

요구조건 식별서

구분		내용		비고
사업명		의료 쇼핑을 마케팅 전략을 위한 데이터 인프라 구축		
사업목표		<ul style="list-style-type: none"> - 고가용성, 탄력성, 안정성 있는 인프라 구축 - 자동화된 서버 배포 파이프라인 구축 - 자동 데이터 백업 및 비용 효율적인 장기 보관 시스템 구축 - 체계적인 데이터 수집 및 분석 파이프라인 구축 		
사업 범위	기간		계약일로부터 60일	
	범위	인프라 요구조건 식별	<ul style="list-style-type: none"> - 기존 2-Tier 인프라에서 AWS의 3Tier 환경으로 재구축 - 고가용성을 위해 멀티티어 환경 구축 - 갑자기 증가하는 트래픽에도 서비스를 지속적으로 제공할 수 있는 탄력성 있는 환경 구축 - 장애가 발생해도 빠른 대응이 가능한 고가용성 환경 구축 - 오래된 데이터는 비용 효율적으로 보관할 수 있는 환경 구축 	
		CI/CD 요구조건 식별	지속적으로 서버 업데이트가 가능한 자동화된 서버 배포 파이프라인 필요	
		데이터 분석 요구조건 식별	<ul style="list-style-type: none"> - 데이터베이스에 저장된 사용자 정보와 구매 목록, 상품 카테고리 등의 상관 관계를 파악할 수 있는 분석 필요 - 예민한 데이터의 경우 분석 전 처리 - 데이터 분석 전 올바른 데이터인지 검사해 자원 절약 - 매 분기마다 진행될 설문 조사 결과 분석을 통하여 광고 유입 경로 추적 - 특정 시간에 얼마나 많은 사용자들이 해당 쇼핑물을 사용하는지 로그를 통해 분석 	<ul style="list-style-type: none"> - 연령대별 유입 경로 - 카테고리 별 판매 수 - 연령대별 카테고리 판매량 - 카테고리별 판매율과 리뷰의 상관관계 - 시간대 별 클릭 스트림 수
		백업 요구조건 식별	<ul style="list-style-type: none"> - 일본 법적 규정사항(전자상거래법), 법적 분쟁 대응, 고객 문의 대응 등으로 인해 데이터베이스의 데이터를 지속적으로 백업 및 보관할 수 있는 환경 필요 - 일본의 법적 규정사항으로 인해 구매, 고객 데이터의 5년간 장기 보관 필요 - 법적 분쟁 대응을 위해 리뷰 데이터의 5년간 장기 보관 필요 - 비용 효율적인 백업 장기 보관 환경 필요 	
		모니터링 및 알람 요구조건 식별	<ul style="list-style-type: none"> - Web, WAS 서버의 지표를 모니터링할 수 있는 환경 구축 필요 - 이벤트 발생 시 담당자에게 즉시 알림이 전송되는 환경 필요 	
		보안 요구조건 식별	자른 웹 공격에 대응할 수 있는 보안 솔루션 도입 필요	
사업 요구사항	시스템 구축 요구사항	Infrastructure	<ul style="list-style-type: none"> - 웹 애플리케이션(3-Tier) 환경 구성 - private subnet에 Web, WAS, DB 서버 구축 - 장애 대응을 위한 DB 이중화 구현 	
		Server	<ul style="list-style-type: none"> - Web 서버: EC2 - 갑자기 증가하는 트래픽에도 유연하게 대응할 수 있도록 Auto Scaling 도입 - 도메인으로 접속 가능한 서버 환경 구축 - WAS 서버: EC2 - 갑자기 증가하는 트래픽에도 유연하게 대응할 수 있도록 Auto Scaling 도입 - 도메인으로 접속 가능한 서버 환경 구축 - DB 서버: RDS(MySQL) - 고객 table, 구매 table, 리뷰 table의 데이터 - 도메인으로 접속 가능한 서버 환경 구축 	
		Network	<ul style="list-style-type: none"> - 일본 고객을 타겟으로 한 온라인 쇼핑몰로, 도쿄(ap-northeast-1) 리전에 환경 구축 - 가용영역 2개(az-a, az-c) - 서브넷 - public subnet: 2개 - private subnet: 6개 - 외부에서 바로 Web, WAS, DB 서버에 접근할 수 없도록 서브넷 분리 	
	CI/CD	서버 배포 파이프라인 자동화	<ul style="list-style-type: none"> - 지속적으로 업데이트된 서버 배포가 가능한 CI/CD 환경 구축 - AWS CodeCommit, AWS CodePipeline, AWS CodeBuild, AWS CodeDeploy를 활용한 코드 파이프라인 구축 - 서버가 자동적으로 배포될 수 있도록 자동화 환경 구축 	
	백업	데이터베이스 백업	<ul style="list-style-type: none"> - 데이터베이스의 데이터를 매일 자동 백업할 수 있는 환경 구축 - 백업 데이터 장기 저장 시 비용 절감을 위해 스토리지 클래스 분리 필요 - Amazon S3 Standard: ~1개월 데이터 보관 - Amazon S3 Standard-IA: 1개월~3개월(1분기) 데이터 보관 - Amazon S3 Glacier Flexible Retrieval: 3개월(1분기)~2년 데이터 보관 - Amazon S3 Glacier Deep Archive: 2년~5년 - 스토리지 클래스 전환 및 데이터 삭제(5년 후 자동 삭제) 시 수행 주기 정책 활용 	
	보안 구축 요구사항	시스템 보안	<ul style="list-style-type: none"> - 인스턴스 단위로 보안 그룹 설정 - 도메인 접속 시 https로 접속 	
		보안솔루션	<ul style="list-style-type: none"> - SQL 인젝션 차단 등이 가능한 AWS WAF 도입 - Route 53에서 도메인으로 인증서를 발급하고 ALB와 연동해 HTTPS 통신 보장 	
	모니터링	Metric 정의	CPU 사용량	
		모니터링 구축	이벤트 및 장애 발생 여부를 모니터링하기 위해 CloudWatch를 통한 모니터링 환경 구축	
		Alarm 전달 방법	<ul style="list-style-type: none"> - 이벤트 발생 시 운영자에게 Slack을 통해 알람 전달 - 이벤트: CPU 75% 이상 	
	데이터 분석	데이터 추출	<ul style="list-style-type: none"> - 데이터 분석을 위한 데이터베이스 테이블 추출 - 분석을 위한 분기 별 설문 조사 데이터 변경 - 실시간 클릭 로그 스트림 추출 	- RDS 데이터 추출 시 민감 데이터 제거 후 S3 이동
		데이터 전처리	<ul style="list-style-type: none"> - 빠른 분석을 위해 parquet로 변환 - 설문조사 데이터의 경우 더 정확한 분석을 위해 파싱 진행 - 설문조사 데이터 파싱 진행 후 해당 Lambda에서 바로 parquet 변환 - 로그 스트림의 경우 Amazon Kinesis의 동적 파티셔닝으로 파티셔닝 설정 	<ul style="list-style-type: none"> - 설문조사 파싱 및 파케이 변환을 위한 Lambda 코드는 AWS Wrangler 사용 - Pandas 등 외부 라이브러리 추가할 필요가 없을 뿐만 아니라 AWS Glue와 같은 서비스와도 통합되어 사용
		데이터 분석	- AWS Athena를 사용한 데이터 분석	<ul style="list-style-type: none"> - AWS Athena 실행 전 AWS Crawler를 통해 메타 데이터 추출 - 파티셔닝의 경우 데이터가 저장된 S3의 Key를 기준으로 진행
		데이터 시각화	<ul style="list-style-type: none"> - Quick Sight을 사용해 분석된 데이터 시각화 - 비용 절감과 데이터베이스의 부하도 감소를 위해 spice 방식을 활용하여 데이터 가져오기 	
		데이터 파이프라인 자동화	<ul style="list-style-type: none"> - 설문조사 데이터를 S3에 저장 시 자동으로 파싱 및 parquet 형식으로 변환 후 분석을 위해 S3에 저장 - Glue Workflow를 사용해 데이터 분석 흐름 자동화 	
		데이터 파이프라인 알람	<ul style="list-style-type: none"> - AWS Glue 작업 상태 변경 시 담당자에게 Slack 알람 제공 - Glue 작업 상태 변경을 확인하기 위해 AWS EventBridge 사용 - 알람 제공 기준 - Glue Job 실패 - Glue Job 성공했으나, 기준 시간 초과시 	<ul style="list-style-type: none"> - 성공했으나 알람을 담당자에게 제공하는 이유 - AWS Glue의 경우 진행되는 시간에 따라 비용 책정 - 때문에 기준치를 초과해 진행된 AWS Glue는 최적화 등을 통해 비용 줄임
기타 요구 사항	기술 내재화		정기 점검 가이드, 비정기 점검 가이드, 변경 관리 가이드, 백업 관리 가이드, 보안 관리 가이드, 장애 관리 가이드, 문제 관리 가이드	
	서비스 확장		추후 AWS 환경에서 다른 서비스와 통합 및 확장 가능성 존제	

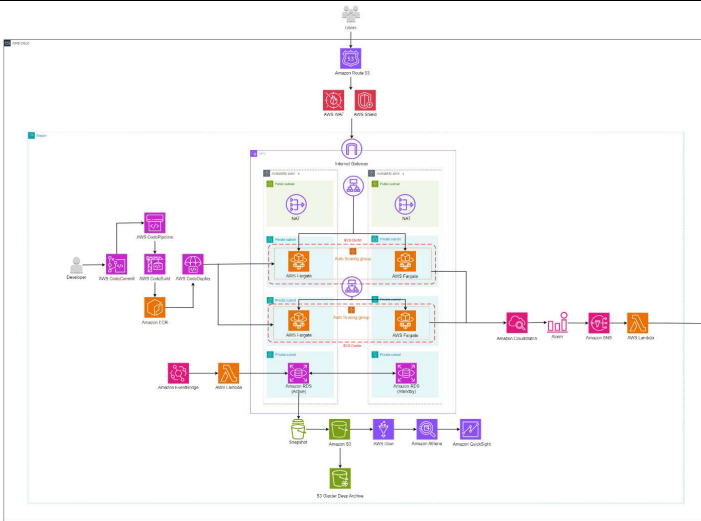
프로젝트 계획서

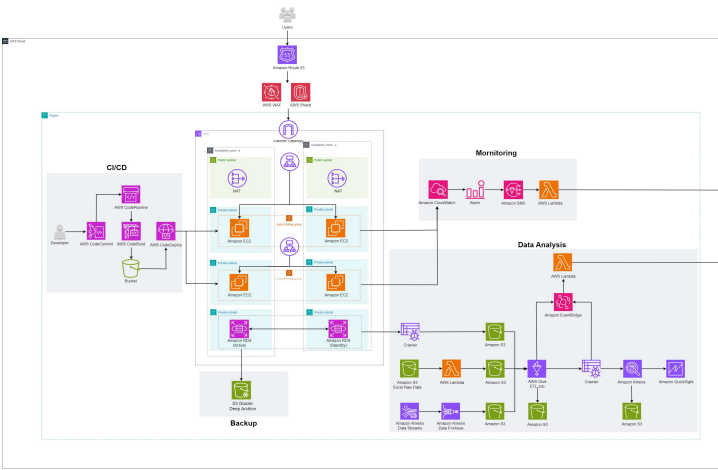
No	Tasks		담당자	비고	Schedule			산출물
					Start	Finish	M/D	
분석								
1	프로젝트 투입		김민경		24.04.24	24.05.15	22 M / D	요구조건 식별서
	요구조건 식별				24.04.24	24.04.24	22 M / D	
		구축 요구조건 식별		3-Tier로 계층 분리 필요 고가용성, 안정성, 탄력성 필요	24.04.24	24.04.24	22 M / D	
		백업 요구조건 식별		데이터베이스 백업 구축 필요	24.04.24	24.04.24	22 M / D	
		CI/CD 요구조건 식별		자동화된 서버 배포 파이프라인 필요 아티팩트 저장 환경 필요	24.04.25	24.04.26	22 M / D	
		데이터 분석 요구조건 식별		데이터베이스, 설문조사, 로그 스트림 데이터별 수집 경로 체계화 필요 데이터베이스, 설문조사 데이터 품질 검사 필요 수집된 데이터 정제 작업 필요 정제된 데이터 분석 작업 필요 분석된 데이터 시각화 작업 필요 데이터 파이프라인 이벤트 발생 시 알람 환경 필요	24.04.29	24.05.15	22 M / D	
		모니터링 및 알람 요구조건 식별		Web, WAS 서버 지표의 모니터링 환경 필요 이벤트 발생 시 담당자에게 알람 전송 환경 필요	24.04.29	24.04.29	22 M / D	
		보안 요구조건 식별		웹 공격에 대응할 수 있는 보안 솔루션 도입 필요	24.04.29	24.04.29	22 M / D	
설계								
2	아키텍처		김민경		24.04.26	24.06.12	22 M / D	아키텍처(구성도)
	AWS Resource 구성안			3-Tier 환경 설계	24.04.26	24.04.26	22 M / D	
	CI/CD구성안			Web, WAS 서버 별 자동 배포 환경 설계	24.04.26	24.04.26	22 M / D	
		데이터 분석 파이프라인 구성안		데이터베이스, 설문조사, 로그 스트림 데이터별 수집 환경 필요 데이터베이스, 설문조사 데이터의 품질 검사 환경 필요 데이터 정제 환경 필요 데이터 분석 및 시각화 환경 필요 데이터 파이프라인 자동화 환경 필요 데이터 파이프라인 알람 환경 필요	24.04.29	24.05.15	22 M / D	
	모니터링 및 알람 구성안			Web, WAS 서버의 CPU 지표 모니터링 및 알람 환경 필요	24.04.29	24.05.15	22 M / D	
	백업 구성안			자동화된 데이터 베이스 데이터의 백업 환경 필요	24.04.29	24.05.01	22 M / D	
	보안 구성안			웹 보안 환경 필요	24.04.29	24.05.01	22 M / D	
구축								
3	구축		팀원 전체		24.05.02	24.06.11	22 M / D	구축된 인프라 결과물
	전체 계획과 수행 범위		김민경		24.05.02	24.06.04	22 M / D	
		네트워크 환경 구성		테라폼을 통한 네트워크 환경 구성 - VPC, 서브넷, 서브넷 라우팅 테이블, IGW, EIP, Route 53, ACM 구축	24.05.02	24.05.06	22 M / D	
		서버 환경 구성		테라폼을 통한 네트워크 환경 구성 - Web&WAS 서버, 시작 템플릿, ALB, ALB SG, TG, ASG, ASG SG, RDS, DB SG, DB Subnet Group구축, SSM 연결	24.05.02	24.05.10	22 M / D	
		CI/CD 환경 구성 및 배포 테스트		AWS CodeSeries 구축 자동화된 서버 배포 파이프라인 구축	24.05.23	24.06.07	22 M / D	
		데이터 분석 파이프라인 구성	팀원 전체	데이터 추출 - Glue를 사용한 RDS 데이터 추출(RDS에서 추출한 데이터(CSV 형식)) - 기존의 설문 엑셀 데이터 수집 (기존에 실시간 설문조사 데이터(CSV 형식)) - 로그 스트림 데이터 추출(로그 데이터(JSON 형식)) 데이터 품질 검사 규칙 설정 - DB 품질 검사를 위한 규칙 설정 - 설문조사 품질 검사를 위한 규칙 설정 데이터 품질 검사 구축 - DB 품질 검사 구축(AWS Glue Data Quality 스크린 샷) - 설문조사 품질 검사 구축(AWS Glue Data Quality 스크린 샷) 데이터 정제 - RDS 데이터 정제 - 엑셀 데이터 파싱 후 Parquet 변환 (파싱 및 parquet 변환을 위한 람다 코드, 변환을 진행한 parquet 파일) - 로그 스트리밍 파티셔닝 데이터 분석 및 시각화 - SQL 작성(분석을 위한 SQL 문) - Athena에서 SQL 테스트(SQL 실행을 통한 Athena 결과물(S3에 저장)) - SQL 결과 저장(SQL 실행을 통한 Athena 결과물(S3에 저장)) - 분석 대상 데이터 선정 - SQL 결과 기반 데이터 시각화 - 시각화된 결과 대시보드 구축(Quick Sight를 사용한 대시 보드) 데이터 파이프라인 자동화 - 데이터 파이프라인 자동화 데이터 파이프라인 이벤트 시 알람 구축 - 데이터 파이프라인 알람 지표 구성 - 데이터 파이프라인 알람 구축	24.05.15	24.06.11	22 M / D	
		백업 구성	김민경	RDS 스냅샷 자동화 - Amazon EventBridge를 통해 매일 새벽 2시에 람다 함수 실행 - 람다 함수를 통해 RDS 스냅샷 생성 및 S3 Standard에 월, 일 별 디렉토리 생성 후 해당 디렉토리에 월, 일 별 데이터 저장 데이터 생성 기간별 스토리지 클래스 분리 - Amazon S3 Standard: ~1개월 데이터 보관 - Amazon S3 Standard-IA: 1개월~3개월 데이터 보관 - Amazon S3 Glacier Flexible Retrieval: 3개월~2년 데이터 보관 - Amazon S3 Glacier Deep Archive: 2년~5년 수명 주기 정책을 활용하여 스토리지 클래스 전환 및 5년 후 데이터 자동 삭제	24.06.10	24.06.14	22 M / D	
		보안		SQL 인젝션 차단 등 웹 공격 대응이 가능한 AWS WAF 도입	24.06.10	24.06.10	22 M / D	
		모니터링		Web, WAS 서버 모니터링 환경 구축 이벤트 발생 시 담당자에게 알람 전송 환경 구축	24.06.05	24.06.11	22 M / D	
모니터링								
4	모니터링 계획		김민경		24.06.10	24.06.14	22 M / D	모니터링 결과물
	모니터링 계획과 수행 범위				24.06.10	24.06.14	22 M / D	
		3-Tier 전체 모니터링		Web, WAS, DB 서버 접속 모니터링 ALB를 통한 Web, WAS 서버의 분산 접속 모니터링	24.06.10	24.06.14	22 M / D	
		데이터 파이프라인 모니터링		데이터 정제 중 이벤트 발생 여부 모니터링	24.06.10	24.06.14	22 M / D	
운영 계획								
5	운영계획		김민경		24.06.10	24.06.14	22 M / D	정기점검 가이드, 비정기점검 가이드, 변경 관리 가이드, 백업 관리 가이드, 보안 관리 가이드, 장애 관리 가이드, 문제 관리 가이드
		정기점검			24.06.10	24.06.14	22 M / D	
		정기적으로 점검해야 할 서비스 점검						
		비정기점검			24.06.10	24.06.14	22 M / D	
		오류 발생 시 점검						
		변경관리			24.06.10	24.06.14	22 M / D	
		변경된 구성 내용 기록						
		백업관리			24.06.10	24.06.14	22 M / D	
		중요 서비스 데이터 백업						
		보안관리			24.06.10	24.06.14	22 M / D	
		ISMS-P 기준에 맞는 계정 관리 등						
		장애관리			24.06.10	24.06.14	22 M / D	
		장애 발생 시 알림 과정 및 결과 점검						
		문제관리			24.06.10	24.06.14	22 M / D	
		장애 재발 방지						

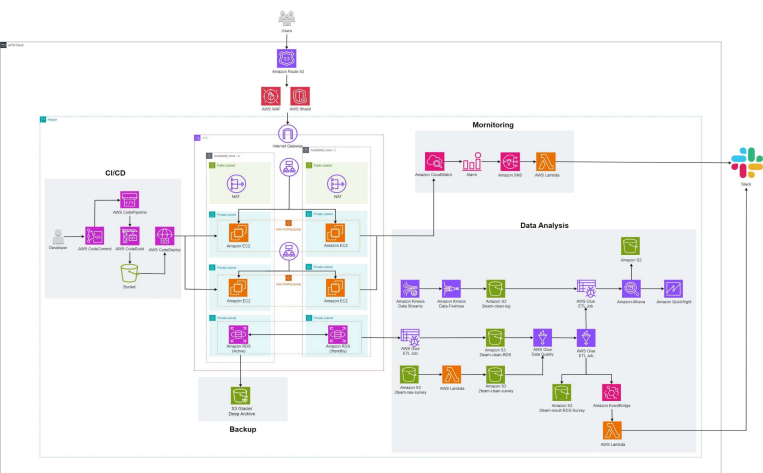
Project Milestone														
진행단계	수행내용	상세수행내용	담당	기간	4월	5월			6월					
					4W(24~28)	5W(29)	1W(1~3)	2W(6~10)	3W(13~17)	4W(20~24)	5W(27~31)	2W(3~7)	3W(10~14)	
1. 분석	프로젝트 투입	구축요구조건식별	김민경	04.24~04.24										
		백업 요구조건식별	김민경	04.24~04.24										
		C/CD 요구조건 식별	김민경	04.25~04.26										
		데이터 분석 요구조건 식별	김민경	04.29~05.15										
		모니터링 및 알람	김민경	04.29~04.29										
2. 설계	여기텍처	보안 요구조건 식별	김민경	04.29~04.29										
		AWS Resource 구성안	김민경	04.26~04.26										
		C/CD 구성안	김민경	04.26~04.26										
		데이터 분석 파이프라인 구성안	김민경	04.29~05.15										
		모니터링 및 알람 구성안	김민경	04.29~05.15										
3. 구축	네트워크	백업 구성안	김민경	04.29~05.01										
		보안 구성안	김민경	04.29~05.01										
		VPC 구축	김민경	05.02~05.06										
		서브넷 구축	김민경	05.02~05.06										
		인터넷 게이트웨이 구축	김민경	05.02~05.06										
	서버	EIP 생성	김민경	05.02~05.06										
		NAT 게이트웨이 구축	김민경	05.02~05.06										
		보안 그룹 구축	김민경	05.02~05.06										
		Web 서버 구축	김민경	05.02~05.10										
		Web ASG 구축	김민경	05.02~05.10										
		Web ALB 구축	김민경	05.02~05.10										
		Web TG 구축	김민경	05.02~05.10										
		App 서버 구축	김민경	05.02~05.10										
		App ASG 구축	김민경	05.02~05.10										
		App ALB 구축	김민경	05.02~05.10										
	C/CD	App TG 구축	김민경	05.02~05.10										
		AWS CodeCommit 연동	김민경	05.23~06.07										
		AWS CodeBuild	김민경	05.27~06.07										
		AWS Bucket 연동	김민경	05.27~06.07										
		AWS CodeDeploy 구축	김민경	05.27~06.07										
	데이터 분석 파이프라인	도메인 및 HTTPS	AWS CodePipeline 구축	김민경	05.27~06.07									
			Route 53를 통한 도메인 접속	김민경	05.02~05.10									
			HTTPS 보안 접속	김민경	05.02~05.10									
			Glue를 사용한 RDS 데이터 추출	팀원 공통	05.27~05.31									
			기존의 설문·엑셀 데이터 수집	팀원 공통	05.27~05.27									
데이터 품질 검사		클릭 스트림 데이터 수집	팀원 공통	05.27~06.07										
		DB	DB 용량 감사를 위한 규칙 설정	팀원 공통	05.28~05.29									
		설문 조사	설문조사 용질 감사를 위한 규칙 설정	팀원 공통	05.28~05.29									
		DB	DB 용질 감사 구축	팀원 공통	05.29~05.30									
		설문 조사	설문조사 용질 감사 구축	팀원 공통	05.29~06.04									
		DB	RDS 데이터 정제	팀원 공통	05.28~06.04									
		설문 조사	엑셀 데이터 파일 후 CSV 변환	팀원 공통	05.29~05.29									
		로그 스트림	후 구조스트리밍 파티셔닝	팀원 공통	05.29~06.05									
		데이터 분석, 시각화	SQL 작성	팀원 공통	05.27~06.11									
			데이터 분석	Athena에서 SQL 테스트	팀원 공통	05.27~06.11								
SQL 결과 차장	팀원 공통		05.27~06.11											
분석 대상 데이터 선정	팀원 공통(김민경)		05.22~06.11											
데이터 시각화	SQL 결과 기반 데이터 시각화		팀원 공통	05.23~06.11										
모니터링, 알람	알람 구축	시각화된 결과 대시보드 구축	팀원 공통(김민경)	05.15~06.11										
		데이터 파이프라인 자동화	팀원 공통	06.03~06.11										
		알람 지표 구성	팀원 공통	05.27~06.03										
		탐침 구축	팀원 공통	05.30~06.12										
		데이터 파이프라인 알람 구축	팀원 공통	06.10~06.14										
	백업 보안	데이터베이스 백업 구축	김민경	06.10~06.14										
		웹 공간 대응 할당 구축	김민경	06.10~06.10										
		Amazon CloudWatch 연동	김민경	06.05~06.11										
		Amazon SNS 연동	김민경	06.05~06.11										
		Slack 연동	김민경	06.05~06.11										
4. 모니터링	수행	3-Tier 전체 모니터링	김민경	06.10~06.14										
		데이터 파이프라인 모니터링	김민경	06.10~06.14										
5. 운영 계획	보고서 작성	정기점검 가이드 작성	김민경	06.10~06.14										
		비정기 점검 가이드 작성	김민경	06.10~06.14										
		변경 관리 가이드 작성	김민경	06.10~06.14										
		백업 관리 가이드 작성	김민경	06.10~06.14										
		보안 관리 가이드 작성	김민경	06.10~06.14										
		장애 관리 가이드 작성	김민경	06.10~06.14										
		문제 관리 가이드 작성	김민경	06.10~06.14										
		최종 경과를 제출 및 발표	김민경	06.10~06.14										

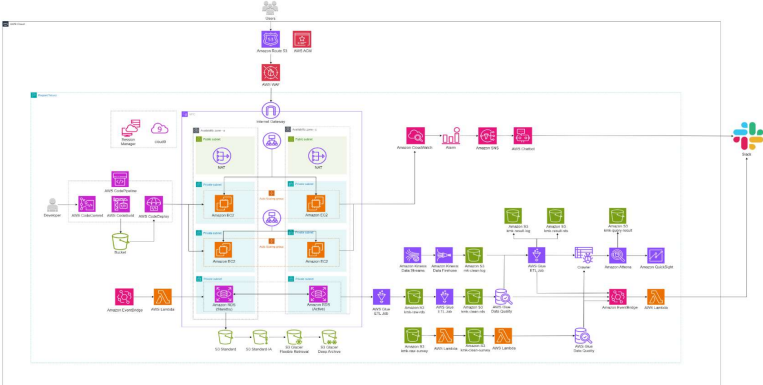
Ver	Architecture file name	Descriptions
1.0		<p>참고) 주제 확정하기 전의 3-Tier 설계</p> <p>비용 절감을 위해 S3 정적 웹 호스팅을 통한 Web 서버 구축</p> <p>캐시 서버로 사용자와 s3 사이에서 속도 증가시키기 위해 CloudFront 사용</p> <p>고가용성을 위해 DB의 Active-Standby 구축</p> <p>빠른 업데이트 반영을 위해 WAS 서버 CI/CD 환경 구축</p> <p>Web, WAS 서버의 CPU 등 특정 임계치가 넘어갈 시 담당자에게 slack으로 알림 전송</p>

[illegible]

Ver	Architecture file name	Descriptions
1.2		<p>기본 웹 보안 솔루션인 AWS WAF 뿐 만이 아니라, 웹 7계층까지 보호하기 위한 AWS Shield 사용</p> <p>실시간 데이터 분석이 아닌 '갑작스럽게 증가하는 트래픽에 대응할 수 있는 인프라를 구축하는 것'으로 시나리오 재설정. 따라서 Kinesis 서비스 제거</p> <p>Amazon RDS 스냅샷을 S3에 저장한 후 Glue, Athena를 통해 분석 분석한 데이터는 Quicksight를 통해 시각화 정제되지 않은 원본 데이터는 S3 Glacier Deep Archive에 보관</p> <p>Amazon RDS는 자동 백업으로 진행</p>

Ver	Architecture file name	Descriptions
1.3		<p>컨테이너 기반 대신 EC2 인스턴스를 사용한 인프라로 대체. 기존 쇼핑몰 인프라는 컨테이너를 사용하지 않았기 때문에, 컨테이너로 구축 시 운영 및 관리 측면에서 비용이 많이 들 예정. 가상 머신으로도 충분히 서비스를 운영할 수 있기 때문에 컨테이너 대신 가상 머신을 사용하는 것으로 변경</p> <p>Fargate의 자동확장 기능 대신 EC2 AutoScaling 기능을 활용하여 급작스럽게 증가한 트래픽에도 탄력성 있게 대응하여 지속적인 서비스를 제공할 수 있도록 구축</p> <p>컨테이너 기반이 아니기 때문에 build 후 생성되는 아티팩트는 ECR 대신 S3 버킷에 저장하는 것으로 변경</p> <p>데이터의 종류에 따라 수집 경로 체계화</p> <ul style="list-style-type: none"> - 데이터베이스, 설문조사, 클릭 스트림 데이터 개별 수집 <p>여러 경로로 수집된 데이터 통합 처리 및 분석 파이프라인 구축</p>

Ver	Architecture file name	Descriptions
1.4		<p>클릭 스트림을 로그 스트림 데이터로 변경.</p> <p>로그 스트림의 데이터 분석 정제 및 분석 방법 변경</p> <ul style="list-style-type: none"> - 로그 스트림 데이터는 수집한 후 품질 검사를 진행하지 않고 바로 ETL 작업 진행 - 데이터 베이스와 설문조사 데이터는 개별 수집한 후 품질 검사를 진행 그 후에 ETL 작업을 진행

Ver	Architecture file name	Descriptions
1.5		<p>데이터 백업 구체화</p> <ul style="list-style-type: none"> - 비용 절감을 위해 데이터가 생성된 기간별로 Amazon S3 스토리지 클래스를 분리하여 저장 <ol style="list-style-type: none"> Amazon S3 Standard <ul style="list-style-type: none"> - 데이터베이스 데이터를 매일 스냅샷으로 찍고, 자동으로 parquet 형식으로 바꾸어 저장 Amazon S3 Standard-IA <ul style="list-style-type: none"> - Amazon S3 Standard 데이터 중 한 달이 지난 데이터는 수명주기를 통해 Amazon S3 Standard-IA로 이동 및 저장하여 비용 절감 - 1분기 데이터는 종종 액세스할 가능성이 있기에 밀리초 단위로 액세스 가능한 Amazon S3 Standard-IA로 전환 Amazon S3 Glacier Flexible Retrieval <ul style="list-style-type: none"> - 2년 간은 법적 분쟁 발생 가능성이 종종 있음 - 급할 시 비교적 빨리 데이터를 확인할 수 있어야 하기 때문에 검색 옵션이 1분부터 12시간까지인 S3 Glacier Flexible Retrieval로 전환 - 3개월 ~ 2년 데이터 저장 Amazon S3 Glacier Deep Archive <ul style="list-style-type: none"> - Amazon S3 Glacier Flexible Retrieval에서 2년 이상된 데이터는 수명주기를 통해 Amazon S3 Glacier Deep Archive로 이동 - 5년 간 장기 보관 (일본 법적 규정사항과 법적 분쟁 대응을 위해 구매, 고객, 리유 데이터 약 5년 간 장기 보관 필요) <ul style="list-style-type: none"> - 수명주기 정책을 활용하여 스토리지 클래스 전환 - 5년 이상된 데이터는 자동적으로 삭제 <p>Amazon Athena 결과물을 저장할 S3 버킷 이름 구체화</p> <p>Web 서버에 HTTPS 접속을 하기 위해 AWS Certificate Manager와 ALB 연동</p> <p>IaC를 활용하여 인프라 구축</p> <p>SSM을 활용하여 서버에 접속</p>