

Developing and optimizing applications for the Hadoop environment

Thursday, 13 October 2016 15:15 (15)

This contribution is about sharing our recent experiences of building Hadoop based application. Hadoop ecosystem now offers myriad of tools which can overwhelm new users, yet there are successful ways these tools can be leveraged to solve problems. We look at factors to consider when using Hadoop to model and store data, best practices for moving data in and out of the system and common processing patterns, at each stage relating with the real world experience gained while developing such application. We share many of the design choices, tools developed and how to profile a distributed application which can be applied for other scenarios as well. In conclusion, the goal of the presentation is to provide guidance to architect Hadoop based application and share some of the reusable components developed in this process.

Primary Keyword (Mandatory)

Data processing workflows and frameworks/pipelines

Secondary Keyword (Optional)

Software development process and tools

Tertiary Keyword (Optional)

Primary author(s) : KOTHURI, Prasanth (CERN)

Co-author(s) : LANZA GARCIA, Daniel (Ministere des affaires etrangeres et europeennes (FR)); HERMANS, Joeri (Universiteit Maastricht (NL)); SURDY, Kacper (CERN); BARANOWSKI, Zbigniew (CERN)

Presenter(s) : KOTHURI, Prasanth (CERN)

Session Classification : Track 5: Software Development

Track Classification : Track 5: Software Development