

CLEANSING_RUN_METADATA		
string	run_id	PK
datetime	created_utc	
string	source_db_path	
string	output_db_path	
int	workers	
int	bin_seconds	
int	max_forward_fill_s	
int	stale_after_s	
string	fill_method_mode	
int	dq_enabled	
float	dq_outlier_threshold_pct	
string	exchange_filter_csv	
string	symbol_filter_csv	
int	total_pairs	
int	total_rows	
int	observed_rows	
int	forward_fill_rows	
int	interpolation_rows	
int	missing_rows	
float	runtime_seconds	

aggregates

CLEANSINGPAIRSUMMARY		
bigint	id	PK
string	run_id	FK
string	exchange_id	
string	symbol	
int	source_tick_count	
int	bins_total	
int	bins_observed	
int	bins_forward_fill	
int	bins_interpolation	
int	bins_missing	

summarizes

CLEANSING_DQ_RUN_SUMMARY		
string	run_id	PK
datetime	created_utc	
int	pair_count	
int	pairs_with_any_issue	
int	total_issue_counter_sum	
int	total_negative_value_rows	
int	total_outlier_return_count	

validates

CLEANSING_DQ_SUMMARY		
bigint	id	PK
string	run_id	FK
string	exchange_id	
string	symbol	
int	source_row_count	
int	null_ingestion_ts_count	
int	invalid_ingestion_ts_count	
int	null_effective_price_count	
int	non_positive_effective_price_count	
int	rows_with_negative_values_count	
int	bid_gt_ask_count	
int	duplicate_tick_count	
int	outlier_return_count	
int	issue_counter_sum	

produces

summarizes_pair

CLEANSED_MARKET		
bigint	id	PK
string	run_id	FK
datetime	bucket_start_utc	

quality_context

datetime	bucket_start_utc
bigint	bucket_epoch_s
string	exchange_id
string	symbol
int	bin_seconds
int	max_forward_fill_s
string	fill_method_mode
float	price
datetime	source_ingestion_ts_utc
bigint	source_ingestion_epoch_s
int	tick_count_in_bin
string	fill_method
float	age_seconds
int	is_stale
int	is_missing

