

State S...

Introduct...

G - state s...

G - Princip...

What A...

State S...

Control

library

Data G...

ERDPlus

hend_S...

hend_S...

G - chrome

Extensio...

Big Del...

community.cloud.databricks.com/?o=1037376470547060#table/default/iot_device_data2

AmazonSave to MerdeleySuggested SitesWildTangent GameseBayImported From Ruler-angle, DCM g...MF201a Course Inf...The Secret to Nevel...Inferring diffusion...Pumber 2 - Free O...www.magazines007...Other bookmarks

default.iot_device_data2

Refresh

My Cluster

DetailsHistory

Description:

Created at: 2023-04-28 15:19:06

Last modified: 2023-04-28 15:19:30

Partition columns:

Number of files: 8

Size: 7.75 MB

Schema:

	col_name	data_type	comment
1	battery_level	bigint	null
2	co2_level	bigint	null
3	cca2	string	null
4	cca3	string	null
5	cn	string	null
6	device_id	bigint	null
7	device_name	string	null

Sample Data:

	battery_level	co2_level	cca2	cca3	cn	device_id	device_name	humidity	ip	latitude	lcd	longitude	scale	temp	timest
1	8	868	US	USA	United States	1	meter-gauge-1xbYRNg	51	68.161.225.1	38	green	-87	Celsius	34	145844
2	7	1473	NO	NOR	Norway	2	sensor-pad-2n2Pea	70	213.161.254.1	62.47	red	6.15	Celsius	11	145844
3	2	1556	IT	ITA	Italy	3	device-mac-38TWSKUT	44	88.36.5.1	42.83	red	12.83	Celsius	19	145844
4	6	1080	US	USA	United States	4	sensor-pad-4mzWkz	32	66.39.173.154	44.06	yellow	-121.32	Celsius	28	145844
5	4	931	PH	PHL	Philippines	5	therm-stick-5gimpU188	62	203.82.41.9	14.58	green	120.97	Celsius	25	145844
6	5	1210	US	USA	United States	6	sensor-pad-6al7RTAc6R	51	204.116.105.67	35.93	yellow	-85.46	Celsius	27	145844
7	3	1129	CN	CHN	China	7	meter-gauge-7GeDoanM	26	220.173.179.1	22.82	yellow	108.32	Celsius	38	145844

20°C Mostly sunny

Search

19:14 28/04/2023

State So... Introdu... G: state sp... G: Princip... What A... State So... Control... library... Data G... ERDPlus... hend_S... hend_S... G: chrome... Extensio... Big Del... +

community.cloud.databricks.com/?o=1037376470547060#notebook/4203043310594709/command/4203043310594710

Amazon Save to Mendeley Suggested Sites Wikifangent Games eBay Imported from R Euler angle, DCM, g... MIT201x Course Int... The Secret to Never... Inferring diffusion L... Plumber 2 - Free O... www.magazines007... Other bookmarks

hend_Sprints Python

File Edit View Run Help Last edit was 4 minutes ago Give feedback

Run all Connecting Publish

Cell 1

1 display(dbutils.fs.ls("/databricks-datasets/samples/docs/"))

Python ▶ 1/81 v - ✕

• (3) Spark jobs

Table +

	path	name	size	modificationTime
1	dbfs:/databricks-datasets/samples/docs/README.md	README.md	3137	1465190449000

1 row | 7.49 seconds runtime Refreshed 2 hours ago

Command took 7.49 seconds -- by hend_m_ay@yahoo.com at 4/28/2023, 4:50:27 PM on My Cluster

Cell 2

1 textFile = spark.read.text("/databricks-datasets/samples/docs/README.md")

testFile: pyspark.sql.dataframe.DataFrame = [value: string]

Command took 2.06 seconds -- by hend_m_ay@yahoo.com at 4/28/2023, 4:52:04 PM on My Cluster

Cell 3

1 textFile.count()

• (2) Spark jobs

Out[6]: 65

Command took 4.38 seconds -- by hend_m_ay@yahoo.com at 4/28/2023, 4:54:03 PM on My Cluster

Cell 4

1 scala
2 val lines = sc.textFile("/databricks-datasets/samples/docs/README.md")
3 val counts = lines.flatMap(line => line.split(" ")).map(word => (word, 1)).reduceByKey(_+_)

lines: org.apache.spark.rdd.RDD[String] = /databricks-datasets/samples/docs/README.md MapPartitionsRDD[15] at textFile at command-796742441769341:1
counts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[18] at reduceByKey at command-796742441769341:2

Command took 2.64 seconds -- by hend_m_ay@yahoo.com at 4/28/2023, 5:09:58 PM on My Cluster

Cell 5

1 scala
2 counts.collect()

20°C Mostly sunny

Search

19:16 28/04/2023

State Sp... Introduc... G state sp... G Principi... What A... State Sp... Control... Library... Data Gr... ER EADPlus... hend_Sc... hend_Sc... G chrome... Extensio... Big Data... +

community.cloud.databricks.com/?o=1037376470547060#notebook/4203043310594709/command/4203043310594710

Amazon Save to Mendeley Suggested Sites Wikifongest Games eBay Imported From E Euler angle, DCM, q... MF201x Course Int... The Secret to Never... Infering diffusion L... Plumber 2 - Free O... www.magazines007... Other bookmarks

hend_Sprints Python

File Edit View Run Help Last edit was 5 minutes ago Give feedback

Run all My Cluster Publish

Cell 5

```
1 #scala
2 counts.collect
```

⚡ (1) Spark Jobs

res0: Array[(String, Int)] = Array((So,1), (this,1), (Python,1), (repos="http://cran.stat.ucla.edu/"',1), (SKIP_API=1,1), (have,3), (include,1), (plain,1), (generates,2), (-o,1), (here,1), (locally,1), (while),1), (som e,1), (only,1), (NOTE:,1), (This,2), (Similarly,,1), (markdown,1), (documentation,7), (textfiles,1), (over,2), (read,1), (just,2), (extra,1), (out,2), ('_site',2), (version,3), (documentation,,1), (subprojects,1), (ste p,1), (are,1), (build',1), ((i.e.,,1), (Compiling,1), (can,7), (commands,1), ("install.packages(c("knitr",,1), (build,9), (4086,1), ('docs/',1), (Sphinx,,1), (When,1), (scaladoc,,1), ('build/sbt,2), (into,2), (as,5), (o pposed,1), (Start,1), ((currently,1), (create,1), (readme,1), (yourself?,1), (regardless,1), ((as,1), ((Scaladoc,,1), ({s,3), (API,4), (currently,1), (follows,1), (takes,1), (modify,1), (R,1), (classes,1), (Build,1), (built,1), (with,6), (live,1), ({s,10}, (corresponds,1), (want,,1), (specific,1), (using,2), (We,2), (rest,1), (navigating,1), (directly,1), (captured,1), (Python,,1), (building,4), (##,3), (from,3), (well,1), ('make,1), (sphinx,1), (Generating,1), (learn,1), (SPARK_PROJECT_ROOT,1), (formatted,1), (Read,1), ('SKIP_API=1,1), (containing,1), (scaladoc,2), (uses,1), (unidoc',2), (will,4), (Welcome,1), (get,1), ('""',71), (Why,1), (Documentat ion,2), (generating,1), (wiki,,1), (PySpark,2), (Scala,,2), (so,1), (R,,1), (directory),,1), (included,1), (skip,1), (default,1), (Spark,9), (it,5), (The,3), (revision,2), (definitive,1), (index.html,1), (viewing,1), (s uch,1), (SparkR,1), (through,1), (sudo,4), ({s,3}, (following,1), (documentation,1), ('__init__.py',,1), (automatically,1), (content,1), (which,2), ((which,1), (also,4), (Serve,1), (for,2), (To,2), (along,1), (install, 4), (walk,1), (control,1), (time,1), (gem,2), (called,1), (used,1), (Pygments,1), (PRODUCTION=1,1), (the,38), (compiled,1), (release,2), (includes,1), (github,1), (part,1), (jekyll',,1), (html',1), ("devtools",),1), (en able,1), (site,,1), (SPARK_PROJECT_ROOT/R/create-docs.sh,1), (if,2), (site,3), (be,2), (source,3), (git,,1), (documentation,1), ('jekyll',1), (all,1), (then,1), (relevant,1), (--watch,1), (haven't,1), (compile,1), (g enerated,1), (listed,1), (started,1), (control,,1), (downloaded,,1), (by,4), (suffix,,1), (code,,1), (Markdown,,1), (an,1), (http://spark.apache.org/documentation.html,,1), (SPARK_PROJECT_ROOT/python/docs,1), (copying, 1), (HTML,2), (yourself,,1), (Execute,1), (Jekyll,2), (Docs,1), ('docs',2), (roxygen2),1), (plugin,2), (about,1), (run,5), ('jekyll,1), (text,2), (number,1), (on,3), (evolve,1), (index.md,,1), ([Sphinx](http://sphinx-do c.org/),,1), (You,4), (before,1), (features,1), (Rscript,1), (at,1), (in,4), (docs,8), (versions,1), (jekyll-redirect-from,1), (copy,1), (In,1), (those,1), ((and,1), (take,1), (Skip,1), (directory,,1), (directory,,2), (port,1), (public,1), (way,1), (use,1), (may,1), (page,1), (running,3), (jekyll,7), (find,2), (files,,1), (serve,1), (Prerequisites,1), (directory,4), (you,10), (pip,2), (files,1), (various,1), (that,4), (cd,1), ({s,5}, (or,3), (to,10), (code,2), (docs,,1), (more,1), (of,10), (tools,1), ((recently),1), (hosted,1), (and,5), ("md",1), (checked,2), (whichever,1))

Command took 2.91 seconds — by hend_w_sabayahoo.com at 4/28/2023, 5:04:45 PM on My Cluster

Cell 6

```
1 #python
2 Use the Spark csv file datasource with options specifying:
3 # first line of file is a header
4 #automatically infer the schema of the data
5 #in otherwords, rouse the already exist schema and generate the datatypes
6 data = spark.read.format("csv") \
7 .option("header", "true") \
8 .option("inferSchema", "true") \
9 .load("/databricks-datasets/samples/population-vs-price/data_geo.csv")
```

⚡ (2) Spark Jobs

data: pyspark.sql.dataframe.DataFrame

2014 rank:	integer
City:	string
State:	string
State Code:	string
2014 Population estimate:	integer
2015 median sales price:	double

20°C Mostly sunny

Search

19:16 28/04/2023

State Sp... Introdu... state sp... Princip... What A... State Sp... Confor... Library... Data G... ERDPlus... hend_S... hend_S... chrome... Extens... 6g Del... +

community.cloudatradicks.com/To=1037376470547060?notebook/4203043310594709/command/4203043310594710

Amazon Save to Mendeley Suggested Sites W W Waldington Games eBay Imported From B Sales angle, DCM, q... MF201+ Course Inf... The Secret to Newer... Inferring Diffusion L... Plumber 2 - Free O... www.magazine007... Other bookmarks

hend_Sprints Python

File Edit View Run Help Last edit was 6 minutes ago Give feedback

Run all My Cluster Publish

(1) Spark jobs

Out[14]: [Row(2014 rank=101, City='Birmingham', State='Alabama', State Code='AL', 2014 Population estimate=212247, 2015 median sales price=162.9), Row(2014 rank=125, City='Huntsville', State='Alabama', State Code='AL', 2014 Population estimate=186226, 2015 median sales price=157.7), Row(2014 rank=122, City='Mobile', State='Alabama', State Code='AL', 2014 Population estimate=194675, 2015 median sales price=122.5), Row(2014 rank=114, City='Montgomery', State='Alabama', State Code='AL', 2014 Population estimate=200481, 2015 median sales price=129.0), Row(2014 rank=6, City='Phoenix', State='Arizona', State Code='AZ', 2014 Population estimate=1537058, 2015 median sales price=206.1), Row(2014 rank=33, City='Tucson', State='Arizona', State Code='AZ', 2014 Population estimate=527972, 2015 median sales price=176.1), Row(2014 rank=119, City='Little Rock', State='Arkansas', State Code='AR', 2014 Population estimate=197786, 2015 median sales price=131.8), Row(2014 rank=56, City='Anaheim', State='California', State Code='CA', 2014 Population estimate=346907, 2015 median sales price=685.7), Row(2014 rank=2, City='Los Angeles', State='California', State Code='CA', 2014 Population estimate=3928864, 2015 median sales price=434.7), Row(2014 rank=59, City='Riverside', State='California', State Code='CA', 2014 Population estimate=319584, 2015 median sales price=281.9)]

Command took 8.34 seconds -- by hend_ajayashou.com at 4/28/2023, 5:11:45 PM on My Cluster

Cell 10

1 data.take(5)

(1) Spark jobs

Out[15]: [Row(2014 rank=101, City='Birmingham', State='Alabama', State Code='AL', 2014 Population estimate=212247, 2015 median sales price=162.9), Row(2014 rank=125, City='Huntsville', State='Alabama', State Code='AL', 2014 Population estimate=186226, 2015 median sales price=157.7), Row(2014 rank=122, City='Mobile', State='Alabama', State Code='AL', 2014 Population estimate=194675, 2015 median sales price=122.5), Row(2014 rank=114, City='Montgomery', State='Alabama', State Code='AL', 2014 Population estimate=200481, 2015 median sales price=129.0), Row(2014 rank=6, City='Phoenix', State='Arizona', State Code='AZ', 2014 Population estimate=1537058, 2015 median sales price=206.1)]

Command took 0.10 seconds -- by hend_ajayashou.com at 4/28/2023, 5:11:50 PM on My Cluster

Cell 11

1 display(data)

(1) Spark jobs

Table Visualization 1 Visualization 2 +

	2014 rank	City	State	State Code	2014 Population estimate	2015 median sales price
1	101	Birmingham	Alabama	AL	212247	162.9
2	125	Huntsville	Alabama	AL	186226	157.7
3	122	Mobile	Alabama	AL	194675	122.5
4	114	Montgomery	Alabama	AL	200481	129
5	6	Phoenix	Arizona	AZ	1537058	206.1
6	33	Tucson	Arizona	AZ	527972	176.1

20°C Mostly sunny

Search

19:17 28/04/2023

State Sp... Introdu... state sp... Principl... What A... State Sp... Control... library... Data G... ERDPlus... hend_5... hend_5... chrome... Extensions... Big De...

community.cloud.databricks.com/?o=1037376470547060#notebook/4203043310594709/command/4203043310594710

Amazon Save to Mendeley Suggested Sites WildTangent Games eBay Imported From IE Euler angle, DCM, q... MIT201x Course Int... The Secret to Newer... Inferring diffusion... Plumber 2 - Free O... www.magazines007... Other bookmarks

hend_Sprints Python File Edit View Run Help Last edit was 6 minutes ago Give feedback Run all My Cluster Publish

1 display(data)

(1) Spark Jobs

Table Visualization 1 Visualization 2

State Code	SUM(2014 rank)
AL	450
AK	100
AZ	50
CA	180
CO	350
CT	600
DC	20
FL	1250
GA	50
HI	50
IA	500
ID	100
IL	650
IN	100
KS	250
KY	100
LA	200
MA	300
MO	20
MI	220
MN	50
MO	180
MS	150
NC	650
ND	220
NE	120
NH	250
NI	80
NM	30
NV	120
NY	350
OH	500
OK	50
OR	320
PA	220
RI	120
SD	150
TN	300
TX	680
UT	120
VA	150
WA	120
WI	120

Edit Visualization 109 rows Refreshed 2 hours ago

Command took 0.45 seconds -- by hend_r_p@yahoo.com at 4/18/2023, 5:12:27 PM on My Cluster

Cmd 12

1 data.createOrReplaceTempView("data_geo")

Command took 0.28 seconds -- by hend_r_p@yahoo.com at 4/18/2023, 5:12:10 PM on My Cluster

Cmd 13

1 %sql
2 select City, State from data_geo

(1) Spark Jobs

20°C Mostly sunny Search 19:17 28/04/2023

State Sp... Introdu... G state sp... G Princip... What A... State Sp... Control... library... Data G... ERDPlus... hend... hend... G chrome... Extens... Big Dan... +

community.cloud.databricks.com/?o=1037376470547050#notebook/4203043310594709/command/4203043310594710

Amazon Save to Mendeley Suggested Sites WildTangent Games eBay Imported from R Euler angle, DCM, q... MI 201x Course Int... The Secret to Never... Inferring diffusion L... Number 2 - Free O... www.magazines007... Other bookmarks

hend_Sprints Python

File Edit View Run Help last edit was 7 minutes ago Give feedback

Run all My Cluster Publish

Cell 14

```
1 %sql
2 select * from data_geo where State = 'Alabama'
```

(1) Spark jobs

_sqlidf: pyspark.sql.dataframe.DataFrame = [2014 rank: integer, City: string ... 4 more fields]

Table

	2014 rank	City	State	State Code	2014 Population estimate	2015 median sales price
1	101	Birmingham	Alabama	AL	212247	162.9
2	125	Huntsville	Alabama	AL	188226	157.7
3	122	Mobile	Alabama	AL	194675	122.5
4	114	Montgomery	Alabama	AL	200481	129

4 rows | 0.96 seconds runtime

Refreshed 2 hours ago

Command took 0.16 seconds -- by hend_m_s@yahoo.com at 4/28/2023, 5:48:04 PM on My Cluster

Cell 13

```
1 %scala
2 val range100 = spark.range(100)
3 range100.collect()
```

(1) Spark jobs

range100: org.apache.spark.sql.Dataset[Long] = [id: long]

range100: org.apache.spark.sql.Dataset[Long] = [id: bigint]

res1: Array[Long] = Array(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99)

Command took 2.84 seconds -- by hend_m_s@yahoo.com at 4/28/2023, 5:48:06 PM on My Cluster

Cell 16

```
1 %scala
2 val df = spark.read.json("/databricks-datasets/samples/people/people.json")
```

(1) Spark jobs

df: org.apache.spark.sql.DataFrame

age: long

20°C Mostly sunny

Search

19:18 28/04/2023

State Sc... Introdu... G state sp... G Princip... What A... State Sc... Control... library... Data G... ERDPlus... hend... hend... chrome... Extensi... Big Dan... + -

community.cloud.databricks.com/?o=1037376470547050#notebook/4203043310594709/command/4203043310594710

Amazon Save to Mendeley Suggested Sites WildTangent Games eBay Imported from R Euler angle, DCM, q... MI 201x Course Int... The Secret to Never... Inferring diffusion L... Rumber 2 - Free O... www.magazines007... Other bookmarks

hend_Sprints Python File Edit View Run Help last edit was 7 minutes ago Give feedback Run all My Cluster Publish

(1) Spark Jobs

ds: org.apache.spark.sql.Dataset[Person] = [age: long, name: string]

defined class Person

ds: org.apache.spark.sql.Dataset[Person] = [age: bigint, name: string]

Command took 4.55 seconds -- by hend_m_@yahoo.com at 4/28/2023, 5:56:46 PM on My Cluster

Cell 18

```
1 //scala
2 case class DeviceIoTData (
3   battery_level: Long,
4   co2_level: Long,
5   cca2: String,
6   cn: String,
7   device_id: Long,
8   device_name: String,
9   humidity: Long,
10  ip: String,
11  latitude: Double,
12  longitude: Double,
13  scale: String,
14  temp: Long,
15  timestamp: Long
16 )
17 val ds = spark.read.json("/databricks-datasets/iot/iot_devices.json").as[DeviceIoTData]
```

(1) Spark Jobs

ds: org.apache.spark.sql.Dataset[DeviceIoTData] = [battery_level: long, co2_level: long ... 13 more fields]

defined class DeviceIoTData

ds: org.apache.spark.sql.Dataset[DeviceIoTData] = [battery_level: bigint, co2_level: bigint ... 13 more fields]

Command took 9.42 seconds -- by hend_m_@yahoo.com at 4/28/2023, 5:58:08 PM on My Cluster

Cell 19

```
1 //scala
2 display(ds)
```

(1) Spark Jobs

table +

battery_level co2_level cca2 cca3 cn device_id device_name humidity ip latitude lcd longitude scale temp

20°C Mostly sunny

Search

19:18 28/04/2023

State Sc... Introdu... G state sp... G Principl... What A... State Sc... Control... library... Data G... ERDPlus... hend_S... hend_S... chrome... Extens... Big Dat... +

community.cloud.databricks.com/?o=1057376470547060#notebook/4203043310594709/command/4203043310594710

Amazon Save to Mendeley Suggested Sites WildTangent Games eBay Imported from IE Euler angle, OCM, q... MF201x Course Int... The Secret to Never... Inferring diffusion L... Plumber 2 - Free O... www.magazine007... Other bookmarks

hend_Sprints Python

File Edit View Run Help Last edit was 7 minutes ago Give feedback

1 %scala
2 display(ds)

(1) Spark jobs

Table +

	battery_level	c02_level	cca2	cca3	cn	device_id	device_name	humidity	ip	latitude	lcd	longitude	scale	temp
1	8	860	US	USA	United States	1	meter-gauge-1xbYRYcj	51	68.161.225.1	38	green	-97	Celsius	34
2	7	1473	NO	NOR	Norway	2	sensor-pad-2n2Pea	70	213.161.254.1	62.47	red	6.15	Celsius	11
3	2	1556	IT	ITA	Italy	3	device-mac-36TWSKt	44	88.36.5.1	42.83	red	12.83	Celsius	19
4	6	1080	US	USA	United States	4	sensor-pad-4mzWkz	32	66.39.173.154	44.06	yellow	-121.32	Celsius	28
5	4	931	PH	PHL	Philippines	5	therm-stick-5gimpUr88	62	203.82.41.9	14.58	green	120.97	Celsius	25
6	3	1210	US	USA	United States	6	sensor-pad-6al7RTAobR	51	204.116.105.67	35.93	yellow	-85.46	Celsius	27
7	3	1129	CN	CHN	China	7	meter-gauge-7GeDoanM	28	220.175.179.1	22.82	yellow	108.32	Celsius	18

10,000 rows | Truncated data | 2.31 seconds runtime

Refreshed 1 hour ago

Command took 2.31 seconds -- by hend_s_awayahou.com at 4/28/2023, 6:18:19 PM on My Cluster

Cell 20

1 %scala
2 ds.take(10).foreach(println(_))

(1) Spark jobs

DeviceIoData(8,868,US,United States,1,meter-gauge-1xbYRYcj,51,68.161.225.1,38.8,-87.9,Celsius,34,1458444854893)
DeviceIoData(7,1473,NO,Norway,2,sensor-pad-2n2Pea,70,213.161.254.1,62.47,6.15,Celsius,11,1458444854119)
DeviceIoData(2,1556,IT,Italy,3,device-mac-36TWSKt,44,88.36.5.1,42.83,12.83,Celsius,19,1458444854120)
DeviceIoData(6,1080,US,United States,4,sensor-pad-4mzWkz,32,66.39.173.154,44.06,-121.32,Celsius,28,1458444854121)
DeviceIoData(4,931,PH,Philippines,5,therm-stick-5gimpUr88,62,203.82.41.9,14.58,120.97,Celsius,25,1458444854122)
DeviceIoData(3,1210,US,United States,6,sensor-pad-6al7RTAobR,51,204.116.105.67,35.93,-85.46,Celsius,27,1458444854123)
DeviceIoData(3,1129,CN,China,7,meter-gauge-7GeDoanM,26,220.175.179.1,22.82,108.32,Celsius,18,1458444854123)
DeviceIoData(8,1536,JP,Japan,8,sensor-pad-8xU6pzsQI,35,210.173.177.1,35.69,139.69,Celsius,27,1458444854123)
DeviceIoData(3,887,JP,Japan,9,device-mac-9GcJZ2pw,85,118.23.68.227,35.69,139.69,Celsius,13,1458444854124)
DeviceIoData(7,1470,US,United States,10,sensor-pad-10BsywSVUF,56,208.189.163.218,33.61,-111.89,Celsius,26,1458444854125)

Command took 1.04 seconds -- by hend_s_awayahou.com at 4/28/2023, 6:18:19 PM on My Cluster

Cell 21

1 %scala
2 val sorted_device = ds.select(\$"battery_level", \$"c02_level", \$"device_name").where(\$"battery_level" >= 6).sort(\$"c02_level")

20°C Mostly sunny

Search

19:18 28/04/2023

State Sc... X Introdu... X state sp... X Princip... X What A... X State S... X Control... X library... X Data G... X ERDPlus... X hend... X hend... X G chrome X Extens... X Big Dat... X

community.cloud.databricks.com/?o=1037376470547060#notebook/4203043310594709/command/4203043310594710

Amazon Save to Mendeley Suggested Sites WildTangent Games eBay Imported from Ruler angle, DCNE, q... MF2019 Course Inf... The Secret to Newer... Inferring diffusion L... Number 2 - Food Q... www.magazines007... Other bookmarks

hend_Sprints Python

File Edit View Run Help Last edit was 9 minutes ago Give feedback

Run all My Cluster Publish

Service: Databricks; Path: /Volumes/sensor-pad-165138dUjgFY0; Command took 1.04 seconds -- by hend_n_whyahoo.com at 4/28/2023, 4:08:10 PM on My Cluster

Cmd 23

```
1 %scala
2 val sorted_device = ds.select($"battery_level", $"co2_level", $"device_name").where($"battery_level" >= 6).sort($"co2_level")
3 display(sorted_device)
```

(1) Spark jobs

sorted_device: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [battery_level: long, co2_level: long ... 1 more field]

Table Visualization 1

	battery_level	co2_level	device_name
1	8	800	sensor-pad-165138dUjgFY0
2	6	800	sensor-pad-1636WNHcHvCY
3	8	800	device-mac-53529CvWkrm6
4	7	800	meter-gauge-4673N0x46gviLZ
5	6	800	device-mac-81051JudyRqLSc
6	8	800	sensor-pad-51022PSPs4Qlg
7	8	800	meter-gauge-60509Mz9VApS

10,000 rows | Truncated data | 8.61 seconds runtime Refreshed 1 hour ago

Command took 0.61 seconds -- by hend_n_whyahoo.com at 4/28/2023, 4:08:10 PM on My Cluster

Cmd 23

```
1 %scala
2 ds.createOrReplaceTempView("iot_device_data")
```

Command took 0.09 seconds -- by hend_n_whyahoo.com at 4/28/2023, 4:13:23 PM on My Cluster

Cmd 23

```
1 %sql
2 select cca3, count (distinct device_id) as device_id from iot_device_data group by cca3 order by device_id desc limit 100
```

(3) Spark jobs

_iqidff: pyspark.sql.dataframe.DataFrame = [cca3: string, device_id: long]

Table

20°C Mostly sunny

Search

ENG 19:19 28/04/2023

State Sp: XIntrodu: XG state sp: XG Principi: XWhat 4: XState Sp: XControl: Xlibrary: XData Co: XERDPlus: Xhend_5: Xhend_5: XG chrome: XExtend: XBig Data: X

community.cloud.databricks.com/?o=1037376470547060#notebook/4203043310594709/command/4203043310594710

AmazonSave to MendeleySuggested SitesWildfangers GameseBayImported from iEEuler angle, OCM, q...MF201s Course Inf...The Secret to Never...latent diffusion L...Plumber 2 - Free O...www.magazines00f...Other bookmarks

hend_SprintsPythonFileEditViewRunHelpLast edit was 8 minutes agoGive feedbackRun allMy ClusterPublish

```
1 %sql
2 select cca3, count (distinct device_id) as device_id from iot_device_data group by cca3 order by device_id desc limit 100
```

(3) Spark jobs

_sqlid: pyspark.sql.dataframe.DataFrame = [cca3: string, device_id: long]

Table +

	cca3	device_id
1	USA	70405
2	CHN	14455
3	JPN	12100
4	KOR	11879
5	DEU	7942
6	GBR	6486
7	CAN	6041

100 rows | 6.39 seconds runtimeRefreshed 1 hour ago

Command took 8.28 seconds -- by hend_w_ay@yahoo.com at 4/28/2023, 6:17:09 PM on My Cluster

Cell 24

```
1 %scala
2 ds.write.saveAsTable("iot_device_data2")
```

(5) Spark jobs

Command took 35.14 seconds -- by hend_w_ay@yahoo.com at 4/28/2023, 6:18:06 PM on My Cluster

Cell 25

```
1 %scala
2 ds.write.mode("overwrite").parquet("/tmp/testParquet")
```

(1) Spark jobs

Command took 10.82 seconds -- by hend_w_ay@yahoo.com at 4/28/2023, 7:08:00 PM on My Cluster

Cell 26

```
1 display(dbutils.fs.ls("/tmp/testParquet"))
```

(1) Spark jobs

20°C Mostly sunny

Search

19:19 28/04/2023

State Sp...Introduc...state sp...Principi...What A...State Sp...Control...Library...Data G...ERDPlus...hend_3...hend_3...chrome...Extens...Big Da...+
community.cloud.databricks.com/?o=1037376470547060#notebook/4203043310594709/command/4203043310594710
AmazonSave to MendeleySuggested SitesWWDWitTanger GamesuBayImported From B...Ruler angle, DCM, q...MP201s Course In...The Secret to Never...Inferring diffusion L...Plumber 2 - Free D...www.magnam007...Other bookmarks

hend SprintsPython

FileEditViewRunHelpLast edit was 8 minutes agoGive feedback

Run allMy ClusterPublish

Cell 25

```
1 %scala
2 ds.write.mode("overwrite").parquet("/tmp/testParquet")
```

(1) Spark jobs

Command took 10.81 seconds — by hend_w_ayyaho.com at 4/28/2023, 7:00:05 PM on My Cluster

Cell 26

```
1 display(dbutils.fs.ls("/tmp/testParquet"))
```

(3) Spark jobs

Table +

	path	name	size	modification
1	dbfs:/tmp/testParquet/_SUCCESS	_SUCCESS	0	16826980961
2	dbfs:/tmp/testParquet/_committed_7826455954826555404	_committed_7826455954826555404	824	16826980961
3	dbfs:/tmp/testParquet/_started_7826455954826555404	_started_7826455954826555404	0	16826980988
4	dbfs:/tmp/testParquet/part-00000-fid-7826455954826555404-ca6c29ab-4423-4e00-6e73-de3f8cc364b9-113-1-c000.snappy.parquet	part-00000-fid-7826455954826555404-ca6c29ab-4423-4e00-6e73-de3f8cc364b9-113-1-c000.snappy.parquet	1051856	16826980947
5	dbfs:/tmp/testParquet/part-00001-fid-7826455954826555404-ca6c29ab-4423-4e00-6e73-de3f8cc364b9-114-1-c000.snappy.parquet	part-00001-fid-7826455954826555404-ca6c29ab-4423-4e00-6e73-de3f8cc364b9-114-1-c000.snappy.parquet	1033306	16826980961
6	dbfs:/tmp/testParquet/part-00002-fid-7826455954826555404-ca6c29ab-4423-4e00-6e73-de3f8cc364b9-115-1-c000.snappy.parquet	part-00002-fid-7826455954826555404-ca6c29ab-4423-4e00-6e73-de3f8cc364b9-115-1-c000.snappy.parquet	1032592	16826980951
	dbfs:/tmp/testParquet/part-00003-fid-7826455954826555404-ca6c29ab-4423-4e00-6e73-de3f8cc364b9-116-1-c000.snappy.parquet	part-00003-fid-7826455954826555404-ca6c29ab-4423-4e00-6e73-de3f8cc364b9-116-1-c000.snappy.parquet	1030053	16826980951

11 rows | 1.26 seconds runtime

Refreshed 11 minutes ago

Command took 1.28 seconds — by hend_w_ayyaho.com at 4/28/2023, 7:00:14 PM on My Cluster

Cell 27

```
1 parquet_directory = spark.read.parquet("/tmp/testParquet")
2 display(parquet_directory)
```

(2) Spark jobs

parquet_directory: pyspark.sql.dataframe.DataFrame = [battery_level: long, c02_level: long ... 13 more fields]

battery_levelc02_levelc03_levelc04_levelc05_levelc06_levelc07_levelc08_levelc09_levelc10_levelc11_levelc12_levelc13_levelc14_levelc15_levelc16_levelc17_levelc18_levelc19_levelc20_levelc21_levelc22_levelc23_levelc24_levelc25_levelc26_levelc27_levelc28_levelc29_levelc30_levelc31_levelc32_levelc33_levelc34_levelc35_levelc36_levelc37_levelc38_levelc39_levelc40_levelc41_levelc42_levelc43_levelc44_levelc45_levelc46_levelc47_levelc48_levelc49_levelc50_levelc51_levelc52_levelc53_levelc54_levelc55_levelc56_levelc57_levelc58_levelc59_levelc60_levelc61_levelc62_levelc63_levelc64_levelc65_levelc66_levelc67_levelc68_levelc69_levelc70_levelc71_levelc72_levelc73_levelc74_levelc75_levelc76_levelc77_levelc78_levelc79_levelc80_levelc81_levelc82_levelc83_levelc84_levelc85_levelc86_levelc87_levelc88_levelc89_levelc90_levelc91_levelc92_levelc93_levelc94_levelc95_levelc96_levelc97_levelc98_levelc99_level

20°C
Mostly sunny

Search

ENG

19:19
28/04/2023

State Sp... Introdu... G state sp... G Principi... What A... State Sp... Control... library... Data G... ERDPlus... hend_s... hend_s... G chrome... Extensio... Big Da... +

community.cloud.databricks.com/?o=1037376470547060#notebook/4203043310594709/command/4203043310594710

Amazon Save to Mendeley Suggested Sites W WildIntelligent Games eBay Imported from E... Euler angle, DCM, q... MF201x Course Int... The Secret to Newel... Inferring diffusion L... Planter 2 - Free O... www.magazinesDOT... Other bookmarks

hend_Sprints Python File Edit View Run Help Last edit was 8 minutes ago Give feedback

Run all My Cluster Publish

Table +

parquet_directory: pyspark.sql.dataframe.DataFrame = (battery_level: long, c02_level: long ... 13 more fields)

	battery_level	c02_level	cca2	cca3	cn	device_id	device_name	humidity	ip	latitude	led	longitude	scale	temp
1	8	868	US	USA	United States	1	meter-gauge-1xbYRij	51	68.161.225.1	38	green	-97	Celsius	34
2	7	1473	NO	NOR	Norway	2	sensor-pad-2n2Psa	70	213.161.254.1	62.47	red	6.15	Celsius	11
3	2	1556	IT	ITA	Italy	3	device-mac-36TWSkIT	44	88.36.5.1	42.63	red	12.83	Celsius	19
4	6	1080	US	USA	United States	4	sensor-pad-4mzWkz	32	66.39.173.154	44.06	yellow	-121.32	Celsius	28
5	4	931	PH	PHL	Philippines	5	therm-stick-5gimpUirBE	62	203.82.41.9	14.58	green	120.97	Celsius	25
6	3	1210	US	USA	United States	6	sensor-pad-6slTRTAobR	51	204.116.105.67	35.93	yellow	-85.46	Celsius	27
7	3	1129	CN	CHN	China	7	meter-gauge-TGeDoanM	26	220.173.179.1	22.82	yellow	108.32	Celsius	18

10,000 rows | Truncated data | 5.02 seconds runtime Refreshed 9 minutes ago

Command took 5.02 seconds -- by hend_n_ao@yahoo.com at 4/28/2023, 7:18:31 PM on My Cluster

Cell 28

1: %scala
2: spark.sql("show tables").show()

1 (2) Spark Jobs

database	tableName	isTemporary
default	iot_device_data2	false
	data_geo	true
	iot_device_data	true

Command took 6.68 seconds -- by hend_n_ao@yahoo.com at 4/28/2023, 7:11:44 PM on My Cluster

Cell 29

1:

Cell 30

1:

20°C Mostly sunny

Search

19:20 28/04/2023