Data Management and Uncertainty Quantification for EPSI

R. Moser, M.Greenwald, M. Parashar, S. Klasky, and the EPSI Team

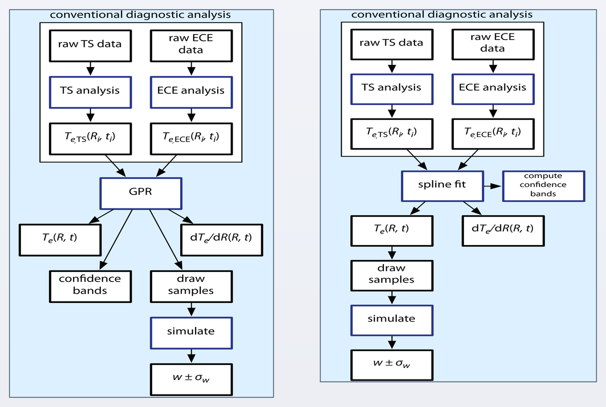
U. Texas, MIT, ORNL, Rutgers

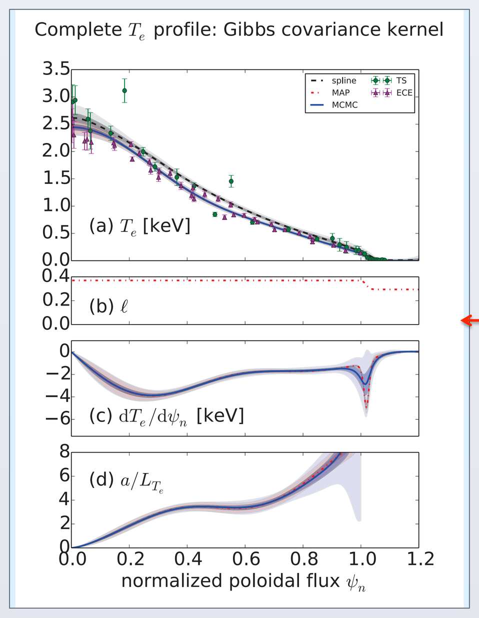
Uncertainty Quantification

Improved Profile Fitting and UQ using Gaussian Process Regression(MIT)[1]

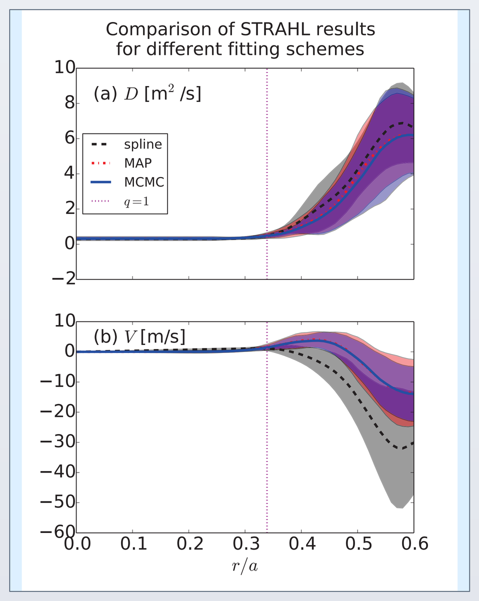
- EPSI Codes need UQ analyzed experimental data for reliable validation
- Fitting of smooth profiles from discrete observations key plasma simulation input
- Existing spline fitting techniques have deficiencies
- Gaussian Process Regression(GPR) captures existing results while delivering
  - Statistically rigorous fits for plasma profiles and gradients(key XGC1 inputs)
  - Uncertainty estimates for key plasma profiles and gradients
  - Increased convergence rate of forward uncertainty propagation
  - Implementation general and widely applicable.

Spline vs GPR fitting Workflows





Profile fitting for T_e(electron temperature) from Thomson Scattering (TS) and electron cyclotron emission(ECE) system data. GPR results shown using maximum a posteriori(MAP) and Markov Chain Monte Carlo(MCMC) approaches. Error bars and dark shading is ±σ, light shading ±3σ.

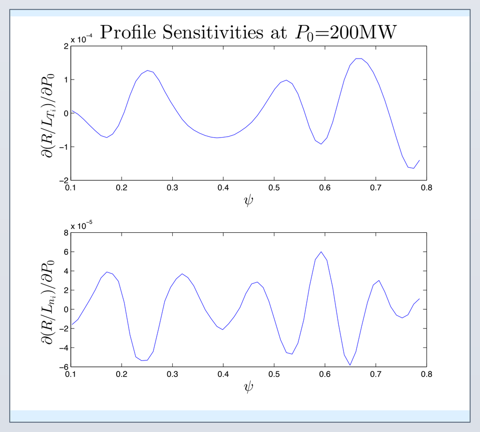


STRAHL takes electron profiles and initial guesses on impurity transport coefficients D and V and iterated to match exp. data. D and V are highly sensitive to profile uncertainties. The uncertainty envelopes are ±σ.

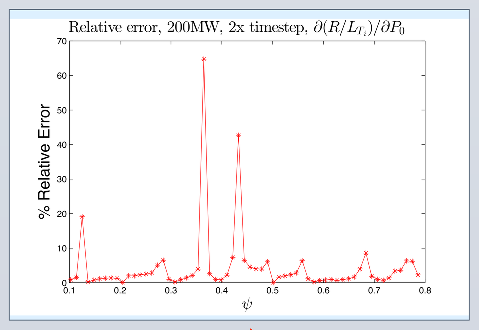
[1] M A Chilenski, M Greenwald, Y Marzouk, N T Howard, A E White, J E Rice, and J R Walk (MIT), submitted to Nuclear Fusion, work supported under cooperative agreement C-Mod: DE-FC02-99ER54512

Sensitivity of Plasma Gradients in XGC1 to Applied Heating(UTexas,PPPL)[2]

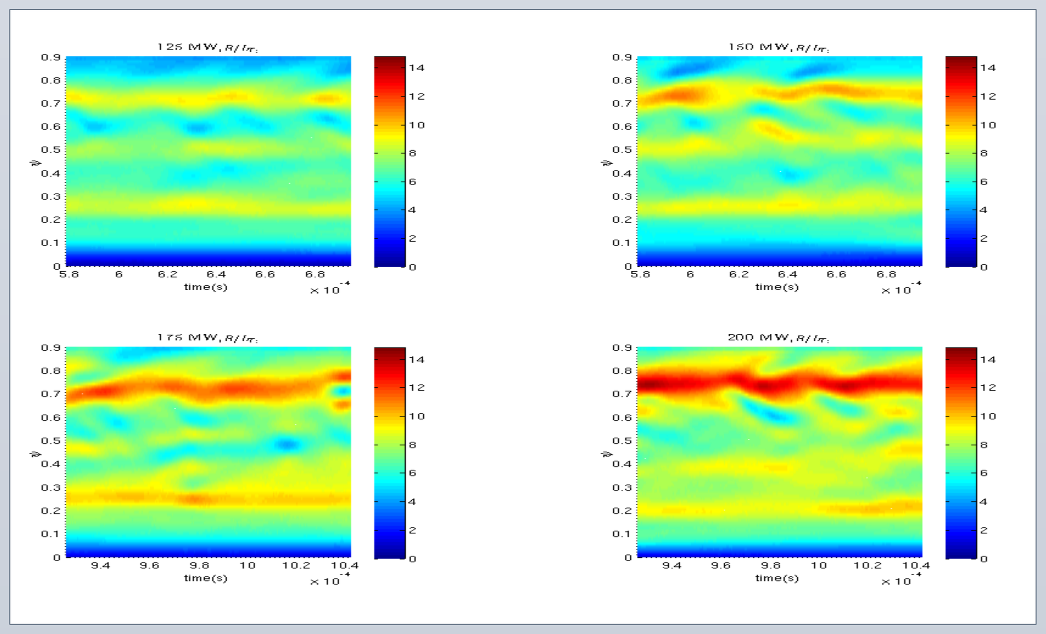
- Complete UQ workflow involves characterization of input uncertainties (see above), forward propagation via XGC1, and comparison of quantity of interest(QoI) (often via synthetic diagnostic) with experimental data.
- Uncertainties in model input parameters (λ) may impact QoI. Many parameters-intractable sampling problem given the cost of global gyrokinetic simulation.
- Goals of Sensitivity Analysis(SA):
  - Potentially reduce size of input space to be sampled by removing unimportant parameters *(Reduce # of simulations for UQ)*
  - Characterize effect of numerical parameters (grid, particle count, timestep) on QoI (plasma profiles) *(Reduce cost of simulations for UQ)*
  - Quantify effects of embedded physics models on QoI *(Reduce cost of simulations)*
- Motivated by our goal of studying L-H transition and nonlocal edge effect on core, we are performing a SA of plasma profiles to a heating model. We are using a combination of sampling approaches to compute dQoI/dλ.
- Moderate case (CYCLONE base case) allows sensitivity computations at lower cost.
- Sensitivity predictions combined with scaling studies will be compared against larger scale, enriched physics XGC1 runs: Use reduced system to project UQ of full system
- Repeat process with enriched physics, increased problem size if extrapolation*(combined with input uncertainties)* is invalidated by new simulation data.



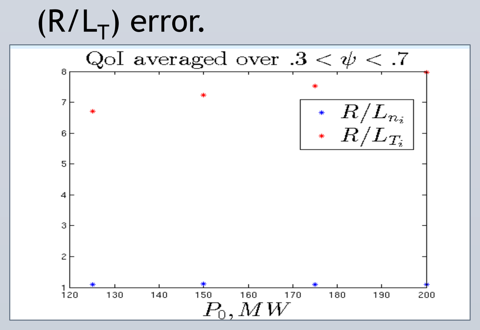
"Steady-state" gradient sensitivities computed at 200MW; temperature (naturally) more sensitive than density.



Attempt to coarsen timestep(2x) (4x timestep unstable at 200MW) led to marginal accuracy results(Mean error=4.42%), with even worse results for profile (R/L_T) error.



Different applied heating powers lead to differing simulation times to steady state.125MW, 150 MW are near steady state, while 175MW, 200MW are not. Expected monotone behavior in P_0 of R/L_T is observed. Quantifying acceptable level of "steady-state" variation is governed by the QoI.



Expected behavior of T, n profiles under P_0 scaling; density QoI stiff and reaches steady state quickly.

[2] Simulations performed at NERSC on Edison.

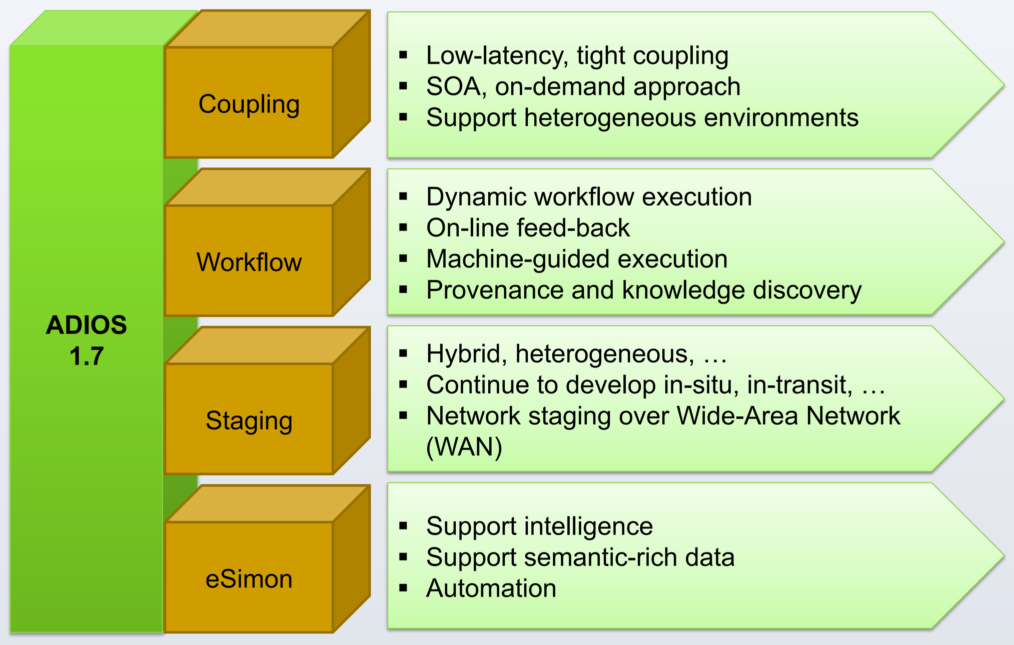
Calibration of XGCa Anomalous Diffusion under Uncertainty(UTEXAS,MIT,PPPL)

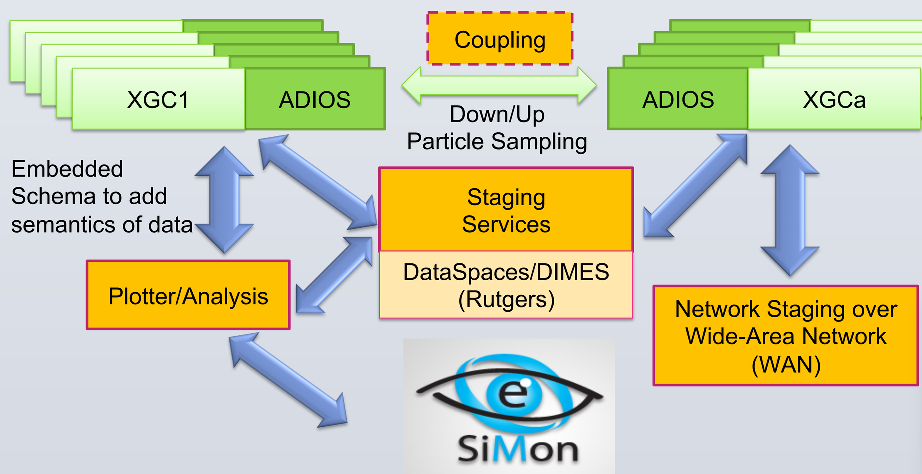
- Following D.Battaglia et al(2014), we are performing a calibration of an anomalous diffusion model.
  - Calibration will be Bayesian, using SCIDAC QUEST center software QUESO.
  - Plan to use Alcator CMOD data to calibrate model in a variety of modes relevant to L-H transition.
  - May use high-fidelity XGC1 simulation results as additional calibration training data.
  - Will examine use of XGCa+calibrated model as a surrogate for XGC1 sensitivity analysis.

Data Management

ADIOS (2013 R&D 100 winner. Current version 1.7) has been integrated with EPSI simulations to support:

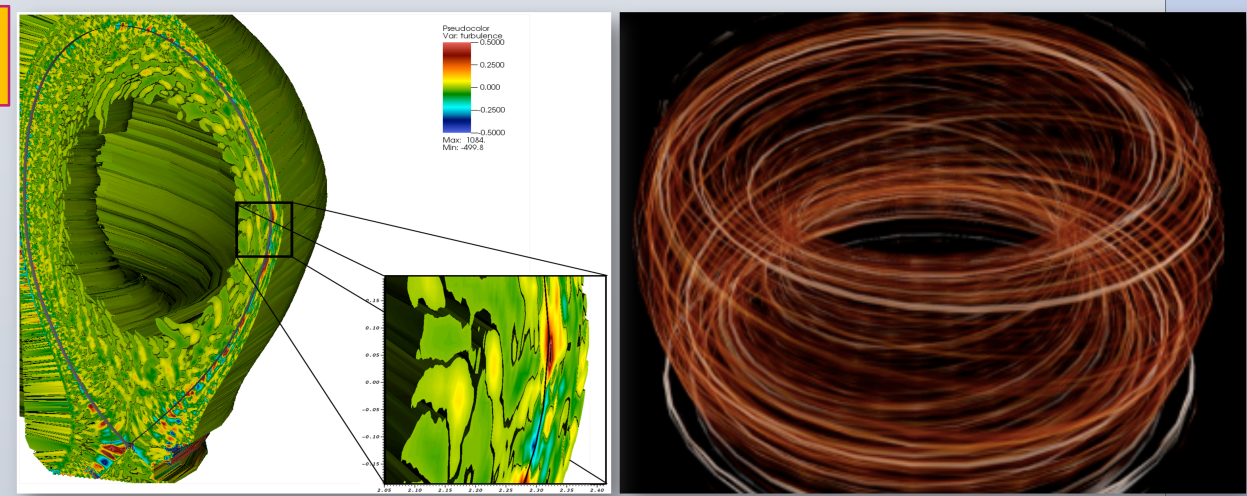
- Data staging with low-latency, tight coupling execution environments through in-memory data exchanges between different codes.
- Service Oriented Architecture (SOA) for on-demand coupling executions with support of dynamic workflow invocation.
- Coupling executions in heterogeneous computing environments with streaming data support. Network staging over Wide Area Network (WAN) under development.
- Selection and chunked reads to enable schedule optimization.





Data-centric integrated execution environment

Our focus is to support EPSI simulation by providing integrated data-centric execution environments for tight code coupling, staged data process, and monitoring system with a support of dynamic workflow system.

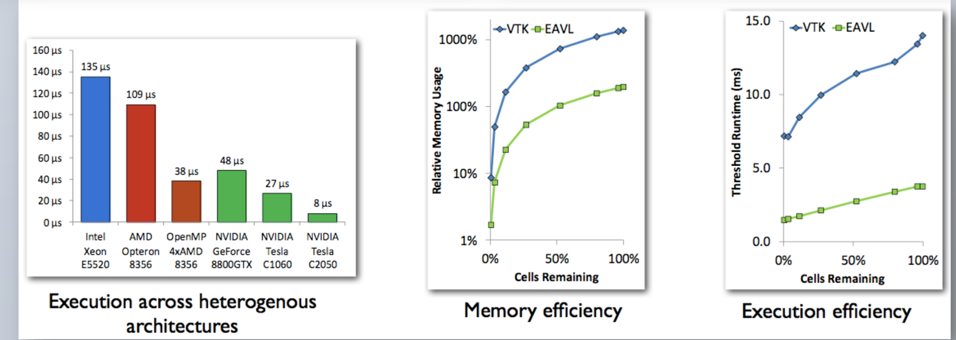


ADIOS Vis Schema

- Create an easy-to-use schema for ALL ADIOS codes.
- Facilitates data sharing without adding code complexity, or slowing down the code.
- Visualization schema: Semantics of the data for the purpose of visualization.
- Describing visualization data for various tools (VTK, Matlab, ParaView).

Future of XGC Reader

- Support for ADIOS Vis Schema.
- Filtering of turbulence.
- Integration with higher quality rendering with visualization software from OLCF.

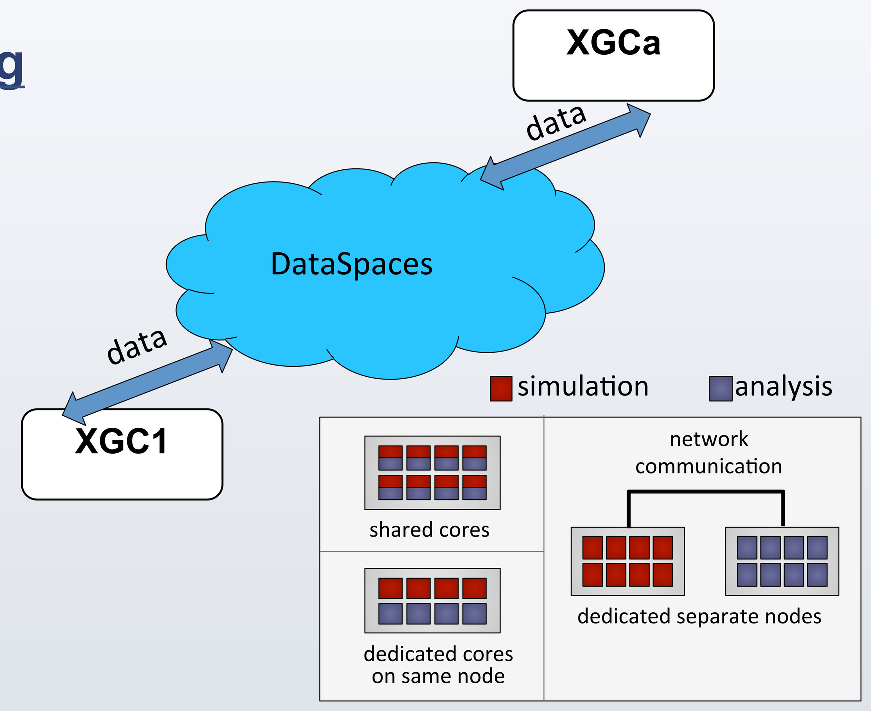


EAVL: Extreme Scale Analysis and Visualization Library

Data Staging

EPSI coupling workflow

- Goal: To enable tightly coupled XGC1 and XGCa workflow using memory to memory coupling for experimental time scale simulation



DataSpaces As a Service

- Persistent staging servers run as a service.
- Coupled applications can join, progress and leave independently.
- New approach targets more complex workflows and provides more flexibility.
- DataSpaces As a Service will improve resource efficiency and increase I/O performance.

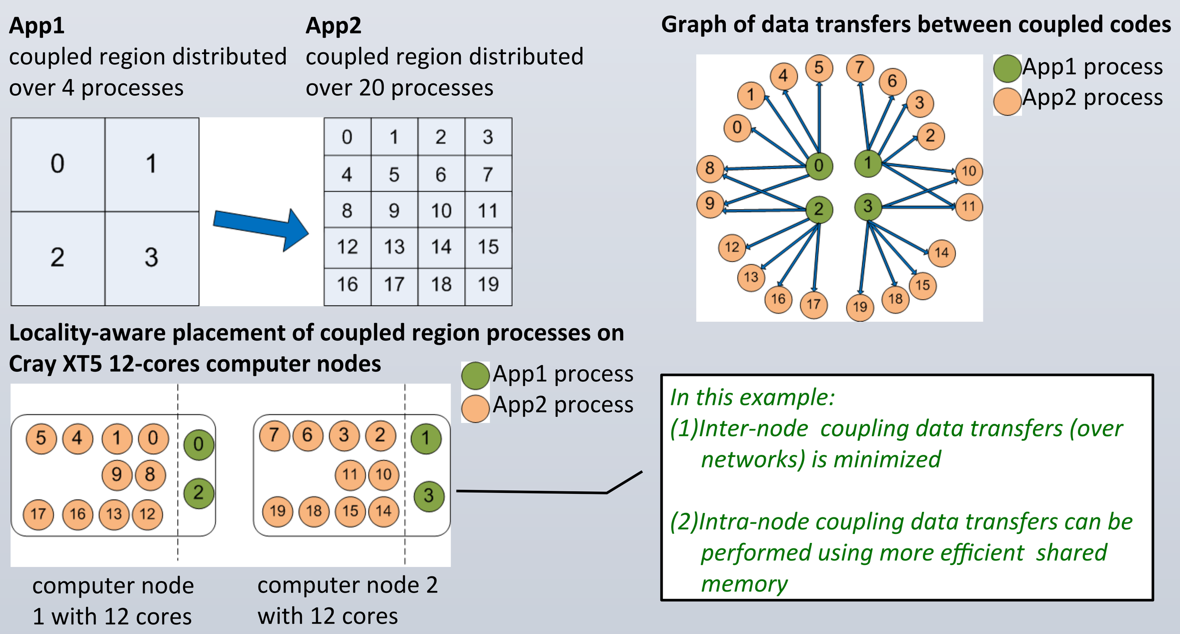
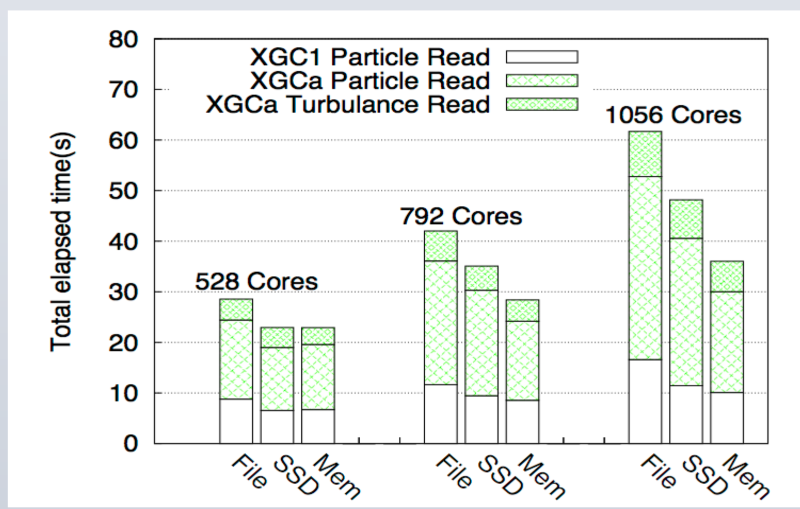


Illustration of the data-centric mapping of the application process for concurrently coupled workflow



Preliminary Result

- Enable exchange data through on-node memory.
- Data locality and core-level parallelism can be exploited to reduce data movement by increasing intra-node data sharing.
- Utilize SSD for data bursting.