



Contribution ID: 11

Type: **not specified**

## Hadoop in Complex Systems Research ( A Review on Tools, Best Practices & Applications )

*Tuesday, August 27, 2013 9:40 AM (40 minutes)*

A Hadoop cluster is the tool of choice for many large scale analytics applications and a large variety of commercial tools is available for Data Warehouses and for typical SQL-like applications.

But how to deal with networks and time series? How to collect data for complex systems studies and what are good practices for working with libraries like Mahout and Giraph?

The sample use case deals with a data set from Wikipedia to illustrate one can combine multiple public data sources with own personal data collections, e.g. from Twitter, intranet servers or even personal mailboxes. Efficient approaches for time series (pre)-processing and time dependent graph analysis will be presented.

**Author:** KÄMPF, Mirko (Cloudera)

**Presenter:** KÄMPF, Mirko (Cloudera)

**Session Classification:** Plenary talks

**Track Classification:** Big Data and Large Storage Systems