Hierarchical Data Prefetching in Multi-Tiered Storage Environments
Hariharan Devarajan, Anthony Koukgas, Xian-He Sun
hdevarajan@hawk.iit.edu, akoukgas@iit.edu, sun@iit.edu

Overview
HFetch is a data-centric prefetching decision engine that utilizes system-generated events, while leveraging the presence of multiple tiers of storage, to perform timely hierarchical data placement. HFetch can boost operations by up to 50%. Compared to other prefetching solutions, HFetch is 10-35% faster.

HFetch Highlights
- Hierarchical
- Data-centric
- Uses System-Push
- Highly Scalable
- Low Application Overhead

Related Work

Challenges
- Resource Utilization
- Support Multiple Tiers
- Application-agnostic

Data Centric

Access Characteristics:
- App A: Sequential
- App B: Strided
- App C: Repeated

Prefetching Approaches:
- Trace-based: KNOWAC
- ML-based: Stacker

Key Observations:
- Application-centric prefetching
- Cache Pollution and Eviction
- Pseudo Hierarchical
- Misses Pipelining
- Misses Concurrency

Demonstrate the inefficiencies in application-centric prefetching

Results
WRF

Montage

<table>
<thead>
<tr>
<th>Time (sec)</th>
<th># Client Cores</th>
</tr>
</thead>
<tbody>
<tr>
<td>320</td>
<td>640</td>
</tr>
<tr>
<td>1280</td>
<td>2560</td>
</tr>
</tbody>
</table>

Disk (client side):