

Published on *CERN openlab* (<http://test-static-05.web.cern.ch>)

[Home](#) > CMS Data-Services Ingestion into CERN's Hadoop Big Data Analytics Infrastructure

---

## CMS Data-Services Ingestion into CERN's Hadoop Big Data Analytics Infrastructure <sup>[1]</sup>

**Date published:**

Tuesday, 1 September, 2015

**Document type:**

Summer student report

**Author(s):**

A. Bose

This document introduces a new data ingestion framework called HLoader, built around Apache Sqoop to perform data ingestion jobs between RDBMS and Hadoop Distributed File System (HDFS). The HLoader framework deployed as a service inside CERN will be used for CMS Data Popularity ingestion into Hadoop clusters. HLoader could also be used for similar use cases like CMS and ATLAS Job Monitoring, ACCLOG databases, etc. The first part of the report describes the architectural details of HLoader, giving some background information about Apache Sqoop. The rest of the report focuses on the HLoader programming API, and is meant to be an entry point for developers describing how HLoader works, and possible directions of extending the framework in future.

**Report on ZENODO:**

[Document on ZENODO](#) <sup>[2]</sup>

- [Visit Us](#)
- [RSS Feeds](#)

DISCLAIMER: This Web page contains pointers to material related to the management of CERN openlab in the Information Technology Department at the European Organization for Nuclear Research (CERN). Their use and distribution are regulated by the [CERN copyright notice](#).



---

**Source URL:** [http://test-static-05.web.cern.ch/publications/technical\\_documents/cms-data-services-ingestion-cern%E2%80%99s-hadoop-big-data-analytics](http://test-static-05.web.cern.ch/publications/technical_documents/cms-data-services-ingestion-cern%E2%80%99s-hadoop-big-data-analytics)

**Links**

[1] [http://test-static-05.web.cern.ch/publications/technical\\_documents/cms-data-services-ingestion-cern%E2%80%99s-hadoop-big-data-analytics](http://test-static-05.web.cern.ch/publications/technical_documents/cms-data-services-ingestion-cern%E2%80%99s-hadoop-big-data-analytics)

[2] <http://zenodo.org/record/31852>