**FileInfo.txt to Json conversion**

**Application 1**

**This code works on marked data .i.e. fileInfo.txt, converts it into key-value with line number being the key and line being the value.**

**Then it extracts the rows that contain metadata, and the rows that have column headings such for different kind of data. Column headings such as objectType, framenumber etc..**

package main.scala

import org.apache.spark.SparkContext

import org.apache.spark.SparkContext.\_

import org.apache.spark.SparkConf

import org.apache.spark.sql.SparkSession

object **SixthApp** extends App{

val conf = new SparkConf().setAppName("HelloSpark").setMaster("local")

val sc = new SparkContext(conf)

println("printing metadata info from fileInfo.txt")

println("printing column headings i.e. fields for different markings")

val x = sc.textFile("C:\\Users\\alay.singhal\\Desktop\\Codes\\After\_markings\_SMPC522A1DC0040\_SE-PGA454\_20171012\_082108\_FileInfo.txt")

val sample1 = x

val newDS = sample1.zipWithIndex().map{case(line,i)=>i.toString + "," + line}

val pairz = newDS.map(x => (x.split(",")(0),x))

//pairz.take(100).foreach(println)

pairz.take(22).foreach(println)

val test1 = pairz.values.filter(line => line.startsWith("22,"))

//val y = test1.flatMap(line => (line.split(",")))

test1.flatMap(line => (line.split(","))).take(54).mkString(", ").foreach(print)

val test2 = pairz.values.filter(line => line.startsWith("241,"))

//val y = test1.flatMap(line => (line.split(",")))

test2.flatMap(line => (line.split(","))).take(54).mkString(", ").foreach(print)

val test3 = pairz.values.filter(line => line.startsWith("1166,"))

//val y = test1.flatMap(line => (line.split(",")))

test3.flatMap(line => (line.split(","))).take(54).mkString(", ").foreach(print)

val test4 = pairz.values.filter(line => line.startsWith("3887,"))

//val y = test1.flatMap(line => (line.split(",")))

test4.flatMap(line => (line.split(","))).take(54).mkString(", ").foreach(print)

sc.stop()

}

**Application 2**

**This code works on fileInfo.txt(before and after marking),**

**Extracts marked data and places that in a folder and then converts the data into JSON format.**

**Subset of marked data has/may have different column headings (derived from Application 1) which will be then added to the JSON format. This marked data in JSON format can then be added to Elastic Search and help us in indexing and searching through marked data and gathering statistics.**

**package** main.scala

**import** org.apache.spark.SparkContext

**import** org.apache.spark.SparkContext.\_

**import** org.apache.spark.SparkConf

**import** org.apache.spark.sql.SparkSession

**import** org.apache.spark.sql

**import** org.apache.spark.sql.types.\_

**import** org.apache.spark.sql.Row

**object** **FifthApp** **extends** App {

**val** conf = **new** SparkConf().setAppName("HelloSpark").setMaster("local")

**val** sc = **new** SparkContext(conf)

**val** x = sc.textFile("C:\\Users\\alay.singhal\\Desktop\\Codes\\After\_markings\_SMPC522A1DC0040\_SE-PGA454\_20171012\_082108\_FileInfo.txt")

**val** sample1 = x

**val** newDS = sample1.zipWithIndex().map{**case**(line,i)=>i.toString + "," + line}

**val** pairx = newDS.map(x => (x.split(",")(0),x))

**val** y = sc.textFile("C:\\Users\\alay.singhal\\Desktop\\Codes\\Before\_Markings\_Sample\_FileInfo.txt")

**val** sample2 = y

**val** newDS2 = sample2.zipWithIndex().map{**case**(line,i)=>i.toString + "," + line}

**val** pairy = newDS2.map(x => (x.split(",")(0),x))

**val** pairz = pairx.subtract(pairy)

pairz.saveAsTextFile("C:\\Users\\alay.singhal\\Desktop\\Codes\\tmp2")

**val** spark = SparkSession.builder.master("local").appName("Spark\_SQL\_basic\_example").getOrCreate()

**val** df = spark.read.format("csv").option("inferSchema","true").load("C:\\Users\\alay.singhal\\Desktop\\Codes\\tmp2\\part-00000")

df.write.format("json").save("C:\\Users\\alay.singhal\\Desktop\\Codes\\JsonConvertedFI3")

}

**Sample Data that is marked in CSV format.**

(2677,2677,236,RggbLogarithmicMasterWindowA,29176,0.741778194904327,0.433953732252121,38.6204626700233,0.0411436069735706,2.83429147188182,0.0727930836576527,-1,-1,2.81762240804963,0.0732522726557229,0.1,-0.0266017553922955,0.174775706913307,-1.51470970061192,0.212380008501002,ibeoLUX,0,46,,,,,,true,4252,V.1.1.0.1,V.1.0.1,,,,,,,,,29176,unknowSmall)

(5075,5075,320,RggbLogarithmicMasterWindowA,29917,car,notSet,normal,none,,,,,,,,,,,,,,,,,,,3.92955994606018,1.5565288066864,0.454251876527515,0.100724988436661,3.19631178960413,0.0462451963664266,-1,-1,0.0459776457067037,0.0250184132748887,1.5,0.00629620455321105,0.370316427552124,0.00267865994870806,0.243787794152984,ibeoLUX,0,45,,,,,,true,1680,V.1.1.0.1,V.1.0.1,29917,,,,,,,,,,,,,,,,,)

(1294,1294,24,RggbLogarithmicMasterWindowA,28354,1.45779728889465,0.463395178318024,28.6771547841576,0.0368880034264654,-16.6008322032612,0.0718778452489615,-1,-1,1.83000849930771,0.0363720594523126,0.1,-0.328236955536249,0.0892096215806666,0.194359594983145,0.340935851230519,ibeoLUX,0,69,,,,,,true,2712,V.1.1.0.1,V.1.0.1,,,,,,,,,28354,unknowBig)

(4680,4680,221,RggbLogarithmicMasterWindowA,29265,car,notSet,normal,none,,,,,,,,,,,,,,,,,,,4.10444021224976,0.722747087478638,23.328740778468,0.0514023864195836,16.193811404249,0.060390012982876,-1,-1,-1.62952913361334,0.0227082802319647,1.5,-0.28043216654672,0.103752200025046,-4.52209559305833,0.304840577248671,ibeoLUX,0,45,,,,,,true,2713,V.1.1.0.1,V.1.0.1,29265,,,,,,,,,,,,,,,,,)

(866,866,310,RggbLogarithmicMasterWindowA,29852,cyclist,notSet,normal,none,,,false,,,,,,,,,,,,,,1.52335667610168,0.54731947183609,29.3282648440982,0.0714310125285752,7.21402685606642,0.0843185466731884,-1,-1,1.59508456429421,0.0548573431532422,1.8,-0.0968693581847845,0.201789111084579,3.69962575595395,0.564226222761021,ibeoLUX,0,65,,,,,,true,2708,V.1.1.0.1,V.1.0.1,29852,,,,)

(356,356,39,RggbLogarithmicMasterWindowA,28410,cyclist,notSet,normal,none,,,false,,,,,,,,,,,,,,1.52335667610168,0.54731947183609,29.5767907270895,0.0365470080298333,-20.3908266584529,0.0672471673973342,-1,-1,1.76351138443632,0.0412207507175833,1.8,0.217709755547791,0.0637120916863834,-0.261886112024575,0.320299936286599,ibeoLUX,0,65,,,,,,true,2708,V.1.1.0.1,V.1.0.1,28410,,,,)

(817,817,298,RggbLogarithmicMasterWindowA,29758,cyclist,notSet,normal,none,,,false,,,,,,,,,,,,,,1.54171359539032,0.439146637916565,29.8190996224529,0.0628491437173229,-21.0302937545423,0.0841537342794404,-1,-1,1.76424157911278,0.0213093943418876,1.8,-1.1915088914576,0.148413376638868,6.2453101553728,0.338470819065281,ibeoLUX,0,64,,,,,,true,2715,V.1.1.0.1,V.1.0.1,29758,,,,)

(5747,5747,492,RggbLogarithmicMasterWindowA,30556,car,notSet,normal,none,,,,,,,,,,,,,,,,,,,2.42684650421143,1.48398494720459,80.2442759393877,0.0703084721339891,-0.479558239513739,0.0655594448950712,-1,-1,-0.0289634592221707,0.0103556388075594,1.5,9.91489299139565,0.324379569694526,-0.223804191822492,0.103154339264838,ibeoLUX,0,47,,,,,,true,2716,V.1.1.0.1,V.1.0.1,30556,,,,,,,,,,,,,,,,,)

(4185,4185,90,RggbLogarithmicMasterWindowA,28468,car,notSet,normal,none,,,,,,,,,,,,,,,,,,,0.314539045095444,1.59242904186249,97.9666336520133,0.0564486553331534,2.83318422478219,0.0600916426337753,-1,-1,0.0201926606479422,0.0205984491620183,1.5,3.36804486096002,0.306990492122593,0.15223338895253,0.0696614661288568,ibeoLUX,0,49,,,,,,true,4245,V.1.1.0.1,V.1.0.1,28468,,,,,,,,,,,,,,,,,)

(763,763,283,RggbLogarithmicMasterWindowA,29643,cyclist,notSet,normal,none,,,false,,,,,,,,,,,,,,1.52335667610168,0.54731947183609,29.29352303071,0.0345629417454338,2.86960724681385,0.0359702547036395,-1,-1,1.52486333803993,0.0150332293750999,1.8,0.154352827214123,0.0527536855661208,3.43514506506132,0.262779182184781,ibeoLUX,0,65,,,,,,true,2708,V.1.1.0.1,V.1.0.1,29643,,,,)

(1151,1151,478,RggbLogarithmicMasterWindowA,30517,cyclist,notSet,normal,none,,,false,,,,,,,,,,,,,,1.54171359539032,0.439146637916565,26.5742157848748,0.0865478414734518,29.5411033442787,0.0833426576846086,-1,-1,1.60136270353028,0.0390052513619353,1.8,-0.213702781090105,0.244403315055836,6.27670934966377,0.404688482319897,ibeoLUX,0,64,,,,,,true,2715,V.1.1.0.1,V.1.0.1,30517,,,,)

(4198,4198,94,RggbLogarithmicMasterWindowA,28525,car,notSet,normal,none,,,,,,,,,,,,,,,,,,,3.92955994606018,1.5565288066864,1.69824568000484,0.0642227063663831,3.04915352264305,0.0274721787290221,-1,-1,-0.048437822548092,0.0240591264674607,1.5,0.0174556726425177,0.315262571869358,0.00253063178874897,0.0152896892992377,ibeoLUX,0,45,,,,,,true,1680,V.1.1.0.1,V.1.0.1,28525,,,,,,,,,,,,,,,,,)

(5302,5302,379,RggbLogarithmicMasterWindowA,30123,car,notSet,normal,none,,,,,,,,,,,,,,,,,,,3.92955994606018,1.5565288066864,0.438842939127599,0.102297049667875,3.20094988905371,0.0478397404670434,-1,-1,0.0437958049122558,0.0254437438000014,1.5,0.151410462504134,0.372189540123598,0.00941171282048467,0.0167845767082142,ibeoLUX,0,45,,,,,,true,1680,V.1.1.0.1,V.1.0.1,30123,,,,,,,,,,,,,,,,,)

(773,773,285,RggbLogarithmicMasterWindowA,29664,cyclist,notSet,normal,none,,,false,,,,,,,,,,,,,,1.52335667610168,0.54731947183609,29.3062386650103,0.0347906613619287,3.18373753448951,0.0360032310720302,-1,-1,1.5273744327543,0.0150217638522292,1.8,0.148293949313977,0.0535205393937405,3.50456720772557,0.262974008594785,ibeoLUX,0,65,,,,,,true,2708,V.1.1.0.1,V.1.0.1,29664,,,,)

(1028,1028,417,RggbLogarithmicMasterWindowA,30283,cyclist,notSet,normal,none,,,false,,,,,,,,,,,,,,1.54171359539032,0.439146637916565,27.2528086551359,0.0337761378776587,12.7938736584576,0.0400806523845522,-1,-1,1.62834378600941,0.0105645464889337,1.8,-0.35708413932978,0.0654699393562632,6.04305710712025,0.273280938227822,ibeoLUX,0,64,,,,,,true,2715,V.1.1.0.1,V.1.0.1,30283,,,,)

(1276,1276,22,RggbLogarithmicMasterWindowA,28337,0.907297134399414,0.477681517601013,29.9609731055158,0.0712543506190182,-18.1479644693719,0.0761693789350247,-1,-1,1.49750475654052,0.0927446147222218,0.1,-0.271849788395853,0.231127564226468,-0.00979750935636602,0.241584413828625,ibeoLUX,0,54,,,,,,true,2710,V.1.1.0.1,V.1.0.1,,,,,,,,,28337,unknowBig)

**FileInfo.txt converted to JSON**

{"\_c0":"(2677","\_c1":2677,"\_c2":"236","\_c3":"RggbLogarithmicMasterWindowA","\_c4":"29176","\_c5":"0.741778194904327","\_c6":"0.433953732252121","\_c7":"38.6204626700233","\_c8":"0.0411436069735706","\_c9":"2.83429147188182","\_c10":"0.0727930836576527","\_c11":"-1","\_c12":"-1","\_c13":"2.81762240804963","\_c14":"0.0732522726557229","\_c15":"0.1","\_c16":"-0.0266017553922955","\_c17":"0.174775706913307","\_c18":"-1.51470970061192","\_c19":"0.212380008501002","\_c20":"ibeoLUX","\_c21":"0","\_c22":"46","\_c28":"true","\_c29":"4252","\_c30":"V.1.1.0.1","\_c31":"V.1.0.1","\_c40":"29176","\_c41":"unknowSmall)"}

{"\_c0":"(5075","\_c1":5075,"\_c2":"320","\_c3":"RggbLogarithmicMasterWindowA","\_c4":"29917","\_c5":"car","\_c6":"notSet","\_c7":"normal","\_c8":"none","\_c27":"3.92955994606018","\_c28":"1.5565288066864","\_c29":"0.454251876527515","\_c30":"0.100724988436661","\_c31":"3.19631178960413","\_c32":"0.0462451963664266","\_c33":"-1","\_c34":"-1","\_c35":"0.0459776457067037","\_c36":"0.0250184132748887","\_c37":"1.5","\_c38":"0.00629620455321105","\_c39":"0.370316427552124","\_c40":"0.00267865994870806","\_c41":"0.243787794152984"}

{"\_c0":"(1294","\_c1":1294,"\_c2":"24","\_c3":"RggbLogarithmicMasterWindowA","\_c4":"28354","\_c5":"1.45779728889465","\_c6":"0.463395178318024","\_c7":"28.6771547841576","\_c8":"0.0368880034264654","\_c9":"-16.6008322032612","\_c10":"0.0718778452489615","\_c11":"-1","\_c12":"-1","\_c13":"1.83000849930771","\_c14":"0.0363720594523126","\_c15":"0.1","\_c16":"-0.328236955536249","\_c17":"0.0892096215806666","\_c18":"0.194359594983145","\_c19":"0.340935851230519","\_c20":"ibeoLUX","\_c21":"0","\_c22":"69","\_c28":"true","\_c29":"2712","\_c30":"V.1.1.0.1","\_c31":"V.1.0.1","\_c40":"28354","\_c41":"unknowBig)"}

{"\_c0":"(4680","\_c1":4680,"\_c2":"221","\_c3":"RggbLogarithmicMasterWindowA","\_c4":"29265","\_c5":"car","\_c6":"notSet","\_c7":"normal","\_c8":"none","\_c27":"4.10444021224976","\_c28":"0.722747087478638","\_c29":"23.328740778468","\_c30":"0.0514023864195836","\_c31":"16.193811404249","\_c32":"0.060390012982876","\_c33":"-1","\_c34":"-1","\_c35":"-1.62952913361334","\_c36":"0.0227082802319647","\_c37":"1.5","\_c38":"-0.28043216654672","\_c39":"0.103752200025046","\_c40":"-4.52209559305833","\_c41":"0.304840577248671"}

{"\_c0":"(866","\_c1":866,"\_c2":"310","\_c3":"RggbLogarithmicMasterWindowA","\_c4":"29852","\_c5":"cyclist","\_c6":"notSet","\_c7":"normal","\_c8":"none","\_c11":"false","\_c25":"1.52335667610168","\_c26":"0.54731947183609","\_c27":"29.3282648440982","\_c28":"0.0714310125285752","\_c29":"7.21402685606642","\_c30":"0.0843185466731884","\_c31":"-1","\_c32":"-1","\_c33":"1.59508456429421","\_c34":"0.0548573431532422","\_c35":"1.8","\_c36":"-0.0968693581847845","\_c37":"0.201789111084579","\_c38":"3.69962575595395","\_c39":"0.564226222761021","\_c40":"ibeoLUX","\_c41":"0"}

{"\_c0":"(356","\_c1":356,"\_c2":"39","\_c3":"RggbLogarithmicMasterWindowA","\_c4":"28410","\_c5":"cyclist","\_c6":"notSet","\_c7":"normal","\_c8":"none","\_c11":"false","\_c25":"1.52335667610168","\_c26":"0.54731947183609","\_c27":"29.5767907270895","\_c28":"0.0365470080298333","\_c29":"-20.3908266584529","\_c30":"0.0672471673973342","\_c31":"-1","\_c32":"-1","\_c33":"1.76351138443632","\_c34":"0.0412207507175833","\_c35":"1.8","\_c36":"0.217709755547791","\_c37":"0.0637120916863834","\_c38":"-0.261886112024575","\_c39":"0.320299936286599","\_c40":"ibeoLUX","\_c41":"0"}

{"\_c0":"(817","\_c1":817,"\_c2":"298","\_c3":"RggbLogarithmicMasterWindowA","\_c4":"29758","\_c5":"cyclist","\_c6":"notSet","\_c7":"normal","\_c8":"none","\_c11":"false","\_c25":"1.54171359539032","\_c26":"0.439146637916565","\_c27":"29.8190996224529","\_c28":"0.0628491437173229","\_c29":"-21.0302937545423","\_c30":"0.0841537342794404","\_c31":"-1","\_c32":"-1","\_c33":"1.76424157911278","\_c34":"0.0213093943418876","\_c35":"1.8","\_c36":"-1.1915088914576","\_c37":"0.148413376638868","\_c38":"6.2453101553728","\_c39":"0.338470819065281","\_c40":"ibeoLUX","\_c41":"0"}

{"\_c0":"(5747","\_c1":5747,"\_c2":"492","\_c3":"RggbLogarithmicMasterWindowA","\_c4":"30556","\_c5":"car","\_c6":"notSet","\_c7":"normal","\_c8":"none","\_c27":"2.42684650421143","\_c28":"1.48398494720459","\_c29":"80.2442759393877","\_c30":"0.0703084721339891","\_c31":"-0.479558239513739","\_c32":"0.0655594448950712","\_c33":"-1","\_c34":"-1","\_c35":"-0.0289634592221707","\_c36":"0.0103556388075594","\_c37":"1.5","\_c38":"9.91489299139565","\_c39":"0.324379569694526","\_c40":"-0.223804191822492","\_c41":"0.103154339264838"}

{"\_c0":"(4185","\_c1":4185,"\_c2":"90","\_c3":"RggbLogarithmicMasterWindowA","\_c4":"28468","\_c5":"car","\_c6":"notSet","\_c7":"normal","\_c8":"none","\_c27":"0.314539045095444","\_c28":"1.59242904186249","\_c29":"97.9666336520133","\_c30":"0.0564486553331534","\_c31":"2.83318422478219","\_c32":"0.0600916426337753","\_c33":"-1","\_c34":"-1","\_c35":"0.0201926606479422","\_c36":"0.0205984491620183","\_c37":"1.5","\_c38":"3.36804486096002","\_c39":"0.306990492122593","\_c40":"0.15223338895253","\_c41":"0.0696614661288568"}

{"\_c0":"(763","\_c1":763,"\_c2":"283","\_c3":"RggbLogarithmicMasterWindowA","\_c4":"29643","\_c5":"cyclist","\_c6":"notSet","\_c7":"normal","\_c8":"none","\_c11":"false","\_c25":"1.52335667610168","\_c26":"0.54731947183609","\_c27":"29.29352303071","\_c28":"0.0345629417454338","\_c29":"2.86960724681385","\_c30":"0.0359702547036395","\_c31":"-1","\_c32":"-1","\_c33":"1.52486333803993","\_c34":"0.0150332293750999","\_c35":"1.8","\_c36":"0.154352827214123","\_c37":"0.0527536855661208","\_c38":"3.43514506506132","\_c39":"0.262779182184781","\_c40":"ibeoLUX","\_c41":"0"}

Coding options can later be changed to Python/Java too.

Ideas : Later “Spark Streaming” or other streaming frameworks can be used to monitor and look for new marked data files and automatically convert them to JSON and further push to data store (Elastic search) or others to search through marked data, index it, categorize it etc..