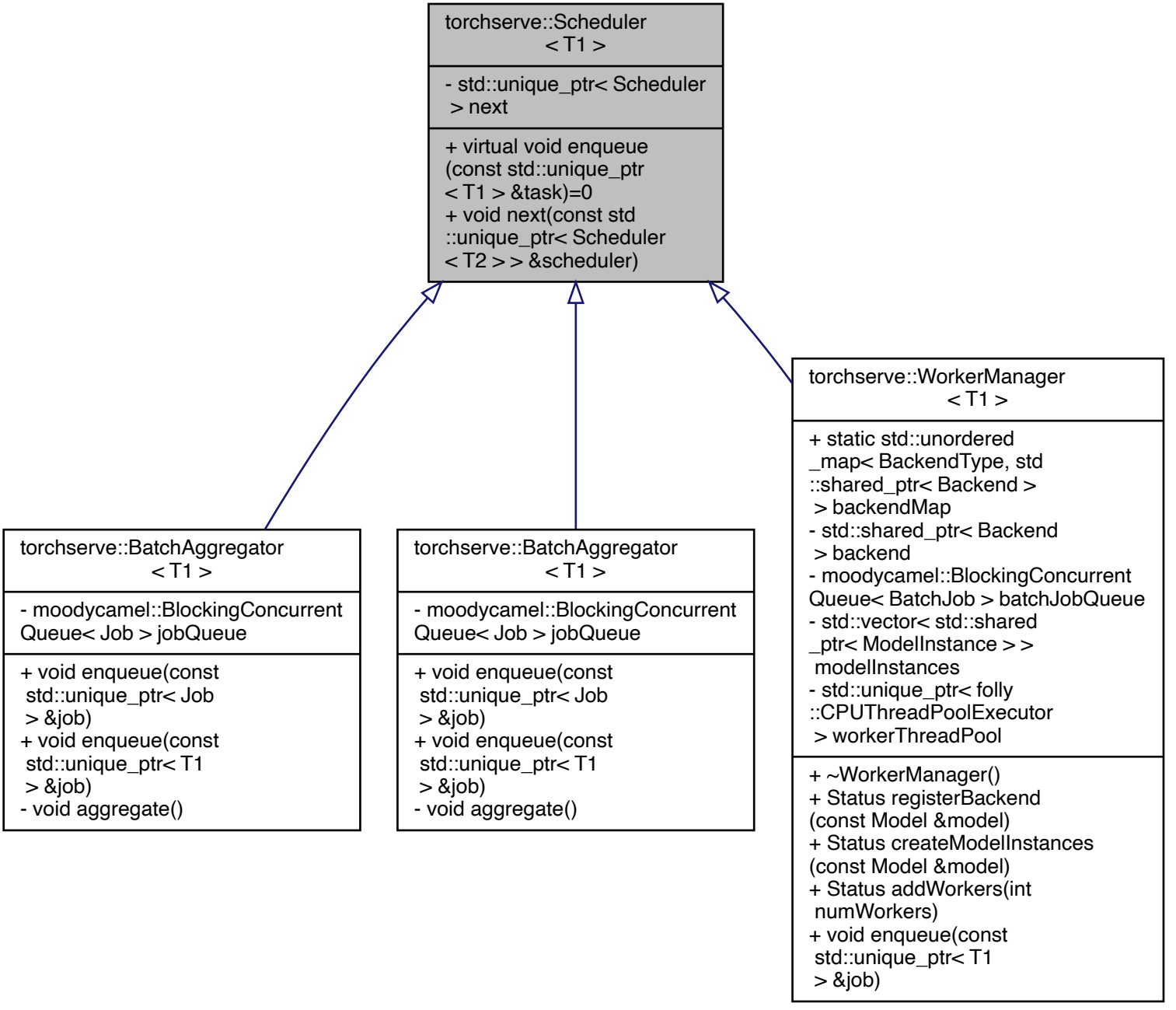
torchserve::Scheduler
< T1 >
- std::unique_ptr< Scheduler
> next
+ virtual void enqueue
(const std::unique_ptr
< T1 > &task)=0
+ void next(const std
::unique_ptr< Scheduler
< T2 > > &scheduler)
torchserve::BatchAggregator
< T1 >
- moodycamel::BlockingConcurrent
Queue< Job > jobQueue
+ void enqueue(const
std::unique_ptr< Job
> &job)
+ void enqueue(const
std::unique_ptr< T1
> &job)
- void aggregate()
torchserve::BatchAggregator
< T1 >
- moodycamel::BlockingConcurrent
Queue< Job > jobQueue
+ void enqueue(const
std::unique_ptr< Job
> &job)
+ void enqueue(const
std::unique_ptr< T1
> &job)
- void aggregate()
torchserve::WorkerManager
< T1 >
+ static std::unordered
_map< BackendType, std
::shared_ptr< Backend >
> backendMap
- std::shared_ptr< Backend
> backend
- moodycamel::BlockingConcurrent
Queue< BatchJob > batchJobQueue
- std::vector< std::shared
_ptr< ModelInstance > >
modelInstances
- std::unique_ptr< folly
::CPUThreadPoolExecutor
> workerThreadPool
+ ~WorkerManager()
+ Status registerBackend
(const Model &model)
+ Status createModelInstances
(const Model &model)
+ Status addWorkers(int
numWorkers)
+ void enqueue(const
std::unique_ptr< T1
> &job)