

## 아키텍처 개요

한국기술마켓 프로젝트는 마이크로서비스 아키텍처로 구성됩니다:

- 컨테이너 오케스트레이션: Amazon ECS
- 서비스 디스커버리: AWS Cloud Map
- 비동기 처리: Amazon SQS
- 데이터베이스: Amazon RDS
- API 관리: Amazon API Gateway
- CI/CD: AWS CodeBuild + ECR

## 권장 모니터링 메트릭

### 1. ECS 서비스별 메트릭

각 마이크로서비스(기술검색, 매칭엔진, 문서처리, 알림)에 대해:

메트릭	설명	권장 임계값
CPUUtilization	CPU 사용률	경고: >70%, 위험: >85%
MemoryUtilization	메모리 사용률	경고: >75%, 위험: >90%
RunningTaskCount	실행 중인 태스크 수	최소 2개 유지 권장
DesiredTaskCount	목표 태스크 수	RunningTaskCount와 일치 확인

알람 설정 권장:

- RunningTaskCount < DesiredTaskCount 시 즉시 알림
- CPU/메모리 임계값 초과 시 Auto Scaling 트리거

### 2. API Gateway

메트릭	설명	권장 임계값
Latency	응답 지연 시간	경고: >1s, 위험: >3s
4XXError	클라이언트 오류율	경고: >5%, 위험: >10%
5XXError	서버 오류율	경고: >1%, 위험: >5%
Count	요청 수	트래픽 패턴 모니터링용

알람 설정 권장:

- 5xx 오류율 5% 초과 시 즉시 알림
- 지연 시간 3초 초과 시 병목 분석

### 3. SQS 문서 처리 큐

메트릭	설명	권장 임계값
ApproximateNumberOfMessagesVisible	대기 메시지 수	경고: >50, 위험: >200
ApproximateAgeOfOldestMessage	최대 대기 시간	경고: >10분, 위험: >30분

NumberOfMessagesSent	전송 메시지 수	처리량 모니터링용
ApproximateNumberOfMessagesNotVisible	처리 중 메시지	동시 처리량 확인

#### DLQ (Dead Letter Queue) 모니터링:

- DLQ 메시지 >0 시 즉시 알림 (처리 실패 감지)

## 4. RDS 데이터베이스

메트릭	설명	권장 임계값
CPUUtilization	CPU 사용률	경고: >70%, 위험: >85%
DatabaseConnections	활성 연결 수	max_connections의 80% 경고
ReadLatency	읽기 지연	경고: >20ms, 위험: >50ms
WriteLatency	쓰기 지연	경고: >20ms, 위험: >50ms
FreeStorageSpace	남은 스토리지	경고: <20%, 위험: <10%

#### 알람 설정 권장:

- 스토리지 10% 미만 시 즉시 확장
- 연결 수 80% 초과 시 연결 풀 점검

## 권장 비즈니스 KPI

기술이전 플랫폼 특성에 맞는 비즈니스 메트릭:

KPI	측정 방법	목표
기술검색 응답 시간	API Gateway Latency (검색 엔드포인트)	<500ms
문서 처리 완료 시간	SQS 메시지 수명	<5분
매칭 알고리즘 응답 시간	매칭엔진 서비스 Latency	<2초
일별 기술검색 요청 수	API Gateway Count	추세 모니터링
문서 처리 성공률	처리 완료/전체 비율	>99%

## CloudWatch 대시보드 구성 권장

Dashboard: KotechMarket-Operations

- API Gateway Overview
  - Request Count (시간별)
  - Latency p50/p90/p99
  - Error Rates (4xx, 5xx)
- ECS Services Health
  - 기술검색 서비스 (CPU, Memory, Tasks)
  - 매칭엔진 서비스 (CPU, Memory, Tasks)

```
|   └── 문서처리 서비스 (CPU, Memory, Tasks)
|   └── 알림 서비스 (CPU, Memory, Tasks)
| └── Document Processing Pipeline
|   ├── SQS Queue Depth
|   ├── Processing Latency
|   └── DLQ Message Count
└── Database Performance
  ├── RDS CPU/Connections
  └── Read/Write Latency
```

## 로그 수집 권장사항

### ECS 컨테이너 로그

각 서비스별 CloudWatch Log Group 구성:

- /ecs/kotech-market/tech-search
- /ecs/kotech-market/matching-engine
- /ecs/kotech-market/doc-processor
- /ecs/kotech-market/notification

### 로그 분석 쿼리 예시

오류 패턴 분석:

```
fields @timestamp, @message
| filter @message like /ERROR/
| stats count() by bin(1h)
```

느린 API 식별:

```
fields @timestamp, endpoint, duration
| filter duration > 1000
| sort duration desc
| limit 20
```