UiO AI hub-node project

Kickoff – Welcome, Goals, Status

November 6th 2018, USIT

Goals of the kickoff

- Define tasks & milestones for the next 6-12 months by
  - revisiting goals of the hub-node project,
  - considering results of the AI workshop in September and the NORFAML proposal.

1. Short overview of AI hub-node, workshop and NORFAML
2. Discussion of tasks & milestones based on use cases & competences
3. Organisation

AI hub-node as it was proposed (April '18)

- The goals of this project are to provide a common platform and service exploiting AI / ML / DL in such scenarios by:
  - providing necessary HW resources in-house preinstalled with state-of-the-art software stacks targeting researchers and students
  - building up competence in using ready-to-use cloud services from external providers (such as Google Edu, IBM IAI, MS Azure, Amazon Web Services, etc)
  - building up competence in advising researchers and students in the use of AI / ML / DL methods and frameworks running on internal or external resources

AI vs HPC

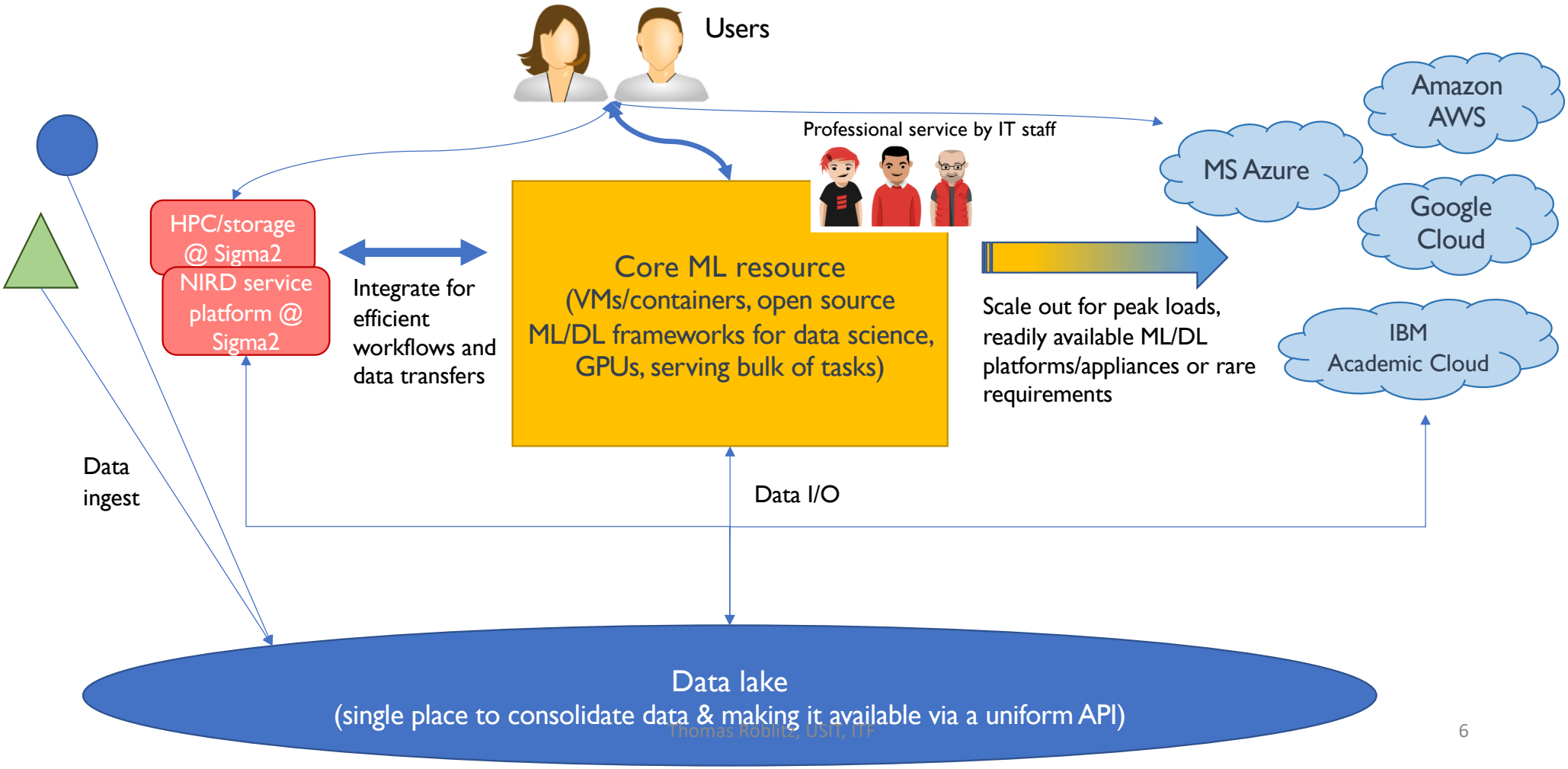
- IT hardware/services, AI frameworks + competence
  - virtually no shared HW available (yet)
  - virtually no software frameworks installed (yet)
  - no access to external Cloud services (yet)
  - very limited competence at ITF
- Compare this to parallel programming/computing (HPC)
  - shared HW available (Abel, Fram, …)
  - all needed SW installed (fabric comm., MPI, OpenMP + applications)
  - strong competence in operations, basic and advanced user support, collaboration in EU projects (PRACE), working on emerging technology

Vision

- Provide IT infrastructure for AI with same maturity as HPC

  - Significant and modern HW base.
  - Flexibility to run any software framework easily.
  - Access to external Cloud services.
  - Strong competence to support users in any AI stage.

Architecture draft



Steps towards vision (1/2)

- AI hub-node project
  - Focus at needs at UiO.
  - Small-scale (HW base, personnel effort).
  - 2-3 year project.
  - Will naturally have a piloting character, still there will be important results.
- NORFAML proposal (Norwegian Facility for Applied Machine Learning)
  - National scope.
  - Medium-scale (HW base, personnel effort).
  - 4 year project (+ 6 years operations).
  - Will establish a production facility similar* to existing HPC facilities.

(*) shared HW, common software stack, basic and advanced user support

Steps towards vision (2/2)

- Forum/arena/network of experts/…
  - Jointly explore this "new" area of service.
  - Bring in external expertise.
  - Start new collaborations.
  - Get inspired.
  - Event series to regularly meet and exchange ideas, learn something new, …
  - 1st event (AI lounge) scheduled for November 20th (Status of AI activities at UiO)
  - Lots of concurrent activities. Collect information, so everyone is aware & can participate …
    - Workshops/Seminars/…: June 14 (library), Sep 7 (USIT+UB), StudyGroup @ UiO, Makerspace @ OsloMet, Bergen, Geilo Winter School (2017, 2018, 2019)
    - Conferences: AIM2North @ OsloMet, Fantastic Futures @ Nat'l Lib.
    - Research Bazaar
    - ...

Constraints of the AI hub-node

- Limited budget for HW → won't satisfy all needs
- Limited budget for people → can only cover competence partially
- Focus on the IT service

→ important to prioritize tasks & milestones

Results workshop September 7th

- AI-based data-driven science, some 40-50 participants, 20 talks
- Diverse HW needs
  - GPUs, CPUs, memory, fast storage
  - hard to quantify ➔ rather aim at cost-efficiency than highest performance
- Diverse support needs
  - HW + maintained SW
  - project support (how to apply ML)
  - introductory courses
- Two modes of use
  - interactively developing models
  - deploying models (in a service or analyzing data)

Results workshop September 7th (cont.)

- Large interest in collaboration user – IT
- Difficult to find out …
  - What resources exist?
  - Who else works with AI?
  - What other activities, events happen?

Proposed tasks & milestones

- Initial procurement of HW
- Work with use cases
- Start event series, offer basic courses

Initial procurement of HW/access to Clouds

- Status hardware
  - Abel: old HW (> 6 years), old OS, some nodes with 1TiB memory, some with GPUs
  - Fram: new HW, new OS, some with 1 TiB memory, no GPUs
  - UH-IaaS: any OS, medium memory, no GPUs *(yet)*
  - apps.sigma2.no: standard "apps" for ML, 128 GB memory, some GPUs (V100)
- Coming @ UiO
  - 3 servers with 128 GiB RAM, 28 cores, 4 NVIDIA RTX 2080 Ti each (about 1/3 of HW budget spent)
  - More HW options in 2018/2019
  - Operations of these servers to be defined: possibly interactive access during the day, batch work in off-working hours.
- Tasks & milestones
  - procure and setup 3 servers (before end of 2018)
  - procure and setup 1-2 HW alternatives {Huawei, AMD,…} (June 2019)
  - establish access to UH-IaaS and apps.sigma2.no (March 2019)
  - establish access to external Clouds (September 2019)

Work with use cases

- Goals
  - Explore HW/SW alternatives
  - Gain experience on basic AI tasks
- Approach
  - Avoid high effort in getting the basics (e.g., TF on Abel)
  - Short (~ 3 months) focused (specific AI task) co-working projects user & IT staff
  - Possibly try implementations on two different platforms (HW1 vs HW2, HW vs Cloud)
- Tasks & milestones
  - Joint projects (continuous activity, first results March 2019)
  - Best practice guides (whenever we have successfully transferred knowledge to 2nd, 3rd project)

Start event series … *AI lounge*

- Once a month, 1-2 hours, NHA: Realfagsbiblotekets undervisningsrom 209
- "Casual atmosphere"
  - talks, discussions, hands-on on a specific topic (AI methods, research, tech, SW, …)
  - anecdotes from participants: What did work? What did not work?
  - announcements
  - results of this or any project
  - external speakers
- November 20th: 14:15 – 15:30, Status of AI activities at UiO (resources, projects, activities + collect info from participants)
- December 11th: 14:15 – 15:30, Topic to be defined
- Courses: could be integrated into AI lounge and/or HPC course week

Vision revisited (in one year)

- Provide IT infrastructure for AI with same maturity as HPC
  - Significant and modern HW base.
  - Flexibility to run ~~any~~ different software framework easily.
  - Access to one external Cloud services.
  - Strong competence to support users in ~~any~~ specific AI stage.
- Established group of "enthusiasts" which run the AI lounge independently.
- Collect and publish information on activities regularly.

Revisit plan for today

- Next
  - Use cases & competences
- Then
  - Organisation, What can we learn from others?, Best working mode? Establishing a forum? Existing hardware and gaps? Legal issues?
  - Wrapping up, follow-up meeting
- Proposed tasks & milestones
  - Initial procurement of HW
  - Work with use cases
  - Start event series, offer basic courses

UiO AI hub-node project

Kickoff – Use cases & competences

November 6th 2018, USIT

Overview competences

- **Use case:** implications of applying AI; ownership of results; regulatory issues; define specific, relevant, attainable, measurable and timely goals
- **Data organisation:** acquire, explore, visualize, clean, enrich, filter, format, normalize, structure, integrate, curate
- **Model learning:** be aware of data scarcity, quality, dimensionality; feature engineering; model design, implementation, training, validation, tuning; manage large parameter spaces; be prepared to explain results
- **Deployment:** integration of AI-based functionality into services; monitor results and compare with goals set by use case

Use case

- implications of applying AI
- ownership of results
- regulatory issues
- define specific, relevant, attainable, measurable and timely goals

Data organisation

- Acquire
- Explore
- Visualize
- Clean
- Enrich
- Filter
- Format
- Normalize
- Structure
- Integrate
- Curate

Model learning

- be aware of data scarcity, quality, dimensionality
- feature engineering
- model design, implementation, training, validation, tuning
- manage large parameter spaces
- be prepared to explain results

Deployment

- integration of AI-based functionality into services
- monitor results and compare with goals set by use case

UiO AI hub-node project

Kickoff – Organisation

November 6th 2018, USIT

Overview

- What can we learn from others (providing assistance in ML, structured way in providing assistance)?
- What is the best working mode (continuous joint work ("sitting together"); just a meeting at start, then separately)?
- For establishing a forum: How do we involve all interested?
- Existing IT hardware and gaps?
- Awareness of legal issues (GDPR)?