Project 1
（Hbase&Phoenix）

Wearable Exam Stress Database

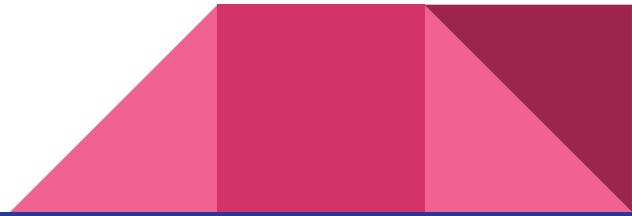
Team 3: Lex Brunet, Carol Jiang, Rob Jones, Olivia Roeger, Manya Singh

Data Introduction

- Source：A Wearable Exam Stress Dataset for Predicting Cognitive Performance in Real World Settings (From PhysioNet)
- Format：The data is organized on different folders for each student, examination and physiological signal.
- Contains：1. Ten Students' grades in Midterm 1, Midterm 2, and Final exam（txt）
  2. Stress Data for each exam and each student includes（csv）:
     - Temperature in celsius
     - Electrodermal activity
     - Blood peripheral circulation
     - Acceleration from -2g to 2g
     - Time between heart beats
     - Heart rate

Goal for data analysis

- The goal is to explore the connection between stress indicators and academic outcomes, offering insights into how stress affects cognitive performance.



Data Preprocessing

- Data Extraction：
  - Extract participants’ grades data from text documents
  - Extract electrodermal activity (EDA) data from csv documents
  - Extract data containing "sid" and "examname" information from file names

- Data Selection：
  - Select grades information about sid examname results
  - Select all electrodermal activity (EDA) data except acc (Acceleration)
    - Because acc data has three dimension without timestamp which is different from others EDA data

- Data Normalization:
  - Assure the grades are normalized for the three exams（out of 100）.
  - Organize the formats for the timestamps to ensure data quality and uniformity of the visualization.
  - Map rowid for each test result record.

Creating the tables in hbase:

Testresult table:

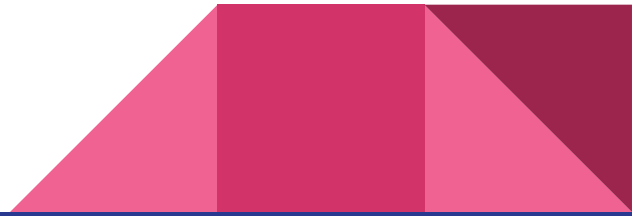
```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv
-Dimporttsv.separator=, -Dimporttsv.columns="\

HBASE_ROW_KEY, \

results:rowkey, \

results:timestamp, \

results:data, \

results:sid, \

results:examname" testresult /tmp/testresult .csv
```

Studentgrades table:

```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv
-Dimporttsv.separator=, -Dimporttsv.columns="\

HBASE_ROW_KEY, \

student:rowkey, \

student:sid, \

student:shipDate, \

student:examname, \

student:result" studentgrades /tmp/studentgrades.csv
```

Hbase Table Design

- StudentGrades: Grades（rowkey, Sid, examname, result）
    - Table used to store the students grade data
- TestResult: result（rowkey, Timestamp, Data, type, sid, examname)
    - Table used to store the students physiological data



Creating the views in Phoenix:

Testresult view:

```
create view "testresult" (
 "row" VARCHAR primary key,
 "results"."rowkey" VARCHAR,
 "results"."timestamp" VARCHAR,
 "results"."data" VARCHAR,

"results"."type" VARCHAR,
 "results"."sid" VARCHAR,
 "results"."examname" VARCHAR);
```

Studentgrades view:

```
create view "studentgrades" (
 "student"."rowkey" VARCHAR primary
key,
"student"."sid" VARCHAR,
 "student"."examname" VARCHAR,
 "student"."result" VARCHAR);
```

Query 1

SELECT * FROM "studentgrades" limit 10;

Output :

'rowkey'	'sid'	'examname'	'result'
'1'	'S01'	'MIDTERM 1'	'78.0'
'10'	'S10'	'MIDTERM 1'	'89.0'
'11'	'S01'	'MIDTERM 2'	'82.0'
'12'	'S02'	'MIDTERM 2'	'85.0'
'13'	'S03'	'MIDTERM 2'	'90.0'
'14'	'S04'	'MIDTERM 2'	'77.0'
'15'	'S05'	'MIDTERM 2'	'77.0'
'16'	'S06'	'MIDTERM 2'	'64.0'
'17'	'S07'	'MIDTERM 2'	'33.0'
'18'	'S08'	'MIDTERM 2'	'88.0'

Query 2

SELECT * FROM "testresult" limit 10;

Output :

'rowkey'	'timestamp'	'data'	'type'	'sid'	'examname'	
'1'	'2018-12-05 16:28:57'	'-0.0'	'BVP'	'S1'	'Final'	
'10'	'2018-12-05 16:28:57.140625'	'-0.0'	'BVP'	'S1'	'Final'	
'100'	'2018-12-05 16:28:58.546875'	'-177.09'	'BVP'	'S1'	'Final'	
'1000'	'2018-12-05 16:29:12.609375'	'-3.6'	'BVP'	'S1'	'Final'	
'10000'	'2018-12-05 16:31:33.234375'	'-0.63'	'BVP'	'S1'	'Final'	
'100000'	'2018-12-05 16:54:59.484375'	'2.1'	'BVP'	'S1'	'Final'	
'1000000'	'2018-12-05 20:49:21.984375'	'-1.25'	'BVP'	'S1'	'Final'	
'1000001'	'2018-12-05 20:49:22'	'1.56'	'BVP'	''	'S1'	'Final'
'1000002'	'2018-12-05 20:49:22.015625'	'4.33'	'BVP'	'S1'	'Final'	
'1000003'	'2018-12-05 20:49:22.031250'	'5.5'	'BVP'	'S1'	'Final'	

Query 3

```
SELECT studentgrades.sid,
 AVG(studentgrades.result) AS Average_Score
FROM studentgrades
WHERE studentgrades.examname IN ('MIDTERM 1', 'MIDTERM 2', 'FINAL')
GROUP BY studentgrades.sid;
```

Output :

sid	averageresult
S01	83.67
S02	85.67
S03	87.0
S04	75.5
S05	74.17
S06	74.17
S07	50.67
S08	90.67
S09	60.67
S10	70.33

Query 4

SELECT * FROM "studentgrades" order by "result" |

Output :

'rowkey'	'sid'	'examname'	'result'
'23'	'S03'	'Final'	'94.0'
'28'	'S08'	'Final'	'92.0'
'8'	'S08'	'MIDTERM 1'	'92.0'
'21'	'S01'	'Final'	'91.0'
'3'	'S03'	'MIDTERM 2'	'90.0'
'22'	'S02'	'Final'	'90.0'
'10'	'S10'	'MIDTERM 1'	'89.0'
'18'	'S08'	'MIDTERM 2'	'88.0'
'26'	'S06'	'Final'	'87.5'
'12'	'S02'	'MIDTERM 2'	'85.0'

Query 5

SELECT * FROM “testresult” WHERE “rowkey” IN (1,4926761,10,4926762,100,4926763,1000,4926764,10000,4926765)

Output :

'rowkey'	'timestamp'	'data'	'type'	'sid'	'examname'
'1'	'2018-12-05 16:28:57'	'-0.0'	'BVP'	'S1'	'Final'
'4926761'	'2018-12-05 22:17:09'	'135.32'	'HR'	'S10'	'Final'
'10'	'2018-12-05 16:28:57.140625'	'-0.0'	'BVP'	'S1'	'Final'
'4926762'	'2018-12-05 22:17:10'	'137.07'	'HR'	'S10'	'Final'
'100'	'2018-12-05 16:28:58.546875'	'-177.09'	'BVP'	'S1'	'Final'
'4926763'	'2018-12-05 22:17:11'	'137.3'	'HR'	'S10'	'Final'
'1000'	'2018-12-05 16:29:12.609375'	'-3.6'	'BVP'	'S1'	'Final'
'4926764'	'2018-12-05 22:17:12'	'139.05'	'HR'	'S10'	'Final'
'10000'	'2018-12-05 16:31:33.234375'	'-0.63'	'BVP'	'S1'	'Final'
'4926765'	'2018-12-05 22:17:13'	'139.32'	'HR'	'S10'	'Final'

Query 6

SELECT * FROM "studentgrades" WHERE "result" >= 80;

Output :

rowkey	sid	examname	result
2	S02	MIDTERM 1	82
8	S08	MIDTERM 1	92
9	S09	MIDTERM 1	80
10	S10	MIDTERM 1	89
11	S01	MIDTERM 2	82
12	S02	MIDTERM 2	85
13	S03	MIDTERM 2	90
18	S08	MIDTERM 2	88
21	S01	Final	91
22	S02	Final	90
23	S03	Final	94
26	S06	Final	87.5
28	S08	Final	92

Query 7

SELECT * FROM "studentgrades" WHERE "examname" = 'MIDTERM 2' AND "result" < 70 LIMIT 4;

Output :

rowkey	sid	examname	result
16	S06	MIDTERM 2	64
17	S07	MIDTERM 2	33
19	S09	MIDTERM 2	39
20	S10	MIDTERM 2	64

Query 8

Query1 : select * from "StudentGrades" where "Sid"='S01';

Output :

'row'	'Sid'	'examname'	'result'
'1'	'S01'	'MIDTERM 1'	'78.0'
'11'	'S01'	'MIDTERM 2'	'82.0'
'21'	'S01'	'Final'	'91.0'

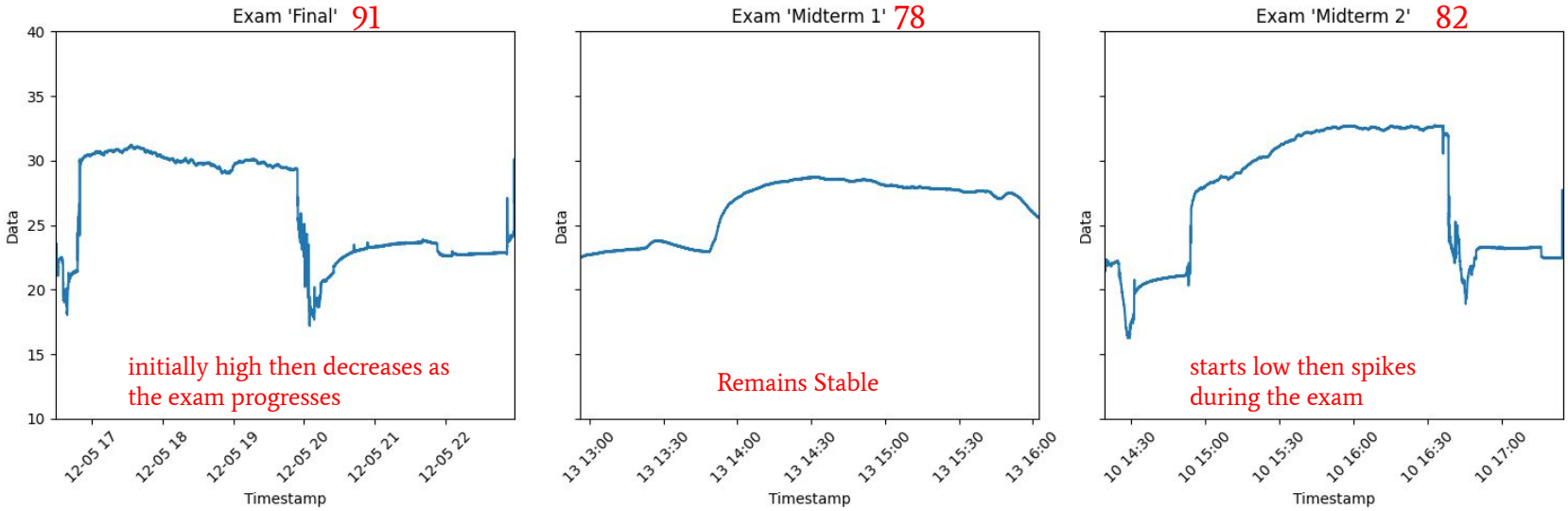
Query 9

Query2 : select * from "testresult" where "Sid" = 'S1' and "type" = 'TEMP' ;

Output :

Saving all output to "/output2.csv". Enter "record" with no arguments to stop it.					
Error: ERROR 601	column 1. (state=42P00	code=601)			
'row'	'Timestamp'	'Data'	'type'	'Sid'	'examname'
'1614345'	'2018-12-05 16:28:57'	'21.89'	'TEMP'	'S1'	'Final'
'3292345'	'2018-11-10 14:19:15'	'21.79'	'TEMP'	'S1'	'Midterm 2'
'1614346'	'2018-12-05 16:28:57.250000'	'21.89'	'TEMP'	'S1'	'Final'
'3292346'	'2018-11-10 14:19:15.250000'	'21.79'	'TEMP'	'S1'	'Midterm 2'
'1614347'	'2018-12-05 16:28:57.500000'	'21.89'	'TEMP'	'S1'	'Final'
'3292347'	'2018-11-10 14:19:15.500000'	'21.79'	'TEMP'	'S1'	'Midterm 2'
'1614348'	'2018-12-05 16:28:57.750000'	'21.89'	'TEMP'	'S1'	'Final'
'3292348'	'2018-11-10 14:19:15.750000'	'21.79'	'TEMP'	'S1'	'Midterm 2'
'1614349'	'2018-12-05 16:28:58'	'21.89'	'TEMP'	'S1'	'Final'
'3292349'	'2018-11-10 14:19:16'	'21.79'	'TEMP'	'S1'	'Midterm 2'
'1614350'	'2018-12-05 16:28:58.250000'	'21.89'	'TEMP'	'S1'	'Final'
'3292350'	'2018-11-10 14:19:16.250000'	'21.79'	'TEMP'	'S1'	'Midterm 2'
'1614351'	'2018-12-05 16:28:58.500000'	'21.89'	'TEMP'	'S1'	'Final'
'3292351'	'2018-11-10 14:19:16.500000'	'21.79'	'TEMP'	'S1'	'Midterm 2'

Analysis for student1 based on TEMP data



- Conclusion: Fluctuations in body temperature may reflect emotional states like nervousness or excitement, potentially affecting exam performance. Heightened excitement may enhance focus and performance, while nervousness might hinder it. The analysis is bias, because based on specific student. We want to try more detailed analysis in the following project according to complex queries.