Contribution ID: **223**   Type: **not specified**

Mempolicy is dead, long live memory policy!

The default global mempolicy is inclusive of all NUMA nodes - where the fallback allocation behavior is typically defined by NUMA distances. Tasks and cgroups are then expected to opt-in to more restrictive policies via set_mempolicy and cpusets interfaces.

This is the opposite of a typical isolation mechanisms - and leads to global resource (such as unmapped page-cache) having poor isolation controls when accessed by tasks with different memory policies.

What if, instead, the default policy was restrictive (DRAM/Default-tier only), and tasks/cgroups were required to opt-in to tiered memory (via a set_mempolicy, cgroup/cpusets, or even new per-file/filesystem policies). We will discuss one design and the implications for various subsystems (reclaim, compaction, migration, pagecache vs anon allocations, cgroups/cpusets controls, etc).

Primary author: PRICE, Gregory (Meta)

Presenter: PRICE, Gregory (Meta)

Session Classification: Kernel Memory Management MC

Track Classification: Kernel Memory Management MC