Experimenting with PyWPS and CUDA

Marco Grimaldi
Calogero Zarbo
Davide Albanese
Gabriele Franch
Envirochange: The Big Goal

Provide **tools/information** to understand the vulnerability to **climate change** of the agricultural environment in **Trentino** (Italian Alps) - 157.9 km² (61 sq mi)

Two kind of data:
- Land *(usage)* data
- Weather data
Weather Data

- Continuous surface for T, P (daily and hourly) 2001-2008 (9800/year)

Meteotrentino:
- More than 50 weather stations (PAT)
  - T Spatial interpolation
  - P Spatial interpolation

- Resolution 100/200m
Land Data

- Digital Terrain Model: 1m - 200m
- Orthophoto 0.5 m
- Administrative spatial data (province, municipalities, etc)
- Land use data

**Agricultural data:**
- Cadastral vineyards ~ 2000 polygons (avg 5 ha)
- Pathogen validation/sampling points
Processing DATA

• Geospatial Processing Services:
  OpenGIS® Web Processing Service (WPS) Interface Standard
    • 52North
    • pyWPS

• High Performance Computing:
  • Computer CLUSTERS
  • GPGPU
GPGPU stands for General-Purpose computation on Graphics Processing Units. Graphics Processing Units (GPUs) are high-performance many-core processors that can be used to accelerate a wide range of applications.

[http://gpgpu.org/]

Different Implementations:

- CUDA (NVIDIA)
- Stream Computing (ATI)
- OpenCL (Open Standard - Khronos)
CUDA

**Compute Unified Device Architecture** is a parallel computing architecture developed by **NVIDIA**. CUDA gives developers access to the **native instruction set** and **memory** of the parallel computational elements in CUDA GPUs. Programmers use 'C for CUDA' (**C with NVIDIA extensions**), to code algorithms for execution on the GPU.

**Plus:**
- Scattered reads
- Shared memory
- More mature than other APIs (e.g.: OPENCL)
- Third party wrappers are available (e.g.: Python, Fortran, Java, ...)

**Cons:**
- Limited support for double precision floats
- Bus bandwidth and latency between the CPU and the GPU
- Threads should be running in groups of at least 32
The Huglin Index: a case study

This index is used to classify viticultural regions in terms of the sum of temperatures required for vine development and grape ripening, (Huglin, 1978). Specifically, it is the sum of mean and maximum temperatures above +10°C – the thermal threshold for vine development.

Different grape varieties are classified according to their minimal thermal requirement for grape ripening.

For example, the HI is 1700 for Chardonnay and 2100 for Syrah. The minimal Huglin index for vine development is 1600.

\[ HI(t) = [(T_{MAX}(t) - 10) + (T_{AVG}(t) - 10)] \cdot k + HI(t - 1) \]
Algorithm Implementation

- The formula for HI has been implemented in **C-CUDA** (Python/C wrapper) and the WPS Implementation of choice is **pyWPS**

```c
static __global__ void huglinK (float* A, float* B, float* C, float* O) {
    ...
}
```

```c
int huglin (float* A, float* B, float* C, float* O, int height, int width, float flag_value) {
    ...
}
```

```python
static PyObject *pycuHuglin_huglin(PyObject *self, PyObject *args, PyObject *keywds) {
    ...
}
```

- Results are compared with
  - **Python/NUMPY** – also running through pyWPS
  - **Java** implementation running on **52North**

- GeoTIFF are stored locally – no GeoServer (ATM)
Results

**Mean Computing Time Huglin Index (365 GeoTIFFs)**

- **Computation only**
- **Python/Numpy – XEON 2.53GHz**
- **Python/CUDA – QuadroFX 3800**
- **Java – XEON 2.67GHz**

**Mean Execution Time WPS Process (365 GeoTIFFs)**

- **Computation and I/O**
- **Python/Numpy – XEON 2.53GHz**
- **Python/CUDA – QuadroFX 3800**
- **Java – XEON 2.67GHz**
Conclusions

- In terms of computing time of the Huglin index, CUDA shows better performance than the Numpy implementation and the Java implementation.
  - Although HI not particularly computational expensive

- In terms of execution time of the “whole” WPS process, the I/O time condition the general efficiency: both CUDA and Numpy show very similar performance, while our Java implementation lags behind.

- In this study case, the real bottleneck is the I/O data access. This has to be taken into account if SPEED is a requirement of the application/WPS service.