

Hadoop MapReduce

Claudio Scheer¹ Gabriell Araujo¹

¹Master's Degree in Computer Science
Pontifical Catholic University of Rio Grande do Sul - PUCRS

High Performance for Big Data Applications

Table of Contents

1 Hadoop

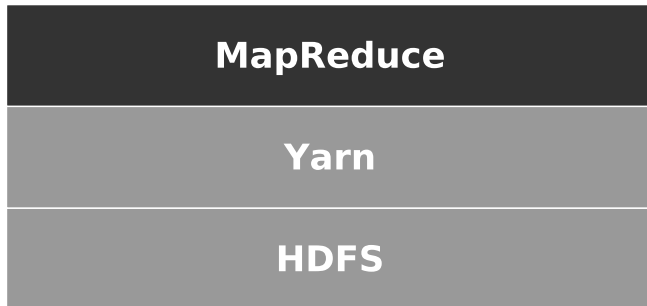
2 Examples

Table of Contents

1 Hadoop

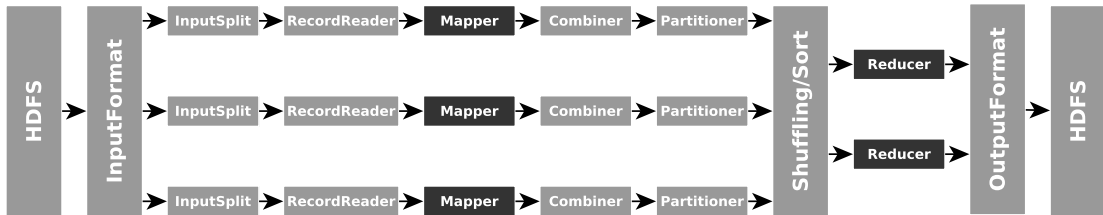
2 Examples

What we mean by Hadoop



<https://data-flair.training/blogs/hadoop-ecosystem-components>

MapReduce execution flow



<https://data-flair.training/blogs/hadoop-ecosystem-components>

Custom data types

- LongWritable = long;
- IntWritable = int;
- Text = String;
- ...

```
public class IntWritable implements WritableComparable<IntWritable> {  
    private int value;  
  
    public IntWritable(int value) { set(value); }  
  
    public void set(int value) { this.value = value; }  
  
    public int get() { return value; }  
  
    @Override  
    public void readFields(DataInput in) throws IOException {  
        value = in.readInt();  
    }  
  
    @Override  
    public void write(DataOutput out) throws IOException {  
        out.writeInt(value);  
    }  
  
    @Override  
    public int compareTo(IntWritable o) {  
        int thisValue = this.value;  
        int thatValue = o.value;  
        return (thisValue < thatValue ? -1 : (thisValue == thatValue ? 0 : 1));  
    }  
}
```

<https://hadoop.apache.org/docs/current/api/org/apache/hadoop/io/package-summary.html>

- TextInputFormat: `<LongWritable, Text>`
- KeyValueTextInputFormat: `<Text, Text>`
 - Key splitted by `\t`;
- NLineInputFormat: `<LongWritable, Text>`
 - `config.setInt("mapreduce.input.lineinputformat.linespermap", 10000);`
- Customs InputFormat must implement 'getSplits' and 'getRecordReader';
 - <https://hadoop.apache.org/docs/stable/api/org/apache/hadoop/mapred/InputFormat.html>

In this slide

In this slide
the text will be partially visible

In this slide
the text will be partially visible
And finally everything will be there

Table of Contents

1 Hadoop

2 Examples

- WordCount; - CountProductsSold; -

Sample frame title

In this slide, some important text will be highlighted because it's important. Please, don't abuse it.

Remark

Sample text

Important theorem

Sample text in red box

Examples

Sample text in green box. The title of the block is "Examples".

This is a text in first column.

$$E = mc^2$$

- First item
- Second item

This text will be in the second column and on a second thought this is a nice looking layout in some cases.