Table 1: MDG Dataset Schema: Attribute Groups and Descriptions (No Missing Fields)

Category	Field	Description
dataset_args	test_size	Proportion of data reserved for testing
	scaler	Scaling method (standard, min-max, none)
	seed	Random seed for reproducibility
	split_strategy	IID or non-IID client assignment
	distribution_param	Dirichlet α, shard count, or quantity skew parameter
clustering_args	run_id	Reference to the experiment run
	clusters_k	Number of clusters
	clustering_init	Initialization method (k-means++, random)
	clustering_n_init	Number of initializations for stability
	clustering_tol	Convergence tolerance threshold
runs	run_id	Unique execution identifier
	dataset	Dataset name used (MNIST, CIFAR-10, etc.)
	task_type	Classification, Regression, or Clustering
	model_type	Model family (CNN, RNN, MLP, MobileNetV2, RF...)
	num_clients	Number of clients simulated
	num_rounds	Training rounds executed
	learning_rate	Optimizer learning rate
	batch_size	Training batch size
	local_epochs	Local update epochs per client
	hidden_layers	Neural network depth
	activation	Activation function name
	optimizer	Optimizer type (SGD, Adam, RMSProp)
rounds	run_id	Reference to run configuration
	round	Round index
	global_loss	Aggregated global loss
	scheduled_clients	Clients selected for the round
	participating_clients	Clients contributing updates
	dropped_clients	Clients removed due to failures or timeout
clients	run_id	Run reference
	client_id	Client/device ID
	participated	Whether the client contributed this round
	data_distribution_json	Client dataset label distribution
	samples_count	Number of data samples on device
qos / metrics	accuracy, f1, precision, recall	Classification performance metrics
	rmse, rmse_original_units	Regression metrics
	silhouette, inertia, ari, nmi	Clustering metrics
	loss	Training loss
	computation_time_s, cpu_time_s	Total compute time usage
	comm_bytes_up, comm_bytes_down	Uplink and downlink communication volume
	memory_util, gpu_util	Hardware utilization over time
	gpu_memory_util/used	GPU memory usage during training
	units	Units associated with metric values (ms, %, bytes)