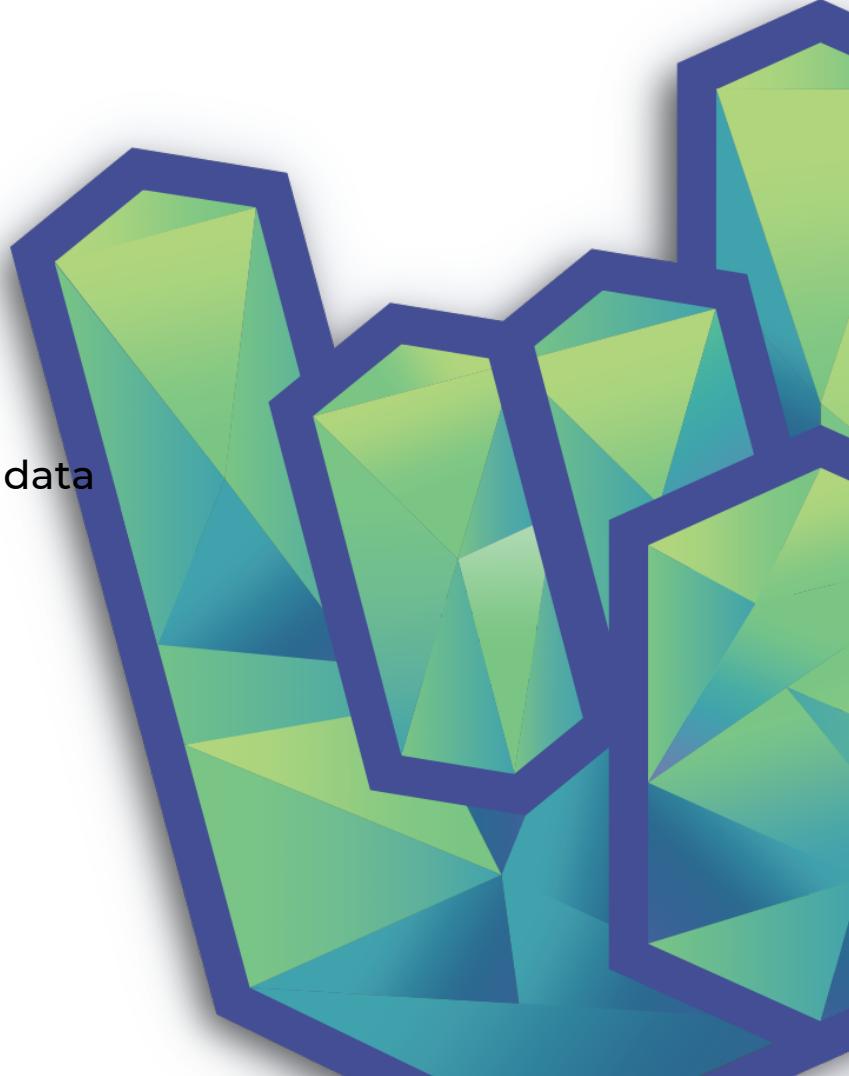


Managing Nulls



Objective

Avoid, overcome and replace nulls and absent data



Nullables

Remember enforcing schemas when reading?

```
val carsSchema = StructType(  
    StructField("Name", StringType, nullable=false)  
    ...  
)  
  
val carsDF = spark.read()  
    .format("json")  
    .schema(carsSchema)  
    .load("path/to/my/cars.json")
```

Non-nullable columns

- are NOT a constraint
- are a marker for Spark to optimize for nulls
- can lead to exceptions or data errors if broken

Spark rocks

