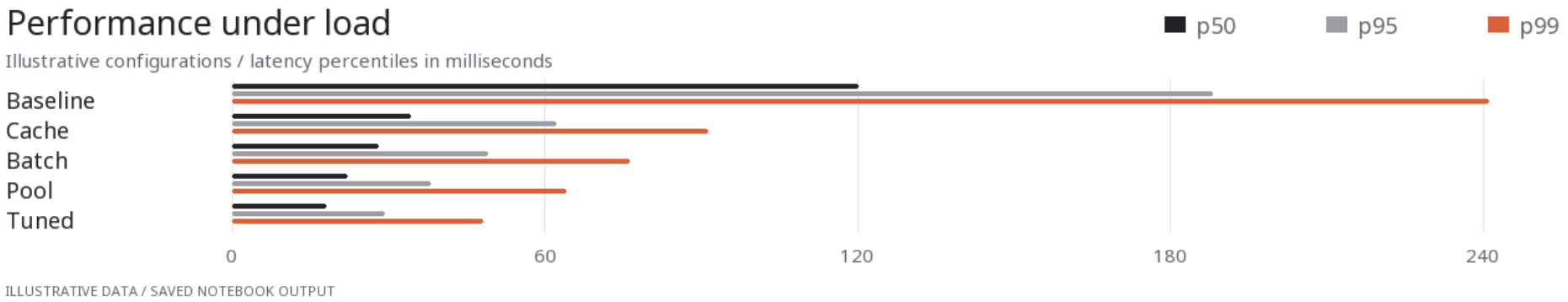
Performance benchmark

Latency percentiles / Five configurations / Illustrative data

Config	Workers	p50 ms	p95 ms	p99 ms	Req/s	RAM MB	Errors
Baseline	1	120	188	241	812	96	0.0%
Cache	1	34	62	91	2890	112	0.0%
Batch	8	28	49	76	3540	128	0.0%
Pool	16	22	38	64	4380	144	0.1%
Tuned	16	18	29	48	5210	152	0.1%



Tradeoff: Faster responses use more memory and show a nonzero error rate. Real benchmarks require repeated measurements. Illustrative data only.