

자동화 된 빅데이터 분석 플랫폼 Acculnsight+

SK 주식회사 C&C

이승환 수석

Table of Contents

1. 개요
2. AccuInsight+ 서비스 – Pipeline
3. AccuInsight+ 서비스 – Modeler
4. AccuInsight+ 서비스 – DCM
5. Demo
6. Contact us

SK C&C의 데이터 분석 플랫폼 그 탄생과 진화

1. 개요

빅데이터 시스템 구축 → Data API 개발 → 빅데이터 플랫폼 개발 → AI devops 플랫폼으로 진화

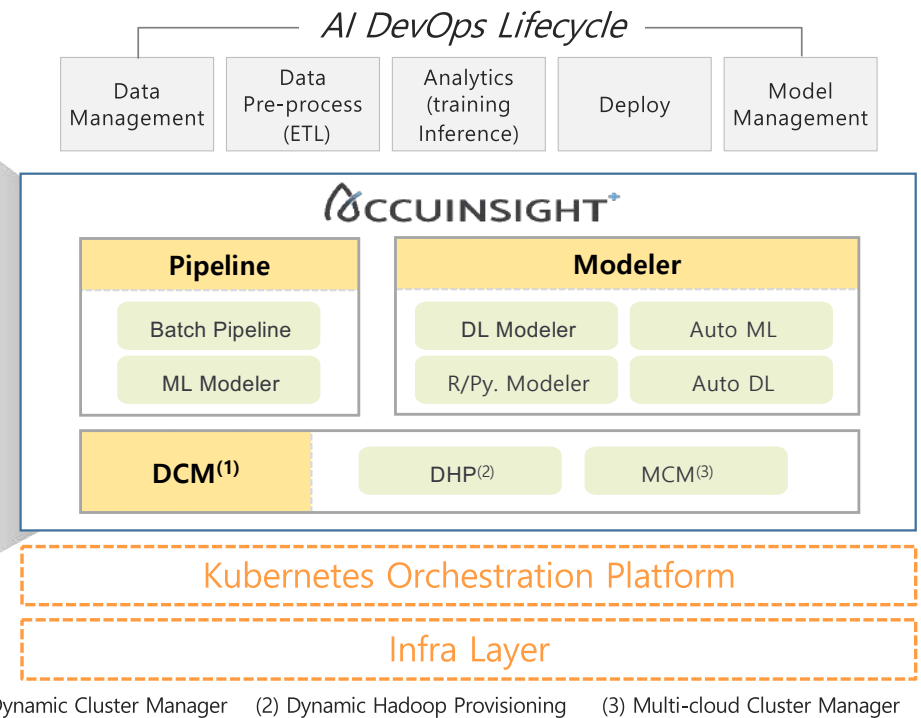
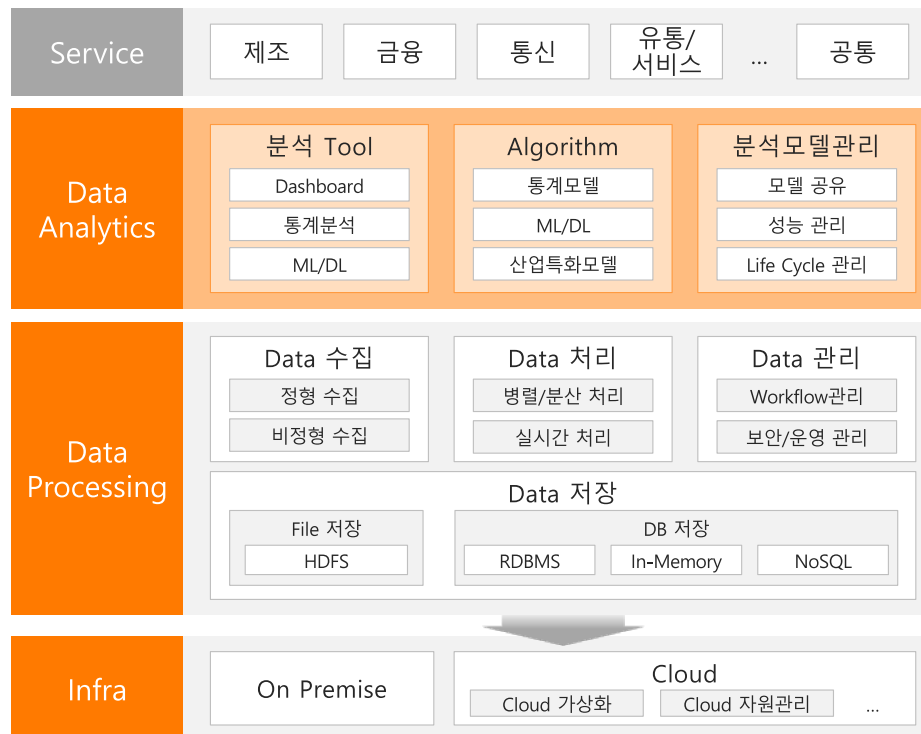
Timeline	기술의 진보	시장은..	나와 우리 팀은..
2012	하둡 빅데이터	하둡 셋업하기 어렵네요.. 구축해주세요	주요 계열사 Hadoop eco system 구축 Map-reduce 코딩
2013	통계 기반 분석	공정 예비감지 해주세요 머신러닝이 뭐예요?	통계능력자가 샘플링, 통계분석 및 수식 디자인. 대용량 처리 가능하도록 map-reduce, hive로 전환
2014	분산 머신러닝	하둡에 데이터 잔뜩 쌓았어요 이제 분석 해주세요	Mahout, spark ML로 yarn cluster에서 전체 데이터 로 학습과 예측. 데이터에 적응하는 모델
2015	구글 Tensorflow 공개	딥러닝이 뭐예요?	통계 기반 분석에서 머신러닝 분석으로
2016	알파고 vs 이세돌	머신러닝, 딥러닝을 전 국민이 알게 됨	반복되는 하둡구축.. 이걸 쉽게 할 수 있도록 내재화 하자 Data API 개발, ML Tool개발
2017	도커, 딥러닝의 부흥	딥러닝 돌리면 다 되는 것 아닌가요?	Data API를 바탕으로 우리만의 서비스를 만들자 컨테이너 기반의 하둡 클러스터 배포
2018	분산 딥러닝	오픈소스 쓰면 비용 없이 간단한 것 아닌가요?	우리 cloud에 서비스를 올려보자 AccuInsight+ 개 발. DL Modeler 개발, GPU 분산학습
2019	분석 플랫폼 시장의 증대 (cloud 3사)	우리 서버 (on-premise) 에 빅데이터 플랫폼 구축 해주세요	딜리버리 할 사람이 없어요 안팔아요.. ???; 삽니다
2020	시민 데이터 과학자	데이터 분석 별거 아니네.. 근데, 퍼블릭 클라우드에서 가능한가요?	이젠 자동화가 살 길이다.. AccuInsight+ 2.0 런칭

AccuInsight+ 는

1. 개요

데이터 전처리부터 분석과 운영, Data Science 전 영역을 지원하는 Multi-Cloud 기반 AI DevOps 플랫폼

Data Science를 지원하는 머신 러닝 플랫폼 서비스이며 Amazon 분석서비스 (AWS) 처럼 다양한 서비스를 구독형 / 설치형으로 지원
 분석 Sandbox 생성 / 데이터 처리 / 모델 학습 및 배포 등 데이터 분석의 모든 활동을 End-to-End로 지원하며, 다수의 사용자에게 다양한 workspace /
 workflow를 제공하여 분석 프로젝트의 성공을 돕는 서비스로 구성



AccuInsight+의 기능 Coverage

1. 개요

AccuInsight+ Pipeline : Drag & Drop UI 기반 데이터 전처리 / 머신 러닝

AccuInsight+ Modeler : 직접 코딩 가능한 Advanced 분석가 대상 모델링 / 운영 배포



AccuInsight+ Pipeline (초급 분석가 대상)

- Spark을 활용한 인터랙티브 전처리
- Oozie 기반의 워크플로우 관리
- 대용량 데이터 분산 병렬 머신러닝 (SparkML)
- 코딩없이 ML Pipeline 구성/실행
- Infra Independent (Cloud, On-Prem Infra 활용)

AccuInsight+ Modeler (고급 분석가 / ML 엔지니어)

- 독립적인 Advanced 개발 환경 제공 (Jupyter, Rstudio)
- DL/ML 모델 라이프 사이클 관리
- Rest API 형태의 모델 배포
- Swagger UI를 통한 단위 테스트 지원
- Model 재 배포를 위한 모니터링/시각화

AccuInsight+ 특징

1. 개요

특징

1 빅데이터 비즈니스에 필요한 전체 서비스 지원

- AccuInsight+는 빅데이터 수집, 처리, 분석, 활용 까지 필요한 모든 서비스 (Pipeline, Modeler, DCM) 를 지원합니다.

2 다양한 기능의 조합을 통한 유연한 서비스 구축

- 상호 보완적인 여러 제품을 다양하게 조합하여 사용자의 비즈니스 모델에 적합한 유연한 서비스 구축이 가능합니다.

3 성공적인 비즈 적용을 위한 단기 집중 교육 서비스 지원

- 맞춤형 교육 서비스 제공을 통해 AccuInsight+ 서비스 적용의 가치를 단기 간에 검증하고 본격적으로 실행에 옮길 수 있습니다.

4 금융, 유통, 제조 등 다양한 사업군 Reference 확보

- 금융사 3곳(S은행, K은행, S증권), 유통사 2곳(L백화점, S면세점), 제조사 1 곳 (H사) 등 사업군 별 Reference 확보

차별화 Point

Data 분석

- 분석 Asset(모델) 공유/협업 및 자산화, 지속갱신 가능
- 대용량 병렬처리 분석 지원으로 분석 시간 단축
- Data Scientist 별 독립된 분석 환경 제공
- 선정된 모델 운영 시스템으로 적용 용이

Platform

- Private/Public Cloud 지원 (AWS, GCP, AZURE*예정)
- On-Premise 인프라 활용 설치 가능
- 검증된 오픈 소스 기술 활용된 플랫폼 제공

Customization (*Option)

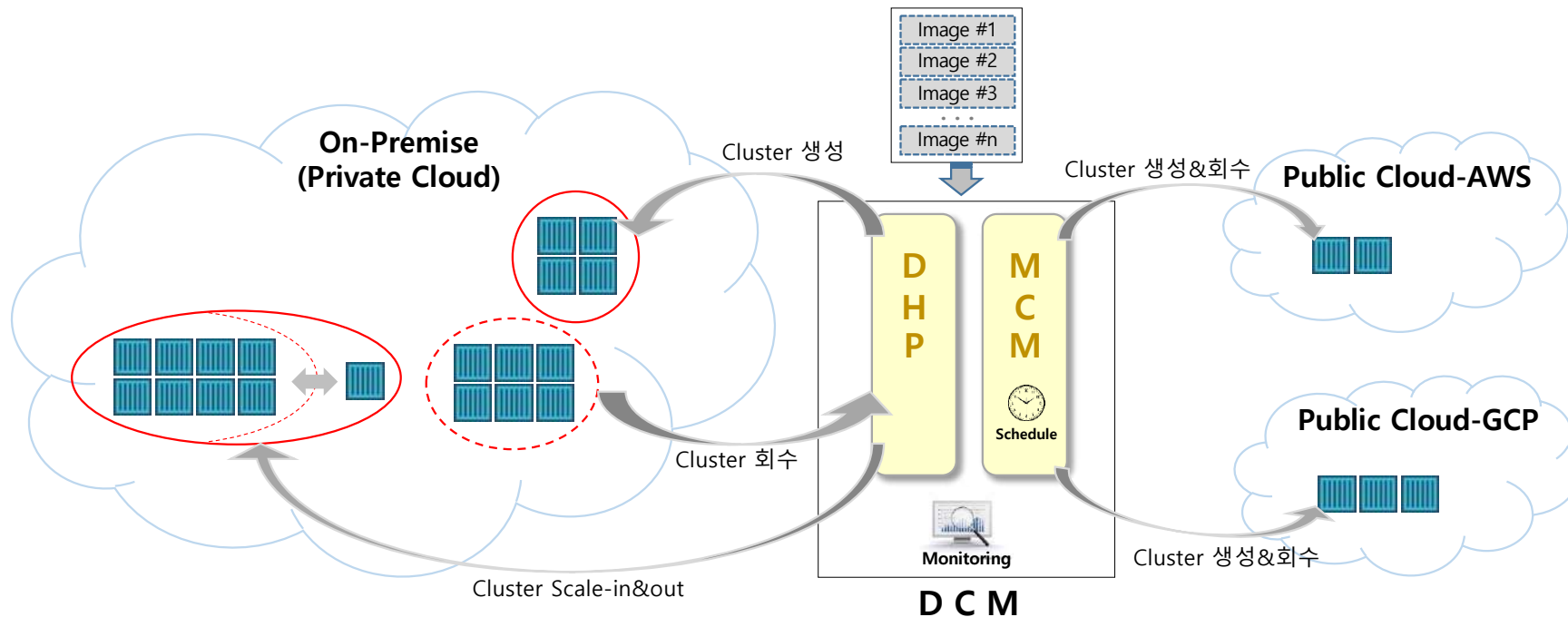
- 데이터 파이프라인 최적화 지원
- 솔루션 성능 최적화 지원
- 고객사 Legacy 연동 파이프라인 커스터 마이징 지원

DCM(Dynamic Cluster Manager) 란?

2. 서비스 - DCM

On-Premise/Public Cloud의 다양한 환경에서 Cluster를 손쉽게 배포 및 관리하여 자원의 효율적인 활용을 지원하는 서비스

- DHP (for On-Premise/Private Cloud) : 인프라/분석 SW 관리 최적화 관리형 Hadoop / Spark 클러스터 자동 배포 서비스
- MCM (for Public Cloud) : Cloud Service Provider 환경에 맞춰 AWS의 EMR, GCP의 Dataproc Cluster 생성 및 관리



AccuInsight+ 서비스 - DCM(Dynamic Cluster Manager)

2. 서비스 - DCM

Hadoop/Spark 클러스터 자동 배포 서비스로 인프라/빅데이터 엔지니어 없이 손쉽게 클러스터를 배포/관리하여 사용자가 데이터 처리/분석에만 집중할 수 있는 최적화된 환경 제공

	선택적 클러스터 배포	자원 활용 효율화	Monitoring	자원 상세 조회
DHP For On-Prem./Private Cloud	Hadoop/Eco Deploy ▪Hadoop, Yarn CPU / Memory 설정 및 클러스터 Spec.에 대한 설정 가능 ▪추가로 설치할 Eco 선택	Node Scaling ▪Hadoop, Yarn, Kafka 노드 Scale out & in 가능 ▪미사용 클러스터는 삭제하여 효율적인 자원 활용	클러스터 자원 모니터링 ▪클러스터의 노드 별 자원 사용률 모니터링 ▪Hadoop 클러스터 및 각각 Eco의 health 체크	History ▪클러스터의 사용중인 자원/Job에 할당된 자원 등 자원 상세 조회 ▪Hadoop 클러스터에서 실행된 job 목록/상태 및 상세 이력 조회
MCM For Public Cloud	Hadoop/Eco Deploy ▪용도별 Custom 클러스터 템플릿 관리 ▪설치 소프트웨어를 리스트에서 선택하여 Custom 추가 설치 가능	Cluster Scheduling ▪스케줄 기반 자동 클러스터 생성/삭제로 효율적인 자원 활용 및 비용 절감 ▪필요 시 워커 노드 증감	클러스터 자원/비용 모니터링 ▪클러스터 리스트/노드 상태 모니터링 ▪사용자 계정 연동을 통해 유형별 (서비스, 사용자, 기간) 비용 조회	

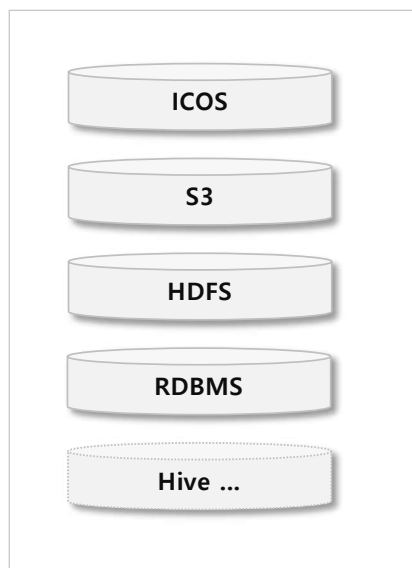
Pipeline 이란?

3. 서비스 - Pipeline

사용자가 GUI 환경에서 분석데이터 수집/처리/저장(Data Pipeline) 및 Spark MLlib기반 머신러닝 모델을 개발/학습(Machine Learning Pipeline)할 수 있도록 지원해 주는 서비스

다양한 Data Source 연동 제공

- 다양한 Data Source 연동 기능 제공으로 도입 시 Legacy 변경 최소화



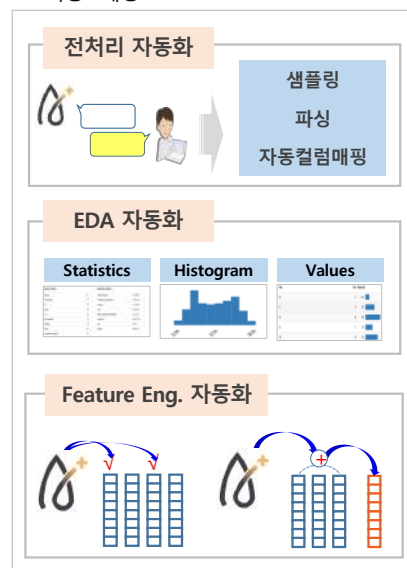
Easy Use

- GUI 기반의 Pipeline 구축 기능 제공
- 데이터 처리 기능 및 Spark MLlib 제공 (GUI 기반 사용)



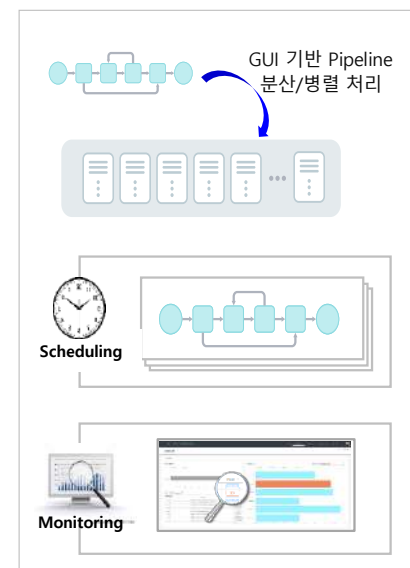
전처리, EDA, Feature Eng. 자동화

- 대화형 I/F기반 전처리 자동화로 수작업 최소화
- EDA, Feature Eng. 자동화 및 지표 시각화 기능 제공



분산/병렬처리 및 스케줄링/모니터링

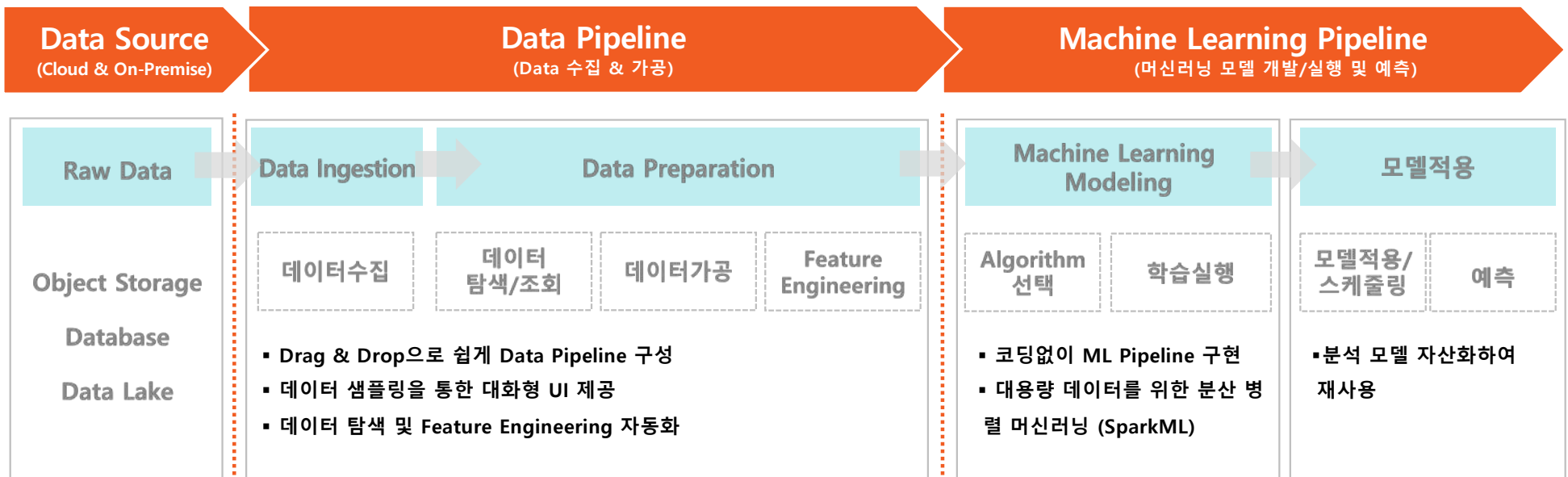
- GUI기반 분산/병렬 처리 제공
- Pipeline 스케줄링 및 Dashboard를 활용한 모니터링 기능 제공



AccuInsight+ 서비스 - Pipeline

3. 서비스 - Pipeline

사용자가 GUI 환경에서 분석데이터 수집/처리/저장(Data Pipeline) 및 Spark MLlib기반 머신러닝 모델을 개발/학습(Machine Learning Pipeline)할 수 있도록 지원해 주는 Pipeline 도구



Modeler 란?

4. 서비스 - Modeler

AI 개발자, Data Engineer, Data Scientist를 위한 머신러닝/딥러닝 모델의 빠른 구축, 학습 및 운영 배포를 지원하는 End-to-End 분석 모델 Lifecycle 관리 서비스

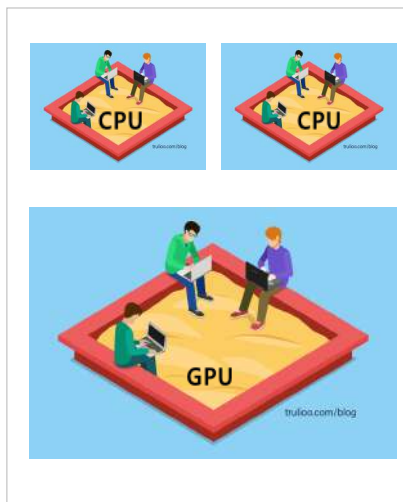
다양한 분석 환경 제공

- 분석자의 역량 수준, 기호별 분석 환경 제공



분석 Sand Box 제공

- 독립적 분석을 위한 Sand Box 제공 및 Sand Box 별 자원 조정 기능 제공
- CPU / GPU 제공



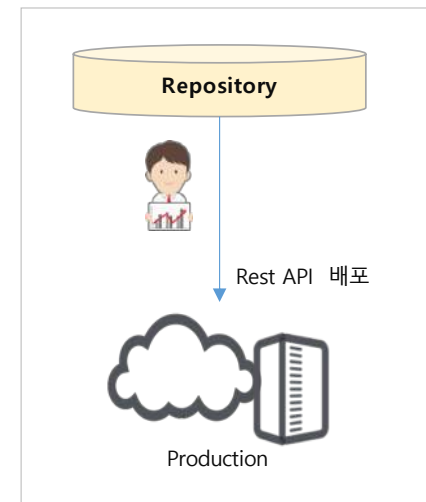
모델 저장, 공유 공간 제공

- 분석 모델 저장, 버전관리, 공유 기능을 제공하여 협업, 재사용 가능



운영 배포 기능 제공

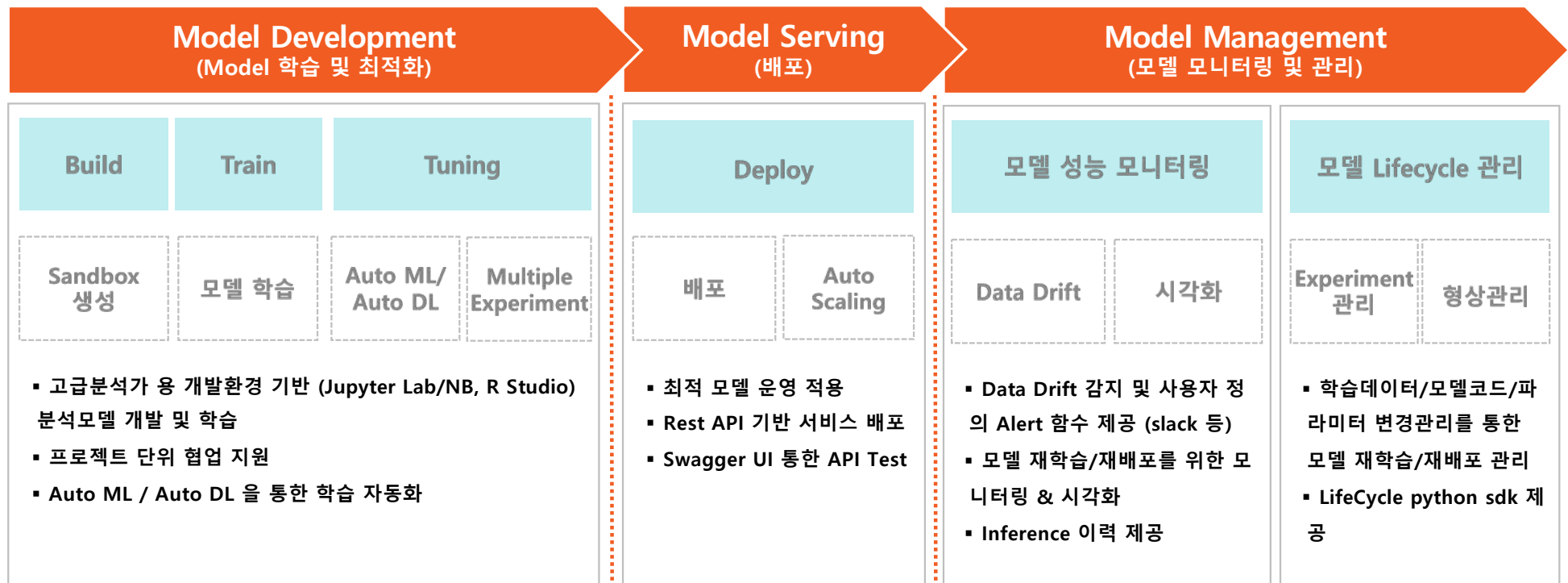
- 분석 모델을 Rest API 방식으로 즉시 배포 가능



AccuInsight+ 서비스 - Modeler

4. 서비스 - Modeler

AI 개발자 및 Data Engineer, Data Scientist를 위한 머신러닝/딥러닝 모델의 빠른 구축, 학습 및 운영 배포를 지원하는 End-to-End 분석모델 Lifecycle 관리 서비스



AccuInsight+ Demo

5. Demo

- DCM : hadoop cluster sandbox 생성
- Pipeline : data workflow 작성
- Modeler : 분석, 모델 배포 및 모니터링

Email

help.accuinsight@sk.com

포탈

<https://accuinsight.cloudz.co.kr>

카카오 채널

http://pf.kakao.com/_xoFxmcb



AccuInsight+ 소개 (V1.10)

감사합니다