

Пример использования Hadoop

Ошнуров Н.А.

ВМИ-115

Необходимые компоненты

- Ubuntu 11.10
- Hadoop 1.0.1
- Java 1.7

Этап 1. Подготовка

- Установка JDK
- Создание пользователя
- Отключение IPv6

Этап 2. Установка

- Загрузка Hadoop
- Распаковка архива
- Установка переменной окружения `$HADOOP_HOME`

Этап 3. Настройка HDFS

- Создание папки в файловой системе
- Редактирование конфигурационных файлов:
 - - `conf/core-site.xml`
 - - `conf/mapred-site.xml`
 - - `conf/hdfs-site.xml`
- Форматирование HDFS

Примеры конфигурационных файлов

- `conf/core-site.xml`
- `<!-- In: conf/core-site.xml -->`
- `<property>`
- `<name>hadoop.tmp.dir</name>`
- `<value>/app/hadoop/tmp</value>`
- `<description>Some description.</description>`
- `</property>`
- `<property>`
- `<name>fs.default.name</name>`
- `<value>hdfs://localhost:54310</value>`
- `<description> Some description</description>`
- `</property>`

Примеры конфигурационных файлов

- `conf/mapred-site.xml`
- `<!-- In: conf/mapred-site.xml -->`
- `<property>`
- `<name>mapred.job.tracker</name>`
- `<value>localhost:54311</value>`
- `<description>Some description`
- `</description>`
- `</property>`

Примеры конфигурационных файлов

- `conf/hdfs-site.xml`
- `<!-- In: conf/hdfs-site.xml -->`
- `<property>`
- `<name>dfs.replication</name>`
- `<value>1</value>`
- `<description>Some description`
- `</description>`
- `</property>`

Форматирование HDFS

- Выполняется с помощью команды `namenode`
- `hduser@ubuntu:~$ /usr/local/hadoop/bin/hadoop namenode -format`

Запуск и остановка Hadoop

- Запуск
 - - start-all.sh
- Остановка
 - - stop-all.sh

Работа с HDFS

- Копирование данных на HDFS
 - - `bin/hadoop dfs -copyFromLocal /tmp/fanatic /user/hduser/fanatic`
- Вывод списка файлов
 - - `bin/hadoop dfs -ls /user/hduser`
- Получение данных с HDFS
 - - `bin/hadoop dfs -getmerge /user/hduser/fan-output /tmp/out`

Запуск Java-задания

- Команда запуска задания
- - `hadoop jar your-job-file-goes-here.jar`

Структура задания

- Ядро задания
 - - реализация интерфейсов Mapper и Reducer
- Mapper
 - - формирует пары ключ/значение
- Reducer
 - - уменьшает набор пар, которые имеют один и тот же ключ в меньший набор значений

Демонстрация

Веб интерфейсы Hadoop

- <http://localhost:50030/> - UI для MapReduce заданий
- <http://localhost:50060/> - UI для заданий
- <http://localhost:50070/> - UI для просмотра статистики HDFS

Ссылки

- <http://hadoop.apache.org>
- <http://hadoop.apache.org/common/releases.html#Download>
- <http://fclose.com/b/cloud-computing>
- <http://www.michael-noll.com>