Job with input and output data

DIRAC Tutorial

Sandboxes

- Sandboxes are managed by a dedicated DIRAC service
  - Not meant to store large data files
  - The total size of a sandbox should not exceed 10 MB
  - Oversized input sandboxes will be rejected
  - Oversized output sandboxes will be stored as Output Data transparently for the user
- JDL example

```
…
InputSandbox = {“mandelbrot”,”config.ini”};
OutputSandbox = {“std.out”,”out.bmp”};
…
```

Sandboxes

- Sandboxes can contain LFNs:

```
…
InputSandbox = {“mandelbrot”,”lib”,
 “LFN:/vo.formation.idgrilles.fr/user/u/usera/data.file”};
OutputSandbox = {“*.bmp”};
…
```

  - In this case the file is not moved while the job submission
- Sandboxes can contain directories
  - The whole directory will be copied to the worker node
    - *lib* will be added to LD_LIBRARY_PATH
- Sandboxes can contain wild cards

Input Data

- Input Data is specified by a list of LFNs

```
…
InputData = {
  “/vo.formation.idgrilles.fr/user/u/usera/data.file1’,
  “/vo.formation.idgrilles.fr/user/u/usera/data.file2’
};
…
```

  - No LFN: prefix needed
  - The data files will be downloaded to the worker nodes before starting the user application

Output Data

- Output Data is specified by a list of files

```
…
OutputData = { “out1.bmp”,”out2.bmp” };
OutputSE = “IN2P3-disk”;
OutputPath = “/mandelbrot1”;
…
```

  - By default the data is stored in directory by JobID
    - /vo.../user/u/usera/12345/12345678/out1.bmp
  - OutputPath is appended to the usual user LFN prefix:
    - /vo.../user/u/usera/mandelbrot1/31234/31234555/out1.bmp
  - The Output Data LFNs can be found in the job parameters

Sandbox vs Data

- The location of the Input Data is taken into account when scheduling jobs
  - Job go to the data
- The location of the LFNs provided in the Input Sandbox is not used in the job scheduling
  - Allows to send job anywhere
  - Can trigger large data transfers over the network

Tutorial

Submit mandelbrot jobs with the output images stored as Output Data

choose OutputSE and OutputPath

Download files to the local machine

Launch jobs with input specified as InputSandbox and as InputData