Student Alcohol Consumption

Melchor Ronquillo, Rob Sisto, Annemarie Andaleon, Jacob Plaza

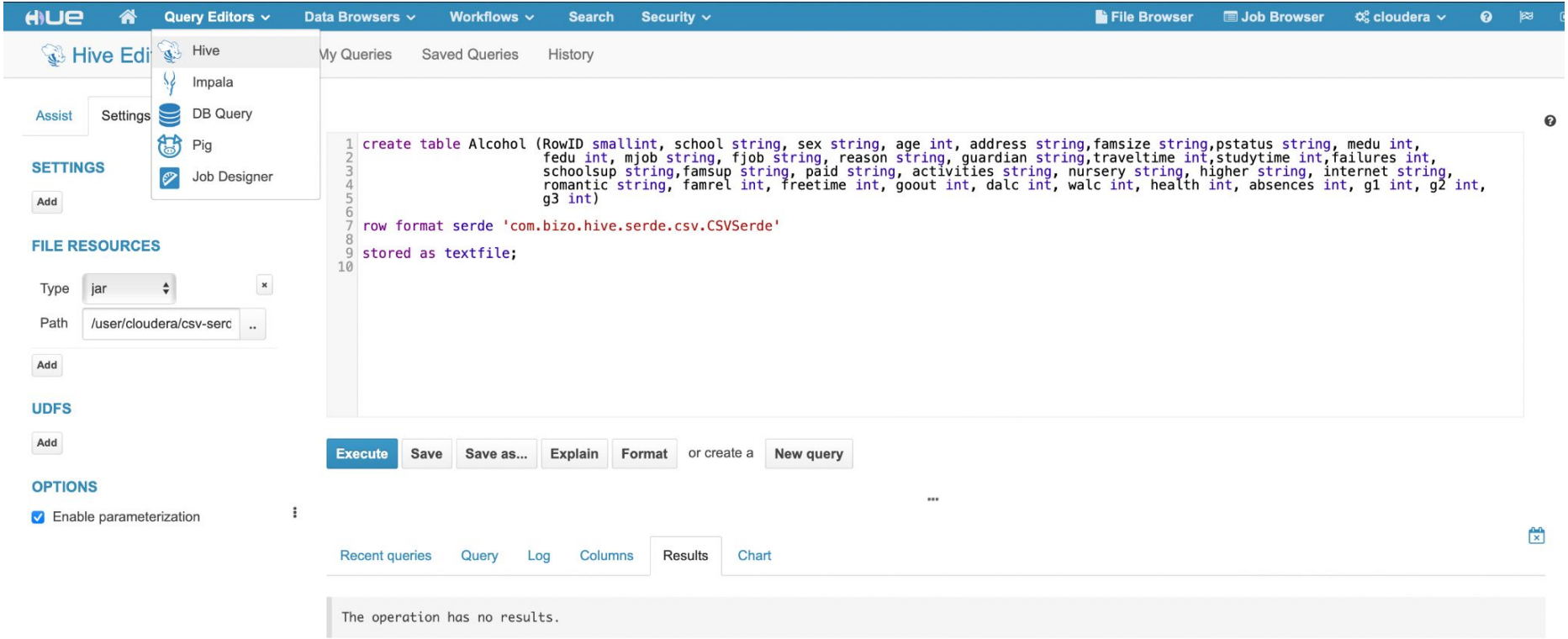
Dataset Description

- Our dataset is from kaggle and uploaded by UCI Machine Learning.
- The data was obtained from a survey of secondary school students in Portugal.
  - 395 rows (students), 30 columns
- Survey primarily focuses on students and their alcohol consumption.
- Alcohol consumption is measured by 2 variables,
  - WALC (Weekend Alcohol Consumption)
  - DALC (Workday Alcohol Consumption).
- Many other variables that contributes to / reflects what is affected by alcohol consumption
  - Demographic (age, sex, address, school, etc)
  - Family information (parents relationship, occupation, education, etc)
  - External information (health condition, time spent traveling to school, going out frequency, internet at home, etc)

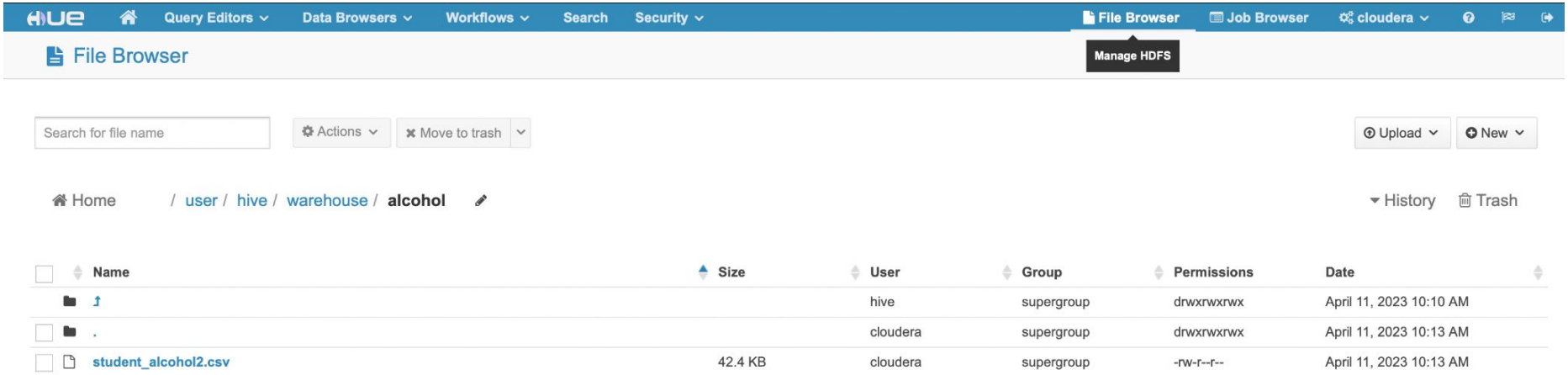
Data Dictionary (30 columns total, only interested in few)

- School: **student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)**
- Age: **student's age (numeric: from 15 to 22)**
- Sex: **student's sex (binary: 'F' - female or 'M' - male)**
- Address: **student's home address type (binary: 'U' - urban or 'R' - rural)**
- Health: **current health status (numeric: from 1 - very bad to 5 - very good)**
- Walc: **weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)**
- Dalc: **weekday alcohol consumption (numeric: from 1 - very low to 5 - very high)**
- Famsize: **family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)**
- Pstatus: **parent's cohabitation status (binary: 'T' - living together or 'A' - apart)**
- Famrel: **quality of family relationships (numeric: from 1 - very bad to 5 - excellent)**
- Medu / Fedu: **mother's / father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)**
- Traveltime: **home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)**
- Failures: **number of past class failures (numeric: n if 1<=n<3, else 4)**
- Higher: **wants to take higher education (binary: yes or no)**
- Internet: **Internet access at home (binary: yes or no)**
- Goout: **going out with friends (numeric: from 1 - very low to 5 - very high)**

First Method - Manual Input



First Method - Manual Input



First Method - Manual Input

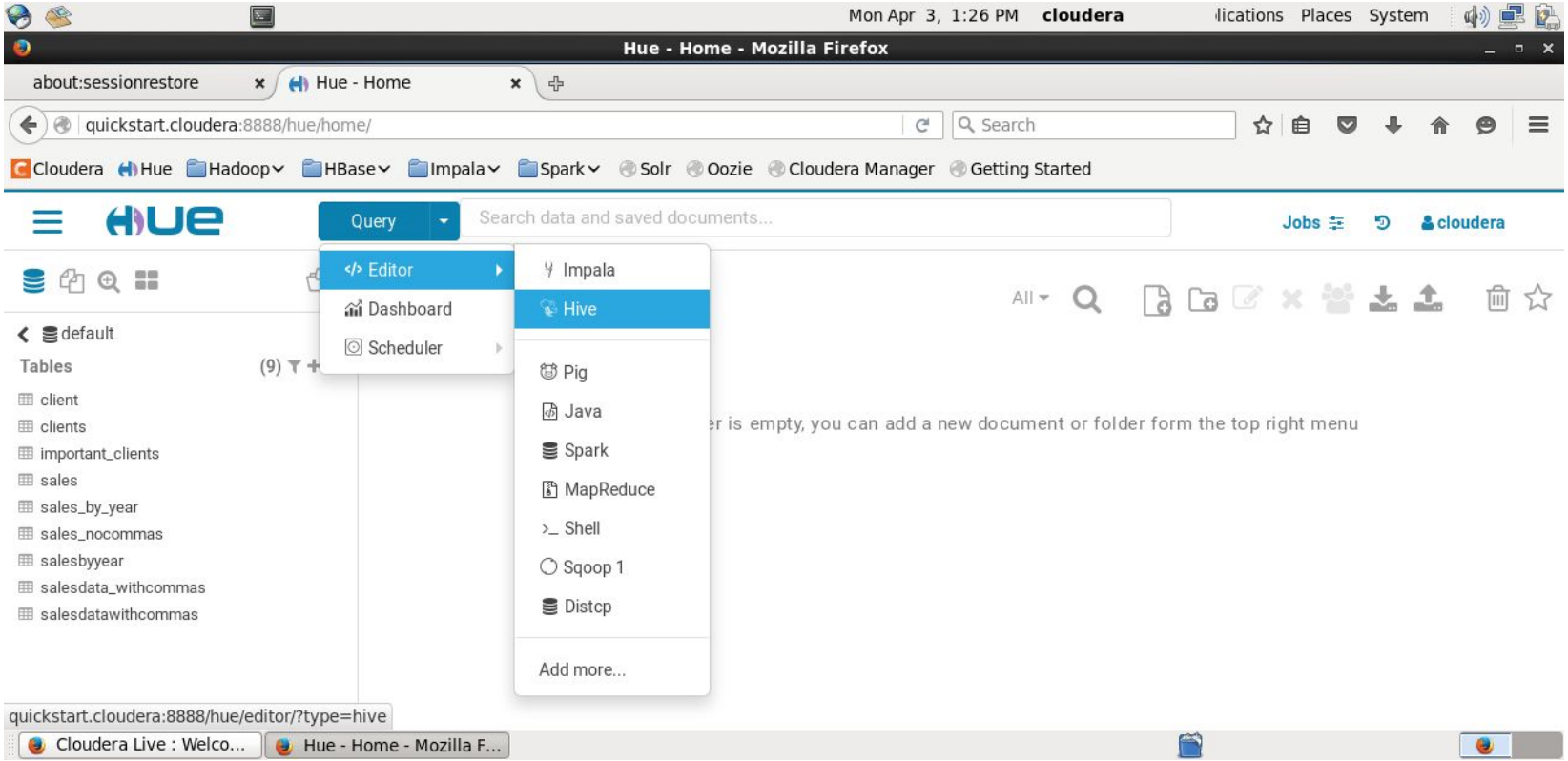
Data sample for alcohol

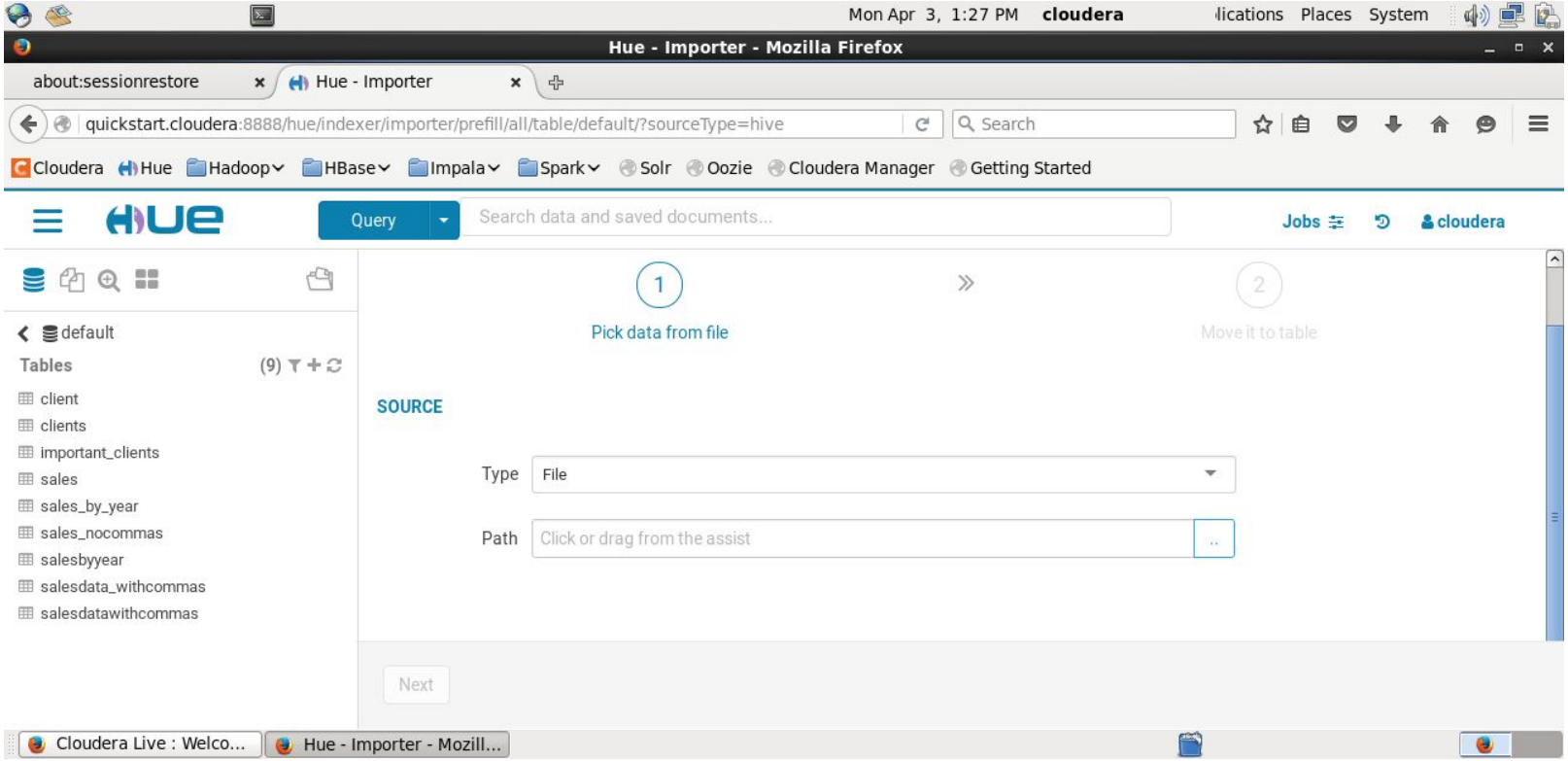
View more...

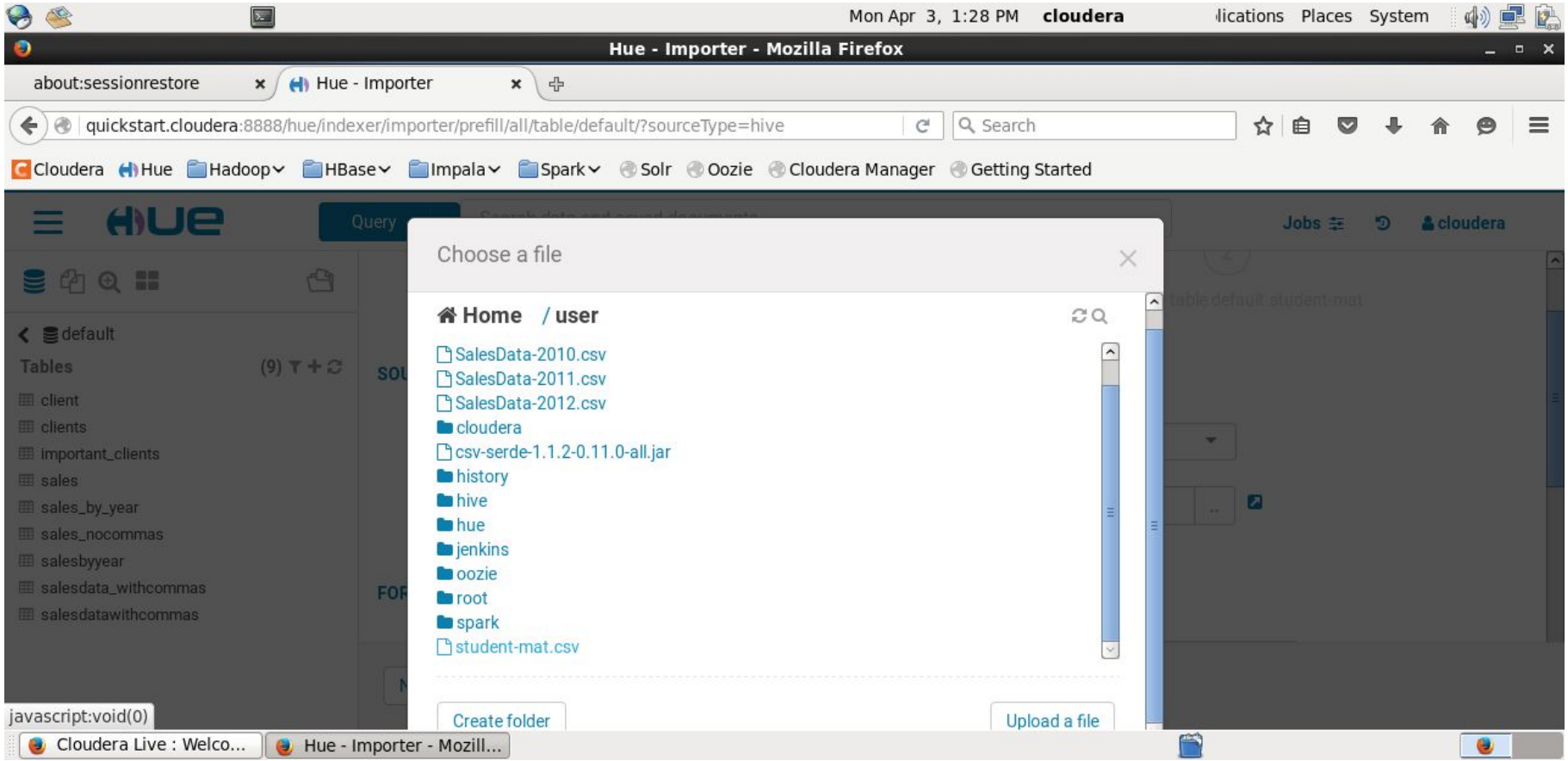
	alcohol.rowid	alcohol.school	alcohol.sex	alcohol.age	alcohol.address	alcohol.famsize	alcohol.pstatus	alcohol.medu	alcohol.fedu	alcohol.mjob	alcohol.fjob	alcohol.reason	alcohol.guardian	alcohol.traveltime
1		school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime
2	0	GP	F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2
3	1	GP	F	17	U	GT3	T	1	1	at_home	other	course	father	1
4	2	GP	F	15	U	LE3	T	1	1	at_home	other	other	mother	1
5	3	GP	F	15	U	GT3	T	4	2	health	services	home	mother	1
6	4	GP	F	16	U	GT3	T	3	3	other	other	home	father	1
7	5	GP	M	16	U	LE3	T	4	3	services	other	reputation	mother	1
8	6	GP	M	16	U	LE3	T	2	2	other	other	home	mother	1
9	7	GP	F	17	U	GT3	A	4	4	other	teacher	home	mother	2
10	8	GP	M	15	U	LE3	A	3	2	services	other	home	mother	1
11	9	GP	M	15	U	GT3	T	3	4	other	other	home	mother	1
12	10	GP	F	15	U	GT3	T	4	4	teacher	health	reputation	mother	1
13	11	GP	F	15	U	GT3	T	2	1	services	other	reputation	father	3

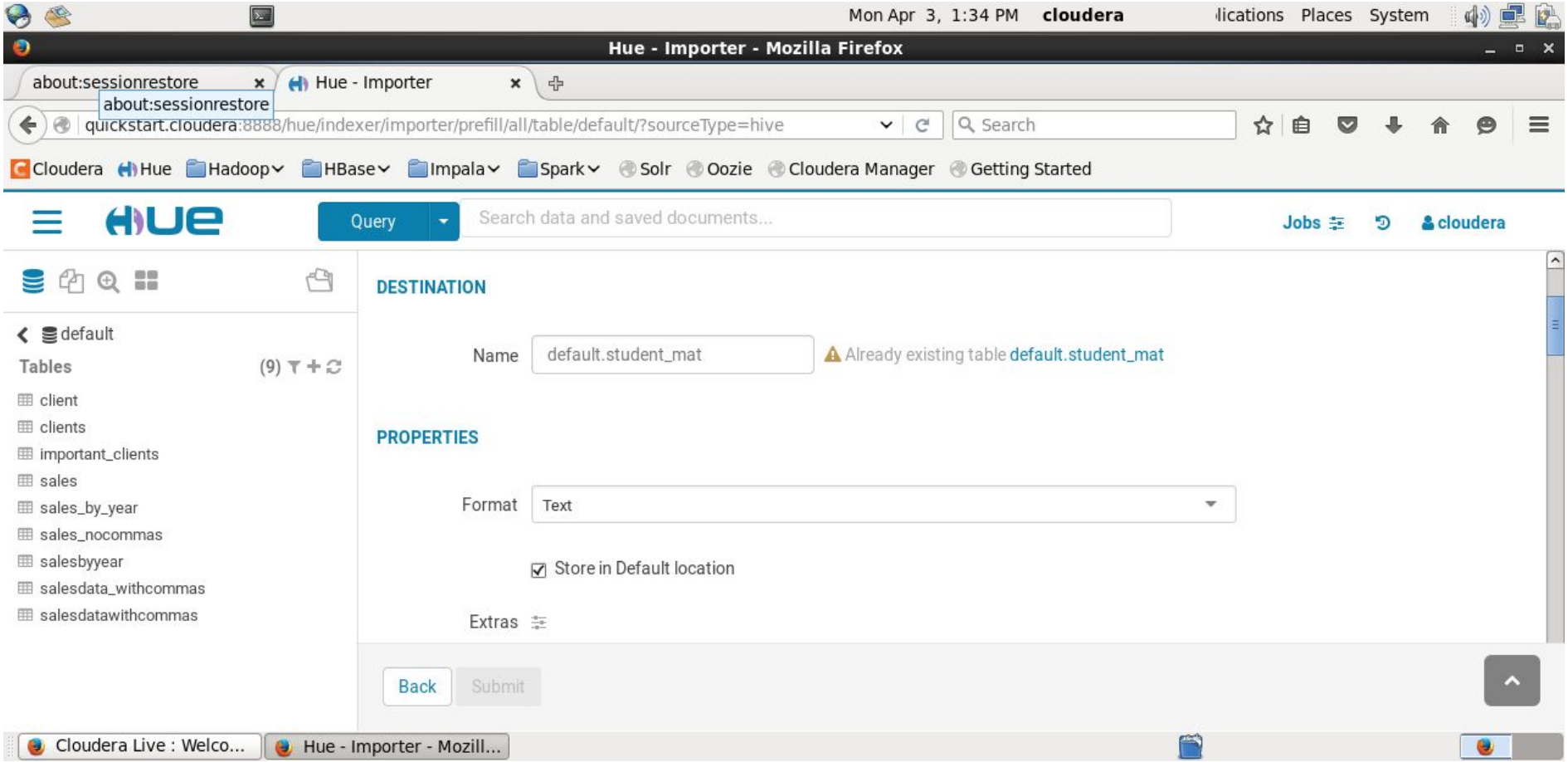
Ok

Second Method - Using the Interface




Mon Apr 3, 1:27 PM cloudera
lications Places System
Hue - Importer - Mozilla Firefox
about:sessionrestore
Hue - Importer
quickstart.cloudera:8888/hue/indexer/importer/prefill/all/table/default/?sourceType=hive
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Query
Search data and saved documents...
Jobs
cloudera
default
Tables (9)
client
clients
important_clients
sales
sales_by_year
sales_nocommas
salesbyyear
salesdata_withcommas
salesdatawithcommas
1
Pick data from file
2
Move it to table
SOURCE
Type
File
Path
Click or drag from the assist
..
Next
Cloudera Live : Welco...
Hue - Importer - Mozill...


Mon Apr 3, 1:28 PM cloudera
Applications Places System
Hue - Importer - Mozilla Firefox
about:sessionrestore
Hue - Importer
quickstart.cloudera:8888/hue/indexer/importer/prefill/all/table/default/?sourceType=hive
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Query
Jobs
cloudera
default
Tables (9)
client
clients
important_clients
sales
sales_by_year
sales_nocommas
salesbyyear
salesdata_withcommas
salesdatawithcommas
Choose a file
Home / user
SalesData-2010.csv
SalesData-2011.csv
SalesData-2012.csv
cloudera
csv-serde-1.1.2-0.11.0-all.jar
history
hive
hue
jenkins
oozie
root
spark
student-mat.csv
Create folder
Upload a file
javascript:void(0)
Cloudera Live : Welco...
Hue - Importer - Mozill...


Mon Apr 3, 1:34 PM cloudera
lications Places System
Hue - Importer - Mozilla Firefox
about:sessionrestore
Hue - Importer
about:sessionrestore
quickstart.cloudera:8888/hue/indexer/importer/prefill/all/table/default/?sourceType=hive
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Query
Search data and saved documents...
Jobs
cloudera
default
Tables (9)
client
clients
important_clients
sales
sales_by_year
sales_nocommas
salesbyyear
salesdata_withcommas
salesdatawithcommas
DESTINATION
Name
default.student_mat
Already existing table default.student_mat
PROPERTIES
Format
Text
Store in Default location
Extras
Back
Submit
Cloudera Live : Welco...
Hue - Importer - Mozill...

Melchor's Queries

- Find the average alcohol level of students (weekend + weekday average) for every possible parent occupation combination (mother job and father job), as well as the amount of students with these parent occupation combination, order by alcohol level

```
SELECT cast(avg(a.walc + a.dalc) as decimal(10,1)) as Average_Alcohol, a.fjob, a.mjob, count(1) as Number_of_kids
FROM alcohol a
GROUP BY a.fjob, a.mjob
ORDER BY Average_Alcohol;
```

Recent queries | Query | Log | Columns | Results | Chart

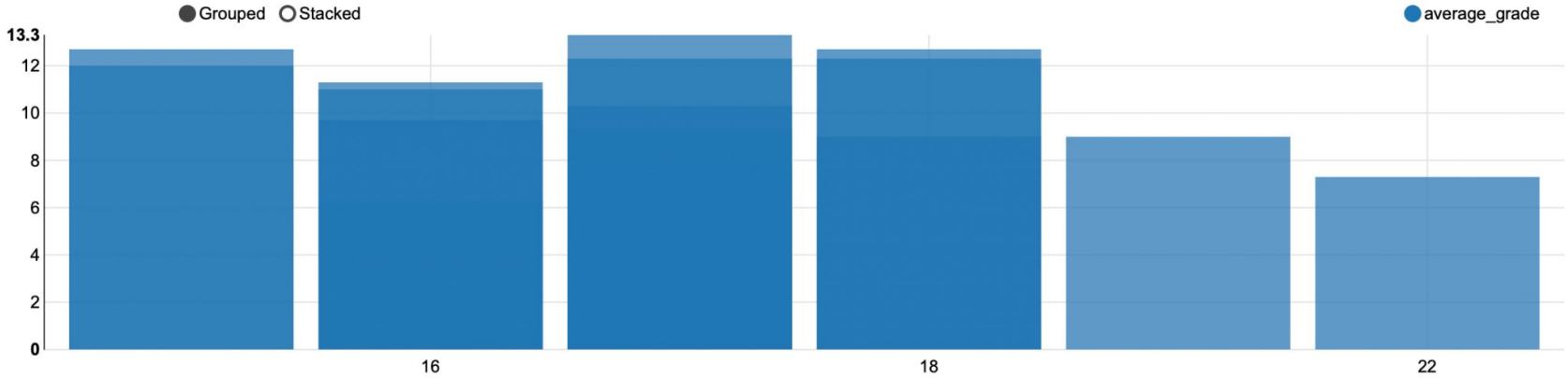
	average_alcohol	a.fjob	a.mjob	number_of_kids
2	2	teacher	at_home	2
3	2	health	at_home	2
4	2.2	at_home	services	6
5	2.5	health	other	2
6	2.8	teacher	services	8
7	3	teacher	other	6
8	3	at_home	teacher	2
9	3	health	health	6
10	3	health	services	4
11	3.2	at_home	other	5
12	3.4	services	health	10
13	3.5	services	teacher	19
14	3.6	other	other	104
15	3.6	at_home	at_home	7
16	3.8	other	services	42
17	3.9	other	health	17
18	3.9	services	at_home	15
19	3.9	other	at_home	33
20	4	teacher	teacher	12
21	4.1	other	teacher	21
22	4.3	services	services	43
23	4.6	services	other	24
24	4.8	health	teacher	4
25	5	teacher	health	1

Melchor's Queries

- Find a student's average grade across 3 periods and their age given both their weekend and weekday alcohol level is 4 or higher, sort by age

```
SELECT cast((a.g1 + a.g2 + a.g3) / 3 as decimal(10,1)) as Average_Grade, a.age
FROM alcohol a
WHERE a.walc >= 4 AND a.dalc >= 4
ORDER BY a.age;
```

	average_grade	a.age
1	12	15
2	12.7	15
3	6.3	16
4	9.7	16
5	11	16
6	11.3	16
7	7.7	17
8	10.3	17
9	13.3	17
10	12.3	17
11	9.3	17
12	12.7	18
13	9	18
14	12.3	18
15	9	20
16	7.3	22



Rob’s Queries

1) Average total alcohol consumption by total parent education level
   a) *select cast(avg(walc + dalc) as decimal(10,2)) as avg_alc, fedu+medu as edu from alcohol group by fedu+medu;*

	avg_alc	edu
1	NULL	NULL
2	2	1
3	4.03	2
4	3.79	3
5	3.93	4
6	3.49	5
7	3.72	6
8	3.55	7
9	3.93	8

Rob’s Queries

1) Average total alcohol consumption by the students school and address
   a) *select cast(avg(walc + dalc) as decimal(10,2)) as avg_alc, school, address, count(school) as num from alcohol group by school, address order by avg_alc;*

	avg_alc	school	address	num
1	NULL	school	address	1
2	3.57	MS	U	21
3	3.66	GP	U	286
4	3.89	GP	R	63
5	4.88	MS	R	25

Annemarie’s Queries

1) The average family relationships of families of certain parental statuses and family sizes
   a) SELECT cast(avg(famrel) as decimal(10,2)) as AvgFamilyRelationship, famsize, pstatus FROM student_mat
      GROUP BY pstatus, famsize
      ORDER BY AvgFamilyRelationship

```
1 SELECT cast(avg(famrel) as decimal(10,2))as AvgFamilyRelationship, count(*), famsize, pstatus FROM student_mat
2 GROUP BY pstatus, famsize
3 ORDER BY AvgFamilyRelationship
```

Query History | Saved Queries | Results (4)

	avgfamilyrelationship	count(*)	famsize	pstatus
1	3.66	21	GT3	A
2	3.87	94	LE3	T
3	3.98	260	GT3	T
4	4.09	20	LE3	A

Annemarie's Queries

1) The failure rate of students based on their health
   a) SELECT health, count(*) AS `Count`, sum(failures) as TotalFailures, cast(avg(failures) as DECIMAL(10,2)) FROM student_mat
      GROUP BY health
      ORDER BY health

```
SELECT health, count(*) as `Count`, sum(failures) as TotalFailures, cast(avg(failures) as DECIMAL(10,2)) as AvgFailures FROM student_mat
GROUP BY health
ORDER BY health
```

Query History | Saved Queries | Results (5)

	health	count	totalfailures	avgfailures
1	1	47	8	0.17
2	2	45	12	0.26
3	3	91	37	0.40
4	4	66	22	0.33
5	5	146	53	0.36

Jacob's Queries

Select the median age of heavy drinkers, and average alcohol consumption above and below the median

```
1 SELECT 'median age' desc, (min(age) + max(age))/2 value from new_mat where walc >= 3
2 union all
3 SELECT 'consumption above median' desc, cast(avg(walc) as decimal(10,2)) value from new_mat where age >= 18.5 and walc >= 3
4 union all
5 SELECT 'consumption below median' desc, cast(avg(walc) as decimal(10,2)) value from new_mat where age < 18.5 and walc >= 3;
```

Query History | Saved Queries | Results (9)

	_u1.desc	_u1.value
1	consumption above median	3.36
2	consumption below median	3.7
3	median age	18.5

Jacob's Queries

Average total alcohol consumption by studytime group

```
9
10 Select studytime, cast(avg(dalc+walc) as decimal (10,2)) avg_alc from new_mat GROUP BY studytime order by studytime;
```

Query History | Saved Queries | Results (4)

	studytime	avg_alc
1	1	4.52
2	2	3.72
3	3	3.02
4	4	3.07

studytime: weekly study time
1 - less than 2 hours
2 - 2 to 5 hours,
3 - 5 to 10 hours
4 - greater than 10 hours

avg_alc:
1 - very low
2 - low
3 - high
4 - very high