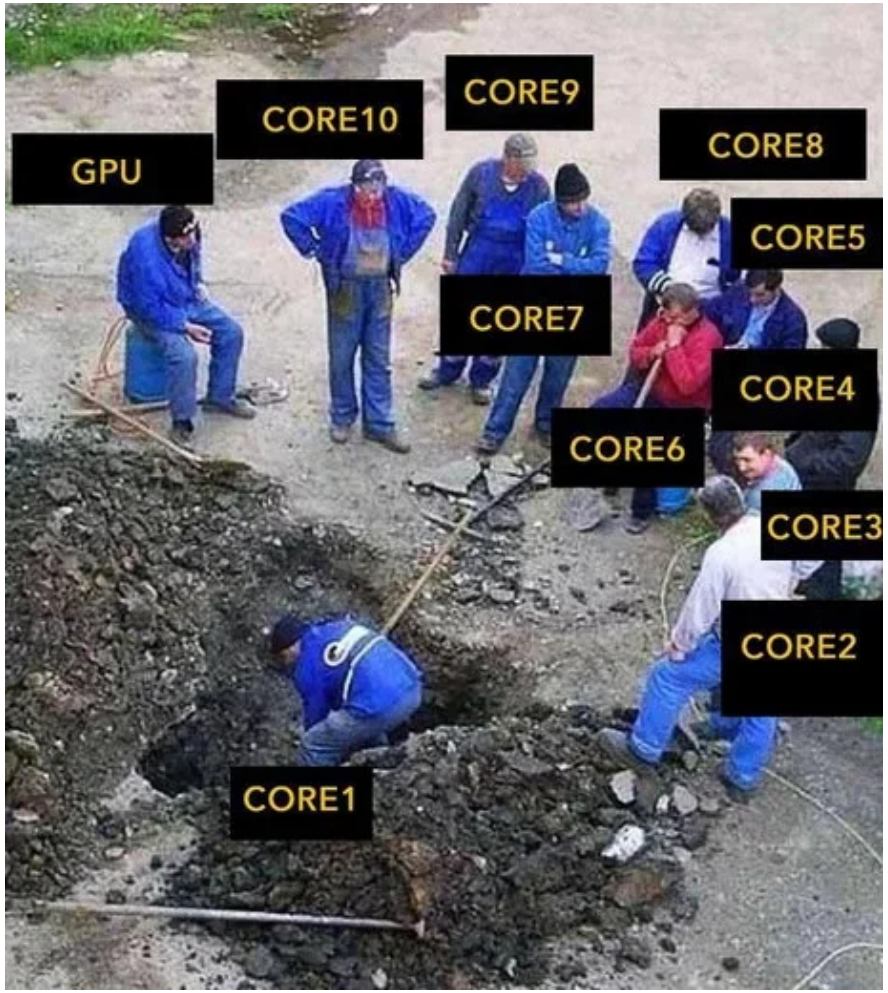
Tuples:
immutable arrays of elements

Streams:
unbounded arrays of tuples

Usecases

Uses:
- Stream processing
- Real time views
- Distributed computation

Areas:
- Financial services
- Retail
- Telecom
- Web
- ...

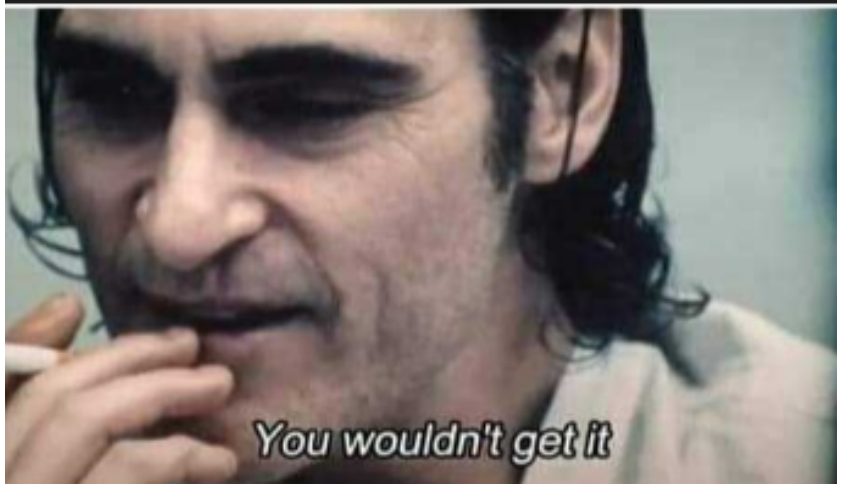


Features

- Simple syntax
- Multilingual support (*use Java tho*)
- Fault tolerant
- CLI tools and daemon commands are Python-based
- Support for local and remote modes

```java
public class Meme
{
    private Joke joke;

    public void setJoke(Joke newJoke)
    {
        this.joke = newJoke;
    }
}
```



Practical Part

https://github.com/Dormant512/itmo_notes/tree/main/storm_local_01

ИТМО

Thank you for attention!