

DX 데이터 거버넌스 운영 표준서

적용 범위: MES/NAS 원천 데이터 → Backend Daemon(1~11) → PostgreSQL → FastAPI → Streamlit → 이메일 알림 체계

기준 시간대: KST(Asia/Seoul)

1. 문서 목적

본 문서는 제조 현장에서 생성되는 테스트/설비 로그 데이터를 표준화하고, 운영 시스템(FastAPI/Streamlit)에서 신뢰 가능한 KPI를 안정적으로 제공하기 위한 **데이터 거버넌스 실행 기준**을 정의한다.

핵심 목표는 다음과 같다.

1. **정확한 KPI 산출:** OEE, 비가동, Fail 반복, 손실시간 지표 신뢰성 확보
 2. **운영 연속성 확보:** DB 단절·지연 상황에서도 무중단에 가깝게 복원
 3. **변경 통제:** 스키마/API 변경 시 영향도 기반 승인·배포 관리
 4. **보안·추적성 확보:** RBAC, 감사로그, 데이터 품질 이슈 이력 관리
-

2. 시스템 범위 및 데이터 플로우

2.1 원천(Source)

- **MES 계층**
 - FCT_test_log
 - Vision_test_log
- **NAS/머신로그 계층**
 - FCT_test_details
 - FCT_Machine_log
 - Vision_Machine_log
 - PLC_Machine_log

2.2 가공(Backend Data-mining Layer)

- a1: history

- a2: FCT_test_sum
- a3: Vision_test_sum
- c1: FCT_test_details
- d1: FCT1~4_Machine_log
- d2: Vision1~2_Machine_log
- d3: PLC_Machine_log
- e1_1: FCT_operation_ct
- e2_1: Vision_operation_ct
- e1_2: FCT_test_ct
- e2_2: Vision_test_ct
- e5: FCT_core_method_status
- e6: FCT_non_pd_worst_case
- e7: FCT_pd_worst_case
- f: FCT_whole_details
- g_c: Operation_criteria
- g_f: FCT_non_operation
- g_v: Vision_non_operation
- h: sparepart_repl_timing

2.3 서비스(Serving Layer)

- PostgreSQL
- FastAPI
- Streamlit
- E-mail System (Shift별 알림/리포트 송부)

2.4 백엔드 서비스(1~11)

- 1_amount
- 2_pass_pct
- 3_FCT_fail_items

- 4_Vision_fail_items
 - 5_master_test
 - 6_FCT_worst
 - 7_assemble_fail
 - 8_MES_fail
 - 9_plan_non_stop_time
 - 10_P_type_chg_timing
 - 11_OEE
-

3. 데이터 표준 정책

3.1 공통 필드 표준

(1) shift_type

- 허용값: day, night
- 금지: 그 외 문자열(대소문자 변형 포함)
- 저장 규칙: 소문자 고정

(2) station

- 허용값(표준 코드):
 - FCT: FCT1, FCT2, FCT3, FCT4
 - Vision: Vision1, Vision2
- 표준 외 값은 사전 매핑 테이블로 변환 후 저장

(3) end_day / prod_day

- 운영 호환성을 위해 **이중 정책** 적용:
 - 원천 호환키: end_day = YYYYMMDD(text)
 - 분석 표준키: prod_day = DATE 또는 YYYY-MM-DD 파생 컬럼
- 리포트/대시보드는 기본적으로 prod_day + shift_type을 주 집계키로 사용

(4) 시간 컬럼

- from_time, to_time, end_time, updated_at 등은 KST 기준 처리
 - 비교·정렬용 내부 키는 timestamp 정규화(초 단위 또는 half-up 규칙 문서화)
-

3.2 네이밍 및 타입 표준

- 네이밍: snake_case
 - 코드형: TEXT
 - 수치형: INTEGER/NUMERIC
 - 시간형: TIMESTAMP(가능 시 timezone 포함)
 - 불리언/플래그: BOOLEAN 또는 0/1 명시 규칙 중 1개를 테이블별 고정
-

4. 증분 처리·중복 방지 표준

4.1 기본 원칙

1. 테이블별 PK/증분 키를 고정 정의
2. last_pk 기반 incremental fetch
3. seen_pk 캐시로 재처리 중복 차단
4. DELETE/TRUNCATE 금지(운영 테이블 초기화 금지)

4.2 대표 증분 키 예시 (운영 문서 기준)

- (end_day, station, from_time)
- (end_day, station, end_time)
- (end_day, end_time, barcode_information)
- (end_day, end_time_norm, md5(contents)) (특수 케이스)

| 실제 적용키는 Backend 1~11 각 서비스 사양서를 기준으로 고정한다.

5. 데이터 품질(DQ) 관리

5.1 DQ KPI 정의

1. **Completeness**

- 식: 필수컬럼 non-null 레코드 / 전체 레코드
- 기준: $\geq 99.5\%$

2. Uniqueness

- 식: 1 - (중복 PK 레코드 / 전체 레코드)
- 기준: 중복률 $\leq 0.1\%$

3. Validity

- 식: 코드/범위/형식 검증 통과 레코드 / 전체 레코드
- 기준: $\geq 99.9\%$

4. Timeliness

- 식: 이벤트 발생~DB 적재 지연(P95)
- 기준: ≤ 5 분

5. Freshness(Heartbeat)

- 식: shift별 누락 heartbeat 건수
- 기준: 0건

5.2 DQ 운영 주기

- 실시간: heartbeat, 적재 지연 감시
- Shift 종료 +10분: shift DQ report 생성
- 주간: DQ 추세/개선과제 리뷰

5.3 DQ 예외 처리

- 완전 NULL 레코드: 즉시 격리(Quarantine)
- 코드 불일치: 매핑/정정 큐로 이동
- 재처리 실패: Incident 등록 + 원인분류(소스/파이프라인/스키마)

6. 운영 안정성·장애 대응 표준

6.1 장애 내성(Resilience)

- DB 접속 실패 시 무한 재시도(블로킹)

- 실행 중 단절 발생 시 재연결 후 이어서 처리
- 연결 풀 최소화 (pool_size = 1, max_overflow=0) 권장
- 세션별 work_mem 강제 설정(기본 4MB 정책)

6.2 로깅 표준

- 공통 로그 레벨/태그: [BOOT] [INFO] [WARN] [ERROR] [RETRY]
- 필수 로그 항목:
 - 마지막 PK
 - 신규 fetch 건수
 - upsert 건수
 - heartbeat 상태
 - 장애 시 stack trace + 복구 결과

6.3 Incident SLA

- **Sev1 (중대)**: 30분 내 탐지/공유, 2시간 내 우회복구, 24시간 내 RCA
 - **Sev2 (일반)**: 4시간 내 복구, 3영업일 내 재발방지 조치
-

7. 보안·권한 정책

7.1 RBAC

- **Admin**: 스키마/정책 변경 승인, 운영계정 관리
- **Engineer**: 백엔드/파이프라인 운영 및 장애조치
- **Viewer(현장작업자/엔지니어)**: 조회·리포트 확인 중심

7.2 데이터 보호

- 개인정보 최소수집 원칙
- 사용자 식별정보 필요 시 마스킹
- 비밀번호/API 토큰은 코드 하드코딩 금지(환경변수/비밀관리 사용)

7.3 감사 로그

- 조회/수정/설정 변경 이력 기록
 - 감사로그 보존기간: 최소 12개월
-

8. 변경관리(Change Management)

8.1 변경 등급

- **Minor:** 컬럼 추가, 선택 파라미터 추가(하위호환 유지)
- **Major:** 컬럼 삭제/의미 변경/필수화(비호환)

8.2 변경 승인 게이트

1. 변경 요청서 작성
2. 영향도 분석(테이블/API/대시보드/리포트)
3. 테스트 증빙(단위·통합·회귀)
4. 승인자 결재
5. 배포 윈도우 적용
6. 배포 후 검증/롤백 확인

8.3 API 버전 정책

- v1 유지 원칙 + v2 병행
 - 비호환 변경은 사전 공지(최소 2주)
-

9. 데이터 보존·파기 정책

- Raw 계층: 13개월
 - Curated/Serving 계층: 24개월
 - Audit log: 12개월 이상
 - 자동삭제는 계층/테이블별 정책에 따라 수행하며,
삭제 전 백업 및 승인 절차를 필수 적용한다.
-

10. RACI (역할 책임 체계)

- **A (Accountable):** DX Lead (최종 책임/승인)
 - **R (Responsible):** Data/Backend Engineer (개발·운영·장애조치)
 - **C (Consulted):** MES 담당, 품질팀, 생산기술, IT 보안
 - **I (Informed):** 현장 리더, 리포트 수신자, 운영 이해관계자
-

11. 공장 확산(Rollout) 표준

1. Pilot Site에서 PoV 검증
2. 공통 템플릿화(스키마/API/DQ/SLA/Runbook)
3. 사이트별 로컬 파라미터 분리(스테이션명, shift 정책, 알림 대상)
4. Site B/C 확산 시 동일 KPI 체계로 비교 운영
 - 원본 데이터를 그대로 파싱하는 경우, 컬럼 end_day를 써서 실제 시간을 입력할 것
 - 파싱된 데이터를 가공하는 테이블의 경우 컬럼 prod_day를 써서 주 야간 생산 shift를 나누는 type_shift라는 컬럼과 같이 사용될 것
 - 날짜는 date형식 또는 문자형 YYYYMMDD + 예외 규칙 허용
 - 시간은 timestamp 형식일 것
 - NULL 이 허용되는 특정 컬럼 이외에는 NULL을 허용하지 말 것
 - 시험 단위는 무시하고 측정값만 기입될 것, 예시 360M/bit > 360
 - 데이터는 컬럼 end_day와 prod_day의 속성값이 현재 시각 12개월 이상인 경우 자동 삭제될 것
- **KPI Definition:** OEE/불량률 같은 지표의 공식/집계 기준 문서
 - $OEE = Availability \times Performance \times Quality$
 - $Run\ time = 12\text{시간} - Planned\ Stop\ Time$
 - $Performance = (Ideal\ Cycle\ Time \times Total\ Count) / Run\ Time$
 - $Quality = Good\ Count / Total\ Count$