

Experiences with the SZTAKI Cloud

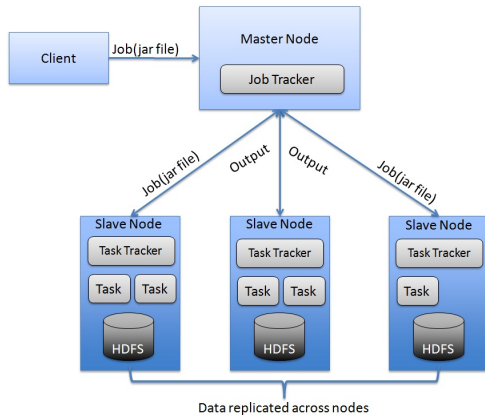
Distributed infrastructure management

András Benczúr & Márton Balassi
Informatics Laboratory¹

¹Computer and Automation Research Institute of the Hungarian Academy of Sciences

June 19, 2014

Motivation



Hadoop architecture

Motivation

Architecture

- ▶ Cloud architecture does not fit the standard Hadoop architecture.
- ▶ Distributed graph processing systems perform decent.
- ▶ Distributed stream processing systems perform almost the same as on bare metal.

Motivation

Architecture

- ▶ Cloud architecture does not fit the standard Hadoop architecture.
- ▶ Distributed graph processing systems perform decent.
- ▶ Distributed stream processing systems perform almost the same as on bare metal.

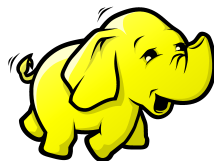
Motivation

Architecture

- ▶ Cloud architecture does not fit the standard Hadoop architecture.
- ▶ Distributed graph processing systems perform decent.
- ▶ Distributed stream processing systems perform almost the same as on bare metal.

Frameworks

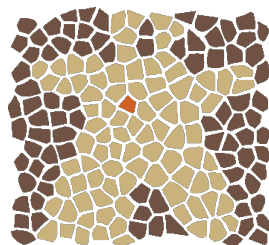
- ▶ Hadoop
- ▶ Giraph
- ▶ Spark & Bagel
- ▶ GraphLab & GraphChi
- ▶ Hama



Hadoop logo

Frameworks

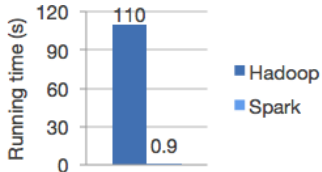
- ▶ Hadoop
- ▶ **Giraph**
- ▶ Spark & Bagel
- ▶ GraphLab & GraphChi
- ▶ Hama



Giraph [logo](#)

Frameworks

- ▶ Hadoop
- ▶ Giraph
- ▶ Spark & Bagel
- ▶ GraphLab & GraphChi
- ▶ Hama



Logistic regression: Spark vs.
Hadoop

Frameworks

- ▶ Hadoop
- ▶ Giraph
- ▶ Spark & Bagel
- ▶ **GraphLab & GraphChi**
- ▶ Hama



GraphLab [logo](#)

Frameworks

- ▶ Hadoop
- ▶ Giraph
- ▶ Spark & Bagel
- ▶ GraphLab & GraphChi
- ▶ **Hama**



Hama [logo](#)

Streaming

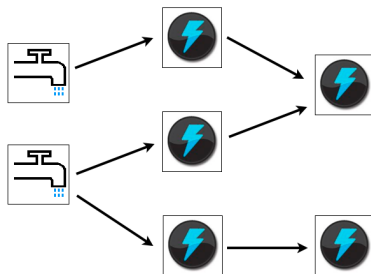
- ▶ S4
- ▶ Storm
- ▶ Trident
- ▶ Akka
- ▶ Samza

S4 *distributed stream
computing platform*

S4 logo

Streaming

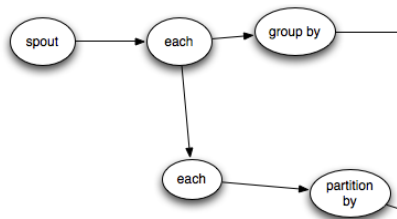
- ▶ S4
- ▶ **Storm**
- ▶ Trident
- ▶ Akka
- ▶ Samza



Storm topology

Streaming

- ▶ S4
- ▶ Storm
- ▶ **Trident**
- ▶ Akka
- ▶ Samza



Trident **topology**

Streaming

- ▶ S4
- ▶ Storm
- ▶ Trident
- ▶ Akka
- ▶ Samza



Akka [logo](#)

Streaming

- ▶ S4
- ▶ Storm
- ▶ Trident
- ▶ Akka
- ▶ **Samza**

Stream Processing



Samza [stack](#)

Demo

Live demo...

Bringing up elastically configurable Hadoop cluster in the cloud and running an example job.

Comments

Things to note...

- ▶ Thanks for providing this service!
- ▶ Additional on demand infrastructure
- ▶ In our experience it is I/O bounded
- ▶ OpenNebula GUI is rather controversial sometimes
- ▶ Update the documentation, please

Comments

Things to note...

- ▶ **Thanks for providing this service!**
- ▶ **Additional on demand infrastructure**
- ▶ In our experience it is I/O bounded
- ▶ OpenNebula GUI is rather controversial sometimes
- ▶ Update the documentation, please

Comments

Things to note...

- ▶ Thanks for providing this service!
- ▶ Additional on demand infrastructure
- ▶ In our experience it is I/O bounded
- ▶ OpenNebula GUI is rather controversial sometimes
- ▶ Update the documentation, please

Comments

Things to note...

- ▶ Thanks for providing this service!
- ▶ Additional on demand infrastructure
- ▶ In our experience it is I/O bounded
- ▶ OpenNebula GUI is rather controversial sometimes
- ▶ Update the documentation, please

Comments

Things to note...

- ▶ Thanks for providing this service!
- ▶ Additional on demand infrastructure
- ▶ In our experience it is I/O bounded
- ▶ OpenNebula GUI is rather controversial sometimes
- ▶ Update the documentation, please