

Symposium of the Center for Network and Storage Enabled Collaborative
Computational Science



Contribution ID: 17

Type: **Presentation**

Bootstrapping Big Data with Spark SQL and Data Frames

Friday, 19 May 2017 10:30 (30 minutes)

Apache Spark, a popular open source big data tool from the Hadoop ecosystem is seeing rapid adoption across industry and academia, yet it is still generally not well known. For this talk we will demonstrate some large scale samples of how easy it is to benefit from spark SQL and Data Frames for Python and R programmers.

Primary author: PALEN, Brock

Presenter: PALEN, Brock

Session Classification: Complementary Technology Solutions

Track Classification: Complementary Technology Solutions