Gentle Introduction - Overview		
Keywork	**Explanation**	**Example**
Spark Session	Create a spark Session	import org.apache.spark.sql.SparkSession val spark = SparkSession .builder .appName("Simple Application") .getOrCreate()
Range	create DF from range	val df = spark.range(100).toDF("numbers")
Where	divisible by 2	df.where("numbers % 2 = 0").show(4)
Read option	read csv	val df = spark .read .option("inferSchema", "True") .option("header", "True") .format("csv") .load("/FileStore/tables/2015_summary-ebaee.csv")
.take(nb)	return an array of the nb first lines	df.take(3)
sort desc	sort by ascending order	df.sort(desc("colName")).show()
explain	physical plan	df.sort(desc("colName")).explain
	create an SQL table to make queries	df.createOrReplaceTempView("dfTable")
select from …	simple query	val sqlWay = spark.sql(""" SELECT DEST_COUNTRY_NAME, count(DEST_COUNTRY_NAME) FROM dfTable GROUP BY DEST_COUNTRY_NAME ORDER BY count(1) DESC LIMIT 5 """)
count groupBy	count each col entry when grouped	val dfWay = df .groupBy("colName") .count() .orderBy(desc("colName"))
max	return the max of a col	spark.sql("SELECT max(colName) from dfTable").take(1)(0) flightDF.select(max("colName")).take(1)
sum of groups	group by a col and sum of other col	df .groupBy("colNameA") .sum("colNameB") .withColumnRenamed("sum(colNameB)", "sumColNameB") .sort(desc("sumColNameB"))