

Datasets, DataFrames, and Spark SQL

```
// Another way, using auto infer schema

// NOTE: In this case make sure that auto inferred data types are matching the case class

import java.sql.Date
import spark.implicits._

case class Flight1(id: String, fldate: Date, month: Long, dowf: Long, carrier: String, src: String, dst: String, crsdephour: Long, crsdeptime: Long, depdelay: Double, crsarrrtime: Long, arrdelay: Double,

// val flightDS_autoinfer = spark.read.format("json").load(z.get("flight-data-2018").toString).as[Flight1]
// same as below

val flightDS_autoinfer = flightDF_autoinfer.as[Flight1]
```

	id	fldate	month	dowf	carrier	src	dst	crsdephour	crsdeptime	depdelay	crsarrrtime	arrdelay	crselapsedtime	dist
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	DL	ATL	BOS	9	850	0.0	1116	0.0	146.0	946.0	
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	DL	ATL	BOS	11	1122	8.0	1349	0.0	147.0	946.0	
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	DL	ATL	BOS	14	1356	9.0	1623	0.0	147.0	946.0	
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	DL	ATL	BOS	16	1620	0.0	1851	3.0	151.0	946.0	
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	DL	ATL	BOS	19	1940	6.0	2210	0.0	150.0	946.0	
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	DL	ATL	BOS	12	1248	0.0	1513	0.0	145.0	946.0	
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	DL	ATL	BOS	22	2215	0.0	39	0.0	144.0	946.0	
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	DL	ATL	BOS	15	1500	21.0	1734	33.0	154.0	946.0	
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	WN	ATL	BOS	15	1500	198.0	1725	208.0	145.0	946.0	
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	WN	ATL	BOS	21	2055	14.0	2330	0.0	155.0	946.0	
[ATL_BOS_2018-01-0...	[2018-01-01]	1	1	WN	ATL	BOS	10	1015	215.0	1250	191.0	155.0	946.0	
[ATL_CLT_2018-01-0...	[2018-01-01]	1	1	AA	ATL	CLT	11	1114	0.0	1238	0.0	84.0	226.0	
[ATL_CLT_2018-01-0...	[2018-01-01]	1	1	AA	ATL	CLT	8	845	0.0	1011	0.0	86.0	226.0	
[ATL_CLT_2018-01-0...	[2018-01-01]	1	1	AA	ATL	CLT	15	1548	0.0	1710	0.0	82.0	226.0	
[ATL_CLT_2018-01-0...	[2018-01-01]	1	1	AA	ATL	CLT	7	705	0.0	821	0.0	76.0	226.0	
[ATL_CLT_2018-01-0...	[2018-01-01]	1	1	AA	ATL	CLT	12	1226	0.0	1347	0.0	81.0	226.0	
[ATL_CLT_2018-01-0...	[2018-01-01]	1	1	AA	ATL	CLT	22	2205	0.0	2319	1.0	74.0	226.0	
[ATL_CLT_2018-01-0...	[2018-01-01]	1	1	DL	ATL	CLT	22	2210	11.0	2324	0.0	74.0	226.0	
[ATL_CLT_2018-01-0...	[2018-01-01]	1	1	DL	ATL	CLT	15	1543	1.0	1659	0.0	76.0	226.0	
[ATL_CLT_2018-01-0...	[2018-01-01]	1	1	DL	ATL	CLT	10	1008	0.0	1124	0.0	76.0	226.0	

only showing top 20 rows

Common Dataset typed transformations “Dataset[T]”

- filter
- map
- groupByKey