Summary of DA-PWC Performance

# Data

Data for this analysis were taken from Region 5 (41870 points total) of the Fungi sequence classification at http://salsahpc.indiana.edu/millionseq/fungi2/fungi2\_index.html

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | # Points | # Clusters | Description |
| A | 12579 | 4 | Region 5 sub cluster 2 (of 10) split into 4 (# clusters) |
| B | 20000 | 10 | Region 5 first 20k points split into 10 clusters |
| C | 40000 | 10 | Region 5 first 40k points split into 10 clusters |

# Parallel Libraries

|  |  |  |
| --- | --- | --- |
| Dataset | C# | Java |
| MPI | Tasks | MPI | Tasks |
| A | MPI.NET v1.0.0 | Microsoft Task Parallel Library .NET v4.0 | OpenMPI v1.7.5rc5 | Habanero Java (HJ-lib) v0.1 |
| B | MPI.NET v1.0.0 | Microsoft Task Parallel Library .NET v4.0 | OpenMPI v1.8.1 | Habanero Java (HJ-lib) v0.1.1 |
| C | MPI.NET v1.0.0 | Microsoft Task Parallel Library .NET v4.0 | OpenMPI v1.8.1 | Habanero Java (HJ-lib) v0.1.1 |

# Parallelisms

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Total | T | P | N | Total | T | P | N | Total | T | P | N |
|  |  |  |

Where,

T - # of concurrent tasks per MPI process

P - # of MPI processes per node

N - # of nodes

|  |  |
| --- | --- |
| Dataset | Parallelisms Tested |
| A | 1,2,4,8,32,64,128,256 |
| B | 64,128,256 |
| C | 64,128,256 |

# System Information

|  |  |
| --- | --- |
| System | Description |
| Tempest | 32 nodes, each has 4 Intel Xeon E7450 CPUs at 2.40GHz with 6 cores, totaling 24 cores per node; 48 GB node memory and 20Gbps Infiniband (IB) network connection. It runs Windows Server 2008 R2 HPC Edition – version 6.1 (Build 7601: Service Pack 1). |
| FutureGrid - India | 128 nodes, each has 2 Intel Xeon X5550 CPUs at 2.66GHz with 4 cores, totaling 8 cores per node; 24GB node memory and 20Gbps IB network connection. It runs Red Hat Enterprise Linux Server release 5.10. |

# Performance Results

## Dataset A – 12579 Points

Note, the TxP patterns that gave the best performance for a given number of node is highlighted in blue rectangle.

The highlighted section of above graphs is presented below.

The above graphs shows 4 different intra-node parallelism patterns – 1x8, 2x4, 4x2, and 8x1. Note how 8x1 performs poorly compared to the other three in OMPI. This happens due to the topology of FutureGrid machines where a socket contains only 4 cores. Therefore running 1 MPI with more than 4 tasks doesn’t yield good performance.

The best pattern for OMPI happens to be 1x8 – highlighted in green.

A different representation of the above graph shows the effect of TxP patterns more clearly.

## Dataset B – 20k Points

// TODO

## Dataset C – 40k Points

Note how 1x8 keeps the best performance still for this data