

ETL KPI Report		
Performance & Throughput KPIs		
KPI	Description	Example / How to Measure
Job Duration	Total time taken for an ETL job to complete.	EndTime – StartTime (Glue job metrics / CloudWatch logs)
Data Volume Processed	Amount of data extracted and loaded.	Input/Output record count or S3 bytes scanned/written
Rows Processed per Second	Processing throughput.	TotalRowsProcessed / JobDurationSeconds
Parallelism / Executors Utilization	Measure of compute efficiency.	Active vs requested executors in Spark UI
Shuffle Size & Spill Metrics	I/O overhead due to repartitioning or skew.	Spark UI / Glue job metrics
S3 I/O Wait Time	Time spent on reading/writing to S3.	CloudWatch + Glue job logs
Data Quality KPIs		
KPI	Description	Example / How to Measure
Record Count Consistency	Difference between source and target record counts.	SourceCount vs TargetCount
Duplicate Rate	% of duplicate rows identified and removed.	(DuplicatesRemoved / TotalRows) × 100
Null Value Ratio	Missing or incomplete field rate.	NullCount(field) / TotalRows
Data Drift	Change in schema, format, or distribution.	Compare column stats (mean, distinct, nulls) over time
Constraint Violations	Violations of PK/FK/business rules.	PyDeequ or custom Spark checks
QC Pass Rate	% of total checks that passed.	(PassedChecks / TotalChecks) × 100
Reliability & Availability KPIs		
KPI	Description	Example / How to Measure
Job Success Rate	% of successful runs vs total runs.	(SuccessfulRuns / TotalRuns) × 100
Failure Rate / Retry Count	Frequency of job failures or retries.	Glue job run metrics / CloudWatch
MTTR (Mean Time to Recovery)	Average time to fix a failed job.	Sum(RecoveryTime)/#incidents
SLA Compliance	% of jobs meeting completion SLAs.	JobsCompletedWithinSLA / TotalJobs
Pipeline Availability	Uptime of the overall ETL system.	(AvailableTime / TotalScheduledTime) × 100
Cost & Resource Efficiency KPIs		
KPI	Description	Example / How to Measure
Cost per GB Processed	Normalized cost metric.	JobCost / DataProcessedGB
DPU Utilization (Glue)	Measure of compute utilization efficiency.	ActiveDPUs / AllocatedDPUs
S3 Read/Write Costs	Track data movement cost.	S3 metrics or Athena bytes scanned
Compute Idle Time	Non-productive time waiting for resources.	Executor Idle Time / Total Duration

KPI	Description	Example / How to Measure
Storage Cost per Partition	Cost efficiency for Iceberg tables.	PartitionSize / TotalPartitions

Metadata & Governance KPIs

KPI	Description	Example / How to Measure
Lineage Completeness	% of datasets with full lineage captured.	LineageCaptured / TotalDatasets
Schema Evolution Rate	How often schema changes occur.	#SchemaChanges / Period
Snapshot Tag Coverage (Iceberg)	% of snapshots tagged with metadata.	TaggedSnapshots / TotalSnapshots
Access Policy Compliance	% datasets aligned with FGAC policies.	DatasetsCompliant / TotalDatasets
Data Retention SLA	Adherence to data expiration policies.	ExpiredSnapshotsDeleted / ExpectedDeletions

User-Facing & Operational KPIs

KPI	Description	Example / How to Measure
Data Latency / Freshness	Time between generation and availability.	IngestionTimestamp – SourceEventTime
Query Performance (Athena)	Average query duration or bytes scanned.	Athena QueryStats
Data Product Readiness	% datasets meeting completeness thresholds.	DatasetsPassingChecks / TotalDatasets
Alert-to-Resolution Time	How quickly alerts are handled.	TimeFromAlertToResolution