torchserve::ModelConfig

+ std::string url
+ uint8_t numInstances
+ uint16_t minWorkers
+ uint16_t maxWorkers
+ uint32_t maxQueueSize
+ uint16_t maxRetries
+ uint16_t batchSize
+ uint32_t maxBatchDelayMSec
+ uint32_t responseTimeoutSec

+modelConfig

torchserve::WorkflowSpec

+ std::unordered_map
< std::string, std::
unique_ptr< ModelConfig
> > subModels
+ std::unordered_map
< std::string, std::
vector< std::string >
> dag

+workflowSpec

torchserve::Model

- std::unique_ptr< Manifest
> manifest
- std::unique_ptr< Scheduler
> schedulers

+ Model(const Manifest
&manifest)
+ Status updateModel
(const ModelConfig &modelConfig,
bool isStartup, bool isCleanup)
+ void addJob(Job &job)