EVEREST
DESIGN ENVIRONMENT
FOR EXTREME-SCALE BIG DATA ANALYTICS
ON HETEROGENEOUS PLATFORMS

Computing Frontiers 2022, BigDaW Session

The EVEREST SDK

Mapping big-data applications onto heterogeneous reconfigurable computing systems

---

CHRISTOPH HAGLEITNER

IBM Research, Zurich; EVEREST Project Coordinator

hle@zurich.ibm.com

Introduction

EVEREST dEsign enVironmEnt foR Extreme-Scale big data analyTics on heterogeneous platforms

SDK System Development Kit = Tools for
- Application description
- Deployment on Target System
- Compilation
- Runtime environment
- Data management and security

EVEREST Consortium

IBM Reseach Lab, Zurich (Switzerland)
Project Administration, Prototype of the target system

Università della Svizzera italiana (Switzerland)
Data security requirements and protection techniques

Centro Internazionale di Monitoraggio Ambientale (Italy) Weather prediction models

Virtual Open Systems (France)
Virtualization techniques, runtime extensions to manage heterogeneous resources

Numtech (France)
Application for monitoring the air quality of industrial sites

Politecnico di Milano (Italy)
Project Administration, High-Level System, Flexbile Memory Manager, Autotuning

TU Dresden (Germany)
Domain-specific extensions, code optimizations and variants

IT4Innovations (Czech Republic)
Exploitation leaders, Large HPC infrastructure, Workflow libraries

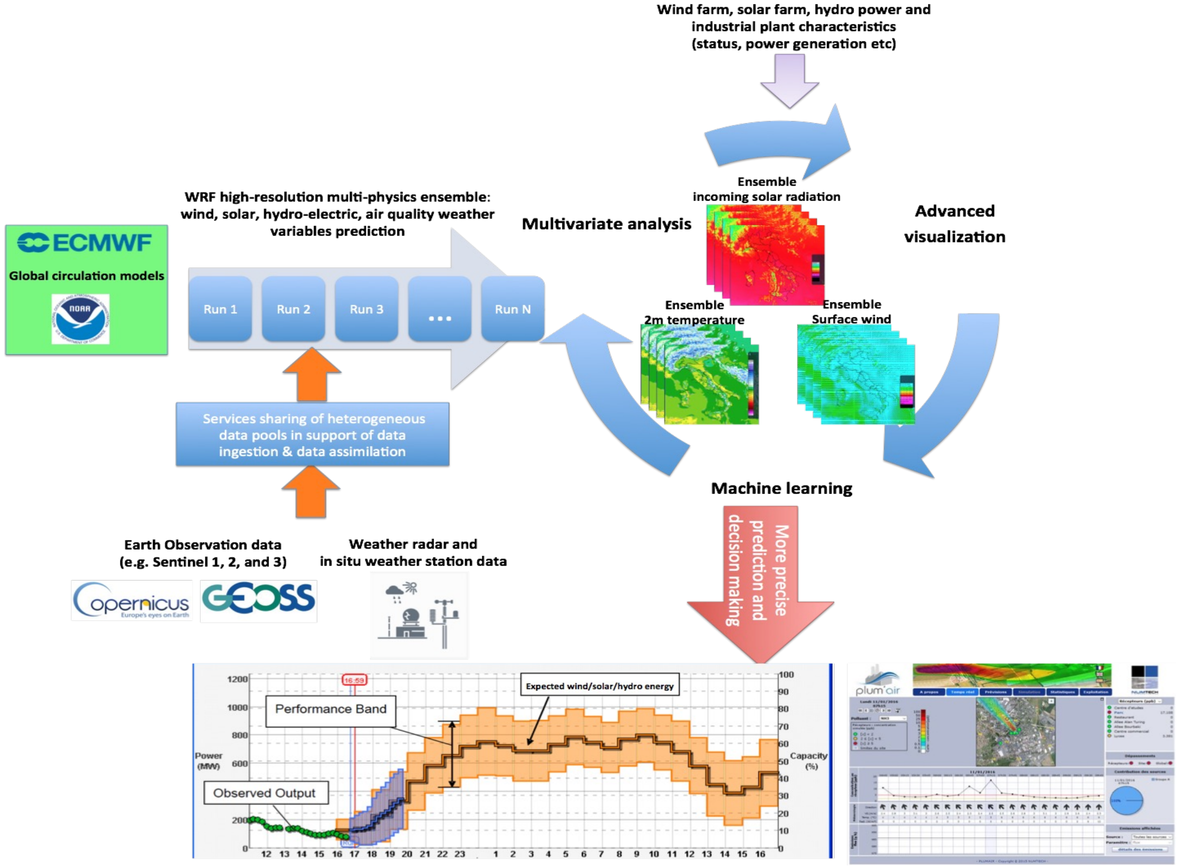
Duferco Energia (Italy)
Application for prediction of renewable energies

Sygic A/S (Slovakia)
Application for intelligent transportation

EVEREST Use Cases

Air-quality modelling in industrial sites

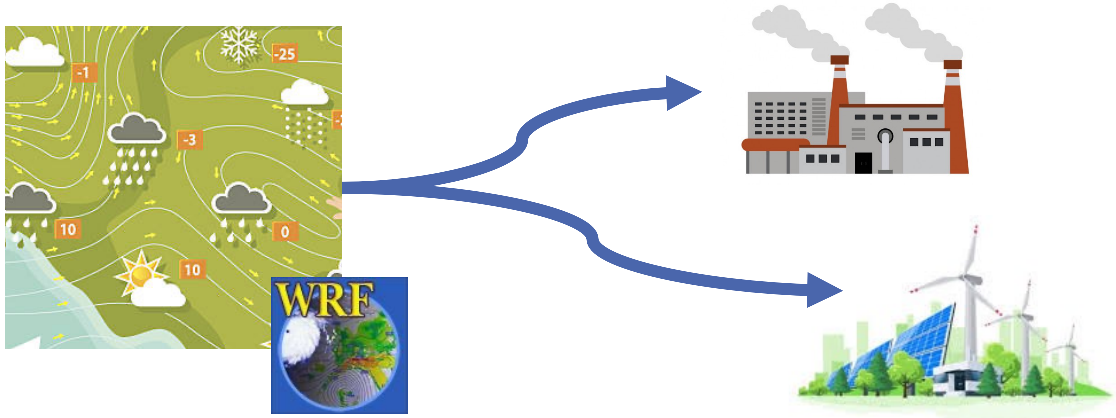
Weather-based prediction of renewable energy production

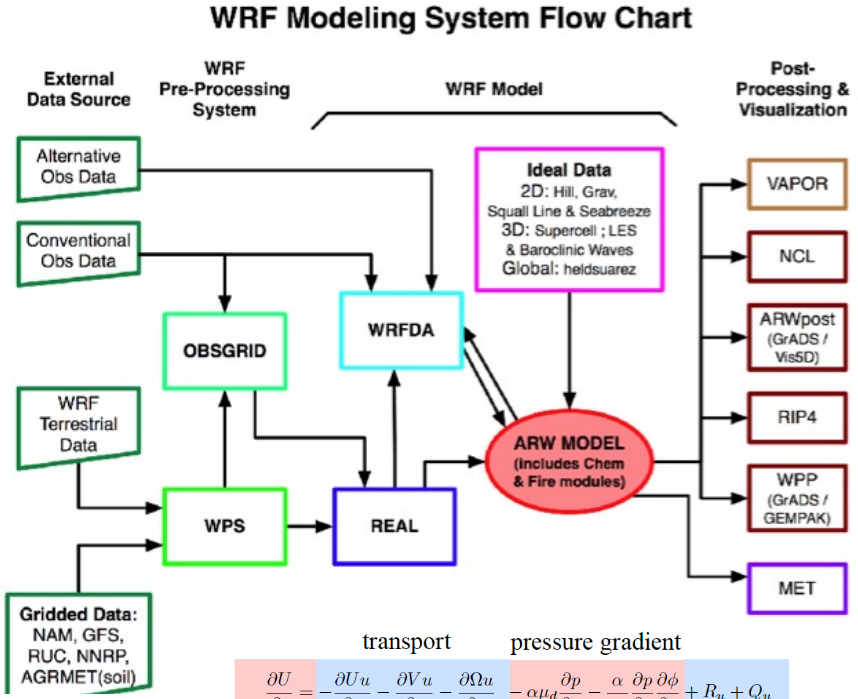


Traffic modelling for intelligent transportation

The WRF Model

First step of two use cases…





transport — pressure gradient

$$\frac{\partial U}{\partial t} = -\frac{\partial U u}{\partial x} - \frac{\partial V u}{\partial y} - \frac{\partial \Omega u}{\partial \eta} - \alpha\mu_d \frac{\partial p}{\partial x} - \frac{\alpha}{\alpha_d}\frac{\partial p}{\partial \eta}\frac{\partial \phi}{\partial x} + R_u + Q_u$$

$$\frac{\partial V}{\partial t} = -\frac{\partial U v}{\partial x} - \frac{\partial V v}{\partial y} - \frac{\partial \Omega v}{\partial \eta} - \alpha\mu_d \frac{\partial p}{\partial y} - \frac{\alpha}{\alpha_d}\frac{\partial p}{\partial \eta}\frac{\partial \phi}{\partial y} + R_v + Q_v$$

$$\frac{\partial W}{\partial t} = -\frac{\partial U w}{\partial x} - \frac{\partial V w}{\partial y} - \frac{\partial \Omega w}{\partial \eta} - g\left(\mu_d - \frac{\alpha}{\alpha_d}\frac{\partial p}{\partial \eta}\right) + R_w + Q_w$$

$$\frac{\partial \mu_d}{\partial t} = -\frac{\partial U}{\partial x} - \frac{\partial V}{\partial y} - \frac{\partial \Omega}{\partial \eta}$$

$$\frac{\partial \Theta}{\partial t} = -\frac{\partial U \theta}{\partial x} - \frac{\partial V \theta}{\partial y} - \frac{\partial \Omega \theta}{\partial \eta} + R_\theta + Q_\theta$$

$$\frac{\partial \mu_d q_j}{\partial t} = -\frac{\partial U q_j}{\partial x} - \frac{\partial V q_j}{\partial y} - \frac{\partial \Omega q_j}{\partial \eta} + R_{q_j} + Q_{q_j}$$

$$\frac{\partial \phi}{\partial t} = -u\frac{\partial \phi}{\partial x} - v\frac{\partial \phi}{\partial y} - \omega\frac{\partial \phi}{\partial \eta} + gw$$

↑ ← numerical filters, physics, projection terms

← geopotential eqn term

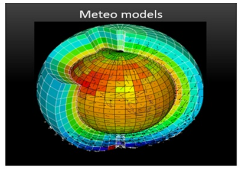
Diagnostic relations: $\frac{\partial \phi}{\partial \eta} = -\alpha_d \mu_d, p = \left(\frac{R_d \Theta_m}{p_o \mu_d \alpha_d}\right)^{\gamma}, \Theta_m = \Theta\left(1 + \frac{R_v}{R_d} q_v\right)$

WRF is an open-source model supported primarily by the US National Center for Atmospheric Research (NCAR), the US National Oceanic and Atmospheric Administration and the US National Center for Environmental Prediction – NCEP

Air-quality use case: Workflows and Challenges

1. WRF Deterministic weather forecast

Global observation

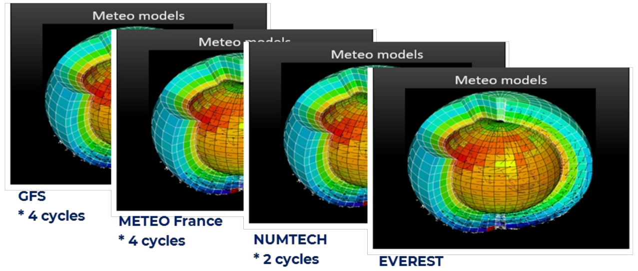


WRF execution

Improve speed to produce forecast

2. Ensemble prediction

N x deterministic weather forecast



\+

Local weather observation on-site

One aggregated weather forecast forced by observation

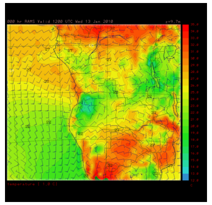
Improve quality of local weather forecast

3. Air-quality dispersion forecast

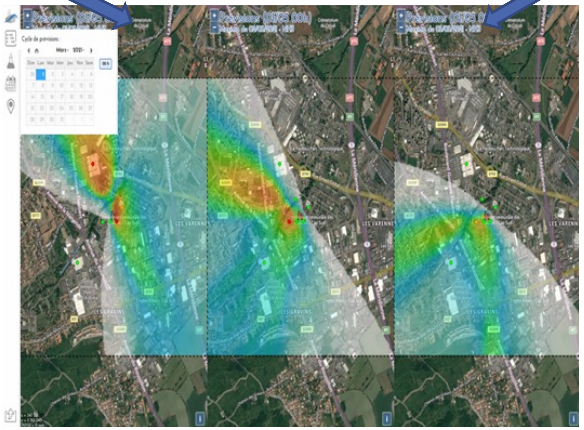
Landuse, Topography

Industrial site data

Emission forecast



Local weather forecast



Air-quality forecast

Improve speed to produce air-quality forecast and its quality

EVEREST

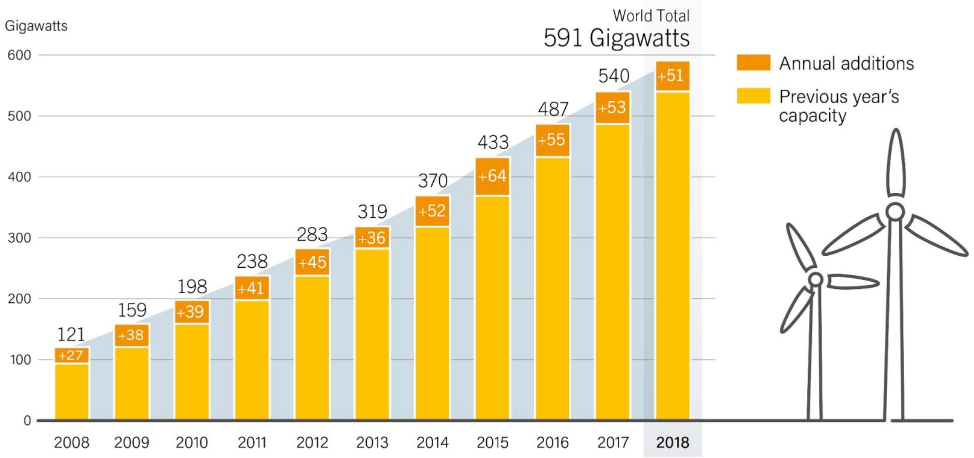
Renewable Energy use case: Context and Challenge

Different challenges due to intermittency of the wind power generation:

- Transmission System Operator (in Italy TERNA) to ensure the balance of grid (very short term horizon: 1s to 1h)
- Traders to forecast the power to sell on energy market, intraday or day ahead (short term horizon: 1h to 24h)
- Wind farm owners to schedule their maintenance programs (long term horizon)

→ **great value of improved wind power forecast accuracy**

Wind Power Global Capacity and Annual Additions, 2008-2018



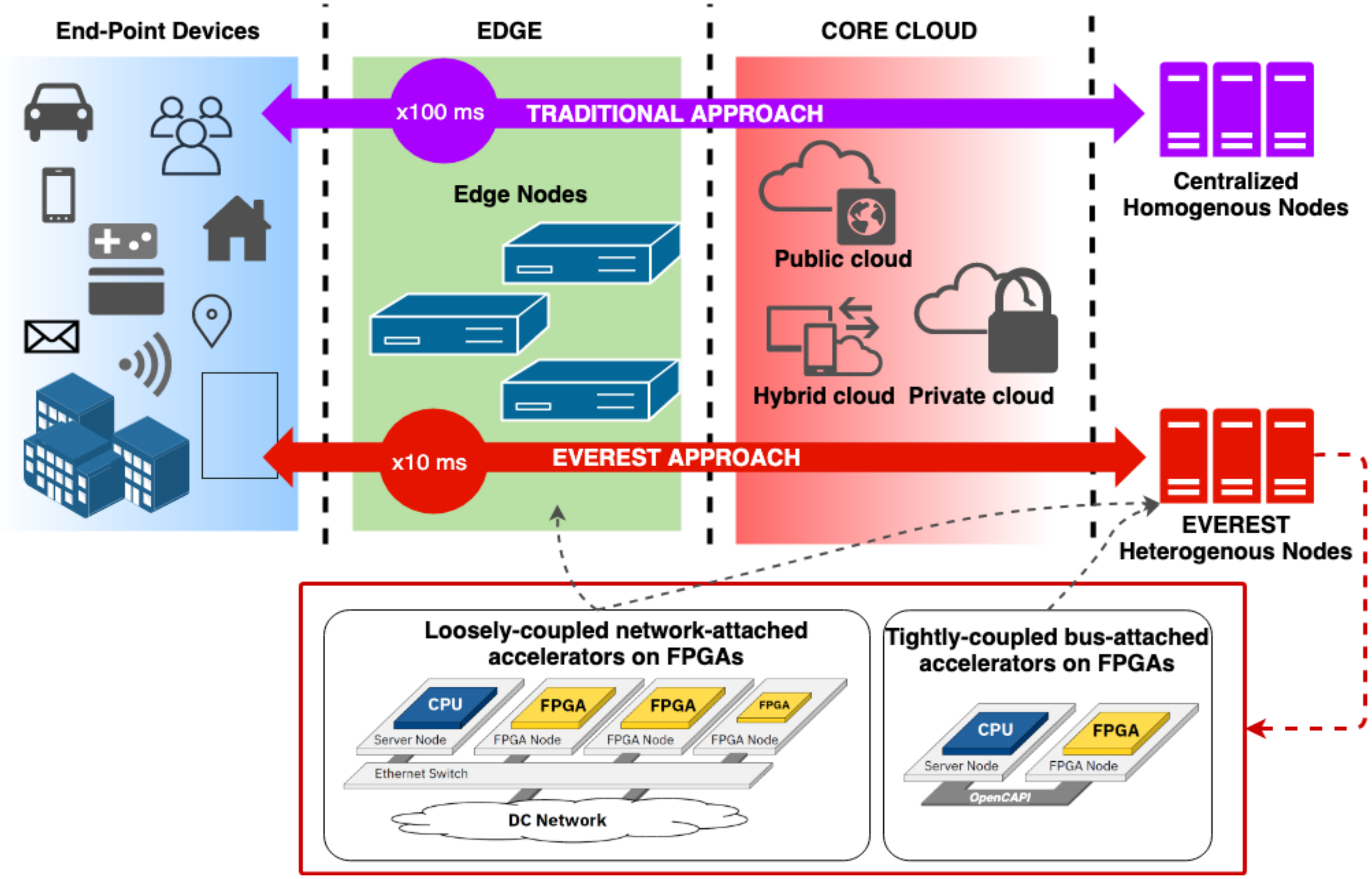
EVEREST

Advanced Traffic Modeling

- Mobility platform supporting cities with advanced traffic modelling

- **Data sources**
  - Historical and real-time **Floating Car Data** (FCD)
  - **Origin-destination matrix** (ODM) defining city
  - **Road network graph** including road restrictions;
  - Historical **weather data** (temperature, precipitation)
- **Traffic services**
  - **What-if analysis** for given scenarios, e.g. road closure;
  - **Intelligent routing** for large amount of vehicles
  - **Traffic prediction** for major road elements of cities

EVEREST Target System



FPGA-based heterogeneous architectures

EVEREST Target System: Brief Overview

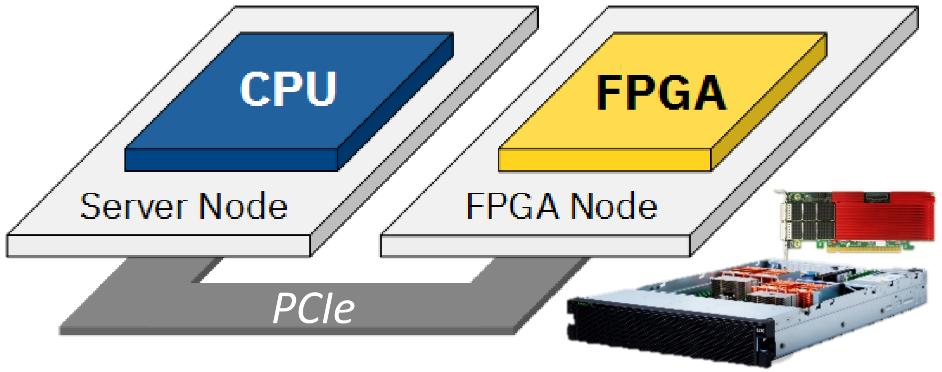
Network-attached and PCIe-attached FPGA nodes

- Off-the-shelf FPGA devices
- User logic can be easily designed and customized with HLS tools

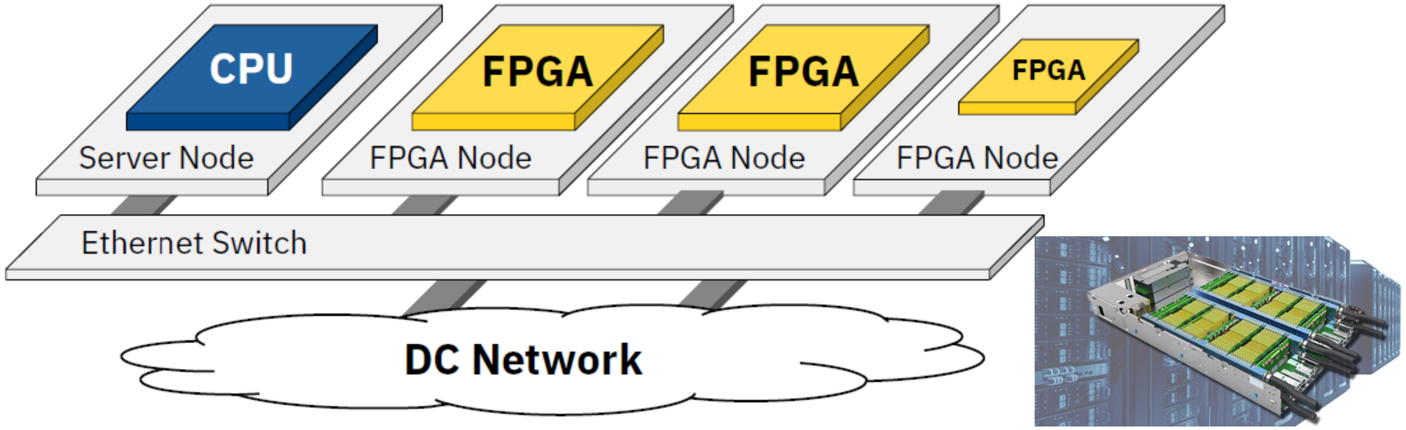
DC infrastructure and Supercomputers

- workflow orchestration
- reference implementation

FPGA as a Co-Processor

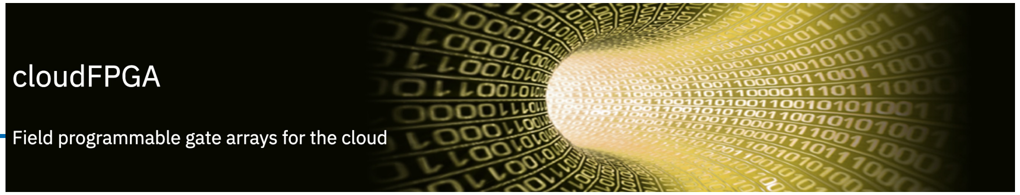


FPGA as a Peer-Processor



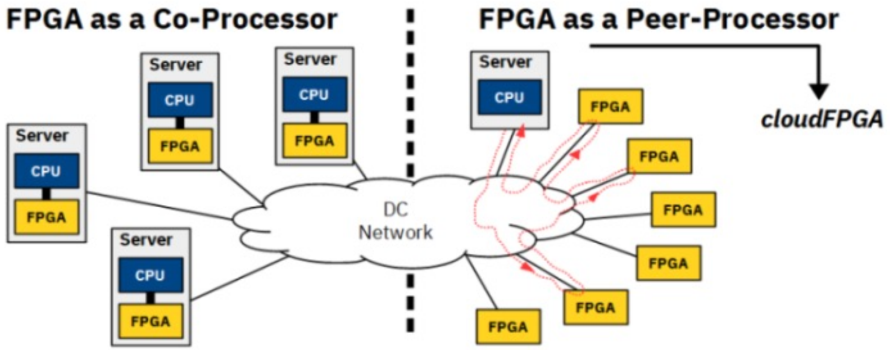
cloudFPGA

IBM



FPGAs as 1st-class citizens within a DC ...

- disaggregated from the server nodes
- connected directly to the DC network for its access and to communicate with CPUs and other FPGAs
- densely packed into DC chassis and racks and distributed across the DC



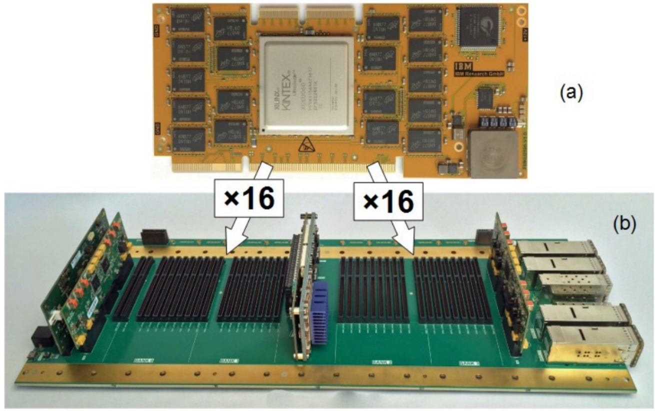
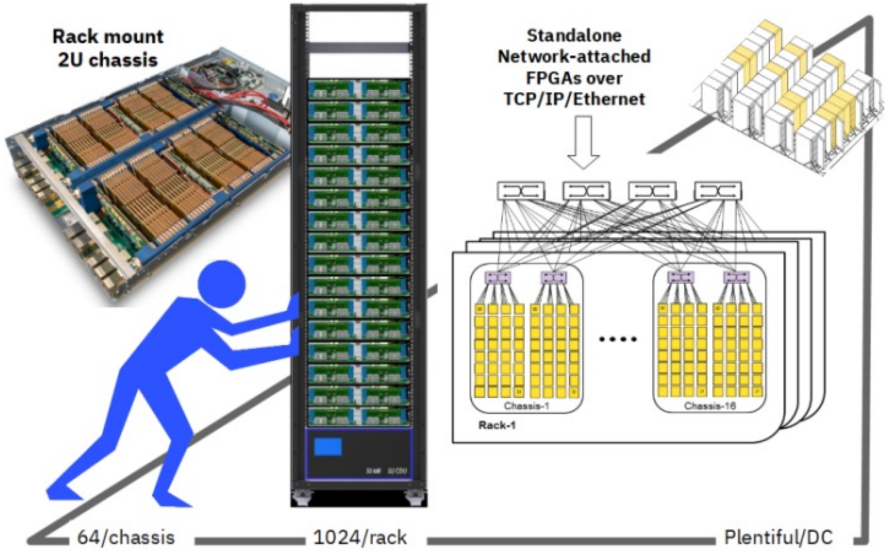


Figure 1: (a) The disaggregated FPGA and (b) the carrier board.

Xilinx Kintex UltraScale XCKU060 FPGA with 2x8GB of DDR4 memory

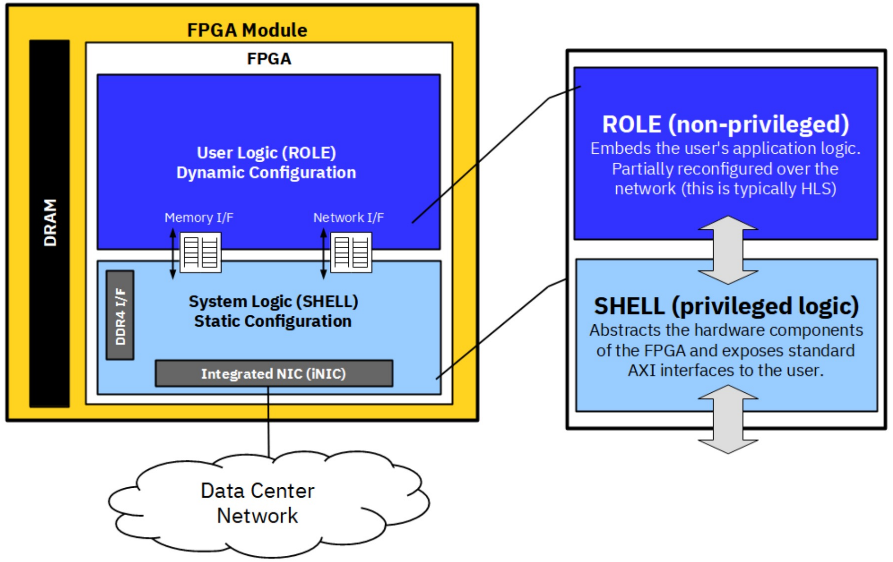


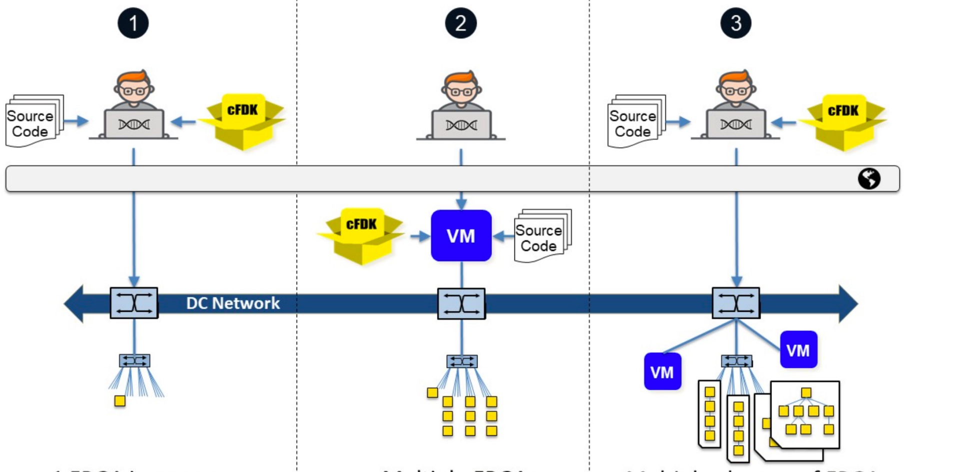
EVEREST

cloudFPGA Development Kit (cFDK)

- network-attached solution composed of:
  - Interface logic already designed (**cF Shell**) to support system integration
    - TPC/UDP communication is managed transparently to the user logic
  - User logic (**ROLE**) that can be easily designed and customized with traditional HLS tools
- application code running on host
  - FPGA accessible through the network
  - Low-level libraries for host-FPGA communication
- create clusters of FPGAs
- IDE incl allocation and mgmt of resources

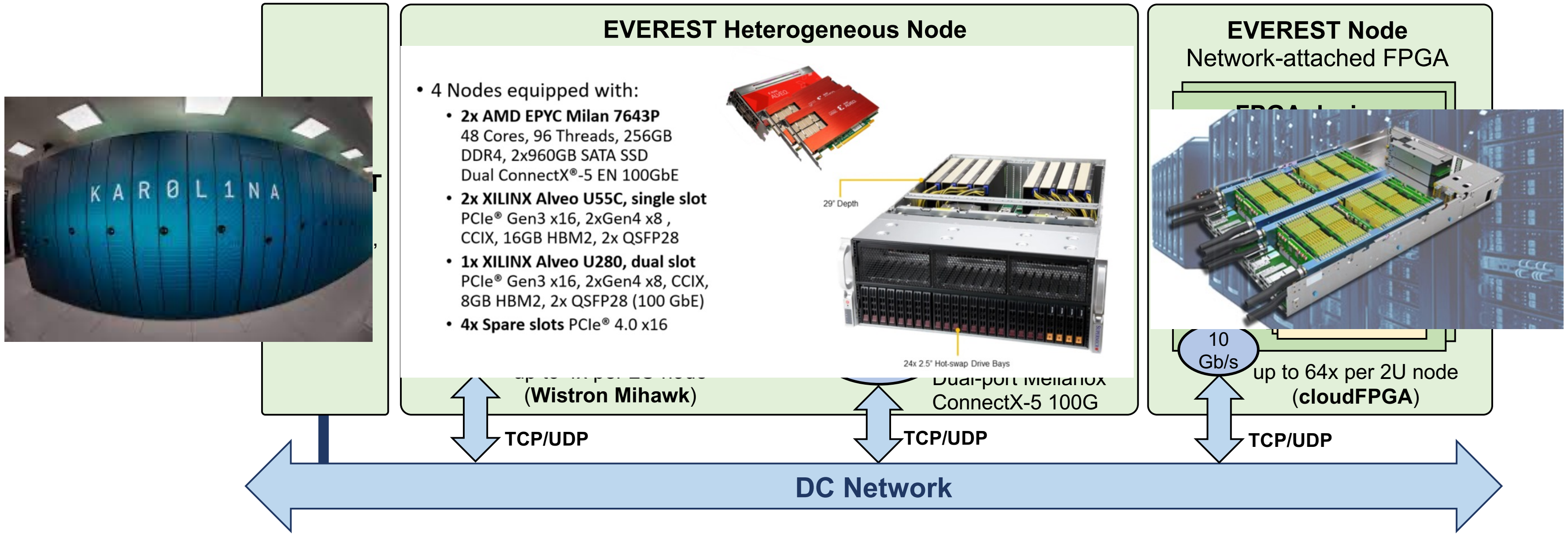
cFDK released at https://github.com/cloudFPGA





Conceptual System Overview

- Envisioned for demonstration purposes
- Multi-node demonstrator based on the technology and the components available during the project's timeline

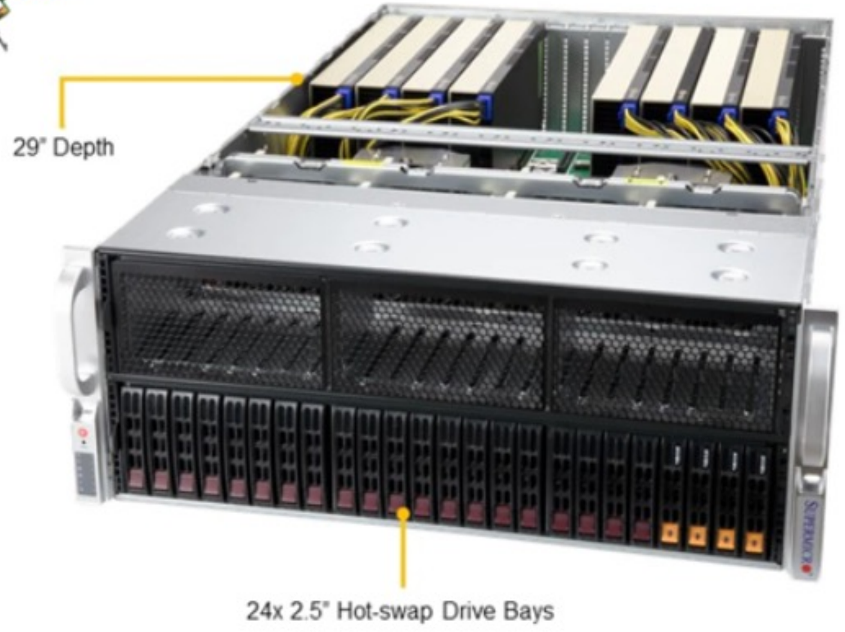


EVEREST

FPGA-accelerated HPC System Overview

- Multi-node demonstrator based on EVEREST SDK

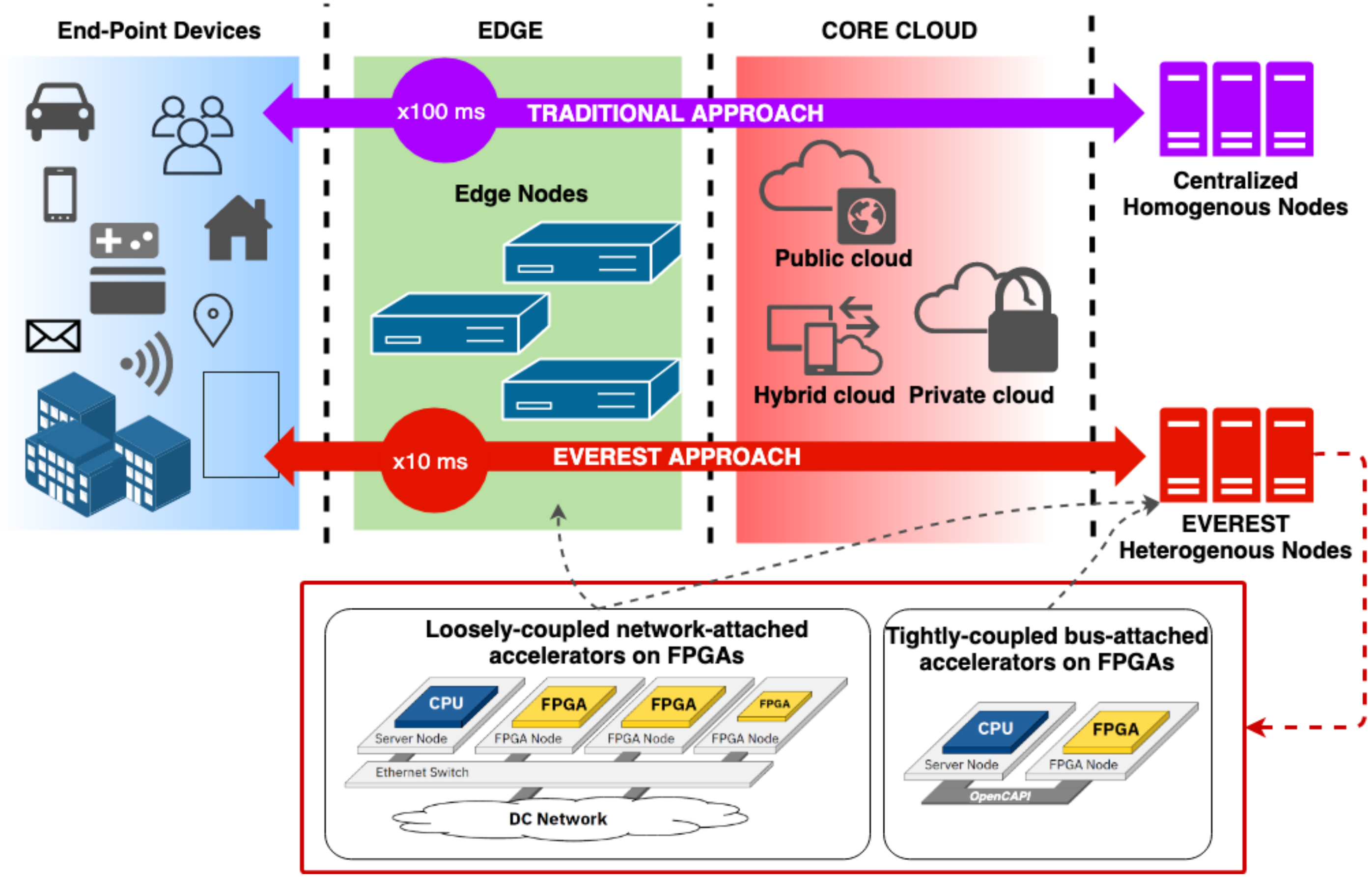
  - 4 Nodes equipped with:
    - **2x AMD EPYC Milan 7643P**
      48 Cores, 96 Threads, 256GB
      DDR4, 2x960GB SATA SSD
      Dual ConnectX®-5 EN 100GbE
    - **2x XILINX Alveo U55C, single slot**
      PCIe® Gen3 x16, 2xGen4 x8 ,
      CCIX, 16GB HBM2, 2x QSFP28
    - **1x XILINX Alveo U280, dual slot**
      PCIe® Gen3 x16, 2xGen4 x8, CCIX,
      8GB HBM2, 2x QSFP28 (100 GbE)
    - **4x Spare slots** PCIe® 4.0 x16



EVEREST

EVEREST Target System

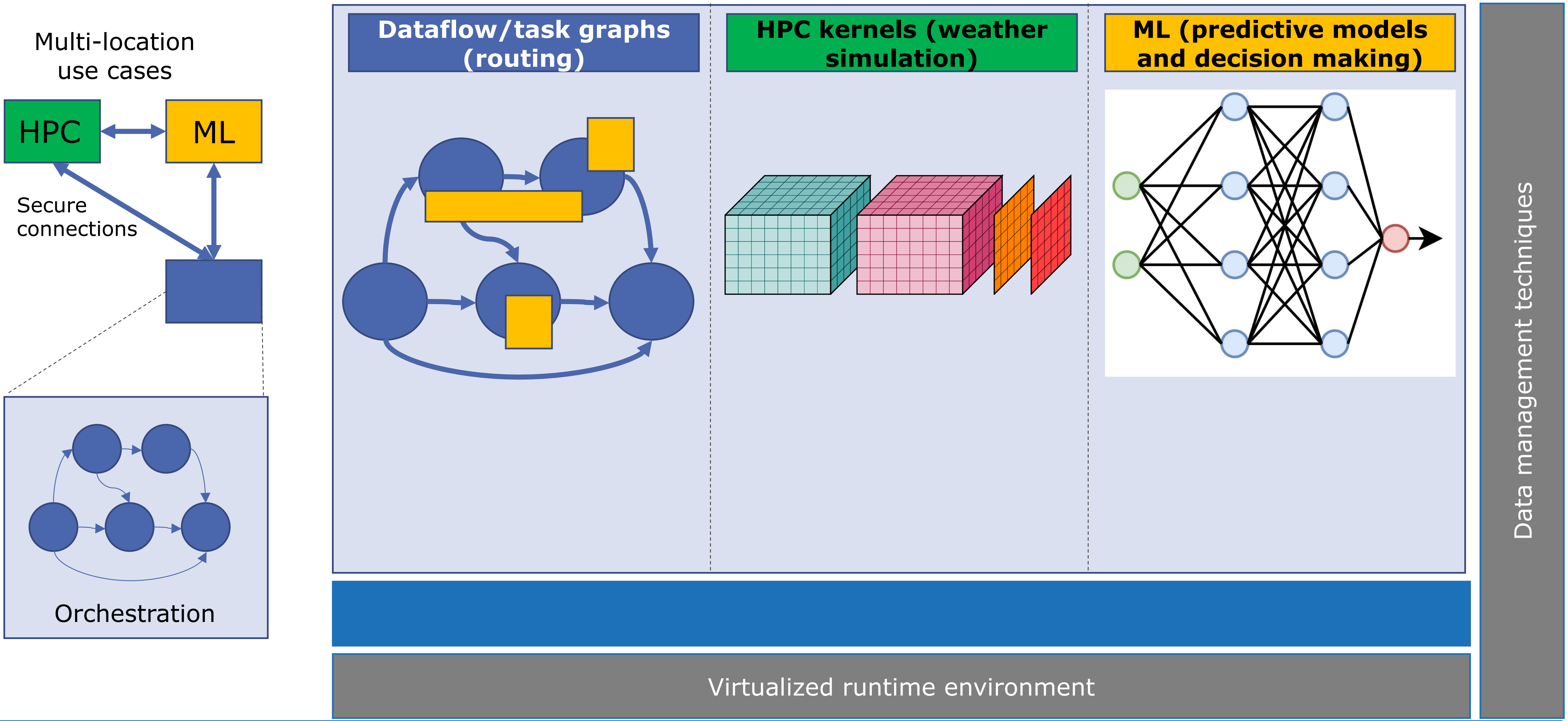
Computing continuum to enable **cloud-to-edge integration**



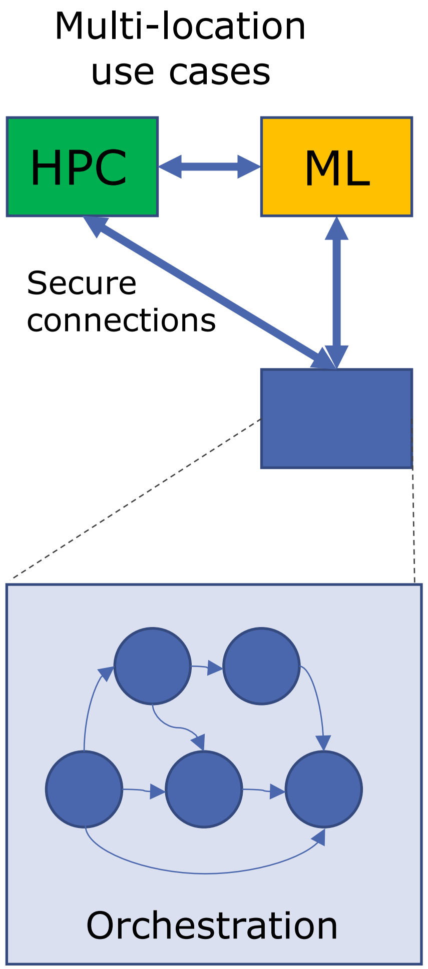
We aim at outperforming centralized/homogeneous solutions

EVEREST

EVEREST Compilation Framework



EVEREST Compilation Framework



Dataflow/task graphs (routing) | **HPC kernels (weather simulation)** | **ML (predictive models and decision making)**

Language/framework support

- Dataflow/task graphs:
  - Sequential syntax, implicit parallelism
  - Deterministic execution
  - Support for shared state
- HPC kernels:
  - DSLs for kernels (numerics, WRF model)
  - Integration in HPC distributed infrastructure
- ML:
  - Interoperability (pytorch, TensorFlow, TVM)
  - Support for exchange formats (ONNX)

High-level source-to-source compilation

- Dataflow/task graphs:
  - Dataflow IR: I/O, batching, pipelining
- HPC kernels:
  - MLIR dialect and polyhedral analysis
- ML:
  - Model partitioning based on TVM/RelayIR

High-level synthesis and memory design flow

- Dataflow/task graphs:
  - HLS support for OpenMP and irregular accesses
- HPC kernels:
  - Generation of memory controllers
- ML:
  - Re-use existing HLS tools and libraries

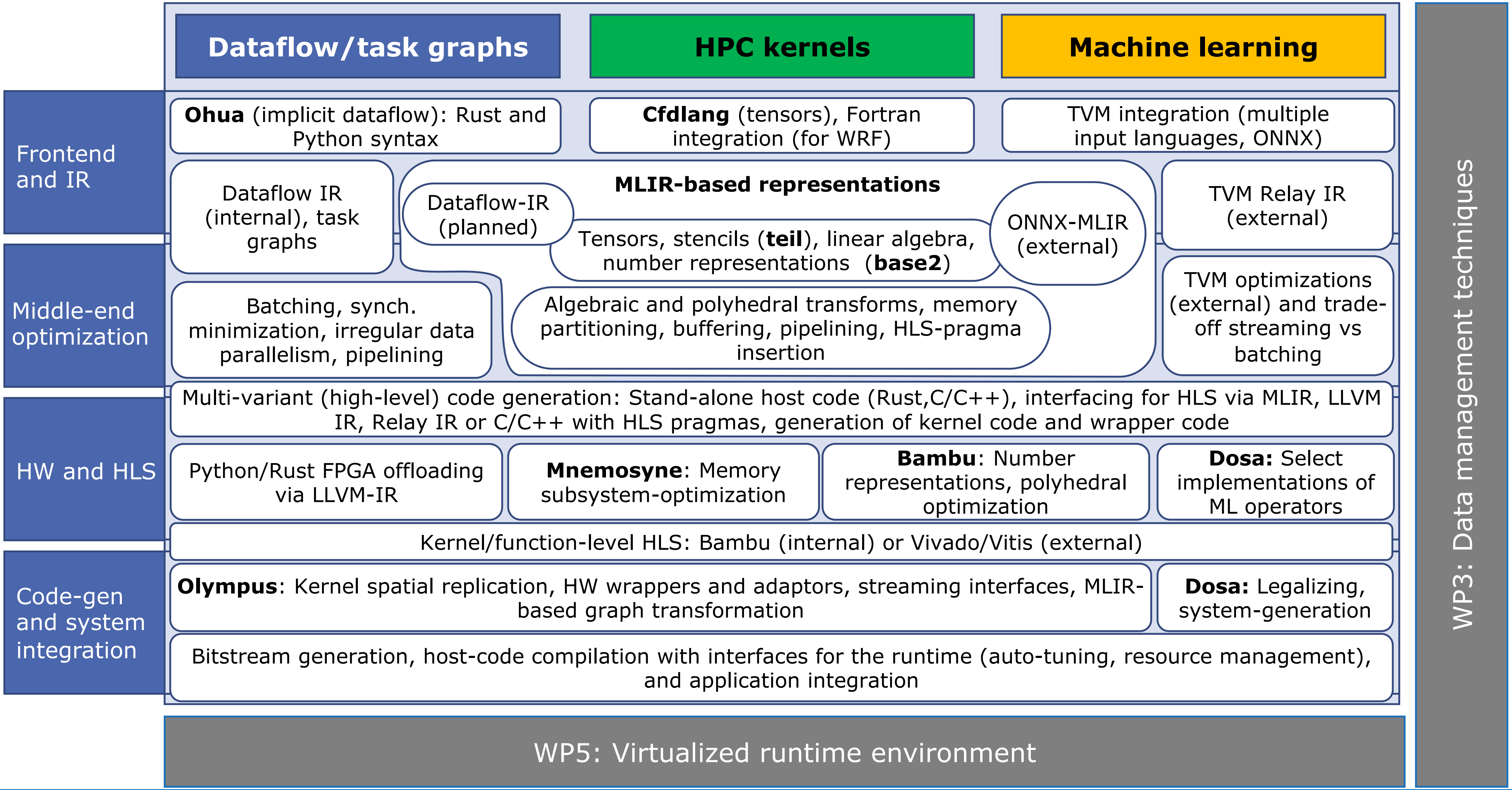
Low-level compilation, bitstream generation, and system integration

WP5: Virtualized runtime environment

WP3: Data management techniques

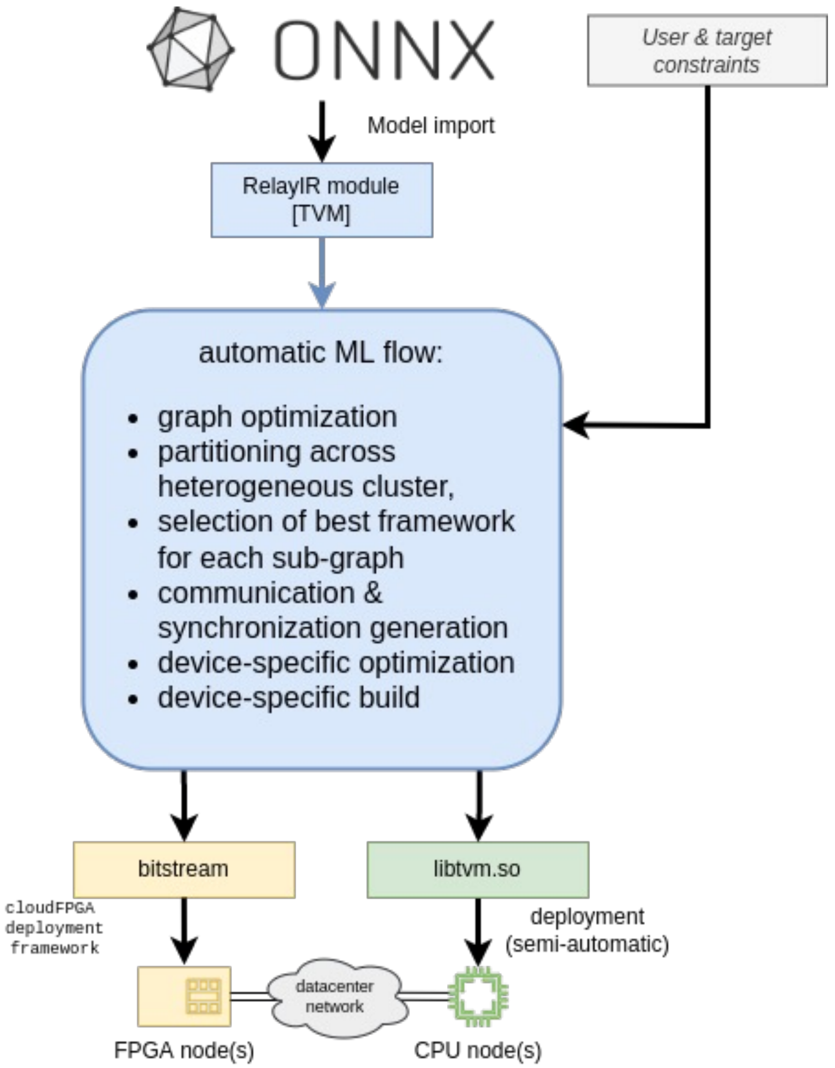
EVEREST

EVEREST Compilation Framework



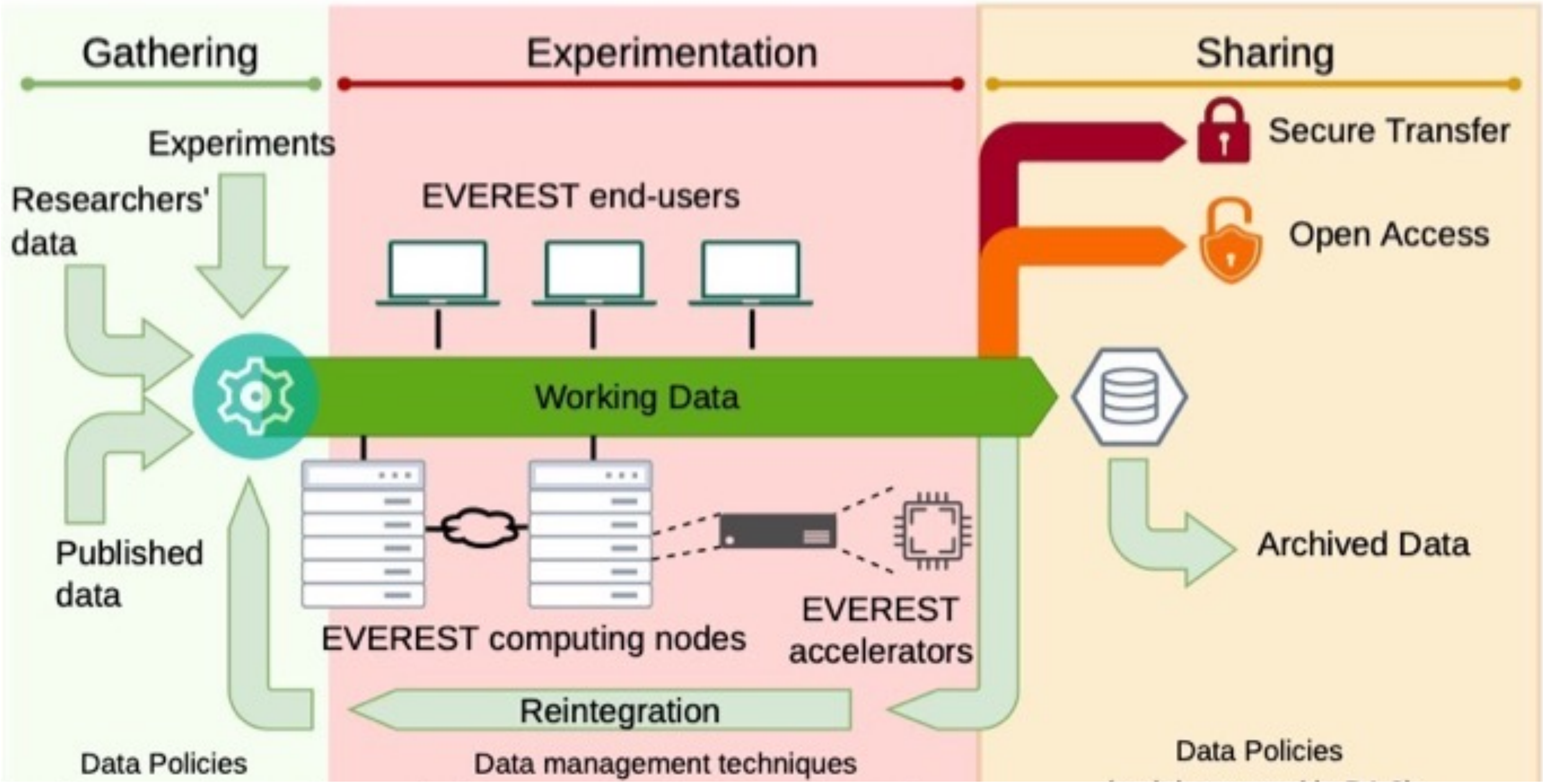
DNN on FPGAs: no yet-another-narrow-framework

- DNN on FPGAs is a highly active area of research
  →chances are that for a particular problem (i.e. some-convolution-to-put-on-FPGA with certain performance) someone has already developed and published a good implementation
  (e.g. haddoc2, FINN, hls4ml, VTA, VitisAI, and other open source frameworks...)
  → Why to re-invent the wheel and not reuse it?
- BUT: Who knows what is the best available implementation for the current problem (I.e. the ONNX input by the user)?
- Standardized way to include all available 3rd-party libraries (including Everest flows) within architecture generation
  - Automatic DSE of best available framework (depending on: operation, precision, target device)
- Frontend currently based TVM, but plan to integrate also MLIR interfaces/modules



EVEREST

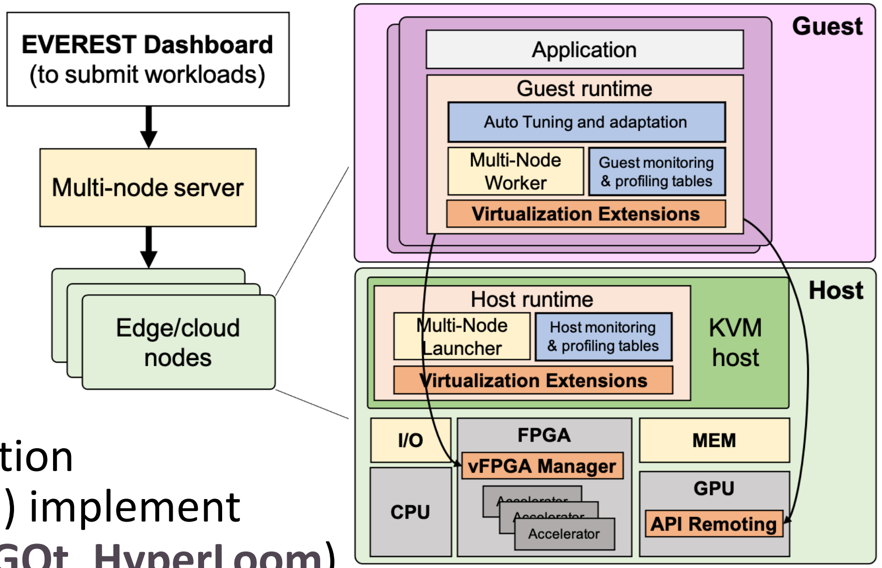
EVEREST Data Management & Protection



EVEREST

EVEREST Runtime Environment

... implements the *selection of "variants"* and the *hardware configuration* based on the *system status*



- **Dynamic adaptation and autotuning** (**mARGOt**)
- **Two-level runtime** for (1) virtualization of hardware resources regardless their distribution and the low-level details of the platforms; (2) implement functional decisions (**VOSYS solutions**, **mARGOt**, **HyperLoom**)

How to collect system status and expose it to the runtime?

Runtime API

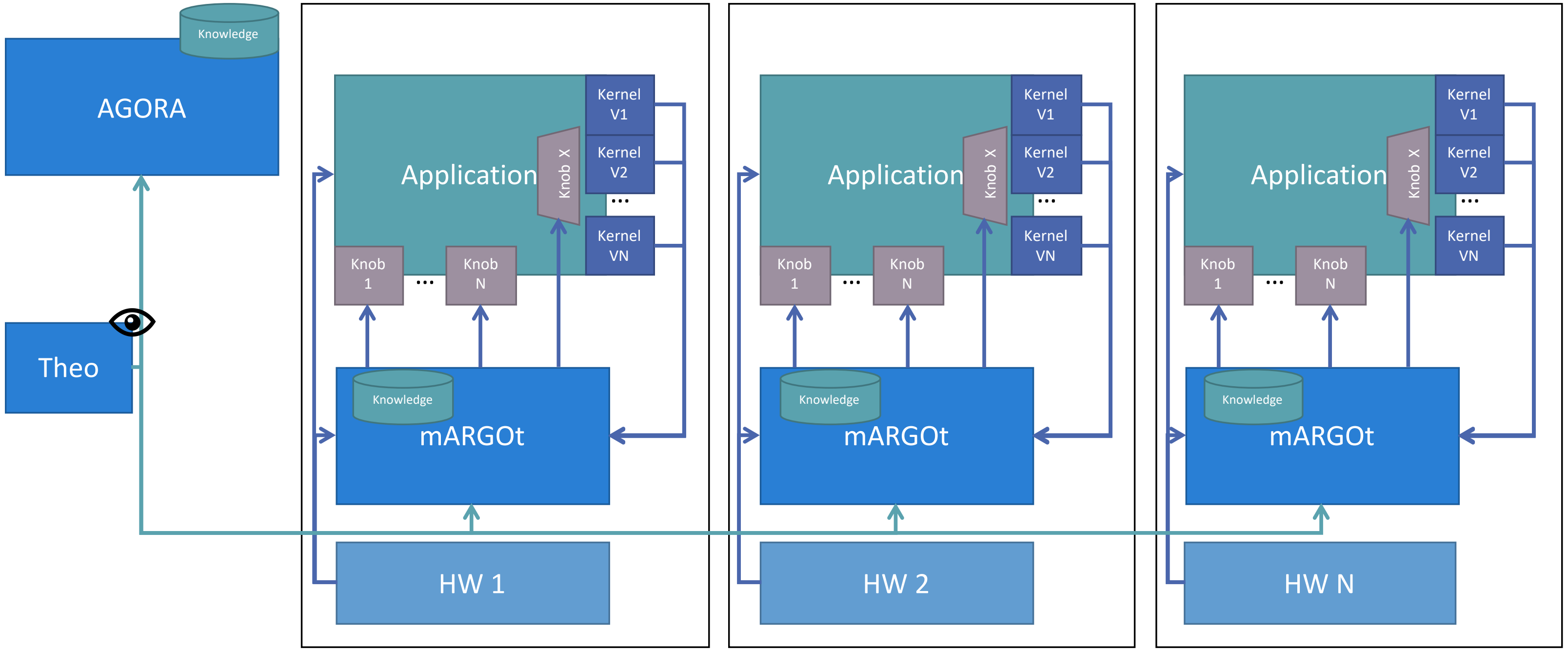
Autotuning API

Hiding communication latency (e.g., prefetching)

Seamless execution when varying the system configuration (resources, nodes, data, etc.)

EVEREST Runtime Environment

The EVEREST FPGA systems include a **monitoring and decision infrastructure** for **dynamic autotuning** based on workload conditions



- **Application variants** (either software or hardware) are generated at design time (compilation and hardware synthesis), and selected at run time based on the actual available hardware resources

Conclusion

- EVEREST is a dEsign enVironmEnt foR Extreme-Scale big data analyTics on heterogeneous platforms
  - built on the assumption that the future of computing is heterogeneous but the current tools do not support it
  - focus on building support for FPGAs
- The work towards an universal IR facilitates the re-use of innovations across the full stack including
  - extensions to new application domains / languages
  - extensions to different accelerator architectures
  - integration with different workflow engines / runtime environments
- Stay tuned ;-) ... https://everest-h2020.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957269

Thank You!

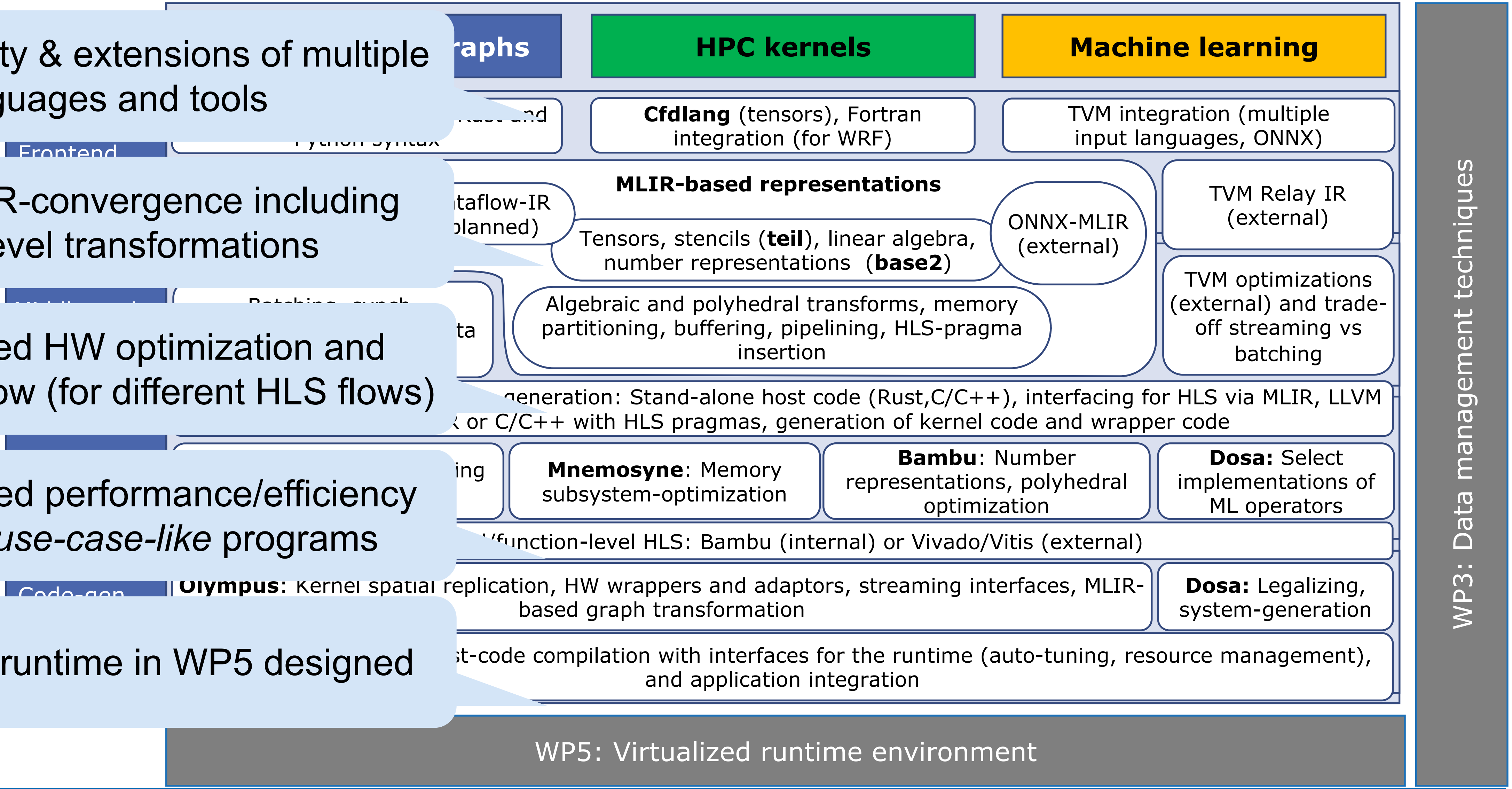
Backups

EVEREST

DESIGN ENVIRONMENT
FOR EXTREME-SCALE BIG DATA ANALYTICS
ON HETEROGENEOUS PLATFORMS

EVEREST Compilation Framework

- Interoperability & extensions of multiple languages and tools
- Partial MLIR-convergence including high-level transformations
- Generalized HW optimization and generation flow (for different HLS flows)
- Demonstrated performance/efficiency gains for *use-case-like* programs
- Interface to runtime in WP5 designed



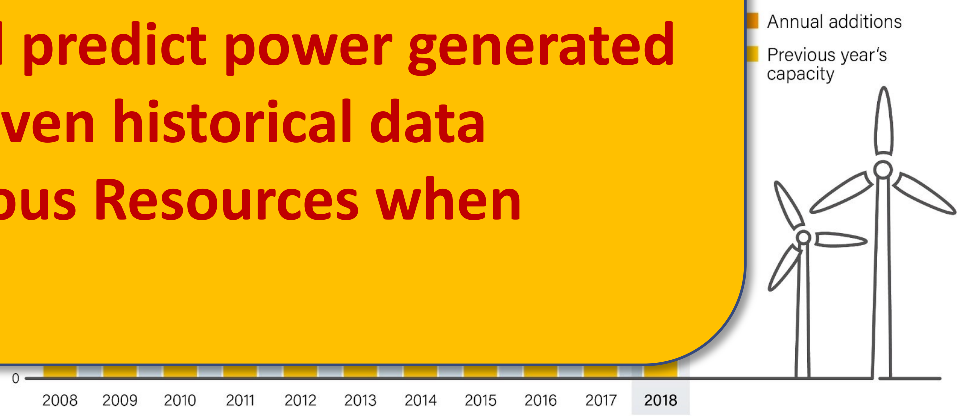
EVEREST

Renewable Energy use case: Context and Challenge

- Different challenges due to intermittency of the wind power generation:
  - Transmissior … e of grid (very short t
  - Traders to fo … y ahead (short term l
  - Wind farm o … horizon)
  - → **great eff**

CHALLENGES:

1. **Integration of the the data coming from sensors available on wind turbine for improving weather forecast model**
2. **Extend the wind speed estimation at 80m instead of the 10m default in WRF to improve prediction capabilities**
3. **Adopt AI models to learn and predict power generated by the specific site/turbine given historical data**
4. **… making use of Heterogeneous Resources when needed**



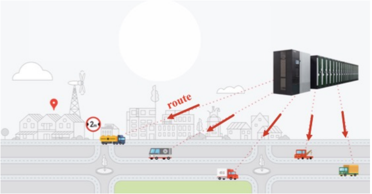
EVEREST

Advance Traffic Modeling for Smart Cities use case

- Mobility platform for supporting cities with advanced traffic modelling
- **Data source**
  - Historical a
    - e.g. GPS
  - **Origin-des**
  - **Road netw**
  - Historical **w**
- **Traffic servi**
  - **What-if ana**
  - **Intelligent routing** for large amount of vehicles
  - **Traffic prediction** for major road elements of cities

CHALLENGES:

1. **Boost the FCD data collection and processing to compute 3D traffic models**
2. **Develop and efficiently deploy a traffic simulator on a multinode architecture to predict traffic behaviour**
3. **Adopt AI techniques to learn the traffic patterns resulting into a traffic prediction service.**
4. **… making use of Heterogeneous Resources**



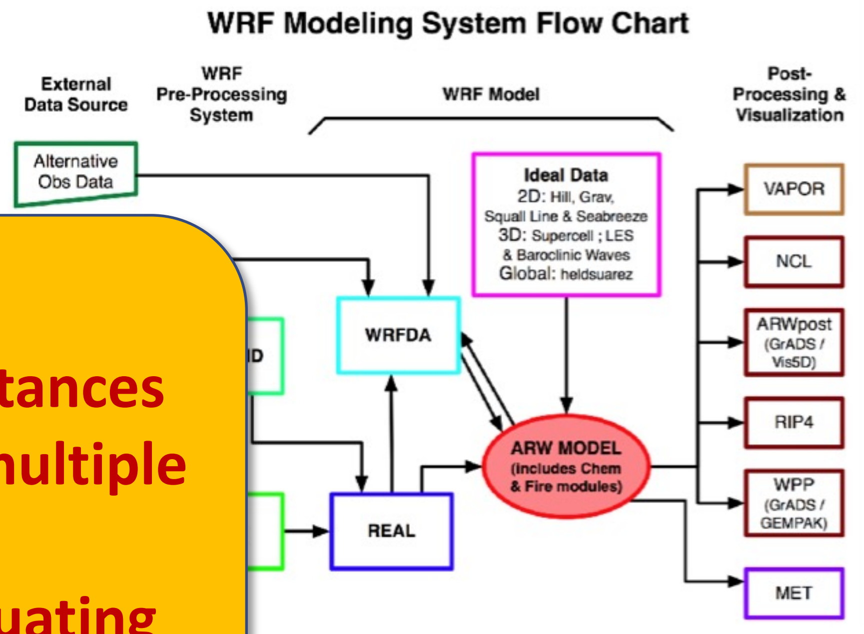
EVERESt

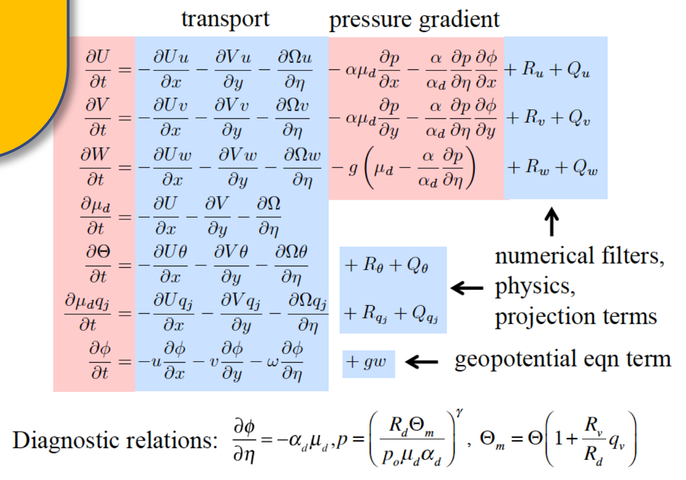
The WRF Model

- First step of two use cases…

CHALLENGES

- **Automatic deployment of several instances of the model in HPC resources with multiple boundaries conditions**
- **Efficient High Resolution output evaluating the usage of FPGA resources**





- WRF is an collab… supported principa… for Atmospheric Research (NCAR), the US National Oceanic and Atmospheric Administration and the US National Center for Environmental Prediction – NCEP)

EVEREST Consortium

[1]IBM Research Europe, Switzerland, [2]Politecnico di Milano, Italy, [3]Università della Svizzera Italiana, Switzerland, [4]Technische Universitat Dresden, Germany, [5]Centro Internazionale di Monitoraggio Ambientale, Italy, [6]IT4Innovations, VSB – Technical University of Ostrava, Czech Republic, [7]Virtual Open System, France, [8]Duferco Energia, Italy, [9]NUMTECH, France, [10]Sygic, Slovakia