Cloud Computing Assignment: MPI on the cloud

Léo Unbekandt

March 7, 2014

Contents

1 Settings

1.1 Parameters

Different sizes of matrix have been chosen in order to study the scalability of MPI on a cloud based cluster. (128, 256, 512, 1024, 1360). The deployed cluster has been composed of 1, 2, 4, 8 or 16 nodes. These values are really convenient as they are all powers of two, so the computation domain can be equally divided among the different virtual machines. The size of cluster is however limited by the quota in vCPUs (32) or Memory (50GB) and finally the different flavors of virtual machines which have been used:

- 1 - m1.tiny (512MB RAM, 1vCPU)
- 2 - m1.small (2GB RAM, 1vCPU)
- 3 - m1.medium (4GB RAM, 2vCPUs)

- 4 - m1.large (8GB RAM, 4vCPUS)

When a virtual machine has more than 1 vCPU, the MPI flag -npernode is used, in order to execute one instance of the application per vCPU, otherwise. Otherwise, it is useless to choose high quality flavors.

1.2 Automation of the execution

These tests have been run using only one command to avoid a maximum of human interactions. A ruby script has been developed to create the nodes cluster and to destroy it. When the cluster has been built, it automatically runs ansible in order to configure the nodes and install Open MPI on them.

Once the platform is ready, another script is able to run the experiments with the different kinds of matrix et get the output result back. (`run_matrix_multiply.sh`)

Finally a last script (`run_experiment.sh`) takes care to:

1. Build a cluster with N nodes and flavor F
2. Run MPI experiments
3. Get the results
4. Shutdown the cluster
5. Restart with other parameters

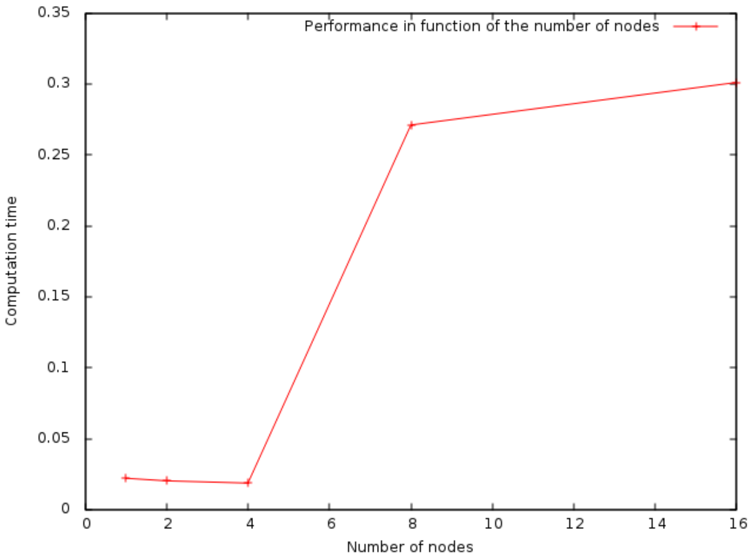
This last program takes care of everything, nothing else has to be done, except reading and interpreting the results! Please read the `README.md`

2 Results

The results in this section come from VMs of flavor 1 because I've been able to execute 16-nodes computation. They have only one CPU and 512MB of RAM.

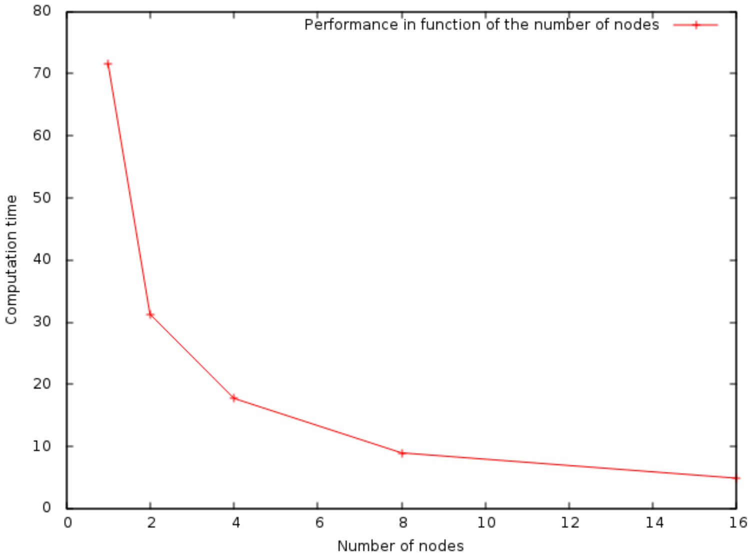
2.1 Small matrices computation

The raw data results can be read in Appendix A. First, for the small matrices (128 rows/columns), it is observable that the communication between the nodes becomes too heavy and even if the computation time is reduced, the overall duration increased hugely. Whatever is the flavor, the performance increases when the number of nodes is low, and then, they get strongly worse.



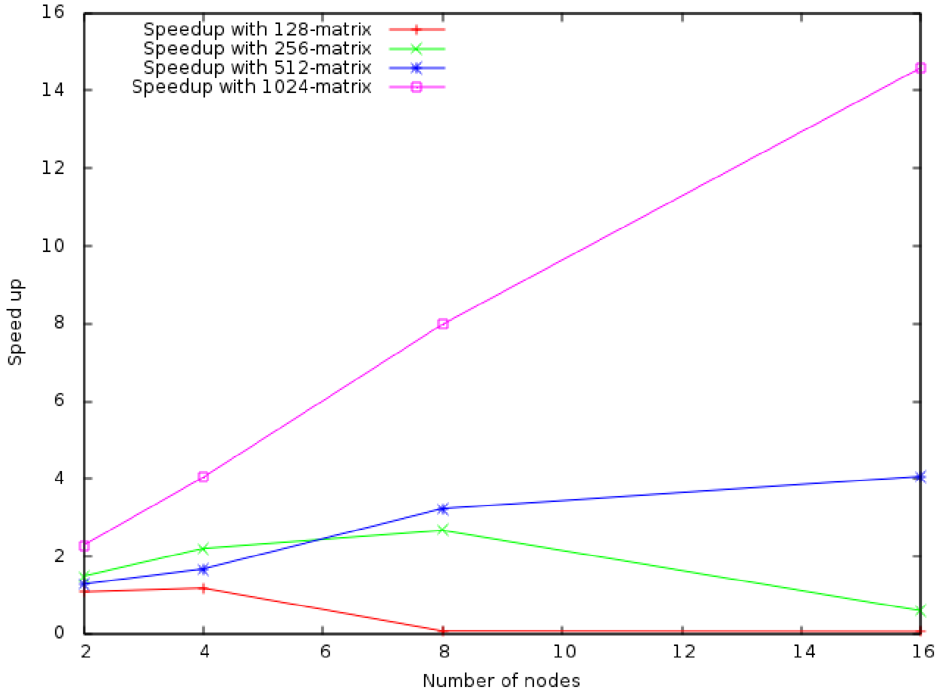
2.2 Large matrices computation

In the case of large matrices computation, the results are closer to what has to be expected, they are better and better but with a speed-up which is not linear.



2.3 Speedup Evolution

The two previous results can be illustrated in the following graph.



With a small matrix, the speed up is getting lower than 1 quickly. It means that the execution time is increasing when the number of nodse is increasing. However, when the matrix get bigger, the speed up increases when the number of VMs increases.

2.4 Cloud irregularity

We have to keep in mind that the hardware is shared among all the user, our tasks don't get dedicated CPUs for example. Consequently, some results may be inadequate. For instance we can se in the measure done with flavor 3 instances, it has been faster to compute the multiplication of a 1360x1360 matrix than a 1024x1024 matrix which is logically abnormal.

3 Pricing

The price for a single instance with one CPU on Amazon EC2 depends of the flavor:

- m1.tiny $0.065 per hour
- m1.small $0.130 per hour
- m1.medium $0.260 per hour
- m1.large $0.520 per hour

For small-scale experiments like these ones, the price would be completely negligible, however for MPI operations, the lesser the communications, the faster the execution. As a result, it is more interesting to take high-CPUs instances. For instance, in the case of intensive computational MPI tasks, one m1.large is more interesting than four m1.t iny. The price per CPU is identical.

Technical difficulties

Only one limitation linked to the OpenStack cluster has been encountered. With the current quota it should have been possible to deploy a 32 nodes cluster. However when booting 32 VMs, only 24 were working correctly, the last 8 got the status `"ERROR"`. Screenshot

Conclusion

To conclude. we can see that MPI on the cloud is working well, however the communication overhead is more important than on a supercomputer and it's important to be aware of it. Furthermore the performance may be unstable, the instances may share their CPUs with other resource-consuming instances.

Appendices

A Results

	128	256	512	1024
1	0.0221	0.1784	1.6523	43.8849
2	0.0186	0.1260	0.9865	29.6860
4	0.2568	0.1083	0.8879	16.5539
8	0.2630	0.0826	0.5524	11.1385
16	0.5145	0.2203	1.1342	11.8693

Table 1: Computation time according to the number of nodes to the size of the matrix for VMs flavor 1

	128	256	512	1024
1	0.0221	0.1787	1.6726	71.6537
2	0.0203	0.1198	1.2978	31.3095
4	0.0187	0.0811	0.9964	17.7449
8	0.2711	0.0669	0.5174	8.9784
16	0.3011	0.2976	0.4117	4.9098

Table 2: Computation time according to the number of nodes to the size of the matrix for VMs flavor 2

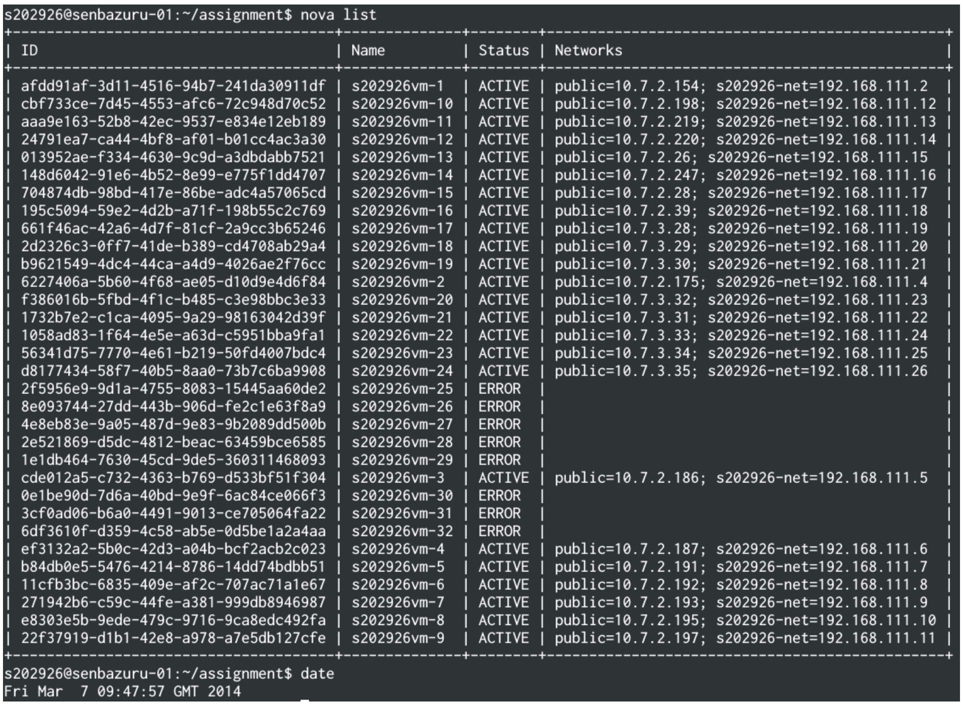
	128	256	512	1024	1360
1	0.0137	0.0937	0.8426	34.9952	21.3736
2	0.0169	0.0758	0.6042	11.7280	11.0668
4	0.2616	0.0698	0.4093	6.0268	6.4729
8	0.3310	0.0868	0.4060	6.0640	6.0757

Table 3: Computation time according to the number of nodes to the size of the matrix for VMs flavor 3

	128	256	512	1024
1	0.0223	0.1798	1.7618	41.7868
2	0.0212	0.1254	1.8515	34.9107
4	0.0213	0.1011	0.7157	18.3058

Table 4: Computation time according to the number of nodes to the size of the matrix for VMs flavor 4

B Errors screenshot



```
s202926@senbazuru-01:~/assignment$ nova list
+--------------------------------------+--------------+--------+-----------------------------------------------+
| ID                                   | Name         | Status | Networks                                      |
+--------------------------------------+--------------+--------+-----------------------------------------------+
| afdd91af-3d11-4516-94b7-241da30911df | s202926vm-1  | ACTIVE | public=10.7.2.154; s202926-net=192.168.111.2  |
| cbf733ce-7d45-4553-afc6-72c948d70c52 | s202926vm-10 | ACTIVE | public=10.7.2.198; s202926-net=192.168.111.12 |
| aaa9e163-52b8-42ec-9537-e834e12eb189 | s202926vm-11 | ACTIVE | public=10.7.2.219; s202926-net=192.168.111.13 |
| 24791ea7-ca44-4bf8-af01-b01cc4ac3a30 | s202926vm-12 | ACTIVE | public=10.7.2.220; s202926-net=192.168.111.14 |
| 013952ae-f334-4630-9c9d-a3dbdabb7521 | s202926vm-13 | ACTIVE | public=10.7.2.26; s202926-net=192.168.111.15  |
| 148d6042-91e6-4b52-8e99-e775f1dd4707 | s202926vm-14 | ACTIVE | public=10.7.2.247; s202926-net=192.168.111.16 |
| 704874db-98bd-417e-86be-adc4a57065cd | s202926vm-15 | ACTIVE | public=10.7.2.28; s202926-net=192.168.111.17  |
| 195c5094-59e2-4d2b-a71f-198b55c2c769 | s202926vm-16 | ACTIVE | public=10.7.2.39; s202926-net=192.168.111.18  |
| 661f46ac-42a6-4d7f-81cf-2a9cc3b65246 | s202926vm-17 | ACTIVE | public=10.7.3.28; s202926-net=192.168.111.19  |
| 2d2326c3-0ff7-41de-b389-cd4708ab29a4 | s202926vm-18 | ACTIVE | public=10.7.3.29; s202926-net=192.168.111.20  |
| b9621549-4dc4-44ca-a4d9-4026ae2f76cc | s202926vm-19 | ACTIVE | public=10.7.3.30; s202926-net=192.168.111.21  |
| 6227406a-5b60-4f68-ae05-d10d9e4d6f84 | s202926vm-2  | ACTIVE | public=10.7.2.175; s202926-net=192.168.111.4  |
| f386016b-5fbd-4f1c-b485-c3e98bbc3e33 | s202926vm-20 | ACTIVE | public=10.7.3.32; s202926-net=192.168.111.23  |
| 1732b7e2-c1ca-4095-9a29-98163042d39f | s202926vm-21 | ACTIVE | public=10.7.3.31; s202926-net=192.168.111.22  |
| 1058ad83-1f64-4e5e-a63d-c5951bba9fa1 | s202926vm-22 | ACTIVE | public=10.7.3.33; s202926-net=192.168.111.24  |
| 56341d75-7770-4e61-b219-50fd4007bdc4 | s202926vm-23 | ACTIVE | public=10.7.3.34; s202926-net=192.168.111.25  |
| d8177434-58f7-40b5-8aa0-73b7c6ba9908 | s202926vm-24 | ACTIVE | public=10.7.3.35; s202926-net=192.168.111.26  |
| 2f5956e9-9d1a-4755-8083-15445aa60de2 | s202926vm-25 | ERROR  |                                               |
| 8e093744-27dd-443b-906d-fe2c1e63f8a9 | s202926vm-26 | ERROR  |                                               |
| 4e8eb83e-9a05-487d-9e83-9b2089dd500b | s202926vm-27 | ERROR  |                                               |
| 2e521869-d5dc-4812-beac-63459bce6585 | s202926vm-28 | ERROR  |                                               |
| 1e1db464-7630-45cd-9de5-360311468093 | s202926vm-29 | ERROR  |                                               |
| cde012a5-c732-4363-b769-d533bf51f304 | s202926vm-3  | ACTIVE | public=10.7.2.186; s202926-net=192.168.111.5  |
| 0e1be90d-7d6a-40bd-9e9f-6ac84ce066f3 | s202926vm-30 | ERROR  |                                               |
| 3cf0ad06-b6a0-4491-9013-ce705064fa22 | s202926vm-31 | ERROR  |                                               |
| 6df3610f-d359-4c58-ab5e-0d5be1a2a4aa | s202926vm-32 | ERROR  |                                               |
| ef3132a2-5b0c-42d3-a04b-bcf2acb2c023 | s202926vm-4  | ACTIVE | public=10.7.2.187; s202926-net=192.168.111.6  |
| b84db0e5-5476-4214-8786-14dd74bdbb51 | s202926vm-5  | ACTIVE | public=10.7.2.191; s202926-net=192.168.111.7  |
| 11cfb3bc-6835-409e-af2c-707ac71a1e67 | s202926vm-6  | ACTIVE | public=10.7.2.192; s202926-net=192.168.111.8  |
| 271942b6-c59c-44fe-a381-999db8946987 | s202926vm-7  | ACTIVE | public=10.7.2.193; s202926-net=192.168.111.9  |
| e8303e5b-9ede-479c-9716-9ca8edc492fa | s202926vm-8  | ACTIVE | public=10.7.2.195; s202926-net=192.168.111.10 |
| 22f37919-d1b1-42e8-a978-a7e5db127cfe | s202926vm-9  | ACTIVE | public=10.7.2.197; s202926-net=192.168.111.11 |
+--------------------------------------+--------------+--------+-----------------------------------------------+
s202926@senbazuru-01:~/assignment$ date
Fri Mar  7 09:47:57 GMT 2014
```