SURFsara Hadoop user documentation Documentation

Release 0.1

SURFsara support team

Sep 28, 2017

Contents

1 General **3**
- 1.1 About the SURFsara Hadoop service . . . 3
- 1.2 Obtaining an account . . . 3
- 1.3 Getting started . . . 3
- 1.4 System specifications . . . 3
- 1.5 Frequently asked questions . . . 3
- 1.6 News . . . 3

2 Processing frameworks (software) **5**
- 2.1 YARN resource scheduler . . . 5
- 2.2 Hadoop MapReduce . . . 5
- 2.3 Apache Spark . . . 5
- 2.4 Apache Flink . . . 5
- 2.5 Cascading . . . 5
- 2.6 Apache Pig . . . 5
- 2.7 Apache Giraph . . . 5

3 Filesystems & databases **7**
- 3.1 Hadoop distributed filesystem (HDFS) . . . 7
- 3.2 Login node local storage . . . 7
- 3.3 Elasticsearch . . . 7
- 3.4 Apache HBase . . . 7
- 3.5 Backup policy . . . 7

4 Indices and tables **9**

Welcome to the SURFsara Hadoop user documentation!

CHAPTER 1

General

About the SURFsara Hadoop service

Obtaining an account

Getting started

System specifications

Frequently asked questions

News

CHAPTER 2

Processing frameworks (software)

YARN resource scheduler

Hadoop MapReduce

Apache Spark

Apache Flink

Cascading

Apache Pig

Apache Giraph

CHAPTER 3

Filesystems & databases

Hadoop distributed filesystem (HDFS)

Login node local storage

Elasticsearch

Apache HBase

Backup policy

CHAPTER 4

Indices and tables

- genindex
- search