Rapport PPD

1. **Partie Pi:**

Le but de ce TP était de voir différentes manières de calcul pour obtenir une valeur qui tend le plus vers le chiffre PI=3.14....

Formule de calcul de PI: 4*Sum(0->m=10^8) = (-1)^i/(2i+1)

Test sur machine en local:
Sur une exécution simple avec stockage dans un fichier on obtient les résultat suivant avec l'aide de la commande:

```
$ time foo real 0m0.003s user 0m0.003s sys 0m0.000s $
```

précision calcul	real	user	sys
10 000	0.004	0.004	0.000
10 00 000	0.019	0.019	0.000
10 000 000	0.036	0.036	0.000
10 00 00 000	3.040	3.040	0.000
10 0 000 000 00	28.390	28.382	0.000

Remarques :

- Le temps système vaut toujours zéro car il n'y a pas d'appels systèmes dans le programme calcpi .
- le temps d'exécution augmente très rapidement car la boucle contient une concaténation de résultat intermédiaire et un calcul de puissance, ce qui consomme davantages de ressources quand l'interval de calcul est grand.

Pi Sur GRID5K:
iIl faut d'abord reservé un noeud sur GRID5K pour exécuter le script avec la commande suivante s:
oarsub -l /nodes=2,walltime=00:03:00 <./script.sh>
Le principe d'utiliser Grid5K est qu'on peut profiter d'un réseau à travers la France contenant des clusters de machines permettant d'exécuter de grand calculs en vitesse accélérée.

précision calcul	real	user	sys
10 000	très petite		
10 00 000	très petite		
10 000 000	0.024		
10 00 00 000	0.856		
10 0 000 000 00	3.178		

Si on compare les deux tableaux de calcul de Pi en Local et sur GRID5K, on remarque un énorme gain de temps, surtout quand les calculs commencent devenir très grand , c'est la on remarque l'avantage de GRID5K et du calcul parallèle.

2. Partie Hadoop:

Test en pseudo distribué:

Anagrammes: dans cette partie ,nous avons implémentés un programme MAP/REDUCE. Le code est disponible dans l'archive et les résultat de l'exécution se trouve dans le dossier outputs.

Code exemple:

```
  //class de mappage hadoop
  public static class TokenizerMapper extends Mapper<Object,Text,Text,Text>{
    private Text word = new Text();
    //methode de mappage
    public void map(Object key,Text value, Context context) throws
IOException,InterruptedException{
      char[] text = value.toString().toLowerCase().toCharArray();
      //tri la chaine par ordre alphabetique
      Arrays.sort(text);
      //la clé => mot Sorté, valeur => mot de base
      context.write(new Text(new String(text)),value);
    }
  }
```

Poste: dans cette partie ,nous avons implémentés un programme MAP/REDUCE. Le code est disponible dans l'archive et les résultat de l'exécution se trouve dans le dossier outputs.

Code exemple:

```
  //class de reduction hadoop
  public static class PosteReducer extends Reducer<Text,Text,Text,Text>{
    // methode de réduction
    public void reduce(Text key,Iterable<Text> values,Context context) throws
IOException,InterruptedException{
      int compt = 0;
      Iterator<Text> iter = values.iterator();
      StringBuilder maker = new StringBuilder();
      maker.append(key);
      while(iter.hasNext()) {
        Text coords = iter.next();
        compt ++;
        maker.append(coords);
        maker.append(";");
      }
        context.write(new Text(key.toString() +"("+ compt +")"),new Text(maker.toString()));
    }
```

Stat: dans cette partie ,nous avons implémentés un programme MAP/REDUCE.
Le code est disponible dans l'archive et les résultat de l'exécution se trouve dans le dossier outputs.
Code exemple:

```
 //class de mappage hadoop
   public static class AgeMapper
   extends Mapper<Object, Text, IntWritable, FloatWritable> {
    //methode de mappage hadoop
    public void map(Object key, Text value, Context context) throws
IOException,InterruptedException {
      String[] line = value.toString().split(",");
      IntWritable age;
      FloatWritable salary;
      //on ignore les libellés dans notre calcul
      if(line[1].equals("age") || line[4].equals("income")) {
        return;
      }
      int age_value = Integer.parseInt(line[1]);
      float salary_value = Float.parseFloat(line[4]);
      age = new IntWritable();
      salary = new FloatWritable();
      age.set(age_value);
      salary.set(salary_value);
      context.write(age, salary);
    }
   }
```

3. Mini-Projet:

Map --> (Clé <Id,Amis>, Valeurs <Autres amis>).
Reduce --> une liste de users qui ont la même clé (deux amis) et il faut regarder si il y a des amis en communs dans les deux listes d'amis.

Test en pseudo distribué:
Dans cette partie ,nous avons implémentés un programme MAP/REDUCE.
Le code est disponible dans l'archive et les résultat de l'exécution se trouve dans le dossier outputs.
Test Sur GRID5K:

Les résultats d'exécution sont les suivants :

```
**********************************************graphe 1**********************************************
Warning: Permanently added 'edel-12.grenoble.grid5000.fr,172.16.17.12' (RSA) to the list of
known hosts.
17/05/02 13:01:29 INFO client.RMProxy: Connecting to ResourceManager at
edel-12.grenoble.grid5000.fr/172.16.17.12:8032
```

17/05/02 13:01:29 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/05/02 13:01:30 INFO input.FileInputFormat: Total input paths to process : 1
17/05/02 13:01:30 INFO mapreduce.JobSubmitter: number of splits:1
17/05/02 13:01:30 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1493729791187_0001
17/05/02 13:01:30 INFO impl.YarnClientImpl: Submitted application application_1493729791187_0001
17/05/02 13:01:30 INFO mapreduce.Job: The url to track the job: http://edel-12.grenoble.grid5000.fr:8088/proxy/application_1493729791187_0001/
17/05/02 13:01:30 INFO mapreduce.Job: Running job: job_1493729791187_0001
17/05/02 13:01:38 INFO mapreduce.Job: Job job_1493729791187_0001 running in uber mode : false
17/05/02 13:01:38 INFO mapreduce.Job: map 0% reduce 0%
17/05/02 13:01:44 INFO mapreduce.Job: map 100% reduce 0%
17/05/02 13:01:52 INFO mapreduce.Job: map 100% reduce 100%
17/05/02 13:01:52 INFO mapreduce.Job: Job job_1493729791187_0001 completed successfully
17/05/02 13:01:52 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=516370
        FILE: Number of bytes written=1226653
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=12176
        HDFS: Number of bytes written=116308
        HDFS: Number of read operations=6
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=4229
        Total time spent by all reduces in occupied slots (ms)=4952
        Total time spent by all map tasks (ms)=4229
        Total time spent by all reduce tasks (ms)=4952
        Total vcore-seconds taken by all map tasks=4229
        Total vcore-seconds taken by all reduce tasks=4952
        Total megabyte-seconds taken by all map tasks=4330496
        Total megabyte-seconds taken by all reduce tasks=5070848
    Map-Reduce Framework
        Map input records=100
        Map output records=4022

```
        Map output bytes=507436
        Map output materialized bytes=516370
        Input split bytes=140
        Combine input records=0
        Combine output records=0
        Reduce input groups=2011
        Reduce shuffle bytes=516370
        Reduce input records=4022
        Reduce output records=2011
        Spilled Records=8044
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=54
        CPU time spent (ms)=3320
        Physical memory (bytes) snapshot=447545344
        Virtual memory (bytes) snapshot=1701883904
        Total committed heap usage (bytes)=402653184
    Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Input Format Counters
        Bytes Read=12036
    File Output Format Counters
        Bytes Written=116308

**********************************************graphe 2**********************************************
Warning: Permanently added 'edel-12.grenoble.grid5000.fr,172.16.17.12' (RSA) to the list of known hosts.
17/05/02 13:04:45 INFO client.RMProxy: Connecting to ResourceManager at edel-12.grenoble.grid5000.fr/172.16.17.12:8032
17/05/02 13:04:45 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/05/02 13:04:45 INFO input.FileInputFormat: Total input paths to process : 1
17/05/02 13:04:45 INFO mapreduce.JobSubmitter: number of splits:1
17/05/02 13:04:46 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1493729791187_0002
17/05/02 13:04:46 INFO impl.YarnClientImpl: Submitted application application_1493729791187_0002
17/05/02 13:04:46 INFO mapreduce.Job: The url to track the job: http://edel-12.grenoble.grid5000.fr:8088/proxy/application_1493729791187_0002/
```

```
17/05/02 13:04:46 INFO mapreduce.Job: Running job: job_1493729791187_0002
17/05/02 13:04:53 INFO mapreduce.Job: Job job_1493729791187_0002 running in uber
mode : false
17/05/02 13:04:53 INFO mapreduce.Job:  map 0% reduce 0%
17/05/02 13:04:58 INFO mapreduce.Job:  map 100% reduce 0%
17/05/02 13:05:04 INFO mapreduce.Job:  map 100% reduce 100%
17/05/02 13:05:04 INFO mapreduce.Job: Job job_1493729791187_0002 completed
successfully
17/05/02 13:05:04 INFO mapreduce.Job: Counters: 49
   File System Counters
        FILE: Number of bytes read=133964
        FILE: Number of bytes written=461843
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=6127
        HDFS: Number of bytes written=55948
        HDFS: Number of read operations=6
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
   Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=3214
        Total time spent by all reduces in occupied slots (ms)=3492
        Total time spent by all map tasks (ms)=3214
        Total time spent by all reduce tasks (ms)=3492
        Total vcore-seconds taken by all map tasks=3214
        Total vcore-seconds taken by all reduce tasks=3492
        Total megabyte-seconds taken by all map tasks=3291136
        Total megabyte-seconds taken by all reduce tasks=3575808
   Map-Reduce Framework
        Map input records=100
        Map output records=1954
        Map output bytes=130050
        Map output materialized bytes=133964
        Input split bytes=141
        Combine input records=0
        Combine output records=0
        Reduce input groups=977
        Reduce shuffle bytes=133964
        Reduce input records=1954
        Reduce output records=977
        Spilled Records=3908
        Shuffled Maps =1
```

```
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=47
        CPU time spent (ms)=2470
        Physical memory (bytes) snapshot=442056704
        Virtual memory (bytes) snapshot=1714876416
        Total committed heap usage (bytes)=402653184
    Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Input Format Counters
        Bytes Read=5986
    File Output Format Counters
        Bytes Written=55948
Connection to edel-12.grenoble.grid5000.fr closed.
Job with id job_1493729791187_0002 finished successfully

**********************************************graphe 3**********************************************
Warning: Permanently added 'edel-12.grenoble.grid5000.fr,172.16.17.12' (RSA) to the list of known hosts.
17/05/02 13:05:40 INFO client.RMProxy: Connecting to ResourceManager at edel-12.grenoble.grid5000.fr/172.16.17.12:8032
17/05/02 13:05:40 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/05/02 13:05:41 INFO input.FileInputFormat: Total input paths to process : 1
17/05/02 13:05:41 INFO mapreduce.JobSubmitter: number of splits:1
17/05/02 13:05:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1493729791187_0003
17/05/02 13:05:41 INFO impl.YarnClientImpl: Submitted application application_1493729791187_0003
17/05/02 13:05:41 INFO mapreduce.Job: The url to track the job: http://edel-12.grenoble.grid5000.fr:8088/proxy/application_1493729791187_0003/
17/05/02 13:05:41 INFO mapreduce.Job: Running job: job_1493729791187_0003
17/05/02 13:05:47 INFO mapreduce.Job: Job job_1493729791187_0003 running in uber mode : false
17/05/02 13:05:47 INFO mapreduce.Job:  map 0% reduce 0%
17/05/02 13:05:52 INFO mapreduce.Job:  map 100% reduce 0%
17/05/02 13:05:59 INFO mapreduce.Job:  map 100% reduce 100%
17/05/02 13:05:59 INFO mapreduce.Job: Job job_1493729791187_0003 completed successfully
17/05/02 13:05:59 INFO mapreduce.Job: Counters: 49
```

File System Counters
	FILE: Number of bytes read=39575
	FILE: Number of bytes written=273065
	FILE: Number of read operations=0
	FILE: Number of large read operations=0
	FILE: Number of write operations=0
	HDFS: Number of bytes read=3317
	HDFS: Number of bytes written=26926
	HDFS: Number of read operations=6
	HDFS: Number of large read operations=0
	HDFS: Number of write operations=2
Job Counters
	Launched map tasks=1
	Launched reduce tasks=1
	Data-local map tasks=1
	Total time spent by all maps in occupied slots (ms)=3118
	Total time spent by all reduces in occupied slots (ms)=3213
	Total time spent by all map tasks (ms)=3118
	Total time spent by all reduce tasks (ms)=3213
	Total vcore-seconds taken by all map tasks=3118
	Total vcore-seconds taken by all reduce tasks=3213
	Total megabyte-seconds taken by all map tasks=3192832
	Total megabyte-seconds taken by all reduce tasks=3290112
Map-Reduce Framework
	Map input records=100
	Map output records=990
	Map output bytes=37589
	Map output materialized bytes=39575
	Input split bytes=141
	Combine input records=0
	Combine output records=0
	Reduce input groups=495
	Reduce shuffle bytes=39575
	Reduce input records=990
	Reduce output records=495
	Spilled Records=1980
	Shuffled Maps =1
	Failed Shuffles=0
	Merged Map outputs=1
	GC time elapsed (ms)=55
	CPU time spent (ms)=1930
	Physical memory (bytes) snapshot=438067200
	Virtual memory (bytes) snapshot=1721044992
	Total committed heap usage (bytes)=402653184
Shuffle Errors
	BAD_ID=0

```
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Input Format Counters
        Bytes Read=3176
    File Output Format Counters
        Bytes Written=26926
Connection to edel-12.grenoble.grid5000.fr closed.
Job with id job_1493729791187_0003 finished successfully
```

Fin