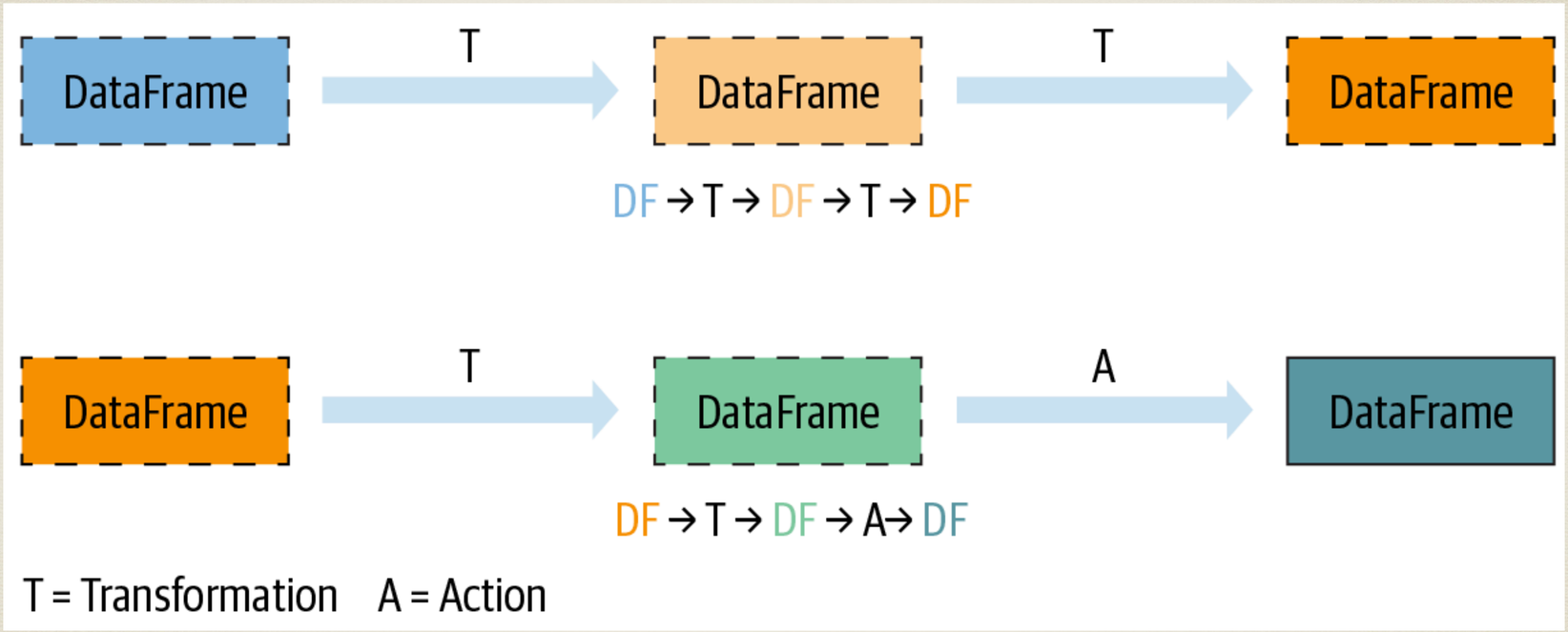
Big Data - Spark, Hadoop e afins.

Leonardo Gloria

Aula 03

Instituto Infnet

Transformations , Actions and Lazy Evaluation


DataFrame
T
DataFrame
T
DataFrame
DF → T → DF → T → DF
DataFrame
T
DataFrame
A
DataFrame
DF → T → DF → A→ DF
T = Transformation A = Action

Transformations	Actions
orderBy()	show()
groupBy()	take()
filter()	count()
select()	collect()

```
strings = spark.read.text('../README.txt')
filtered = strings.filter(strings.value.contains("Spark"))
filtered.count()
```

```
import org.apache.spark.sql.functions._
val strings = spark.read.text("../README.md")
val filtered = strings.filter(col("value").contains("Spark"))
filteres.count()
```

Data frame API

Id (Int)	First (String)	Last (String)	Url (String)	Published (Date)	Hits (Int)	Campaigns (List[Strings])
1	Jules	Damji	https://tinyurl.1	1/4/2016	4535	[twitter, LinkedIn]

Dúvidas?

“Stay hungry, stay foolish.”

–Steve Jobs