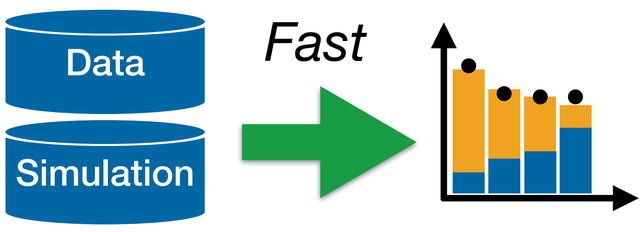
ACAT 2021

Going fast on a small-size computing cluster

Svenja Diekmann, Peter Fackeldey, Niclas Eich, Benjamin Fischer, Martin Erdmann, Dennis Noll

Computing Situation

- Fast turnaround times are required to:
  - Increase and drive scientific output
  - More consolidation studies during peer review processes
  - Interactivity
- Many small institute computing cluster exist outside of the WLCG
- However, cheap solutions are rare



Specifications

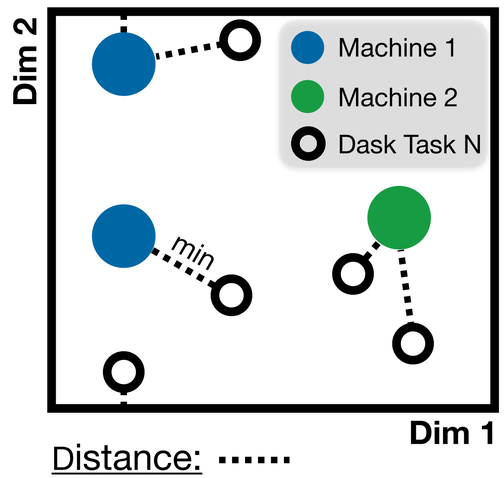
Cluster - VISPA:

- 13 Worker machines, total:
  - 245 threads, 2GB RAM each
  - 22TB SSD (FS-Cache)
- Storage (NFS):
  - 6x12TB HDD (striped)
  - 1TB LVM SSD Cache

Software packages:

- NumPy and python-HEP ecosystem (vectorized operations)
- Dask (dask-jobqueue):
  - Batch submission (HTCondor)
  - On-worker SSD cache affinity

On-Worker SSD Cache



FS-Cache Properties:
- Transparent
- Shared by multiple users

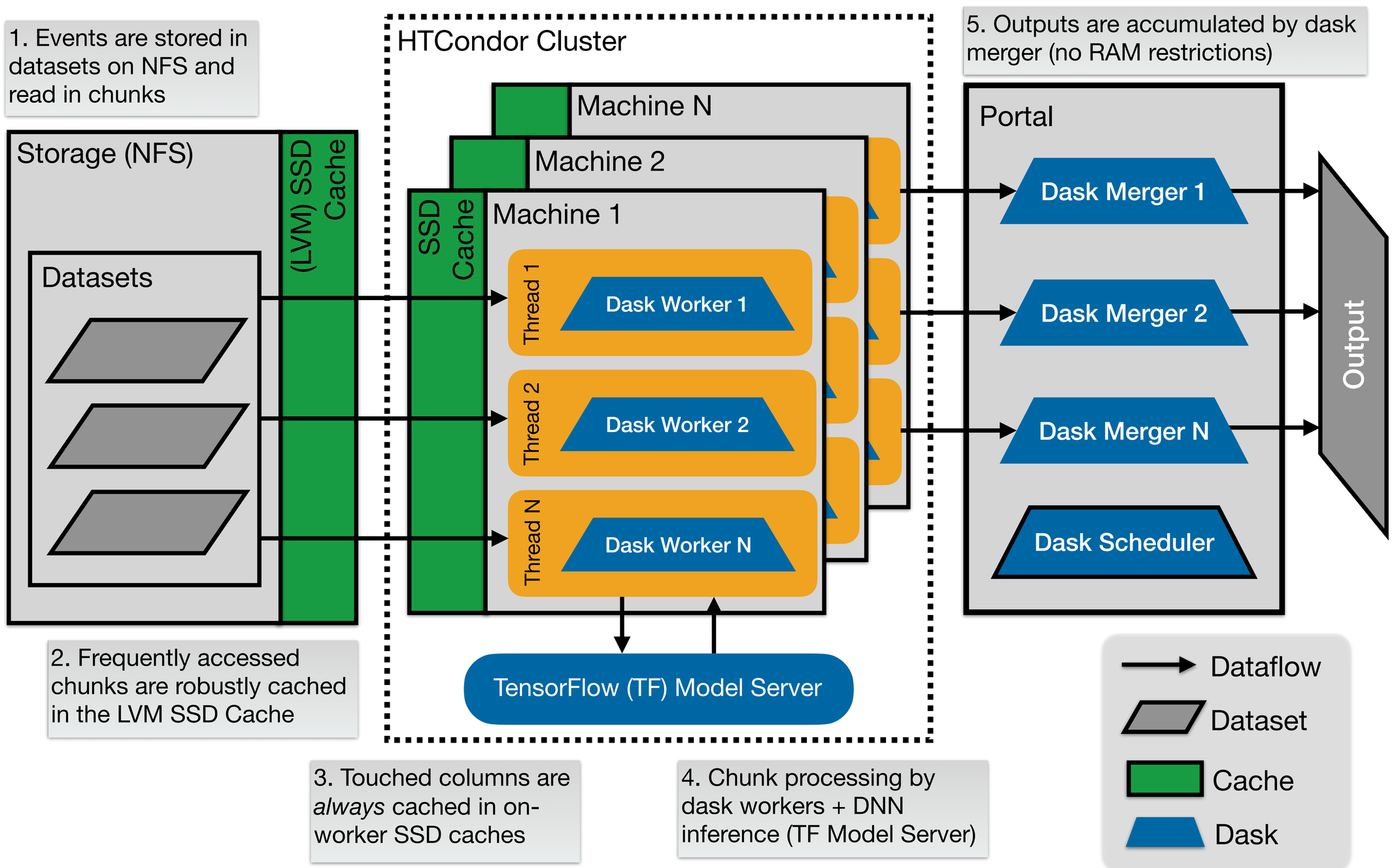
Affine assignment:
- Uses hash distance
- Deterministic
- Smoothly degrading under changes of workers & tasks

Distance: ·····

$$\sum_{i}^{N} min\left(|\bullet - \bullet|, 1 - |\bullet - \bullet|\right)$$

with N: number of dimensions

Data Flow & Cluster Sketch



Results

Datasets:

- O(1) TB NanoAOD CMS Simulation (ROOT file format)
- Touched columns correspond to 386 GB uncompressed data
- Total: $4.18 \cdot 10^9$ events

Run 1 → Run 5:

- Significant runtime reduction, speedup of **1.5x**
- Runtime and read from NFS show almost perfect overlay:
  → Speedup comes from SSD caching only

Data rate [MB/s]	Run 1	Run 2	Run 3	Run 4	Run 5
Total	229	293	336	333	336
Per thread	1.9	3.6	5.0	5.3	5.4

VISPA: vispa.physik.rwth-aachen.de
Contact: peter.fackeldey@cern.ch

RWTH AACHEN UNIVERSITY
Federal Ministry of Education and Research