ry Architecture for Document AI Processi

uted Task Queue with Auto-scaling Kubernetes Workers

Performance Targets
- 1,000+ docs/hour
- Auto-scale 1-50 pods

Auto-scaling Logic
- Queue length > 100
- CPU usage > 70%