Ship Airwake Studies

One of the challenges that helicopter pilots face when landing on a ship is the air turbulence around the ship's superstructure. Professor Gareth Padfield at the University of Liverpool is leading a team of researchers to model this turbulence using the Fluent computational package in order to understand the phenomenon better and to provide a better representation of the turbulence in helicopter training simulators.

Figure 1 - Turbulence issues away from the landing area are understood well.

The detailed modelling of airflows around a ship is computationally demanding and produces large volumes of data. To make this modelling feasible for Prof. Padfield's team, parallelism at several levels must be exploited. Firstly, the Fluent package itself is parallelised, so that parallelism within a model can be exploited. This is essential as each simulation performs 1500 time steps, representing about 15 seconds of simulation time and takes roughly 4500 minutes of real time (just over 3 days) on 32 cores - 8 nodes to complete. The second level of parallelism arises because for a given model, the simulation is repeated to cover 28 different wind angles. Given sufficient computing capacity, multiple wind angles can be studied simultaneously. Approximately 300 GBytes of data are generated for each simulation, so naively storing the raw output requires nearly 9 TBytes of storage.

The NW-GRID, with over 1500 cores on 4 sites, has been approached to provide an effective and efficient solution to resolve the computational challenges of Prof. Padfield's research team. The challenge for the NW-GRID is to deal with the data volume effectively and to reduce the throughput time enabling the scientists to work more efficiently.

Figure 2 - Turbulence near the landing area is less understood and can pose considerable dangers

Data requirements

While the data requirements for a set of simulations are large, the data are not of uniform importance.

Input Data: About 1 GBytes is taken up with input files - these are copied over from an Engineering system and are static during a run

Restart Data: About 6 GBytes of data are generated every few hours to provide a restart basis for the code in the event of a system crash. Copies of this data are useful while the job is running but can be purged upon job completion.

Raw Data: As mentioned earlier, 15 seconds of simulated time generates about 300 GBytes of output, at the rate of 200 MBytes of data every time step. These files are simple text files.

Compressed Raw Data: Once timestep files are created, they are compressed (and then uncompressed as required in the post-processing stage). The compressed raw data needs to be archived in case the post-processed data needs to be recomputed for a different structured mesh (see below) and also for audit purposes. Compression reduces the data volume by a factor of about 2.5.

Post-processed Data: About 50 GBytes of data is generated at the end of each run by interpolating the raw data from a simulation onto a coarse, structured mesh. This data is downloaded onto data storage drives within Engineering and is the essential simulation output. Maintaining a backup copy of this data is highly desirable (so about 1.4 TBytes in total, but built up over time - at most 100 GBytes every 3 days assuming two simulations are run at once).

For the restart data, a separate background job runs every 12 hours to copy the latest restart data, so the restart data files can be overwritten.

Another background script is run hourly to copy over the raw files and to compress them. Once the compressed files have been backed up, they too are deleted, although a formal archive process would be desirable.

If there is sufficient disk space available, the $28 \times (120+50) = 4.8$ TBytes of results data can simply be kept on disk, protected by a backup.

Improving throughput time

With its novel approach of science testbeds, NW-GRID is able to allocate sufficient resources to Professor Padfield's team to complete 28 runs in 6 days, while still meeting NW-GRID commitments to other projects.

Future Plans

A simulated 15 seconds is really the minimum suitable time – it just covers the period as the helicopter is coming down to land. A simulated two minutes (i.e. 8 times longer) allows the full final approach sequence to be modelled. However, this has important implications. Firstly, there is an 8-fold increase in storage requirements – from 5 TBytes storage to just under 40 TBytes of storage. A refinement in what raw data is kept and for how long is necessary. The essential data output would be about 11 TBytes in size, which provides a minimum storage requirement.

Another desire is to reduce the time for a simulation to one day. This requires a 24-fold increase in parallel performance, so that something like 800 cores would be involved in each simulation. This might be possible running Fluent on the Daresbury Blue Gene system (where 4096 cores are available). Feasibility tests using Blue Gene and the 400 cores of Liverpool's NW-GRID High Capability cluster will be conducted in the second half of 2008.