Data Flow: Partitioning, Processing & Merging

Input Dataset (Large File)

Dataset Load & Analysis

Partitioning Engine

Partition 1 (10K samples)	**Partition 2** (10K samples)	**Partition N** (10K samples)

Ray Worker 1
- Load Partition
- Apply Ops
- Save Checkpt
- Log Events

Ray Worker 2
- Load Partition
- Apply Ops
- Save Checkpt
- Log Events

Ray Worker N
- Load Partition
- Apply Ops
- Save Checkpt
- Log Events

Processed P1 + Checkpoints

Processed P2 + Checkpoints

Processed PN + Checkpoints

Result Merger

Final Dataset + Event Logs + Performance