Customized Pipelines for the Processing of Medical Images in Large Population Studies

Henri A. Vrooman
Biomedical Imaging Group Rotterdam (BIGH)
Radiology & Medical Informatics, Erasmus MC
Biomedical Imaging Group Rotterdam

Development of robust, accurate and fully automated tools to perform measurements in MR Brain images
Two focus points:

1. Population Imaging → Understanding disease

2. Computer Assisted Diagnosis → EARLIER and more accurate diagnosis!
Rotterdam Study: Neurologic disease

**Risk factors:**
- Genetic
- Non-genetic
  - Blood pressure
  - Cholesterol
  - Diabetes mellitus
  - Homocysteine
  - Smoking

**Brain changes:**
- Atrophy
- Infarcts
- White matter lesions
- Microbleeds

**Outcome:**
- Stroke
- Cognitive Decline
- Dementia
- Depression

Images of brain scans showing various changes and regions.
Alzheimer’s disease

- Alzheimer's disease (AD), the most common form of dementia, is one of the most devastating healthcare problems facing western society.
- In Europe, the yearly costs associated with AD is estimated to be around 400 billion Euros, exceeding the costs of heart disease, cancer, and diabetes.
- In NL, currently 1 in 60 people with AD; this will double in 2050!
Socioeconomic impact of early diagnosis and preventive strategies

- In Alzheimer’s disease, there are currently 98 molecules in pre-clinical research, and 50 phase 1 trials underway.

- Current preventive therapy can delay the onset of AD symptoms (up to 3.5 years has been reported). The latter would reduce AD prevalence by 1/3.

EARLY DIAGNOSIS
Critical question to be answered

Will the owner of this brain become demented?
Relating diagnosis to reference values
Rotterdam Scan Study (> 8000 participants)

- Tissue quantification
- Lesion assessment
- Microstructural integrity
- Incidental brain findings
- Blood flow quantification
- Cerebral microbleeds

HPC Cloud Day, H.A. Vrooman, BIGR
Generating imaging biomarkers

Biomarkers

0.23
183 cc
17.3% Lesion load
83.12% Connectivity
Thalamus-Cortex
Two frontal lobes

HPC Cloud Day, H.A. Vrooman, BIGR
Bottom line

We need to scan thousands of subjects and analyze the scans!

Population Imaging
Current Problems in Population Imaging

- Initiatives are stuck, scaling-up is hampered
- Every institution has its own data
- Every institution has its own hardware/software infrastructure
- Data formats, scan protocols, processing and validation are not standardized!
- In practice, it’s difficult to share and combine input data and numerical results
- Transfer of very large data sets is tedious
- Closed and protected hospital networks
- No high-quality IT infrastructure present
- Privacy issues
Centralized and standardized archival and processing on the HPC cloud with remote viewing, monitoring and exploration of the results
Benefits of Cloud Computing

- Create and deploy specialized Virtual Nodes on demand
  - Hardware: Memory, Cores (Multi-threading, MPI)
  - Software: OS (Linux, Windows, …), Tools
    (expertise from the pilot in 2011)

- Clone these Virtual Nodes as needed depending on IP task
  - Hybrid Cluster: Use different, tool-specific (node) images for multi-step processing (pipelines)

Implement and deploy cluster nodes lean and mean:
Increases efficiency & flexibility!
Hybrid Cluster Computing

HPC Cloud Day, H.A. Vrooman, BIGR
Benefits of Cloud Infrastructure

- System Memory Size
  - Very large system memory sizes possible
  - Efficient memory & CPU usage

- System Network Infrastructure
  - Advanced Integration of Storage & Computation

To be able to run a variety of processing pipelines in the same ‘hardware/software’ environment means less data transfer, efficient use of resources and in the end faster and more efficient processing.
EYR: Centralized Archival and Processing

Delft
- Remote Viewing

SARA
- Cloud Infrastructure
- Hybrid Cluster
- Image Server
- Viewing Server
- Storage

Leiden
- Workstation
- Scanner

Rotterdam
- Workstation
- Scanner

Image Processing Unit
Enlighten your Research 3
E.D. Peters 10-sep-2011
EYR: Light path Rotterdam, Leiden, Delft → SARA (secure & low latency)
Concluding

How are we going to use the HPC Cloud Infrastructure?
- Creating a Centralized Image Processing Unit (CIPU)

Why do we want to use the HPC Cloud Infrastructure?
- Flexibility
- Efficient use of resources
- Deploy specific nodes based on requests

How important is the HPC Cloud Infrastructure for ‘our’ project?
- A nice pilot to investigate centralized facilities for Population Imaging.
- Trying to have a break-through in scaling up Population Imaging!