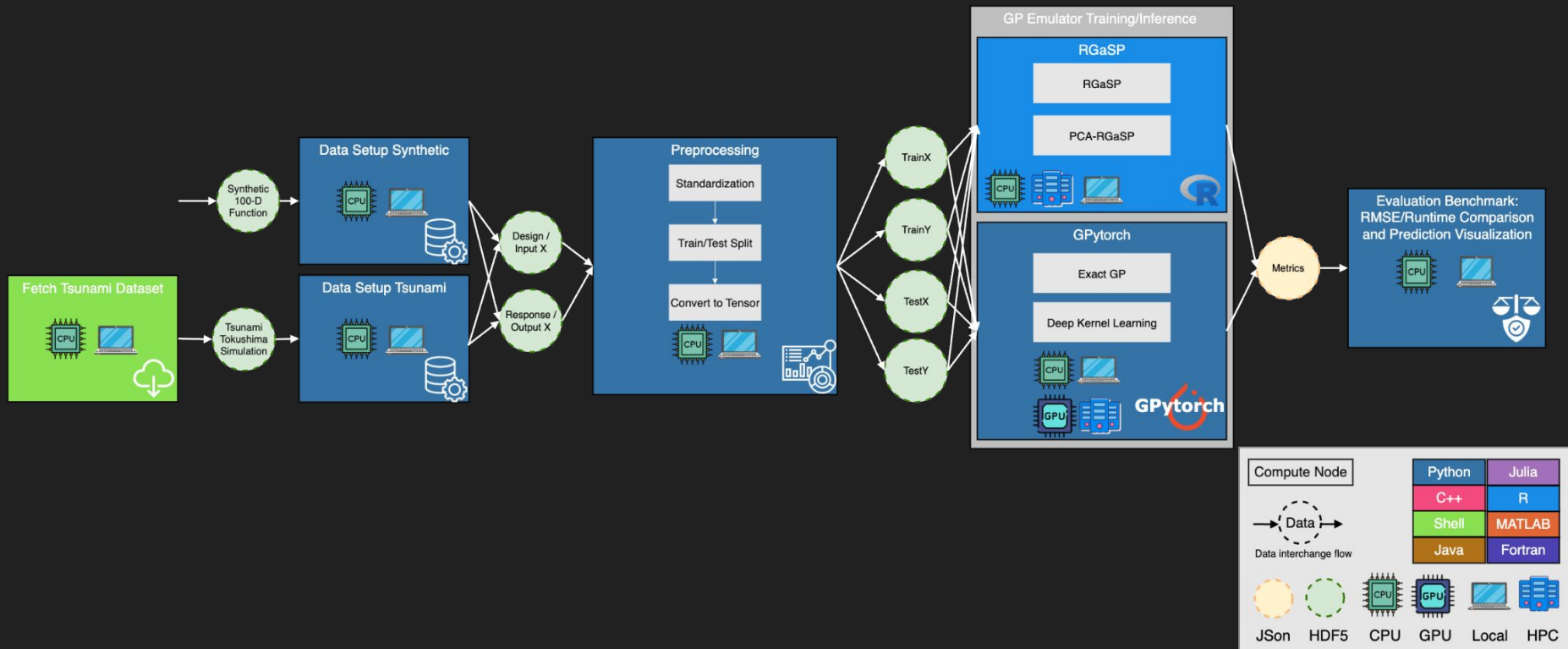


High dimensional Input Problem: Deep Kernel Learning

Hiwi Meeting

Workflow



Nextflow Orchestration

Isolation

- Abstraction
- Modularity
- Reusability

Interoperation

- Data exchange
- Cross-platform

1. Apache Arrow/Feather
2. Apache Parquet
3. **HDF5**
4. CSV/JSON

Orchestration

- Automatic Workflow Integration

modules

- benchmark_metrics.nf
- data_setup.nf
- evaluate_dkl.nf
- evaluate_exactgp.nf
- evaluate_rgasp.nf
- fetch_data.nf
- preprocessing.nf

scripts

- benchmark_metrics.py
- data_setup_synthetic.py
- data_setup_tsunami.py
- evaluate_dkl.py
- evaluate_exactgp.py
- evaluate_rgasp.R
- preprocessing.py

```
workflow {
  println "► Starting pipeline with caseStudy=${params.caseStudy}"

  // 1. Fetch raw inputs (download or generate metadata)
  def raw_ch = fetch_data( params.caseStudy )
  // 2. Convert raw -> X/Y
  def data_ch = data_setup( raw_ch )
  // 3. Standardize, split, save to HDF5
  def tensors_ch = preprocessing( data_ch )
  // 4. Train emulators in parallel based on selected models
  def exactgp_ch = evaluate_exactgp( tensors_ch )
  def dkl_ch = evaluate_dkl( tensors_ch )
  def rgasp_ch = evaluate_rgasp( tensors_ch )
  // 5. Compare metrics and save results
  benchmark_metrics( exactgp_ch, dkl_ch, rgasp_ch )
}
```

NEXTFLOW ~ version 25.04.6

Launching 'workflows/main.nf' [shrivelled_shaw] DSL2 - revision: 1a1f73e3e2

► Starting pipeline with caseStudy=synthetic

```
executor > local (7)
[3f/01bb32] process > fetch_data (synthetic) [100%] 1 of 1 ✓
[53/9d5589] process > data_setup (synthetic) [100%] 1 of 1 ✓
[ad/8ca7b2] process > preprocessing (preprocessing) [100%] 1 of 1 ✓
[1f/417244] process > evaluate_exactgp (ExactGP) [100%] 1 of 1 ✓
[a5/c2dd29] process > evaluate_dkl (DKL) [100%] 1 of 1 ✓
[a2/b31f05] process > evaluate_rgasp (RGASP) [100%] 1 of 1 ✓
[e3/477012] process > benchmark_metrics (benchmark_metrics) [100%] 1 of 1 ✓
Completed at: 29-Jul-2025 12:30:50
Duration : 7m 27s
CPU hours : 0.1
Succeeded : 7
```

Workflow (generated by nextflow)

