Huge pages reduce address translation cost

- Modern applications
  - Large memory footprint
  - Low memory locality
- Huge page TLB coverage increasing
  - Huge pages now useful

Pathologies in current huge page management

- Huge pages increase fragmentation
- Fragmentation happens fast in any size of physical memory

Ingens: Asynchronous and utilization-based huge page promotion

- Memory bloating
- High page fault latency

Ingens design

- Synchronous allocation
- Greedy allocation
- Asynchronous allocation
- Utilization-based allocation

Evaluation

- Throughput (req/s)

Eliminates Redis memory bloating

- Negligible overhead
  - 2% in the worst case