

[오픈테크넷서밋 2018]

KSB 인공지능 프레임워크

2018. 9. 13

Lee Yeonhee

leeyh@etri.re.kr

yeonhee@apache.org

KSB Convergence Research Department, **ETRI**

Agenda

- I. Issus in Developing AI Services**
- II. Recent trend in AI Platforms**
- III. Strategies of KSB Framework**
- IV. Usecases**
- V. Conclusion**

I. Issus in Developing AI Services

AI Opportunity and Challenges

AI Opportunity:

Opportunity to transform their business by implementing sophisticated models for recommendation engines, ads targeting, speech recognition, object recognition, bots, sentiment analysis, predictive analysis, and more.

However, practitioners face multiple challenges when implementing such applications:

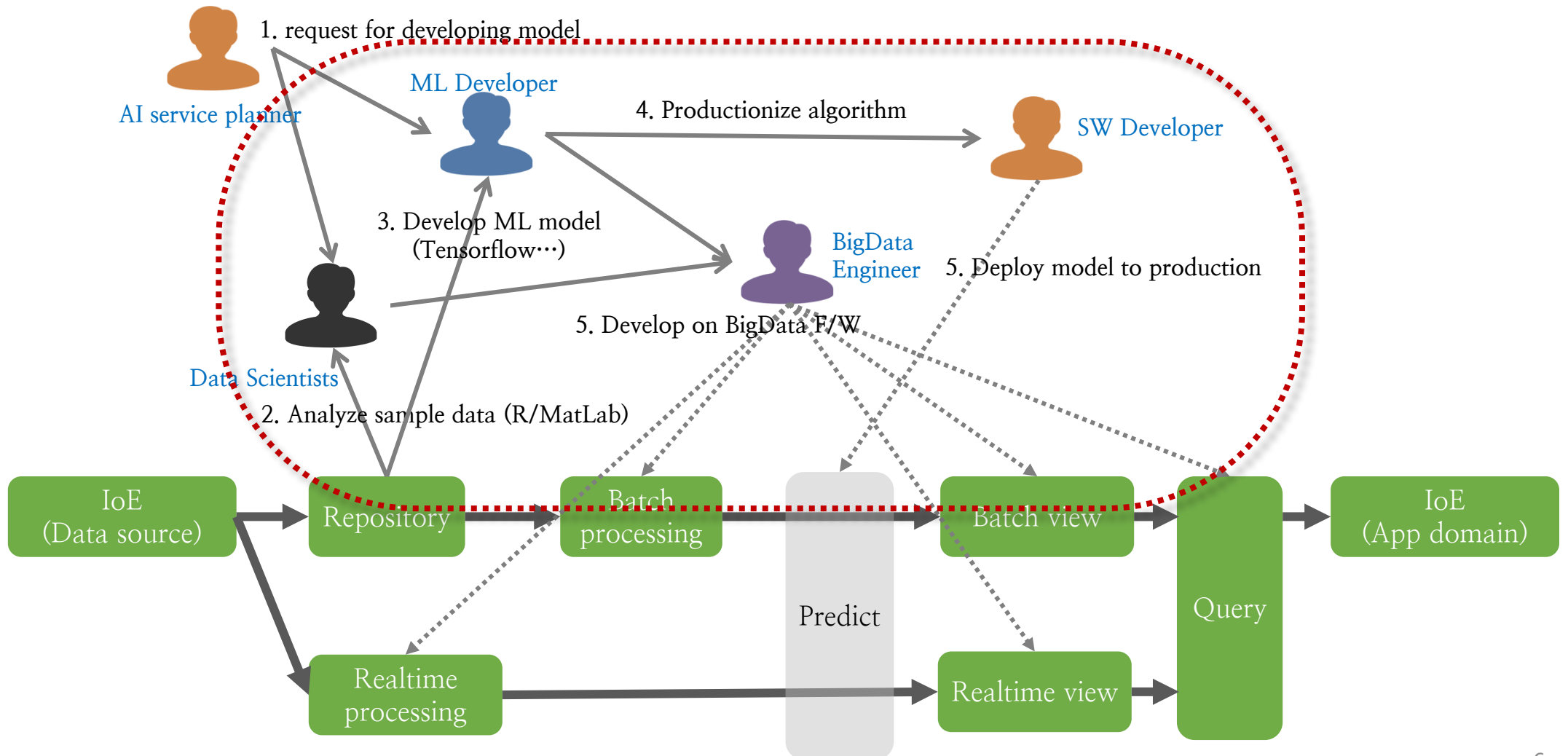
- **Environment Configuration:** Ability to setup and maintain complex environments due to the multitude of Machine Learning frameworks available.
- **Distributed Training:** Ability to train models in a distributed fashion to get results faster. This requires specialized skills and complex code to manage.
- **Compute Power:** Ability to run parallelized workloads on GPUs for maximum performances. This is essential for Deep Learning applications.

** source: <https://databricks.com/blog/2018/06/05/distributed-deep-learning-made-simple.html>*

Era of AI, but still there are many difficulties ...

- Diverse AI frameworks
 - Diverse ML/DL libraries – Tensorflow, MXNet, PyTorch, Caffe2, Torch, theano
- Diverse form of data to the model
- Deliver data to the model on time
- Update model for understanding recent trends
- Interact from models to model for solving one problem
- Cloud dependency
- ...

Journey from Data to AI Application



Machine Learning in Real World

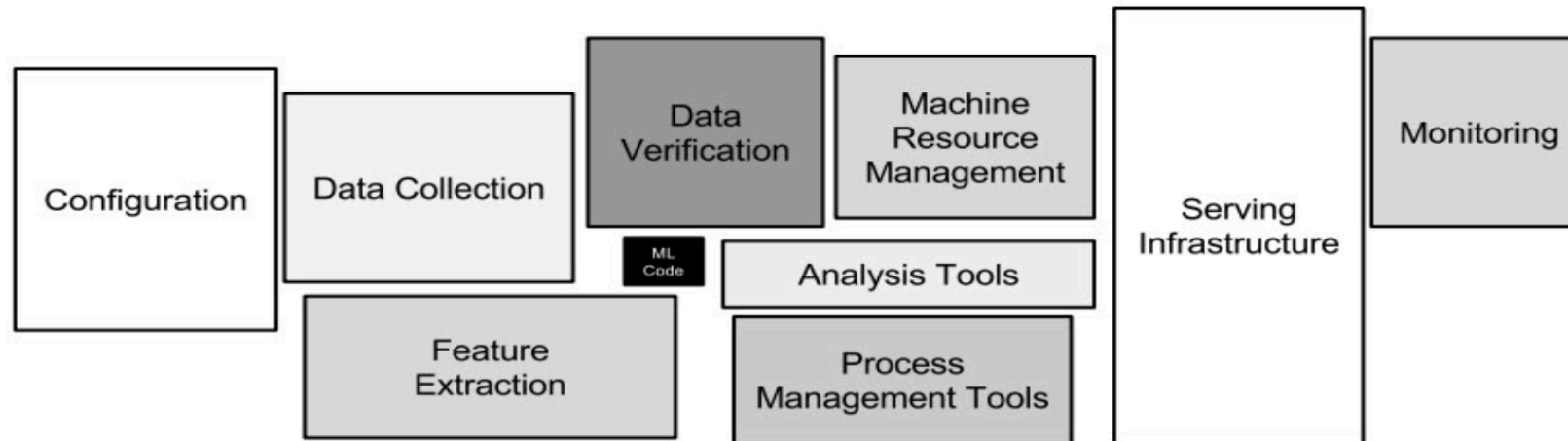
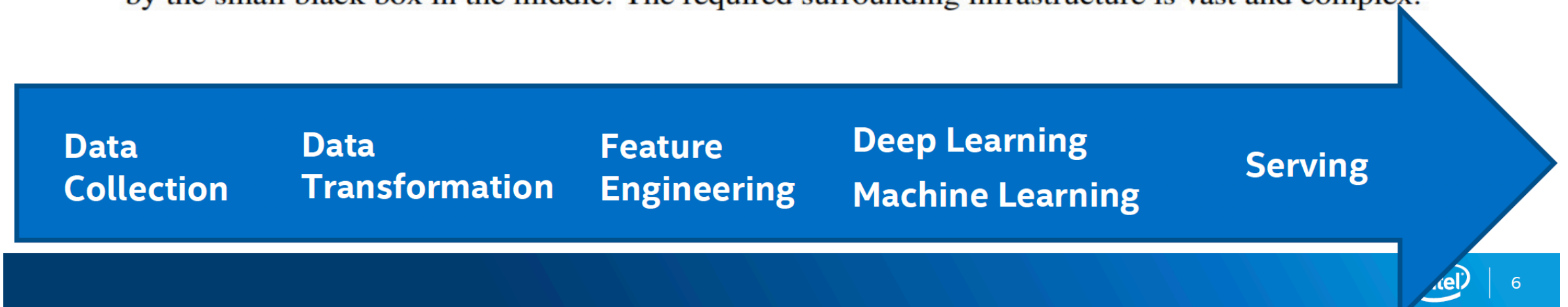


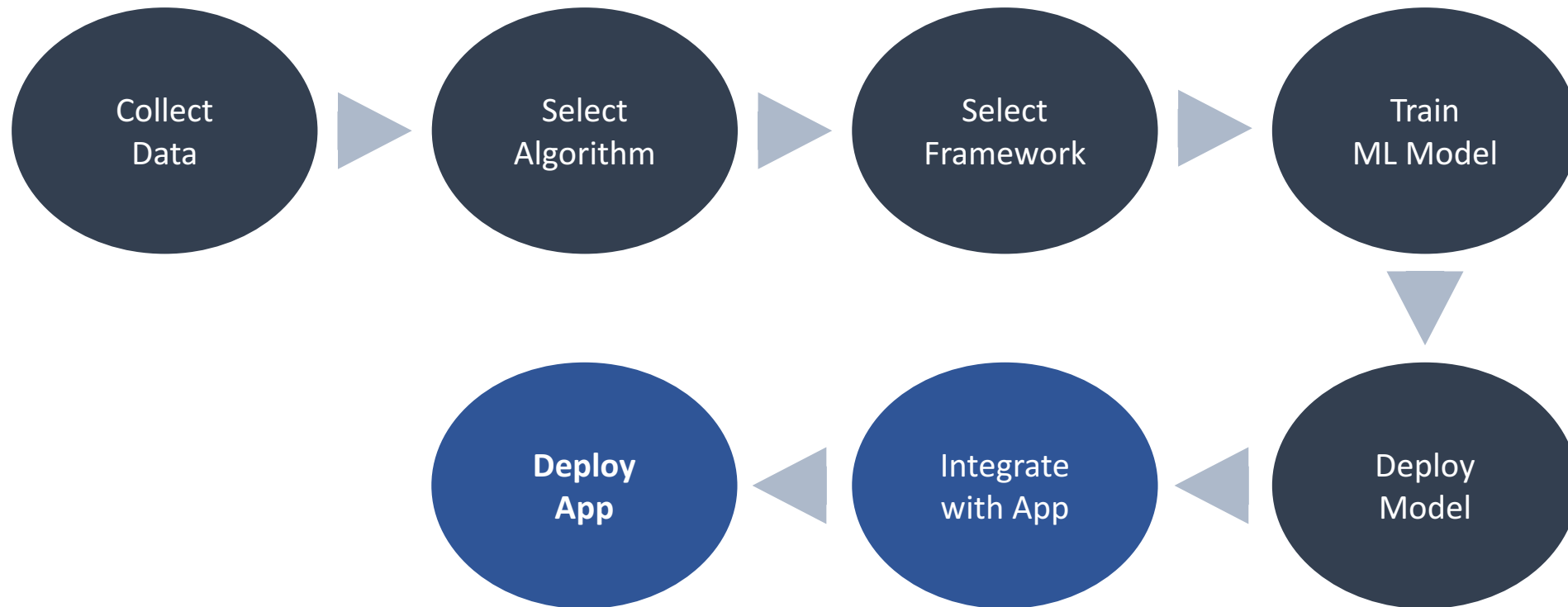
Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.



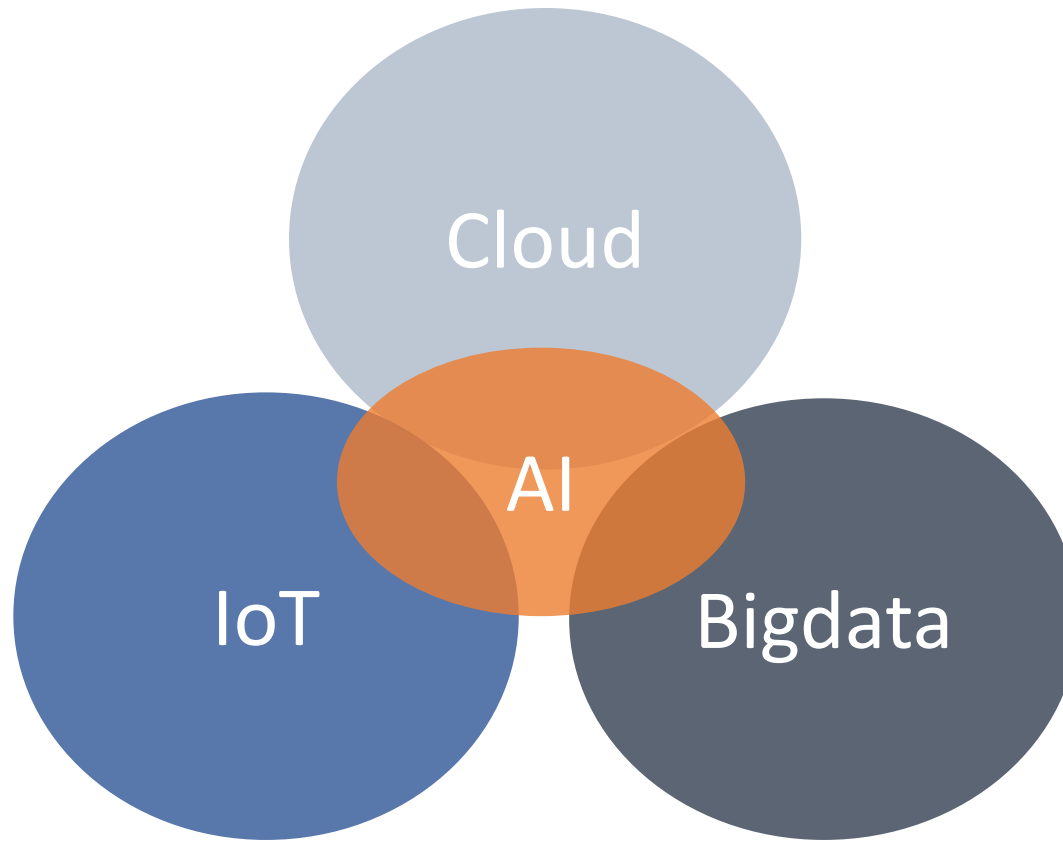
Source: Future of data analytics (Intel, Open Data Science Conference 2017)

II. Recent trend in AI Platforms

Journey from Data to AI Application

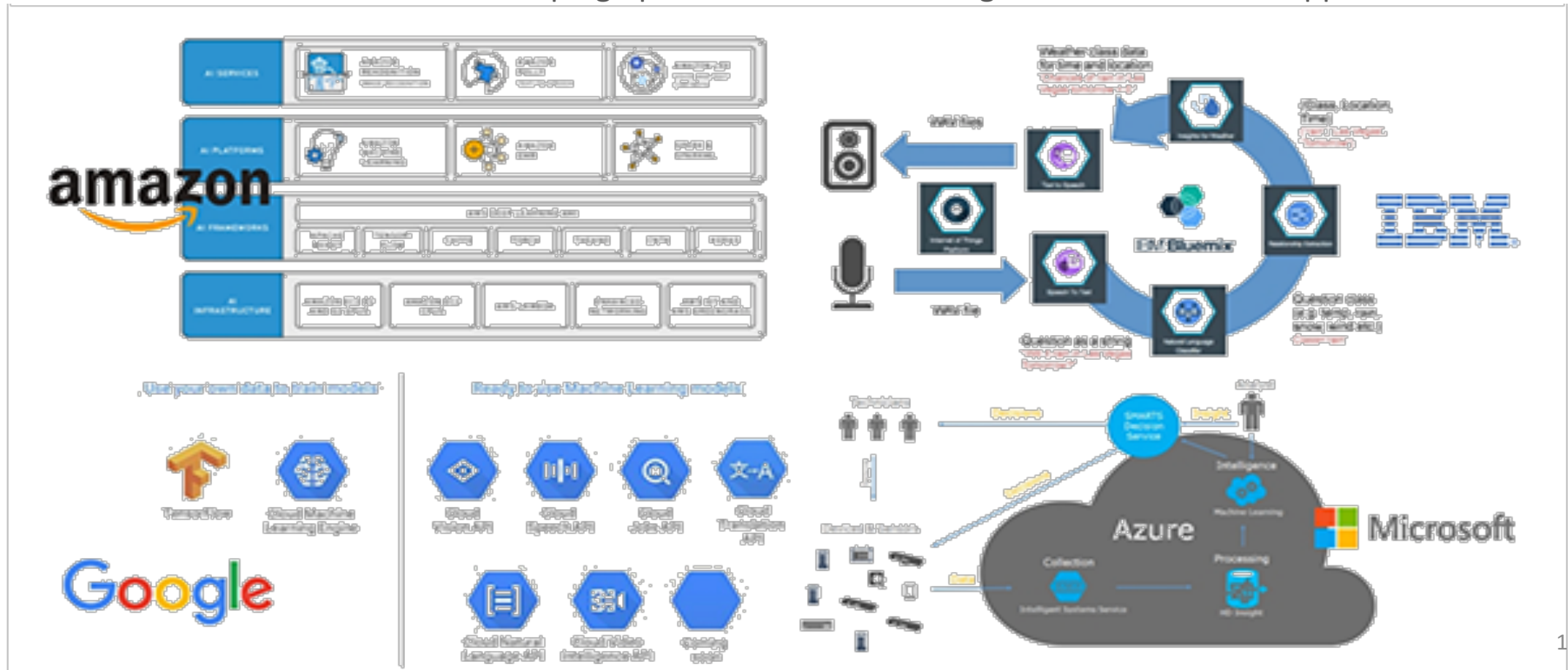


Global IT Companies

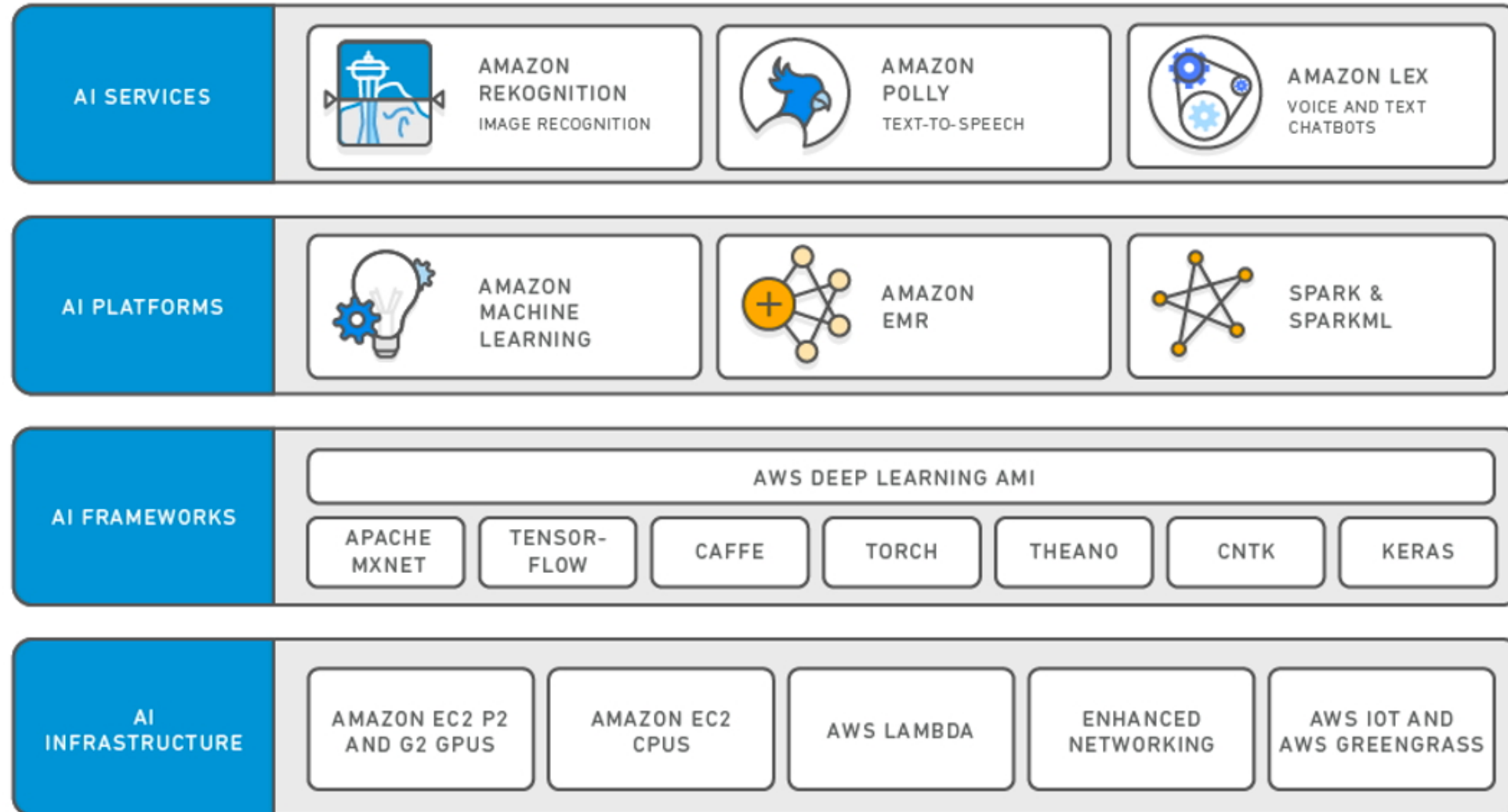


Cloud + AI

- Global IT Companies (Google, Amazon, IBM, MS)
 - Provide cloud computing-based cognitive service and ML service platform
 - Provide foundation for developing open source-based intelligent services and IoT application



Cloud + AI: Amazon AI

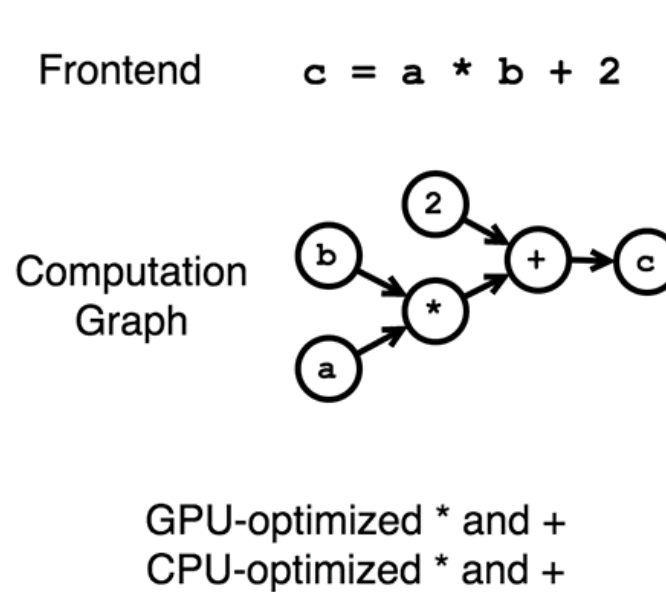


Cloud + AI: Amazon AI

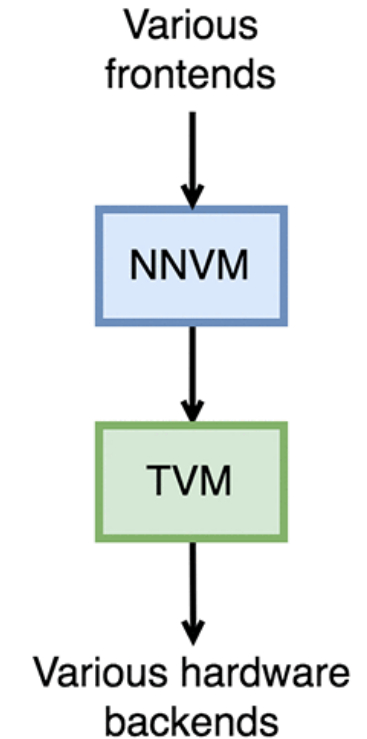
NNVM Compiler: Open Compiler for AI Frameworks

- Three challenges:
 - Switch from one AI framework
 - (e.g. pyTorch -> Caffe2)
 - Multiple backends to guarantee performance
 - Supporting multiple frameworks requires enormous engineering efforts

A typical framework



NNVM Compiler



Source: <https://aws.amazon.com/ko/blogs/machine-learning/introducing-nnvm-compiler-a-new-open-end-to-end-compiler-for-ai-frameworks/>

Cloud + AI: Amazon AI

NNVM Compiler: Open Compiler for AI Frameworks

- an end-to-end compiler based on the TVM stack that compiles workloads directly from various deep learning frontends into optimized machine codes.

Typical Workflow of NNVM Compiler

model from framework

```
graph, params = nnvm.frontend.from_xyz(...)
```

compile

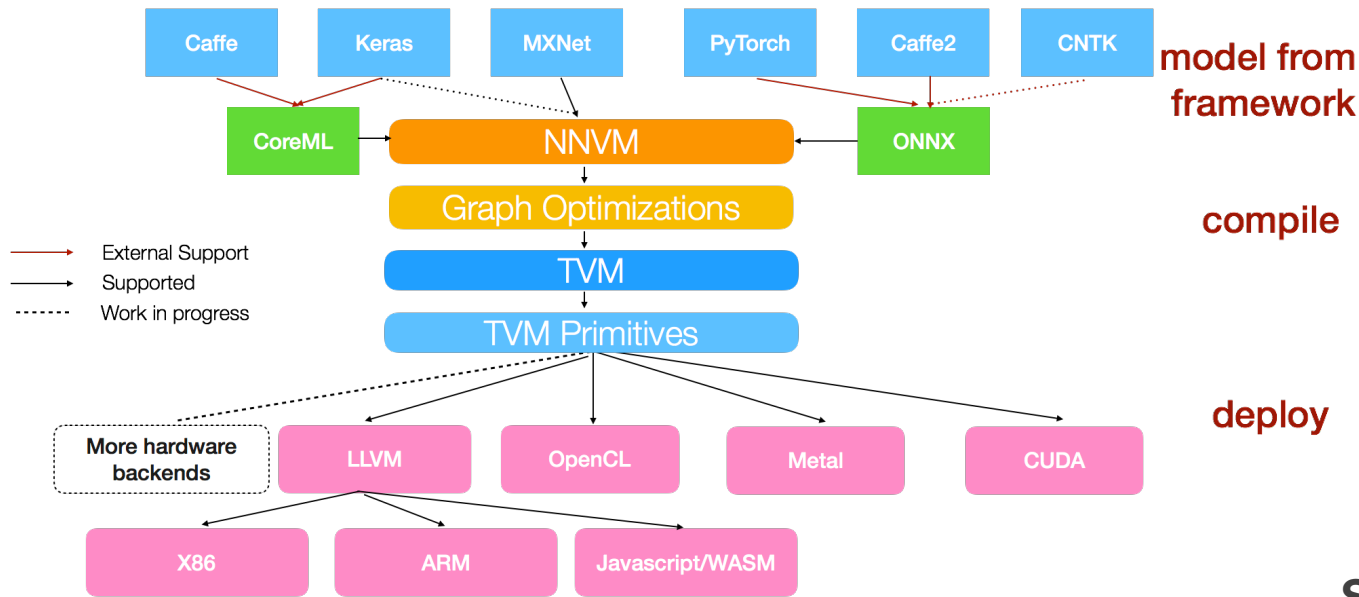
```
graph, lib, params = nnvm.compiler.build(
    graph, target="cuda", {"data", data_shape}, params=params)
```

deploy

```
module = graph_runtime.create(graph, lib, tvn.gpu(0))
module.set_input(**params)
module.run(data=data_array)
output = tvn.nd.empty(out_shape, ctx=tvn.gpu(0))
module.get_output(0, output)
```

Separation of Optimization and Deployment

* Source: NNVM Compiler: Open Compiler for AI Frameworks



Cloud + AI: Google AI

Use your own data to train models



TensorFlow



Cloud Machine Learning Engine

Ready to use Machine Learning models



Cloud Vision API



Cloud Speech API



Cloud Jobs API



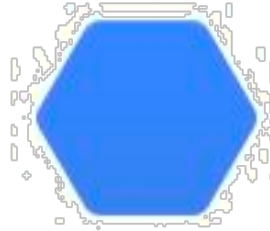
Cloud Translation API



Cloud Natural Language API



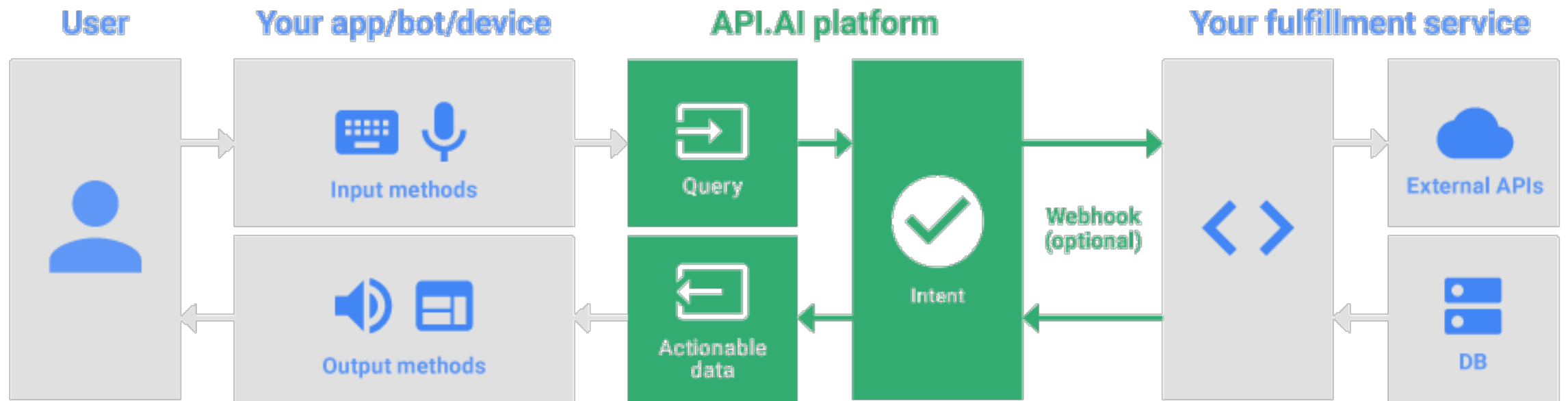
Cloud Video Intelligence API



Coming soon

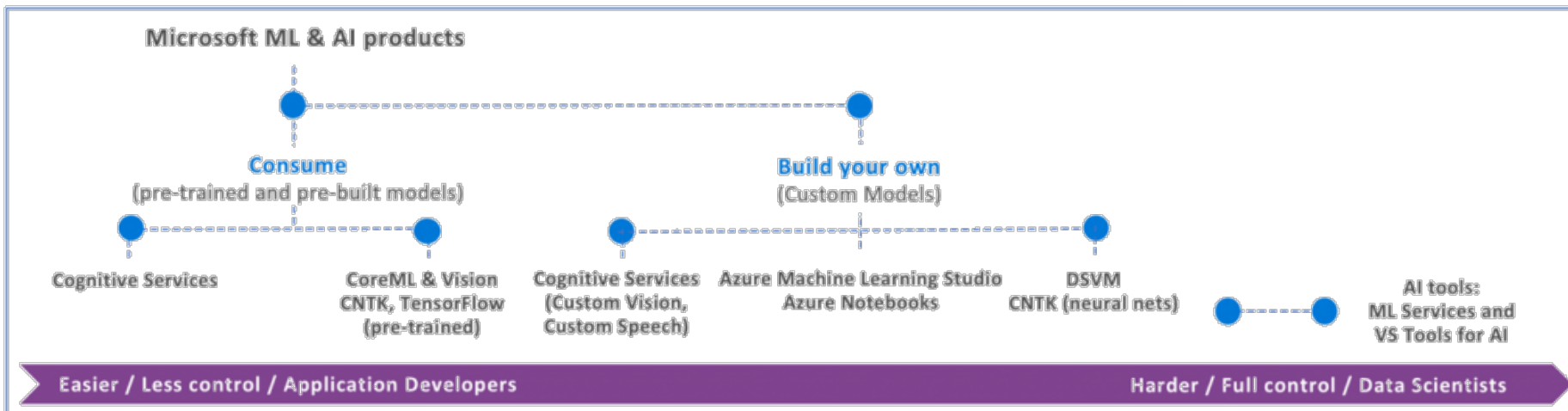
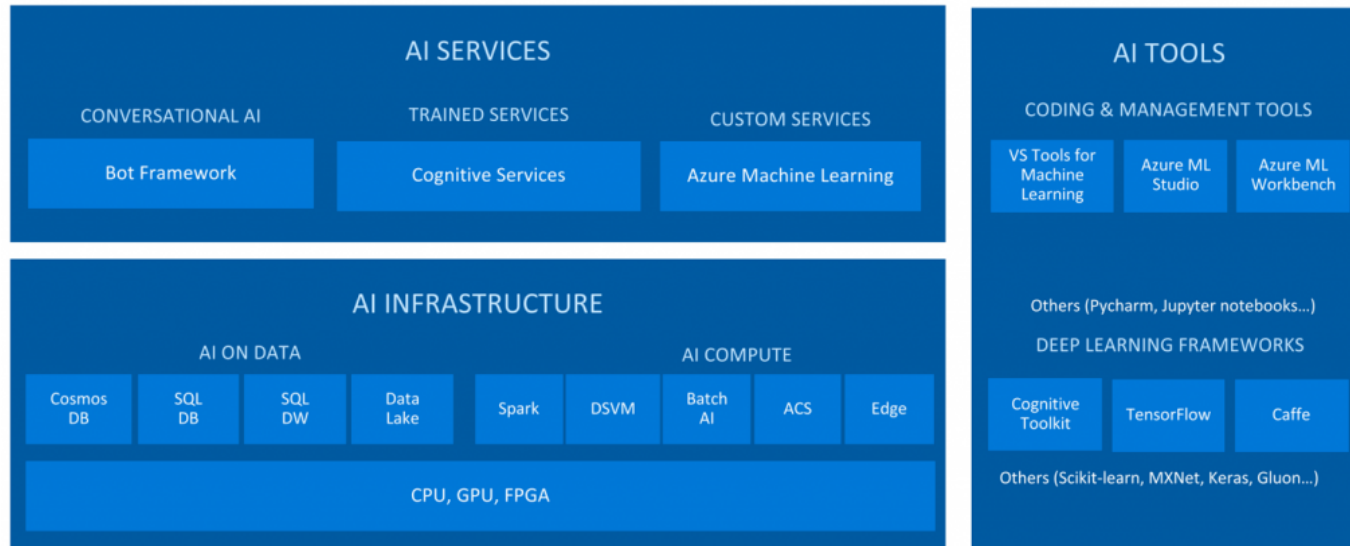
Cloud + AI: Google AI

Dialogflow: Build natural and rich conversational experiences



* source: <https://dialogflow.com>

Cloud + AI: Microsoft AI



Cloud + AI: IBM AI

IBM PowerAI Platform

PowerAI Software Distribution

Deep
Learning
Frameworks

Caffe



Caffe

IBM Caffe



torch



theano



Chainer

Supporting
Libraries

DIGITS

OpenBLAS

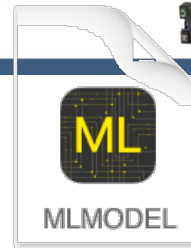
Distributed
Frameworks

Bazel

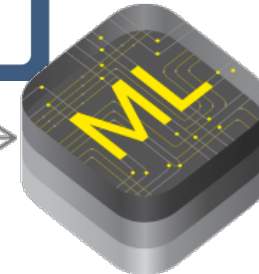
NCCL

IBM Power System for HPC, with NVLink

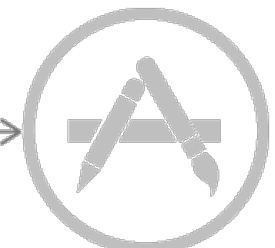
Breakthrough performance for GPU accelerated applications,
Including Deep Learning and Machine Learning



Core ML model



Core ML



Your app

Build a production-ready iOS app

BigData + AI



BigData + AI: Databricks

Open source project announcements:

- **MLflow:** MLflow (currently in alpha) is a cross-cloud open source framework designed to manage the entire machine learning lifecycle and work with any machine learning library.
- **Databricks Runtime for ML:** Based on customer demand, Databricks announced the new native and deep integration of popular ML libraries as part of the Databricks Runtime, including xgboost, scikit-learn, numpy as well as TensorFlow, Keras and Horovod. You can find more information [here](#).
- **Analytics-ready Data with Databricks Delta:** Simplify data reliability and performance of Apache Spark™ with Databricks Delta. Ensure your data is ready for analytics. [Watch the keynote by Apple](#) and demo to learn more.
- **Unified Analytics Platform for Genomics:** Accelerate discovery with a collaborative platform for genomic data processing, tertiary analytics and AI at massive scale. Read the blog and sign-up for our private preview.

AI Opportunity vs. Challenges

AI Opportunity:

Today more than ever, data scientists and Machine Learning practitioners have the opportunity to transform their business by implementing sophisticated models for recommendation engines, ads targeting, speech recognition, object recognition, bots, sentiment analysis, predictive analysis, and more.

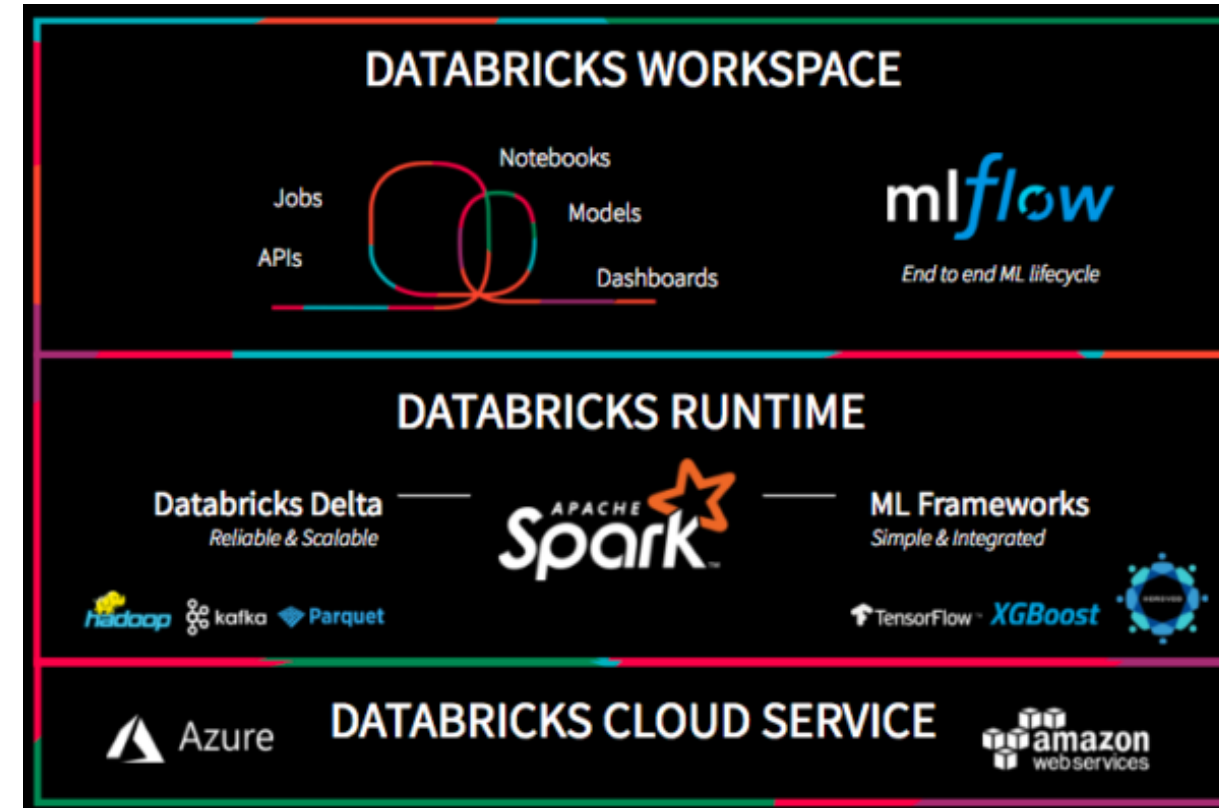
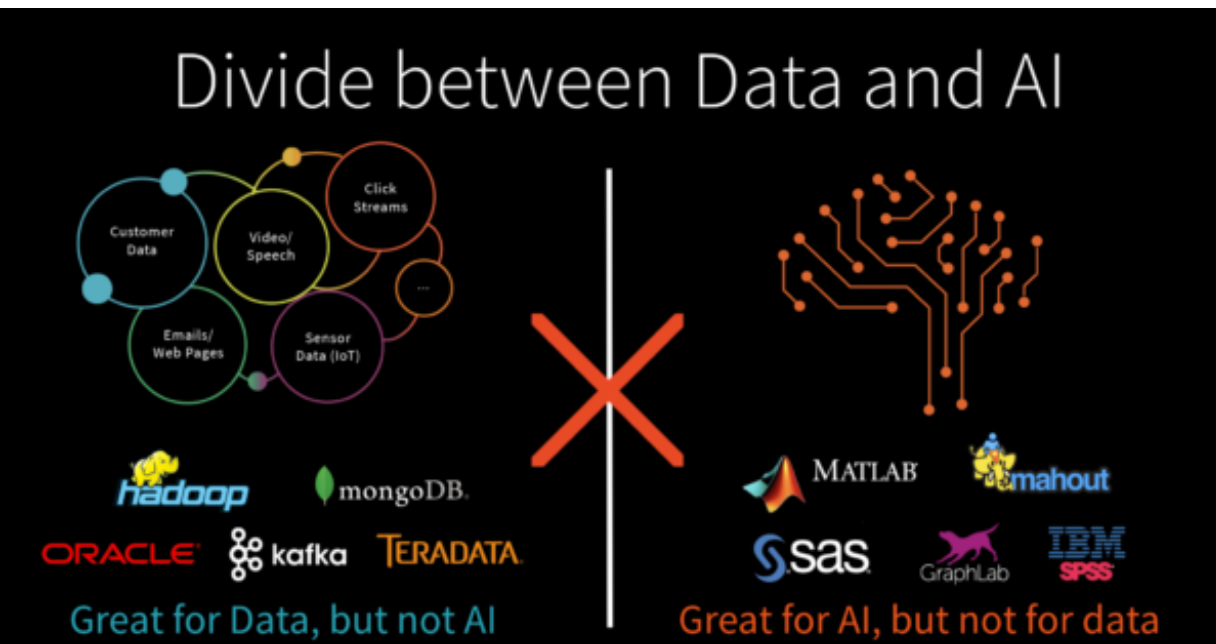
However, practitioners face multiple challenges when implementing such applications:

- **Environment Configuration:** Ability to setup and maintain complex environments due to the multitude of Machine Learning frameworks available.
- **Distributed Training:** Ability to train models in a distributed fashion to get results faster. This requires specialized skills and complex code to manage.
- **Compute Power:** Ability to run parallelized workloads on GPUs for maximum performances. This is essential for Deep Learning applications.

** source: <https://databricks.com/blog/2018/06/05/distributed-deep-learning-made-simple.html>*

BigData + AI: Databricks

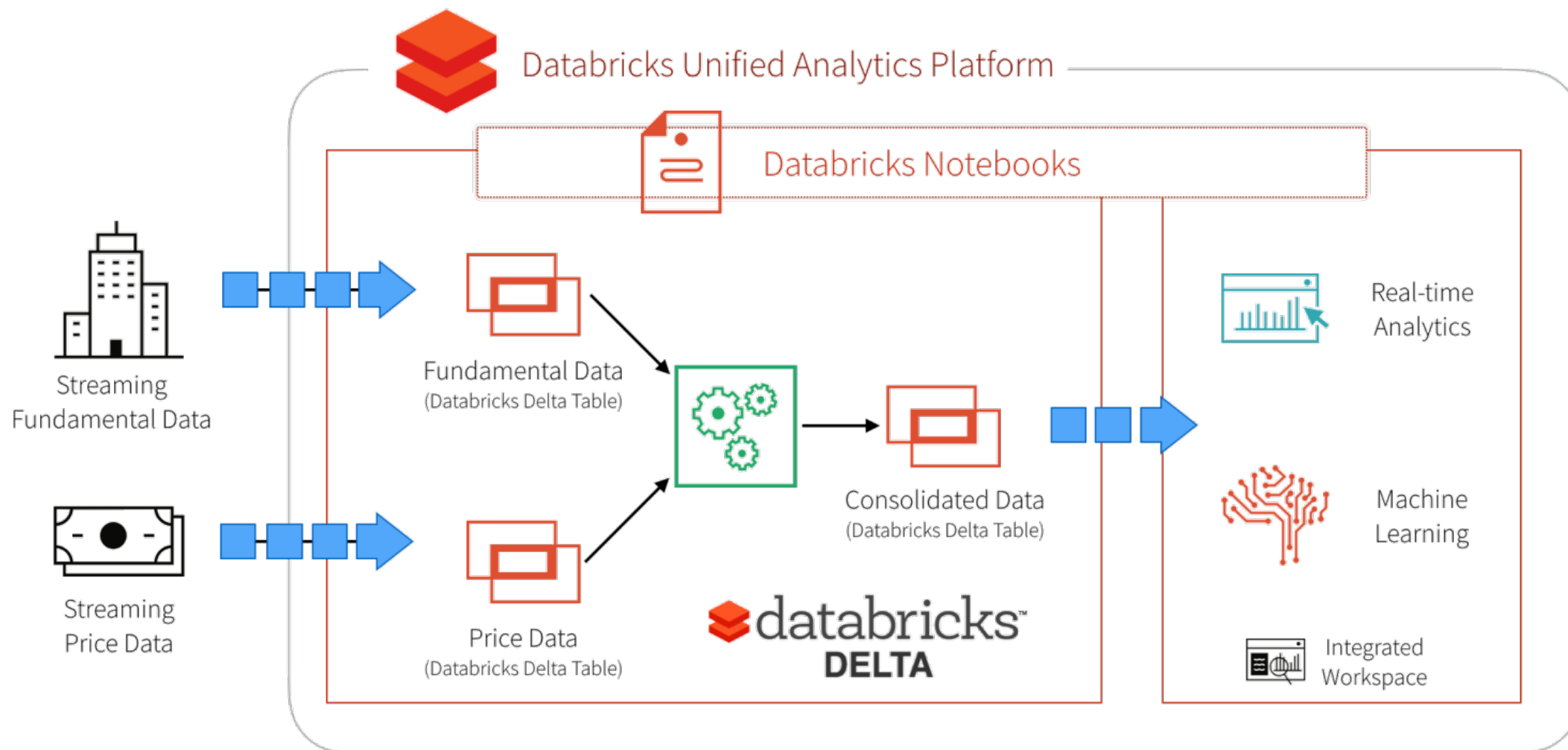
Databricks Unified Analytics Platform



<https://databricks.com/blog/2018/06/05/accelerating-innovation-with-unified-analytics.html>

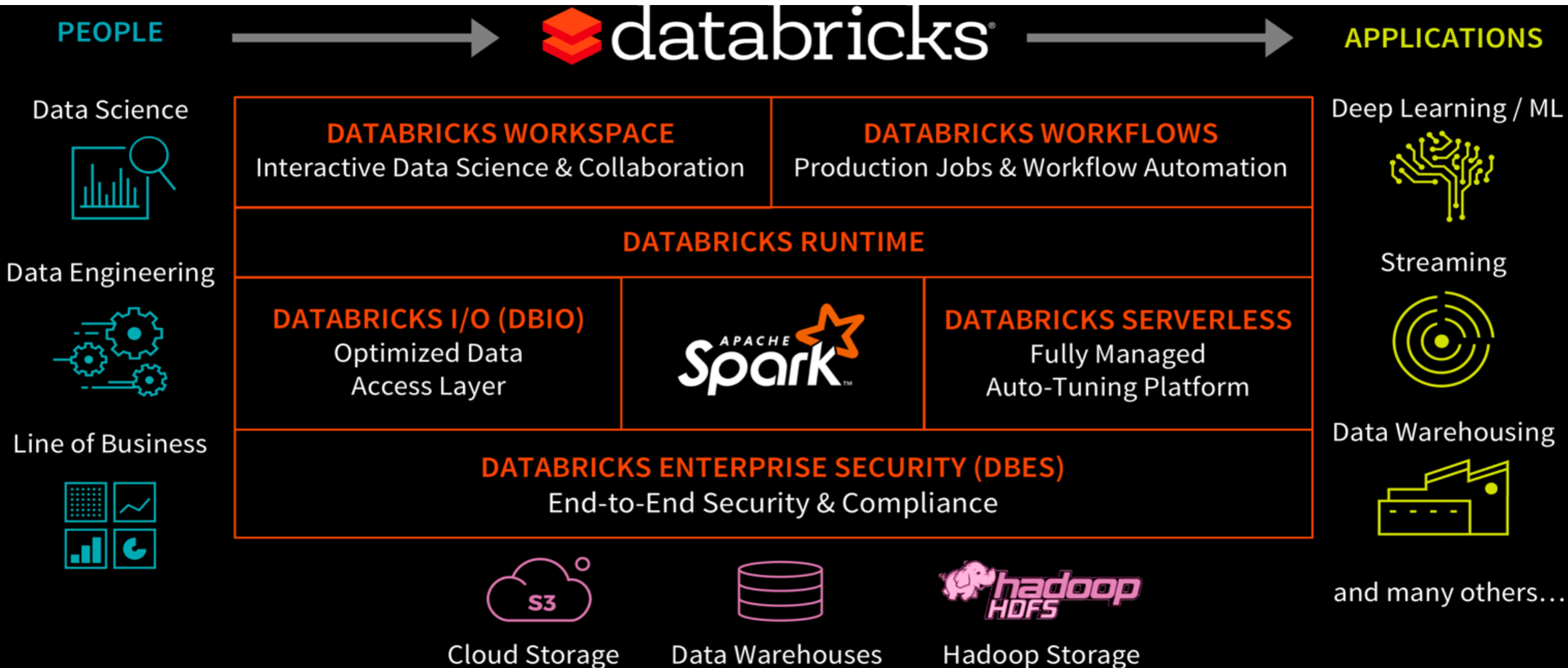
BigData + AI: Databricks

Databricks Unified Analytics Platform

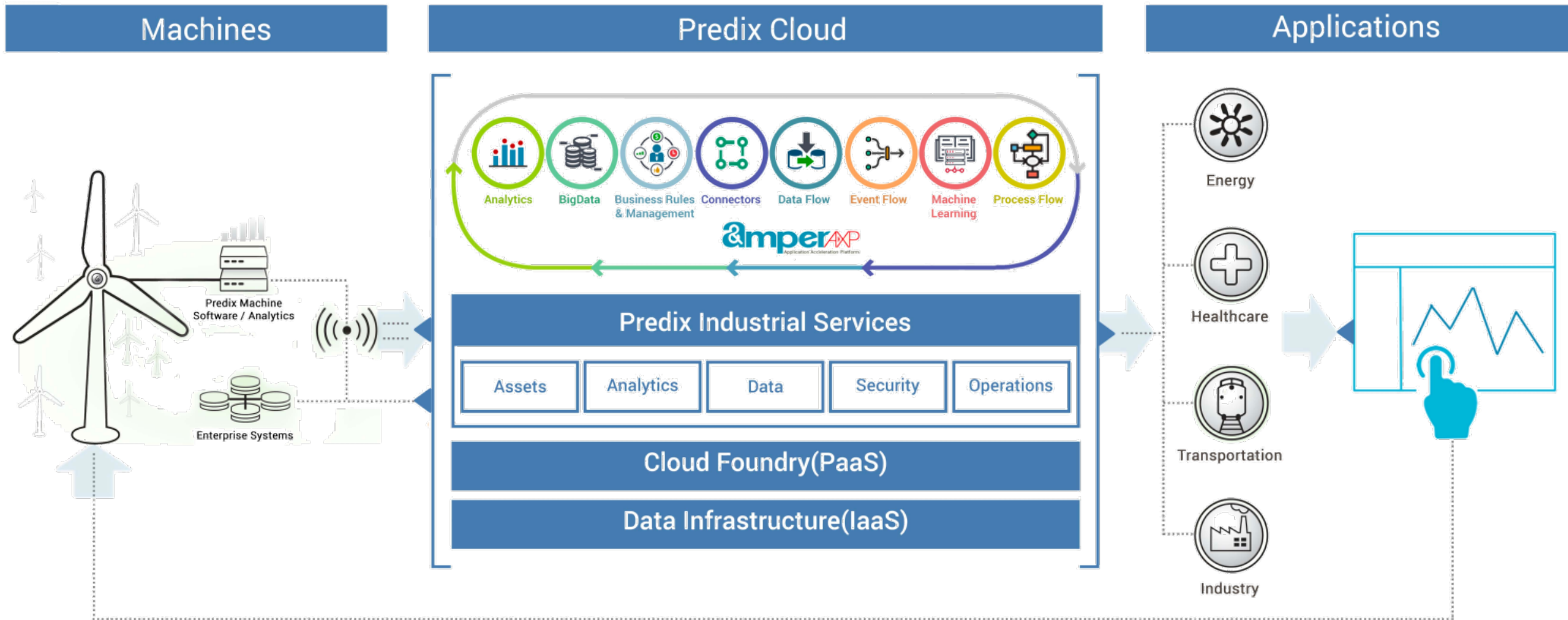


BigData + AI: Databricks

Databricks Unified Analytics Platform



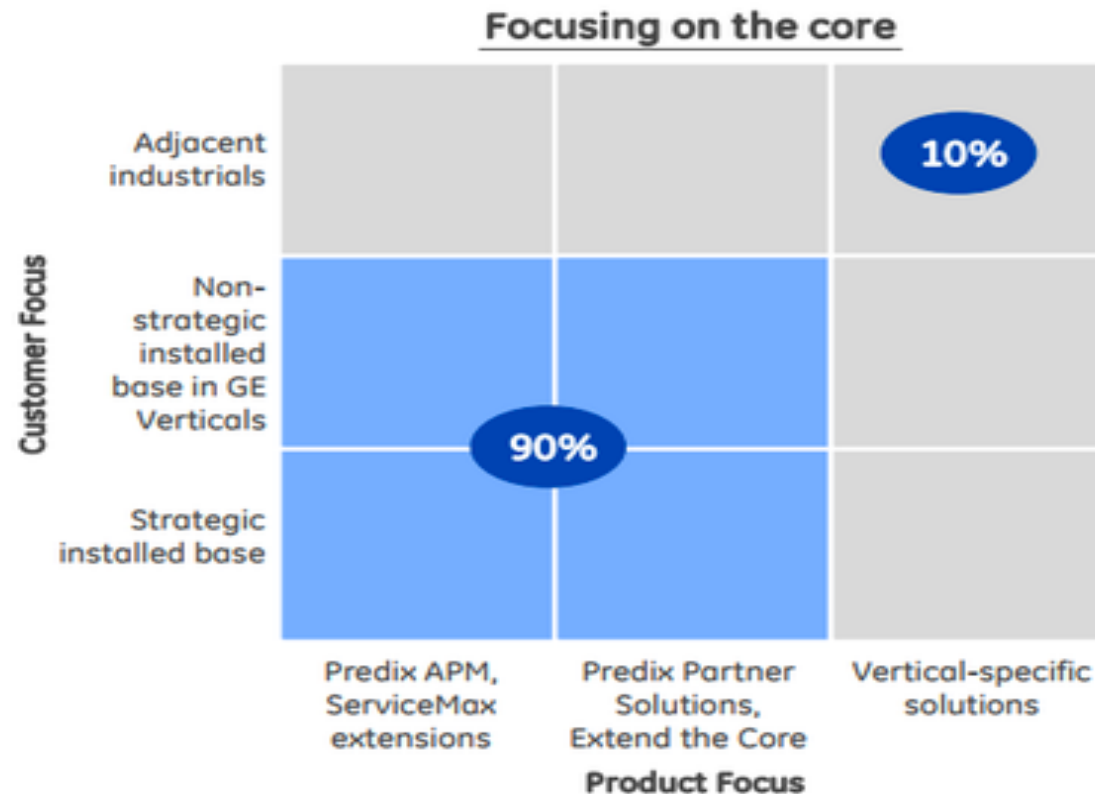
IoT + AI: GE Predix



* source: <http://www.amperxp.com/ge-predix-and-amper/>

IoT + AI: GE Predix

4 Digital



Approach

- 1 Lead with Predix applications that drive customer outcomes: APM, OPM, and ServiceMax
- 2 Focus spend on Predix platform differentiation: asset model, Edge to Cloud, Digital Twin
- 3 Partner for technology that is not differentiated (i.e. Cloud)
- 4 Prioritize go to market around GE business verticals where win rate is ~2x higher

Customer examples



- APM used at 1.3 GW power plant
- 1% efficiency gain on mixed fleet
- ~\$18MM annual customer value



- APM used to improve asset availability
- \$1.4MM saved in production losses
- ~\$1.3MM revenue increase



Targeting \$1B+ Predix-powered revenue and \$0.4B of cost out in 2018

18

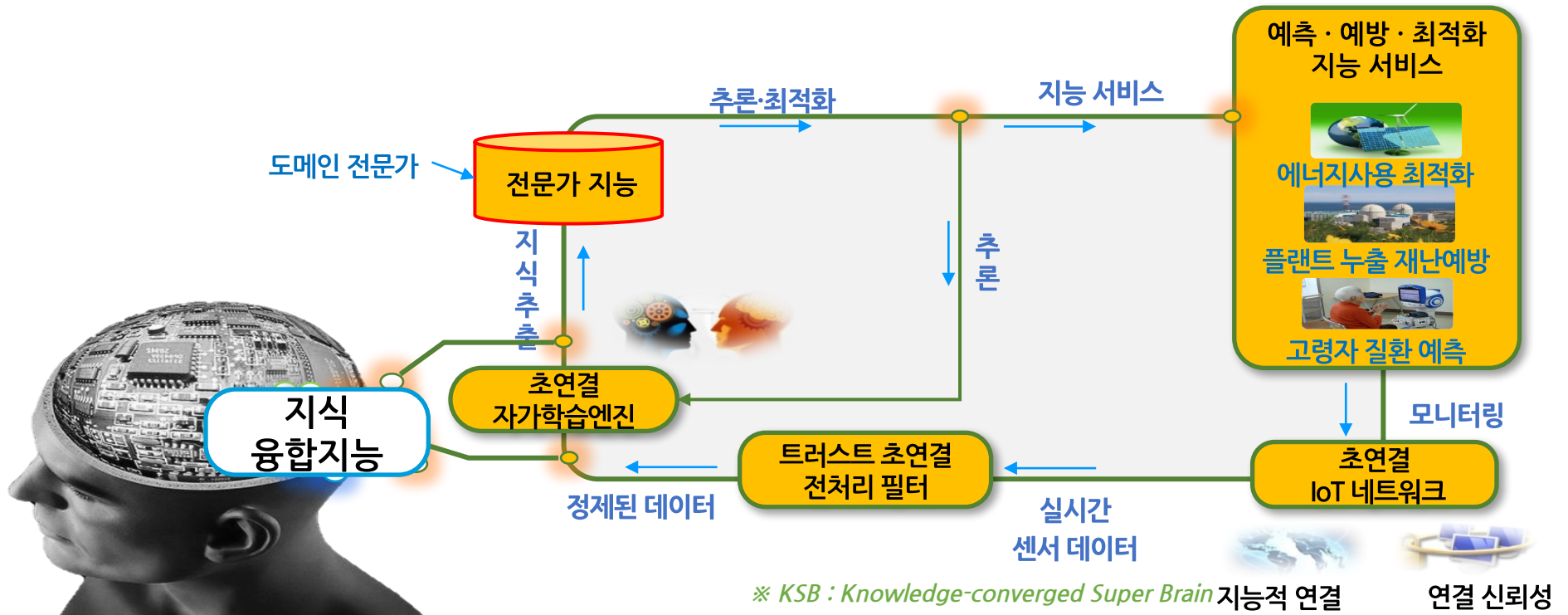
* source: <https://www.zdnet.com/article/ge-to-hone-digital-efforts-leverage-additive-manufacturing-as-it-focuses-on-core-businesses/>

26

III. Strategies of KSB Framework

Knowledge-converged Super Brain

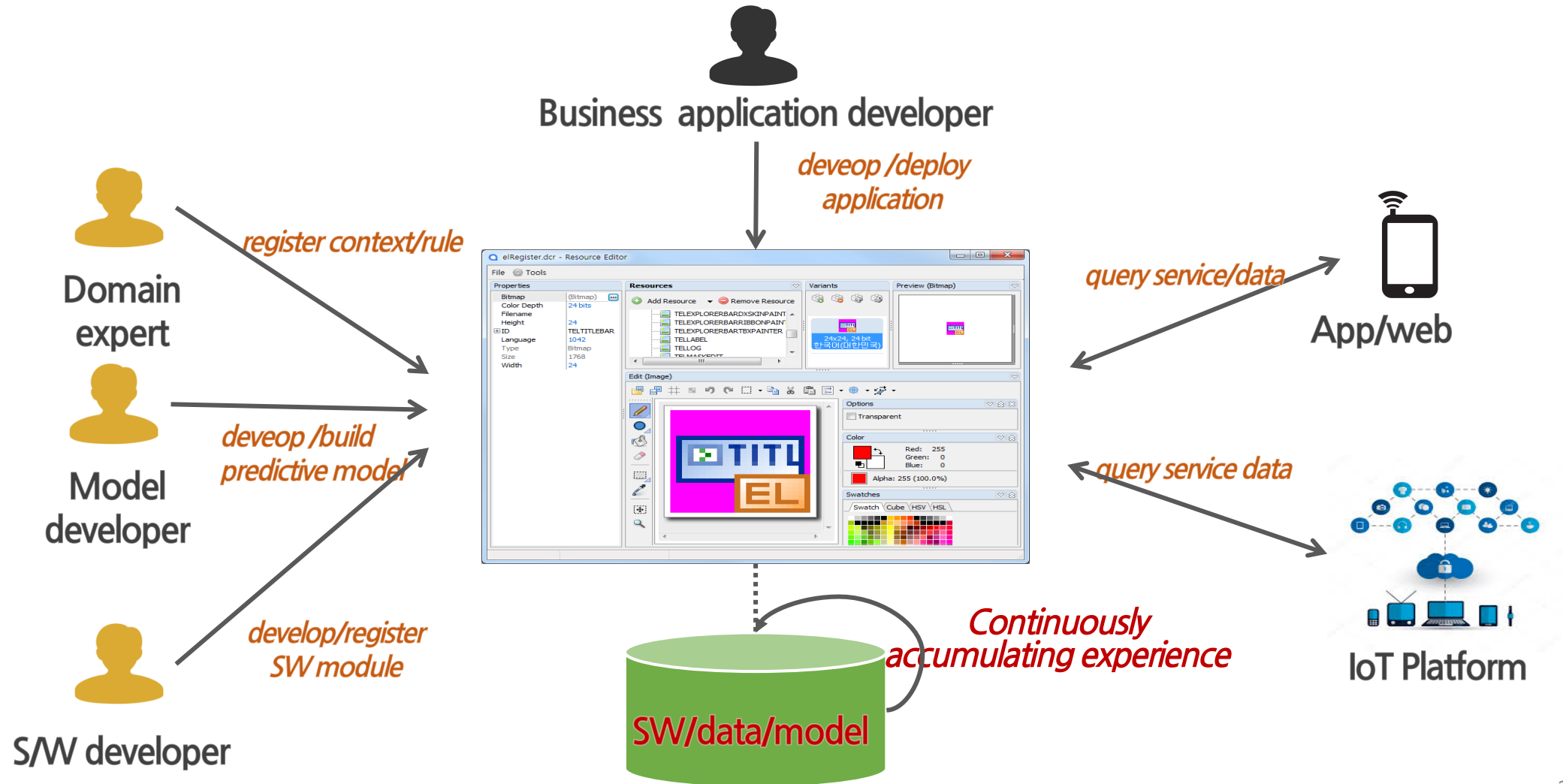
- ① 상태를 모니터링하여 데이터를 수집 ② 수집된 데이터를 정제 ③ 기계학습·지식추출하고
④ 도메인의 전문가지식과 융합하여 추론·최적화함으로써 ⑤ 예측·예방·최적화 지능 서비스 제공



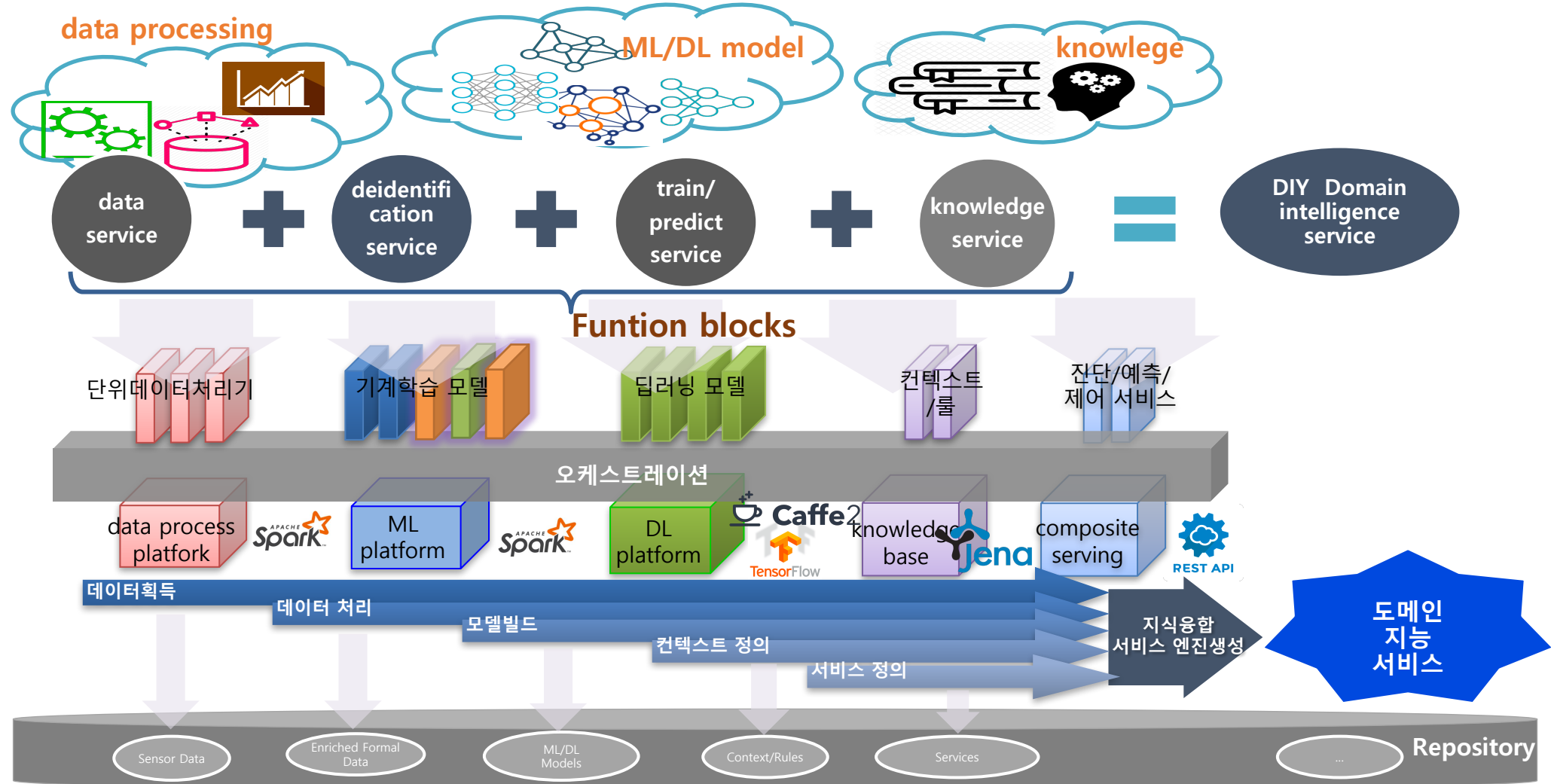
Knowledge-converged Super Brain



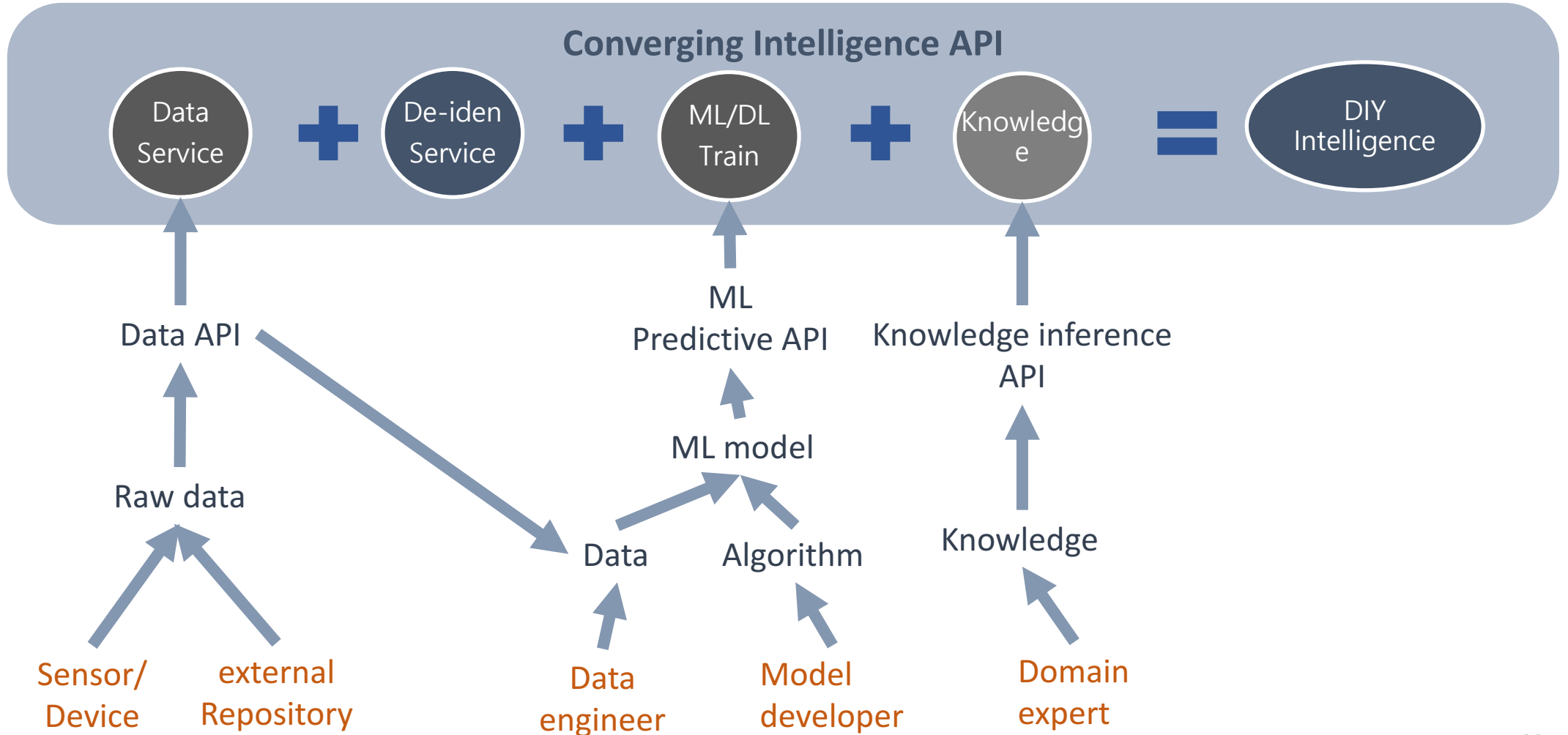
KSB Framework Approach



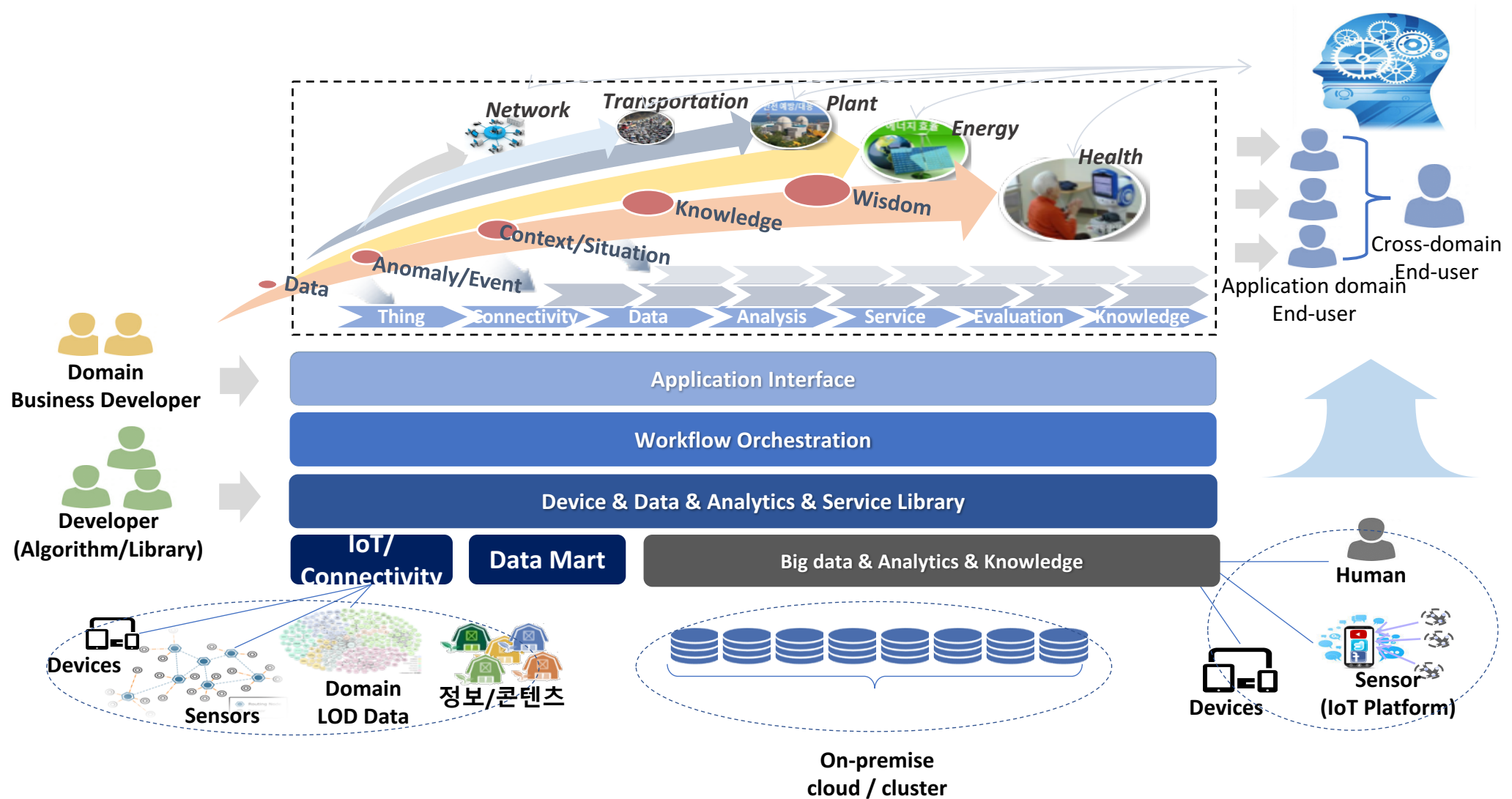
KSB Framework Concept



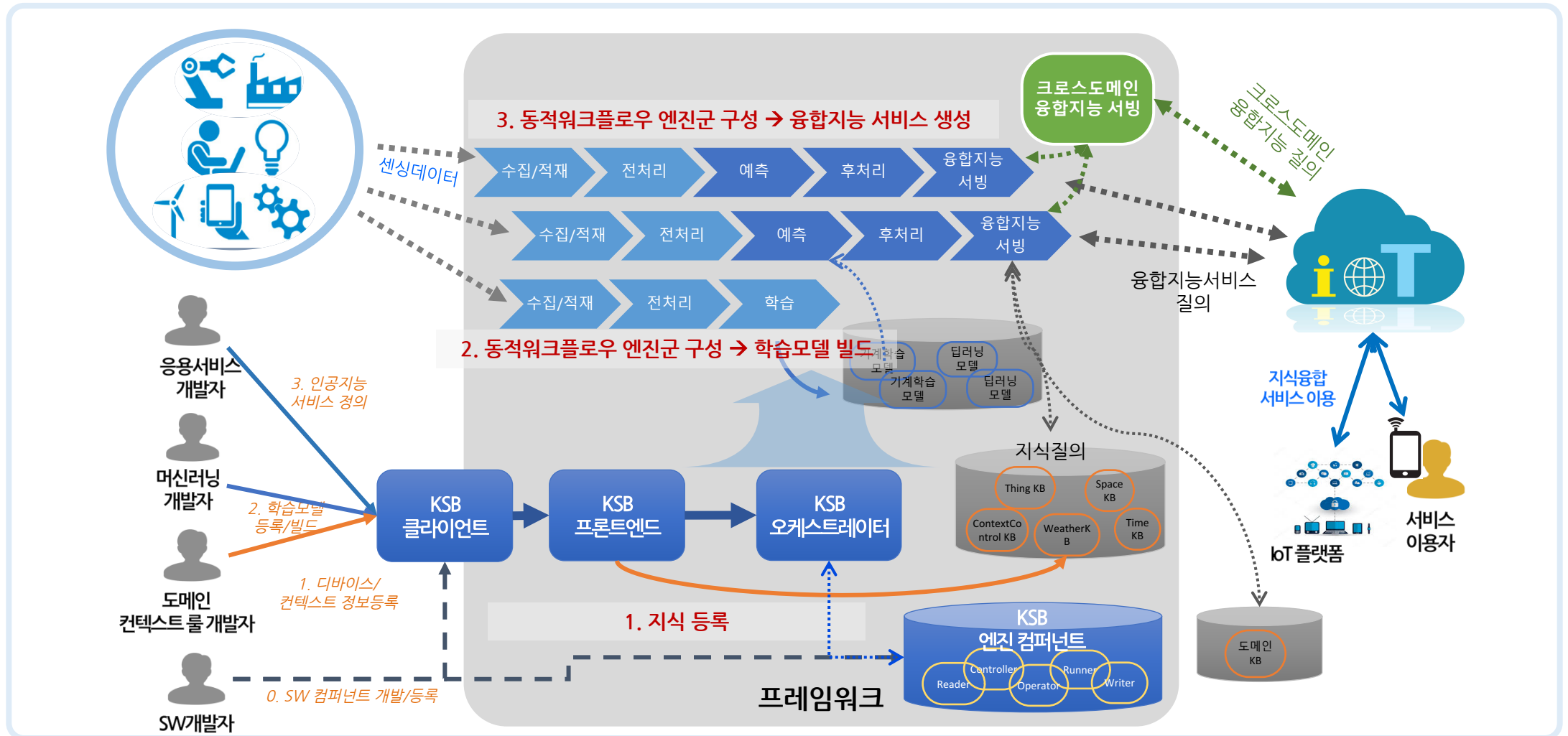
KSB Framework Concept



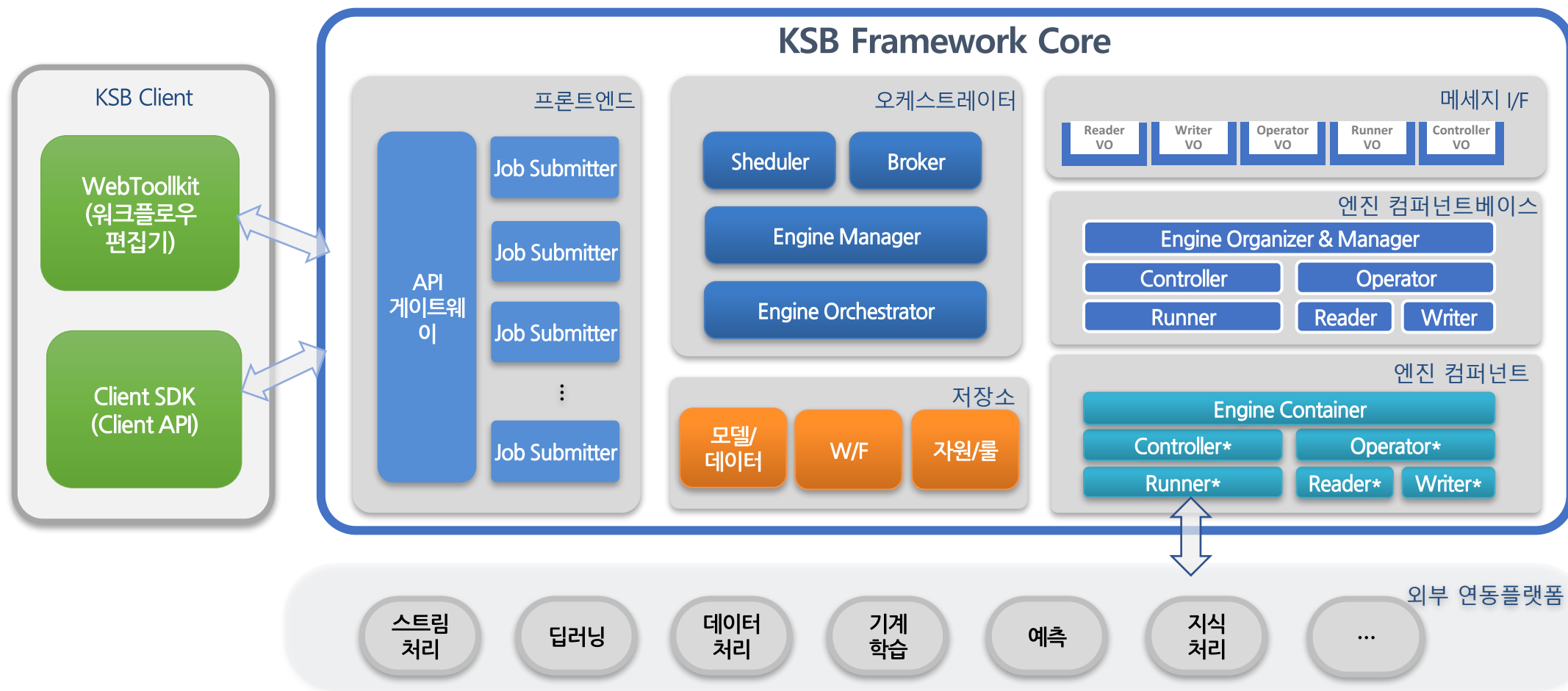
KSB Framework Overview



KSB Framework Overview



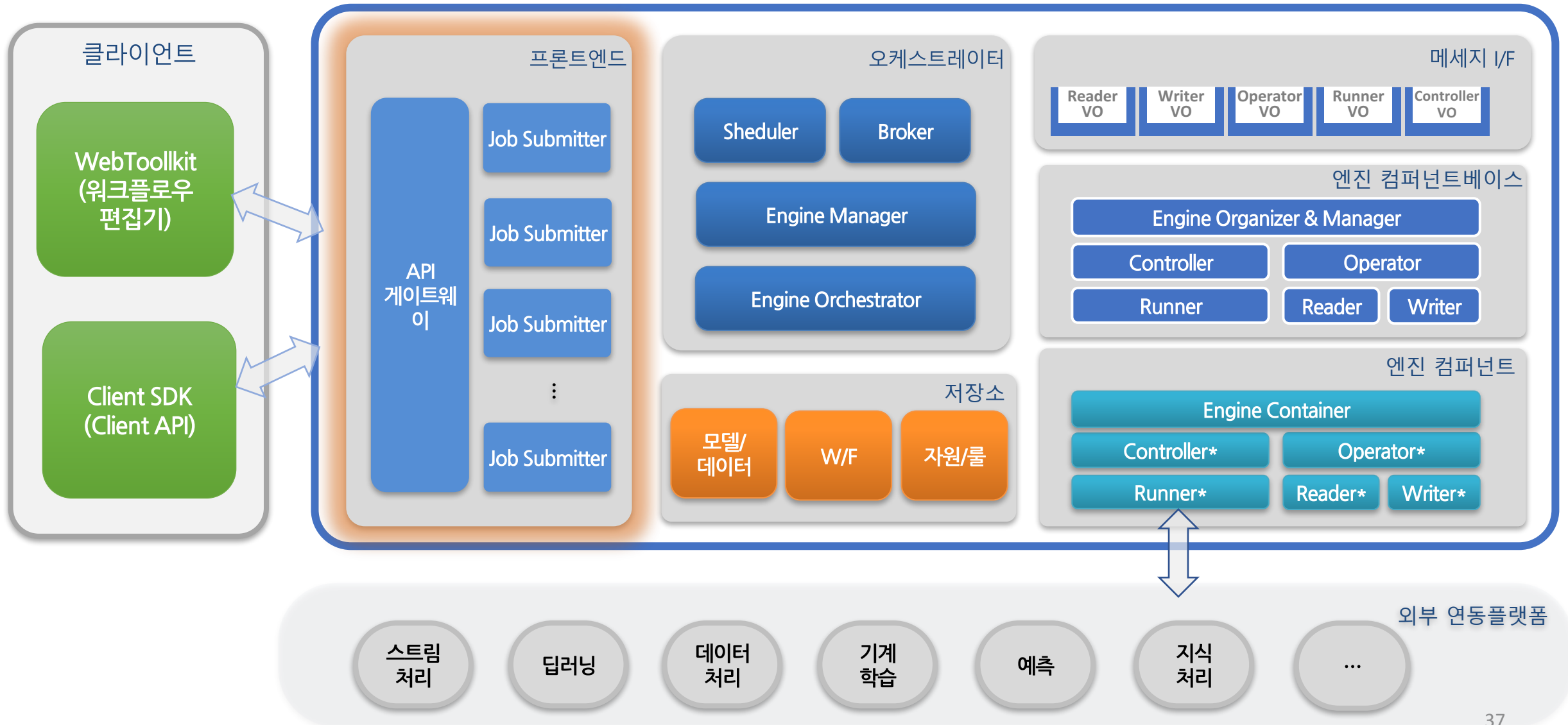
KSB Framework Architecture



KSB 인공지능 프레임워크 Core



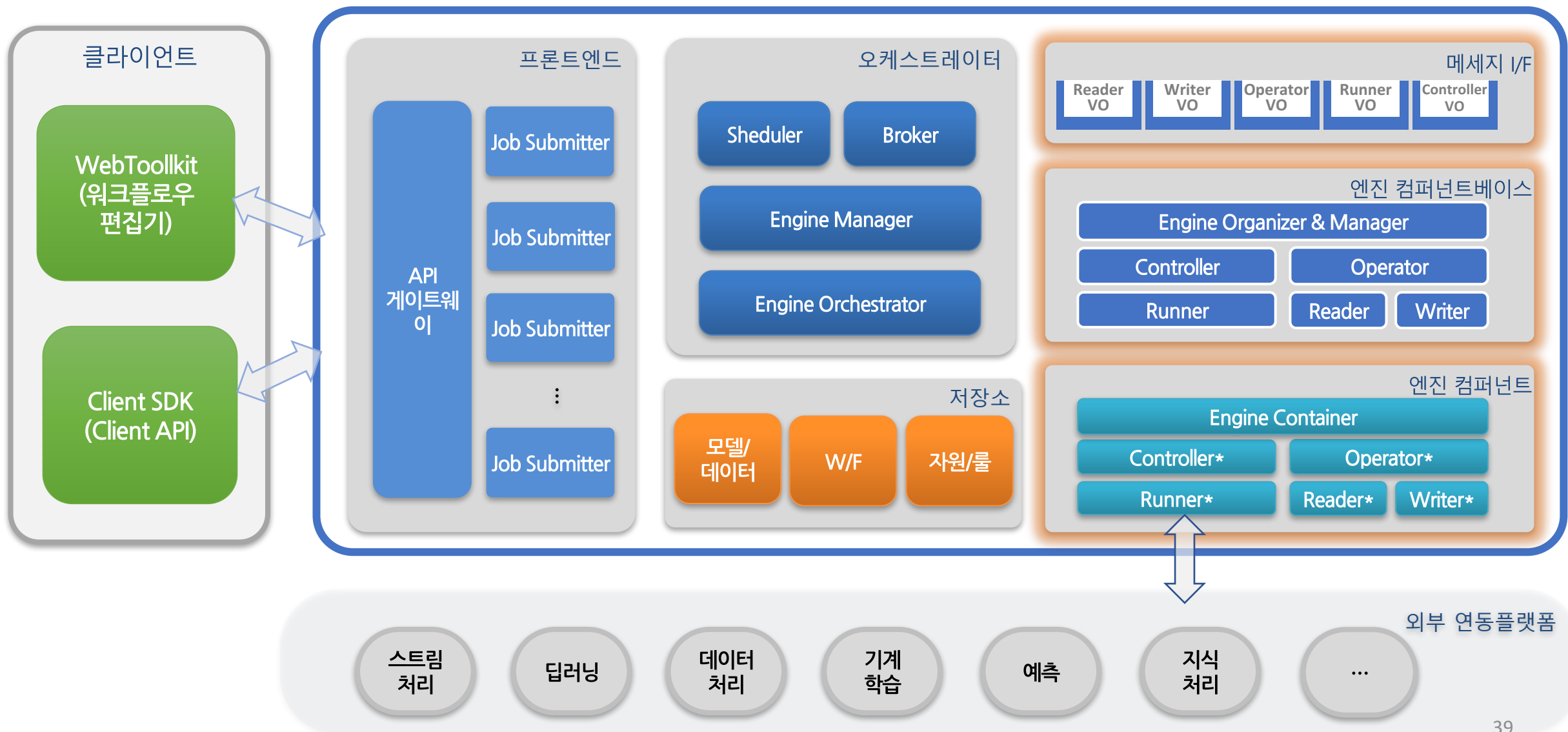
KSB 인공지능 프레임워크 Core



KSB 인공지능 프레임워크 Core



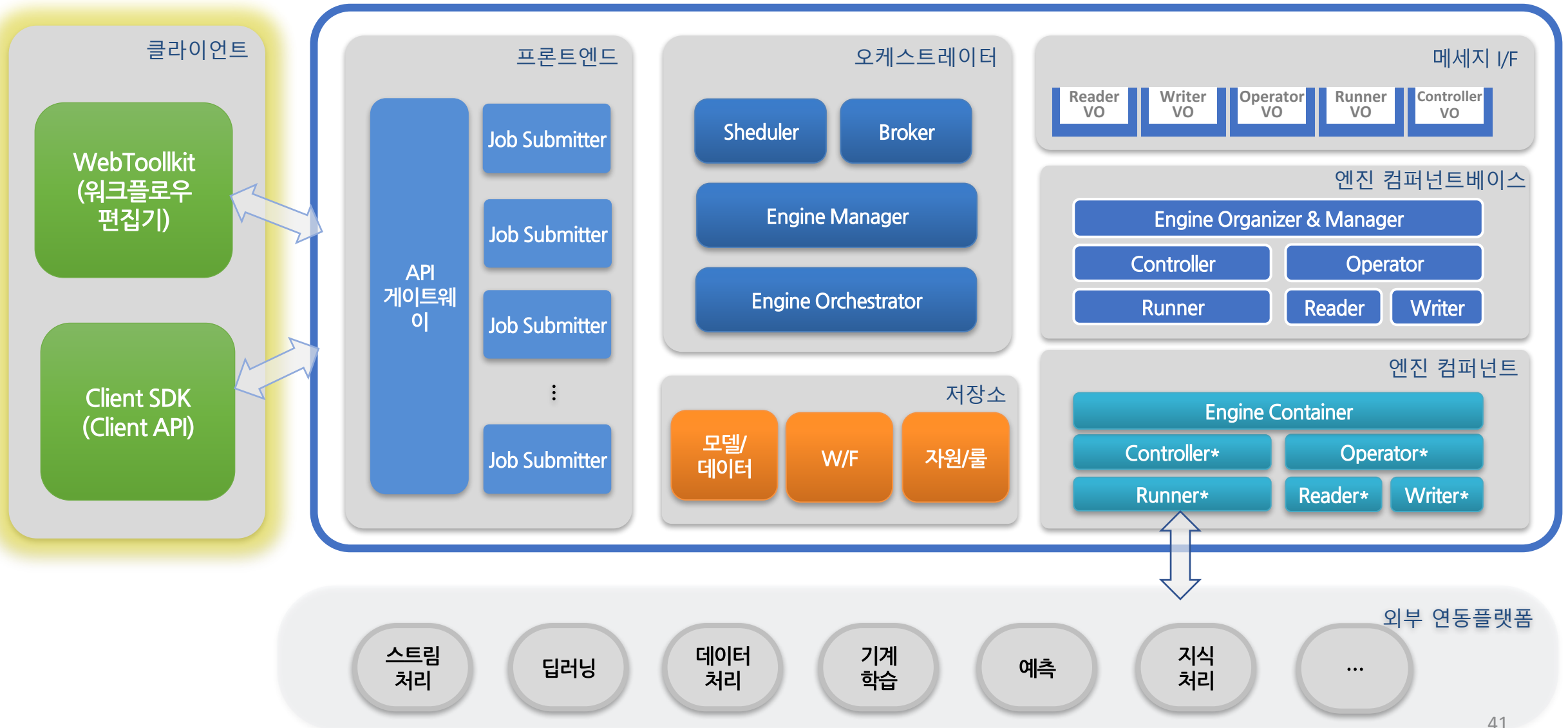
KSB 인공지능 프레임워크 Core



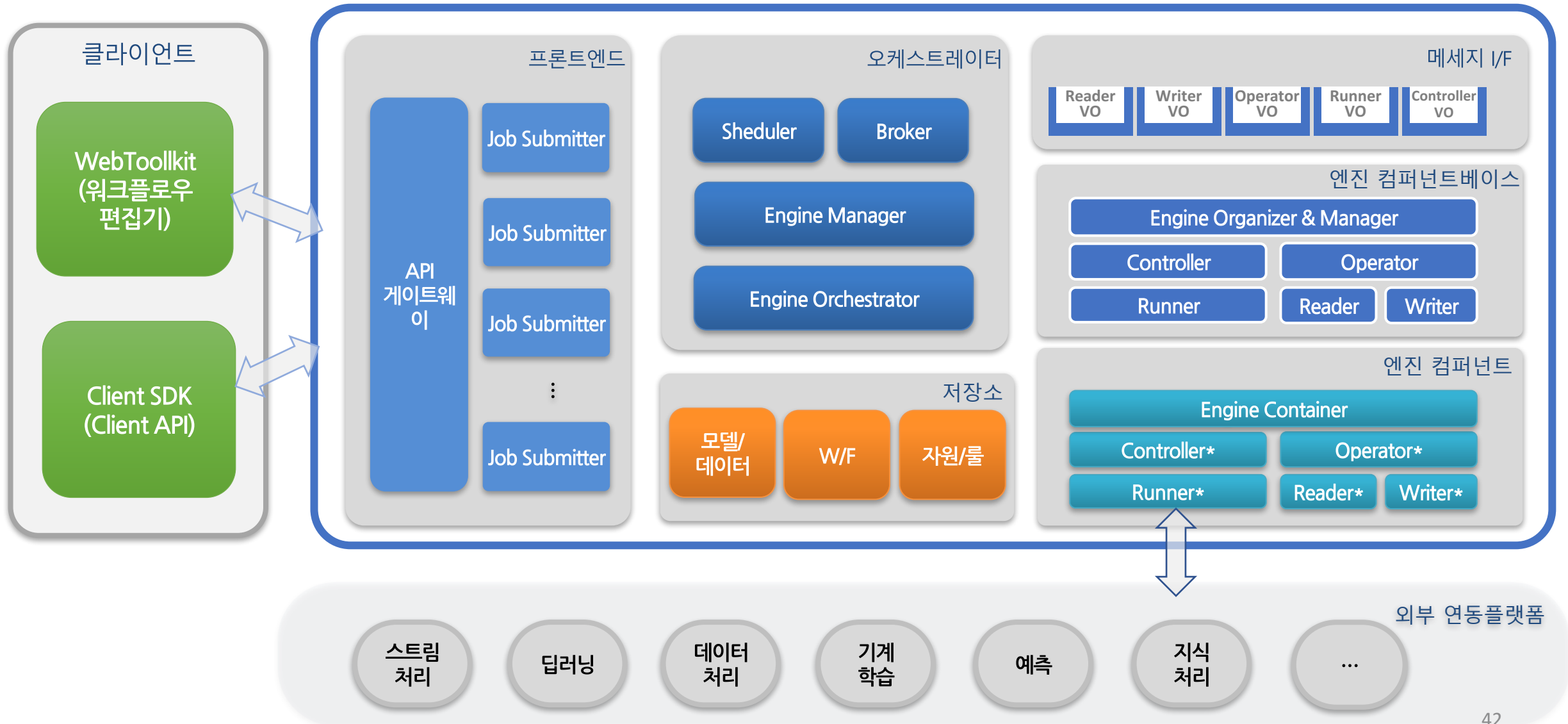
KSB 인공지능 프레임워크 Core



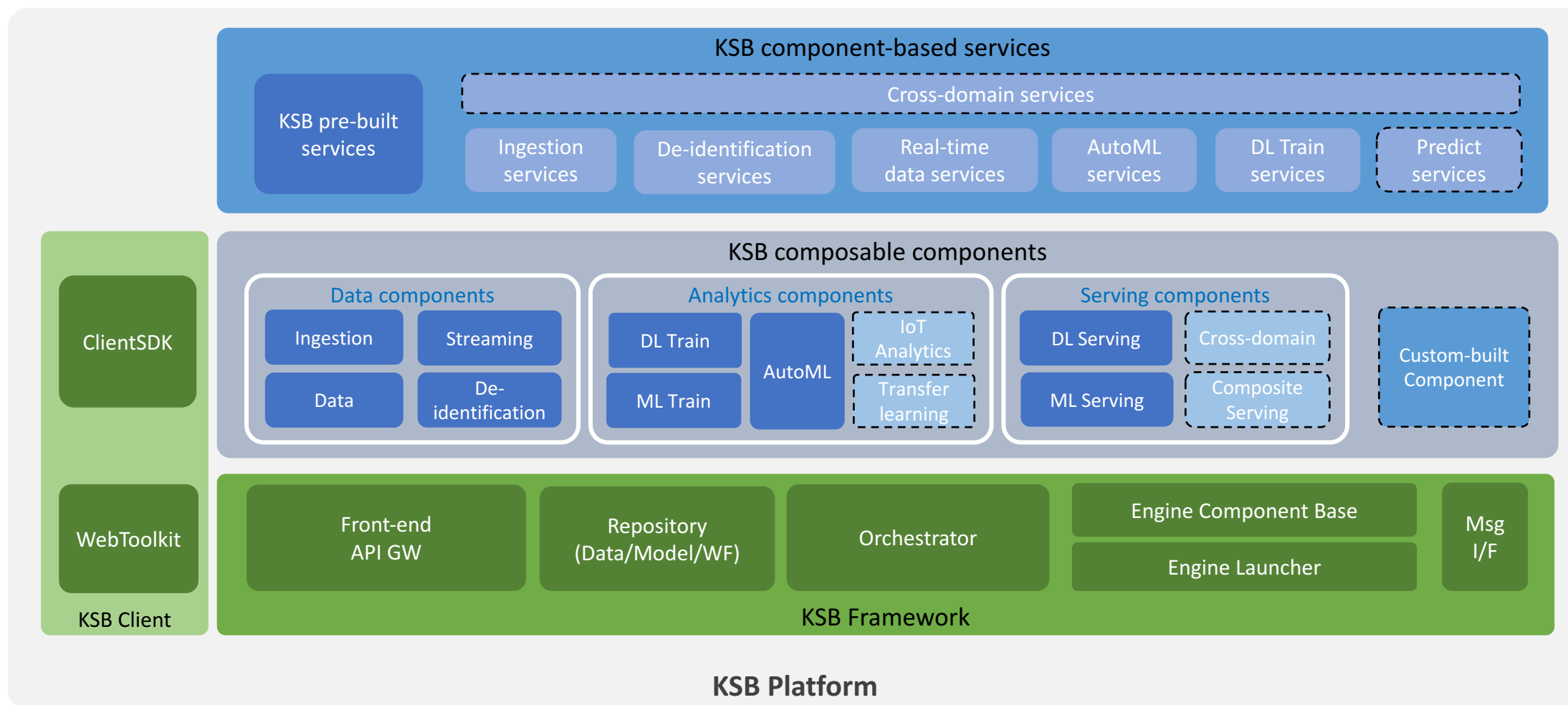
KSB 인공지능 프레임워크 Core

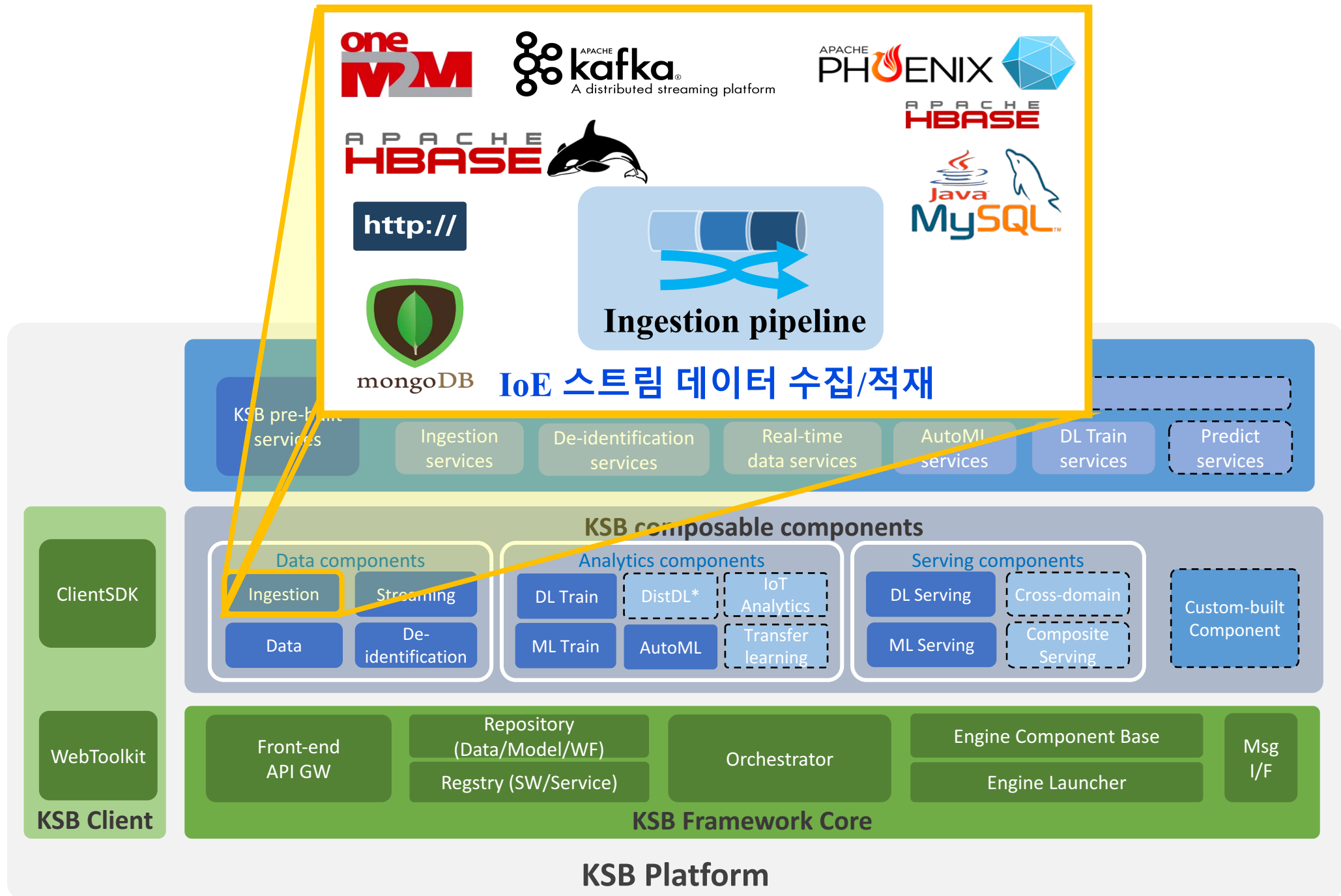


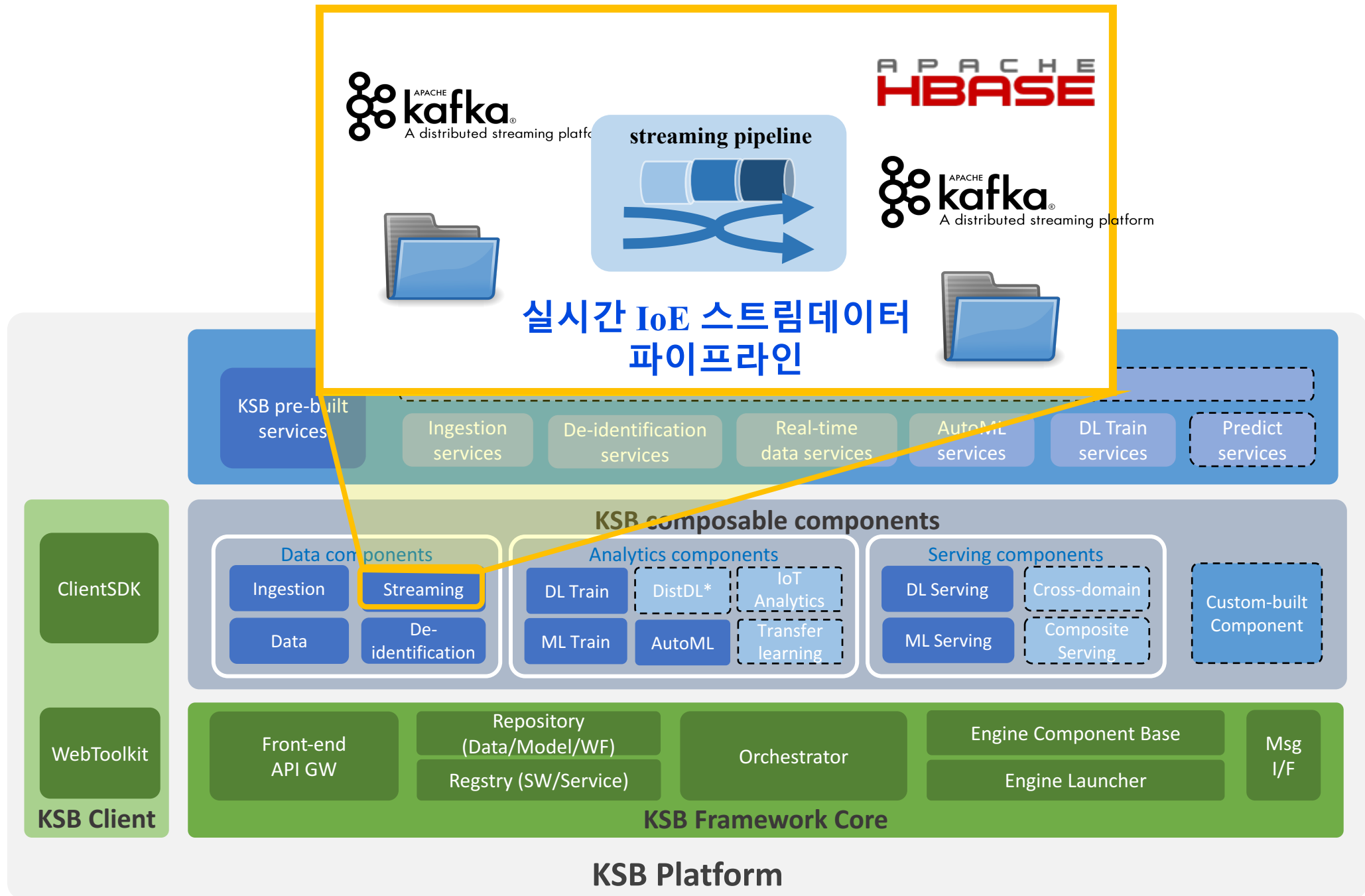
KSB 인공지능 프레임워크 Core

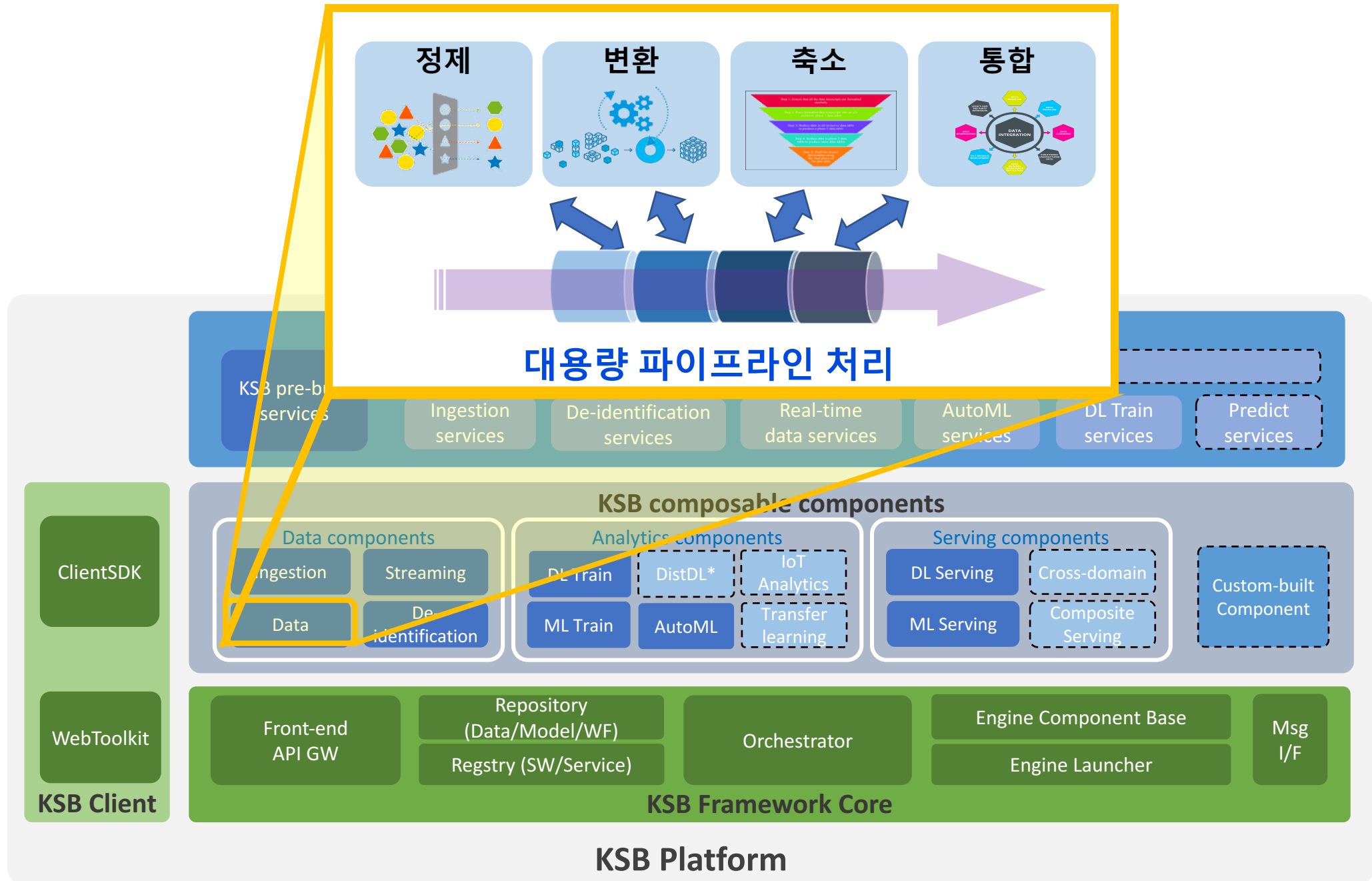


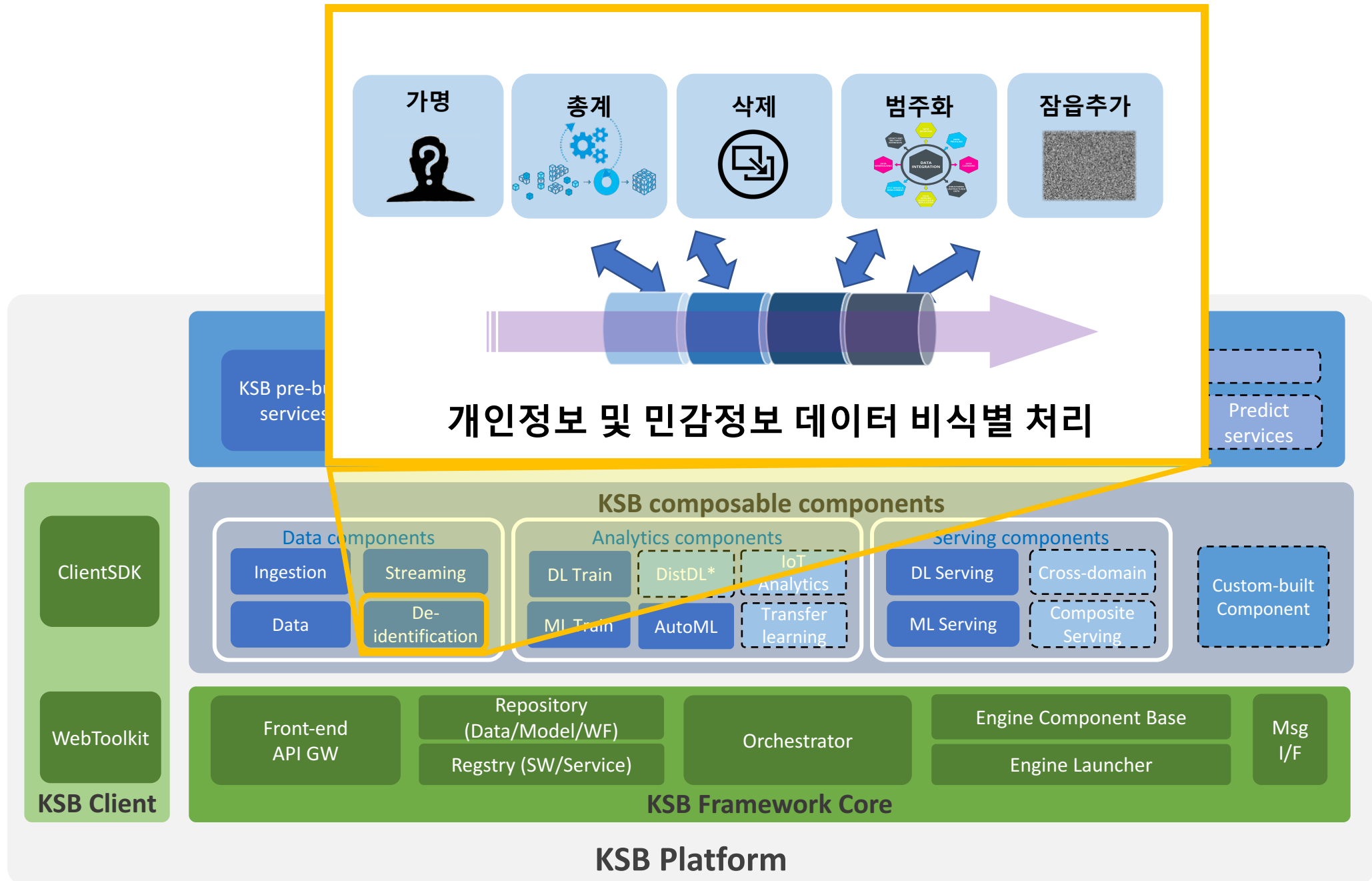
KSB Framework Stack

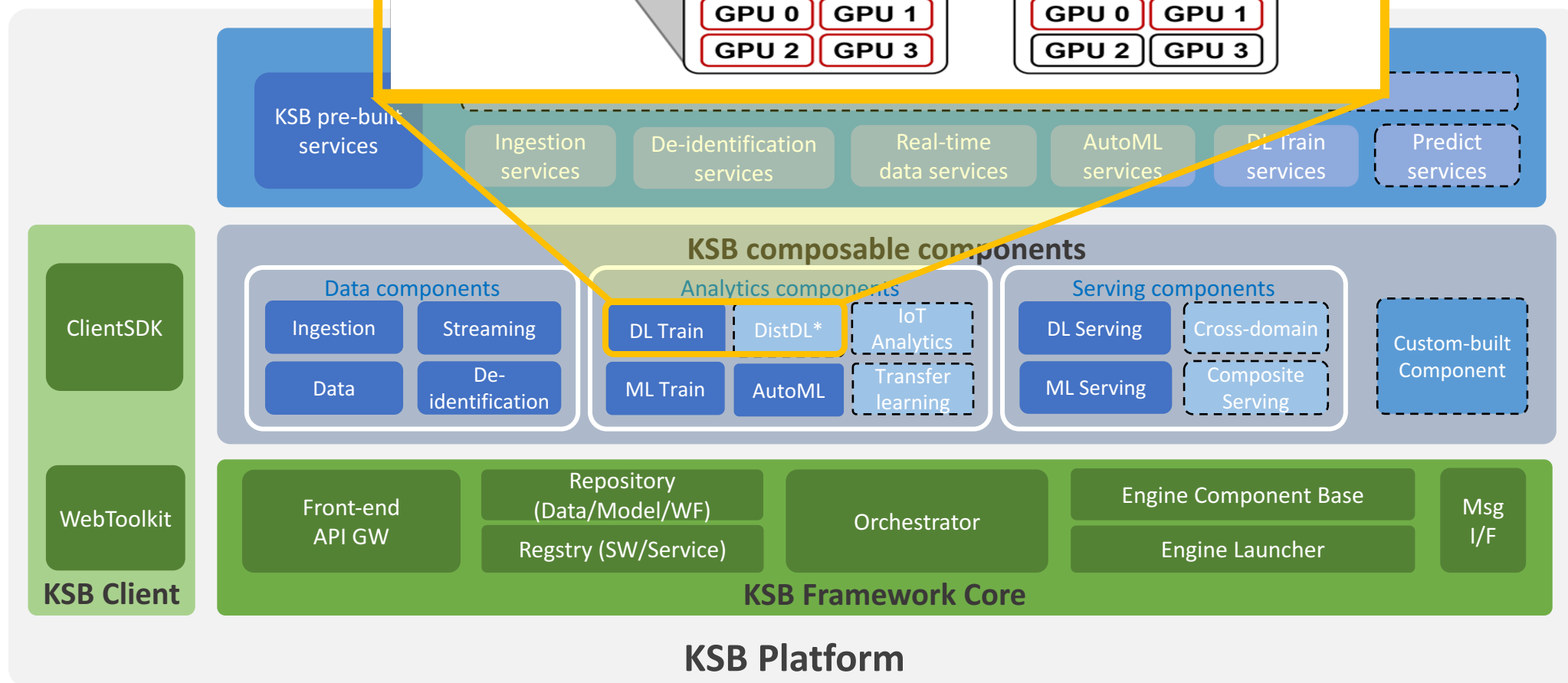
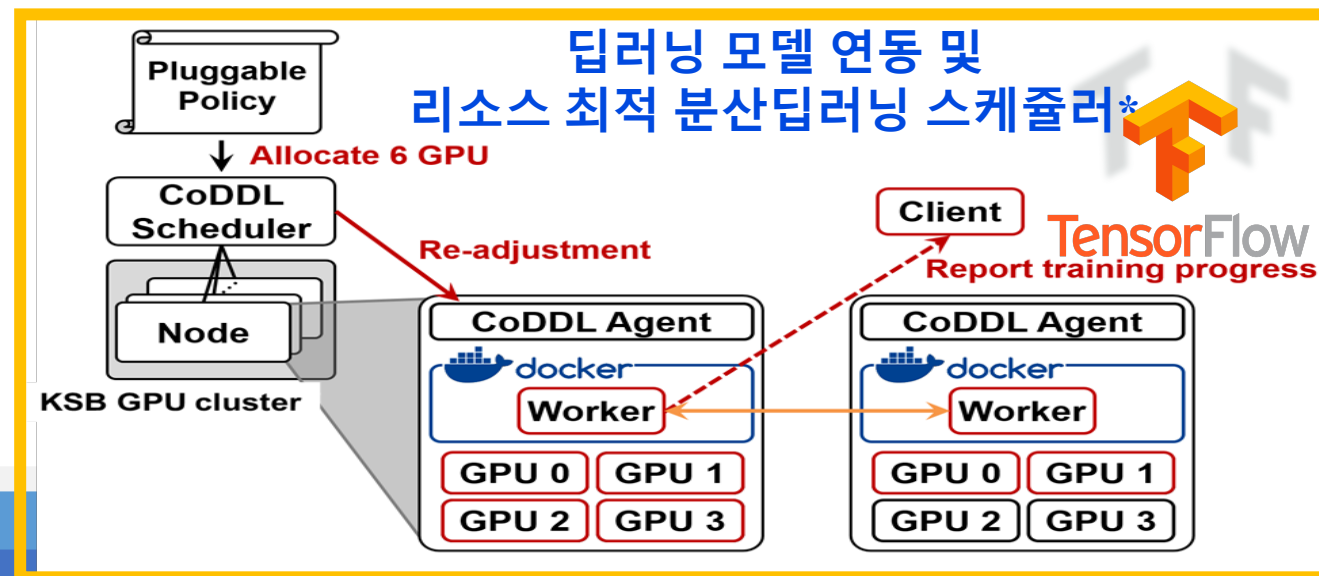




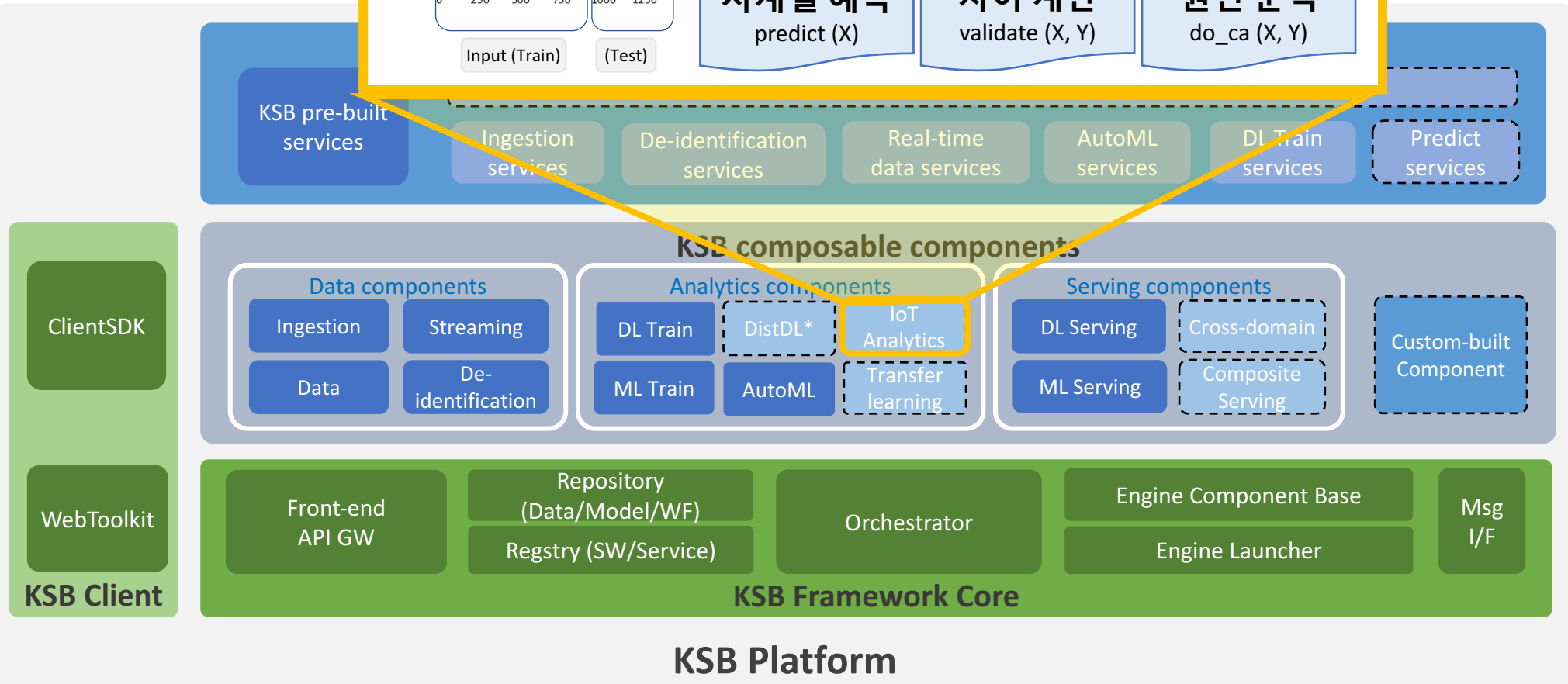
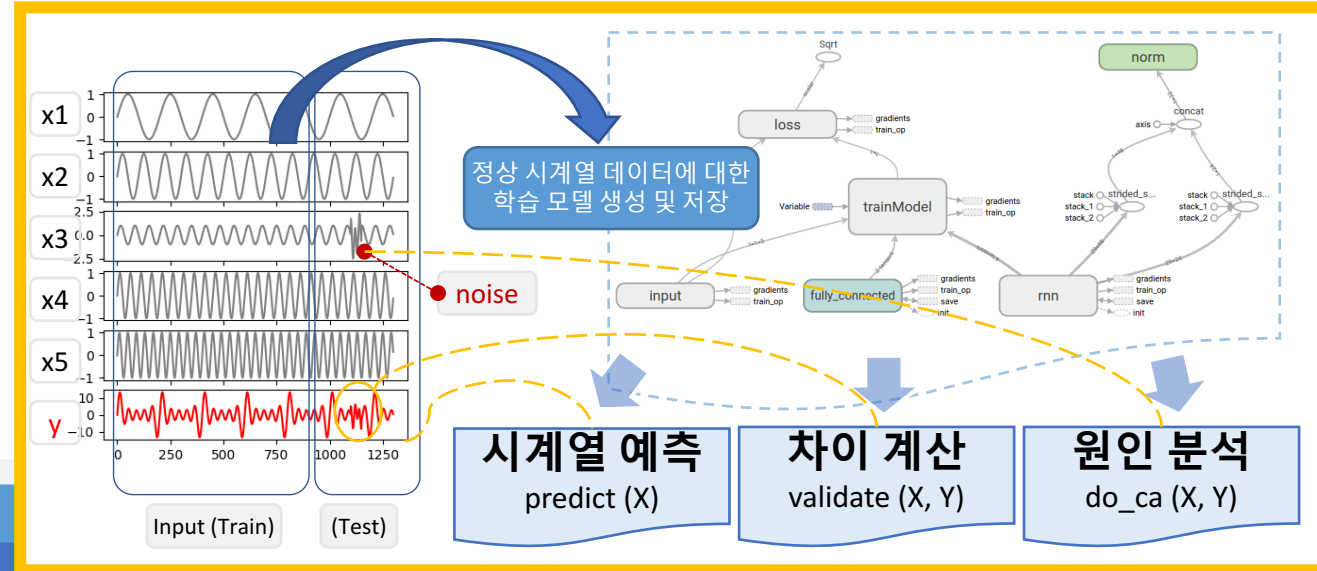


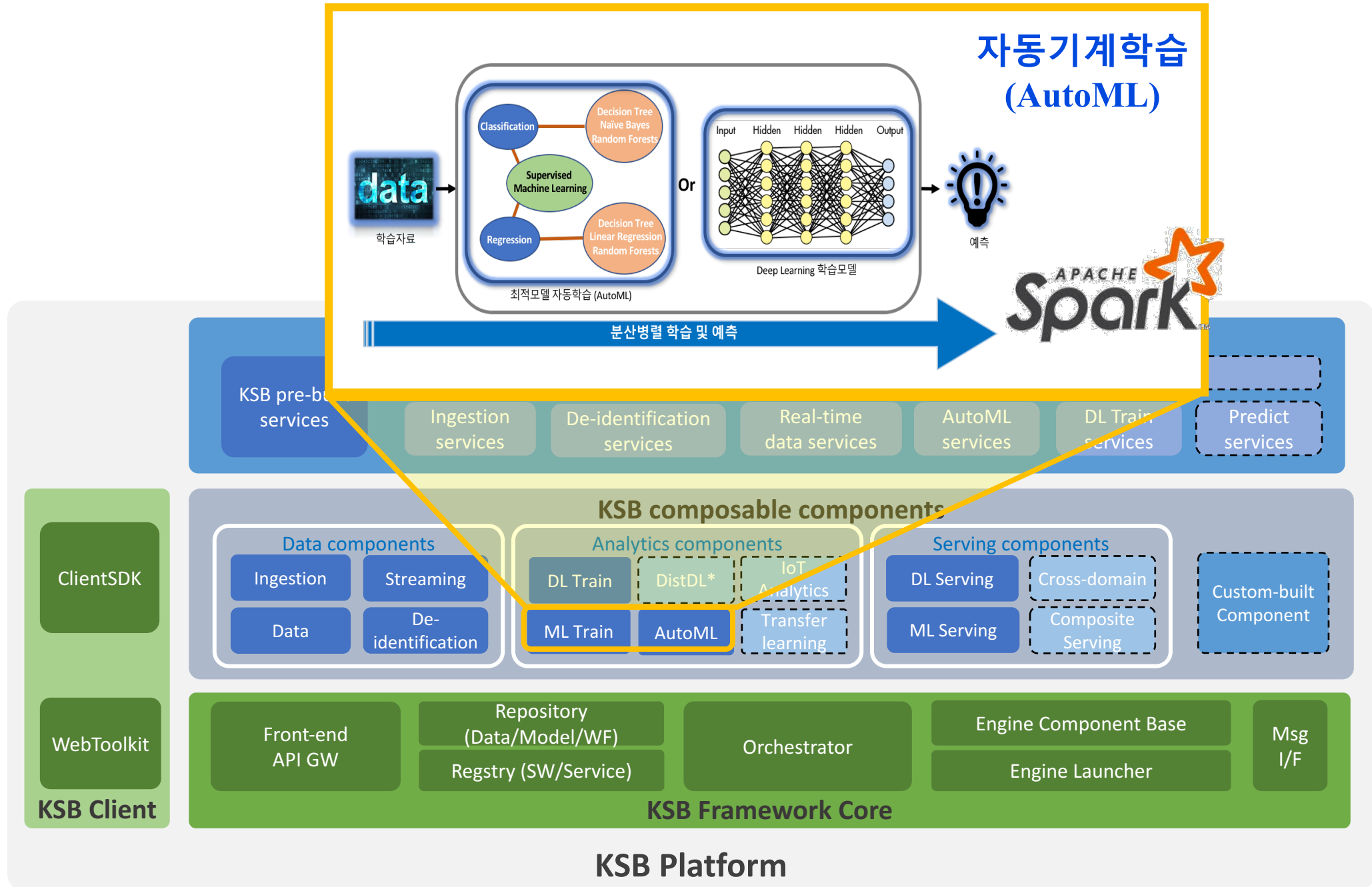


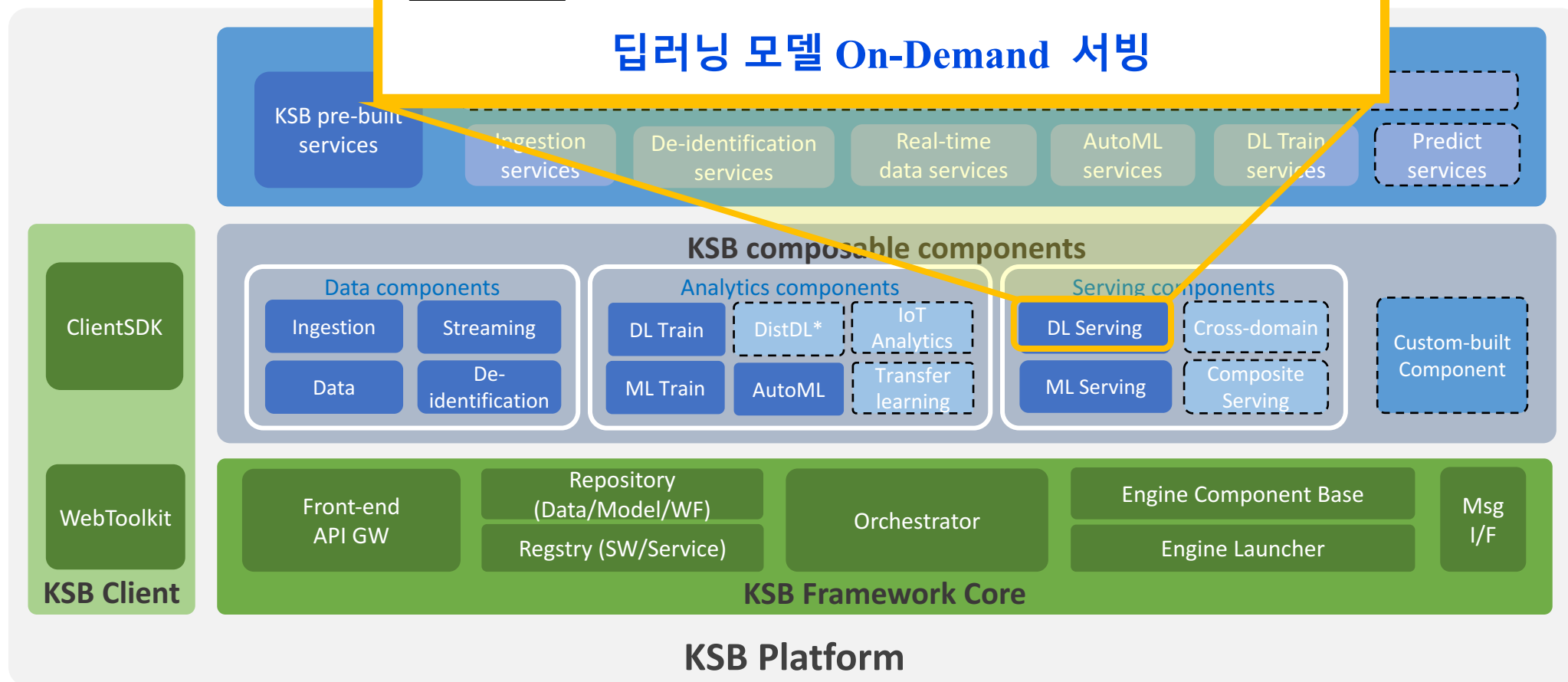
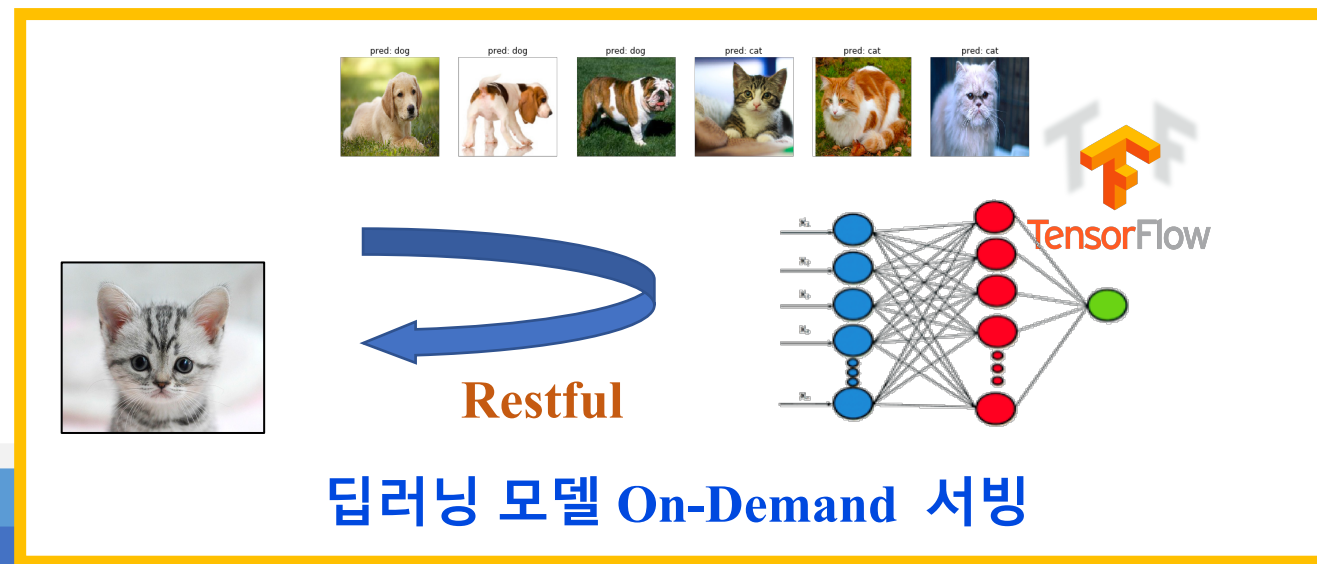


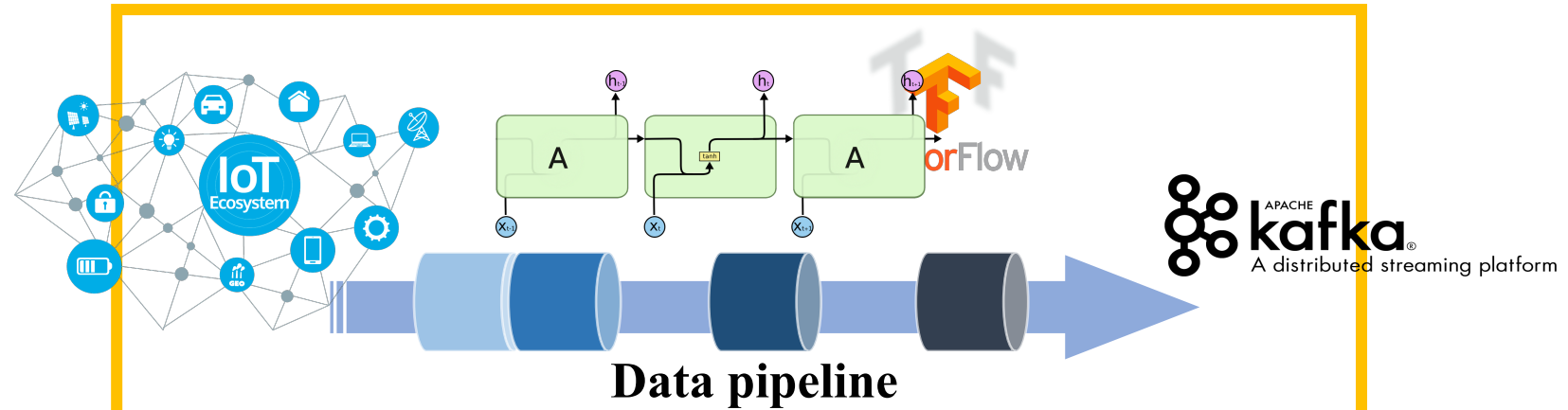


이상감지

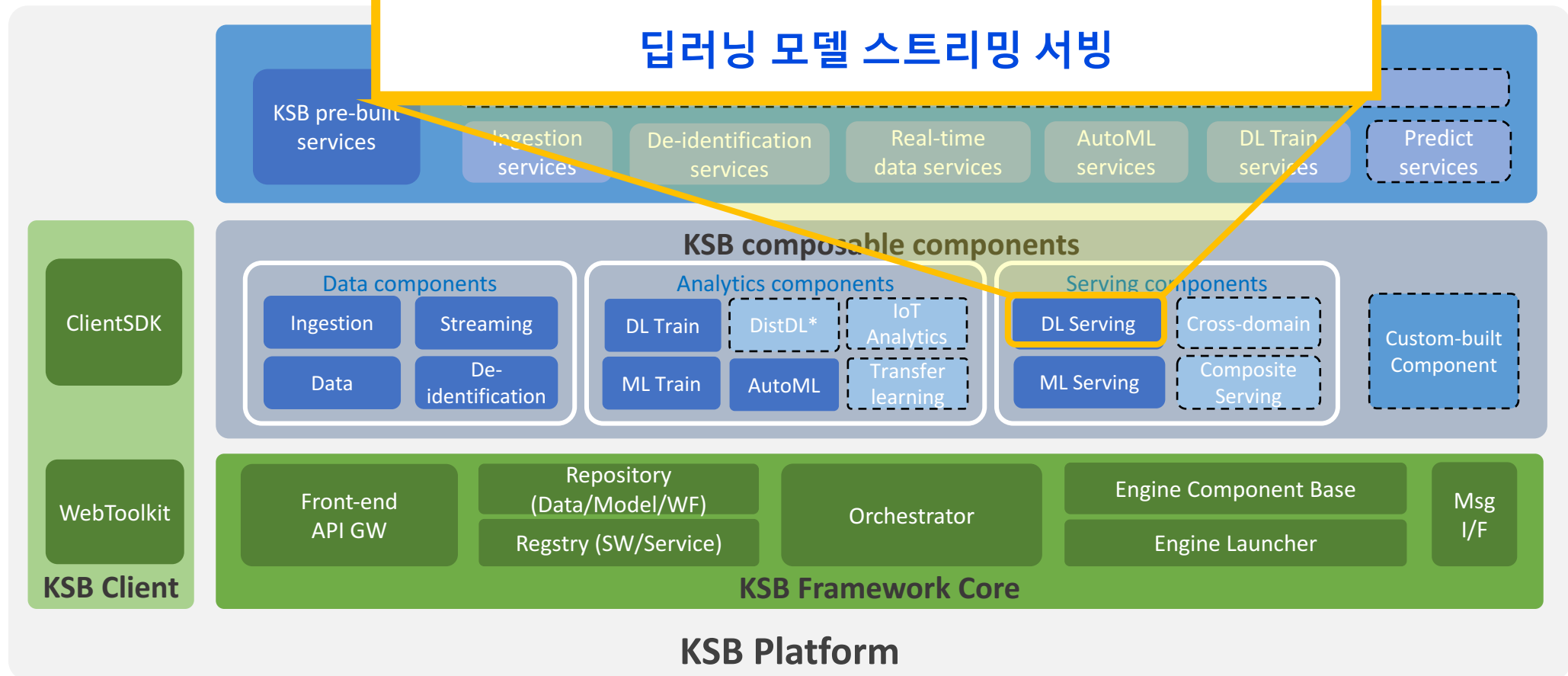


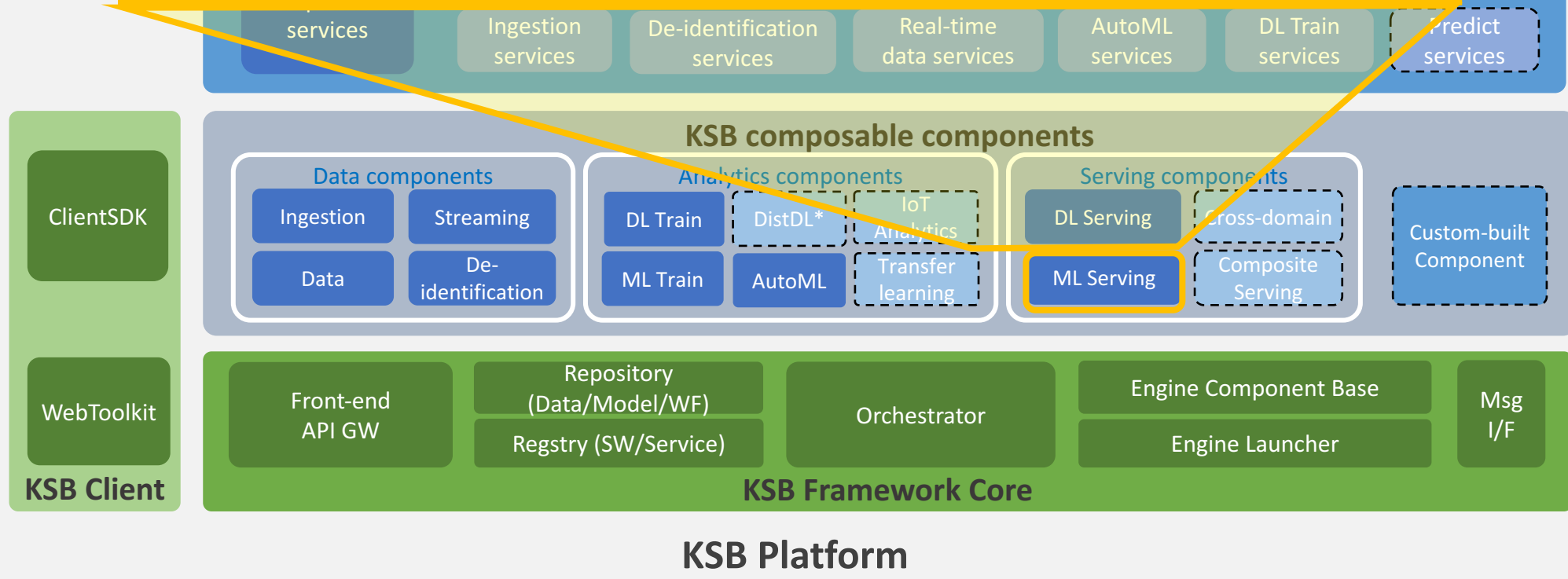
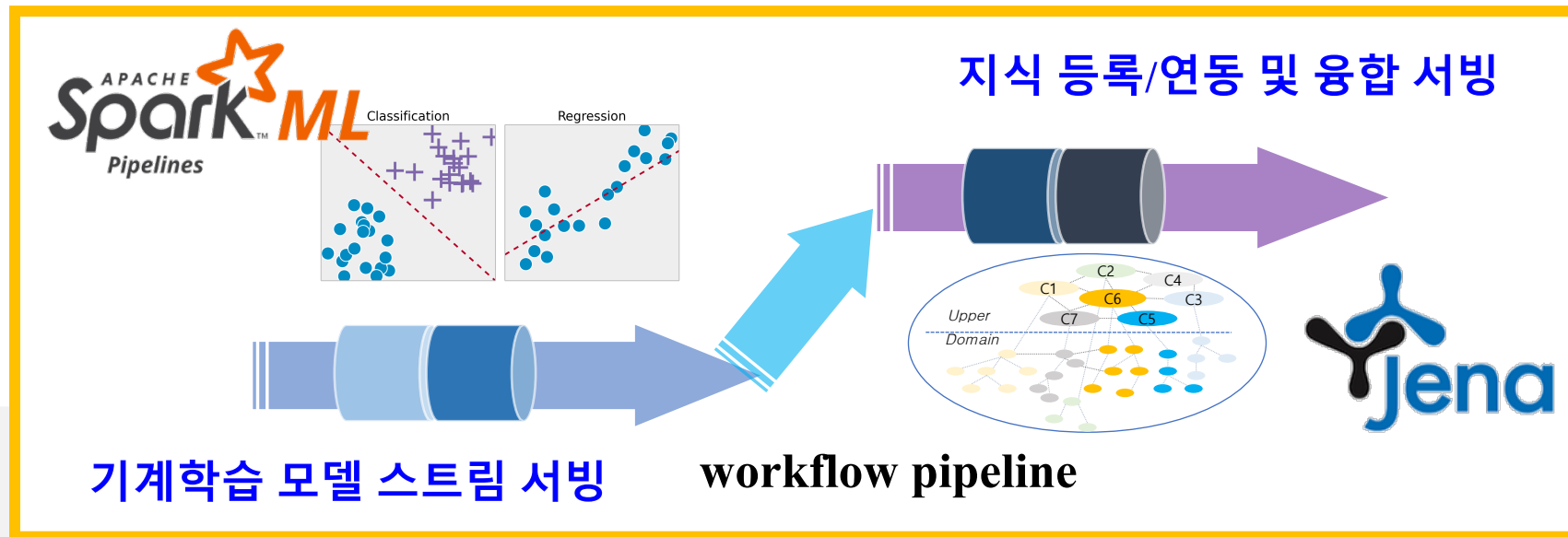


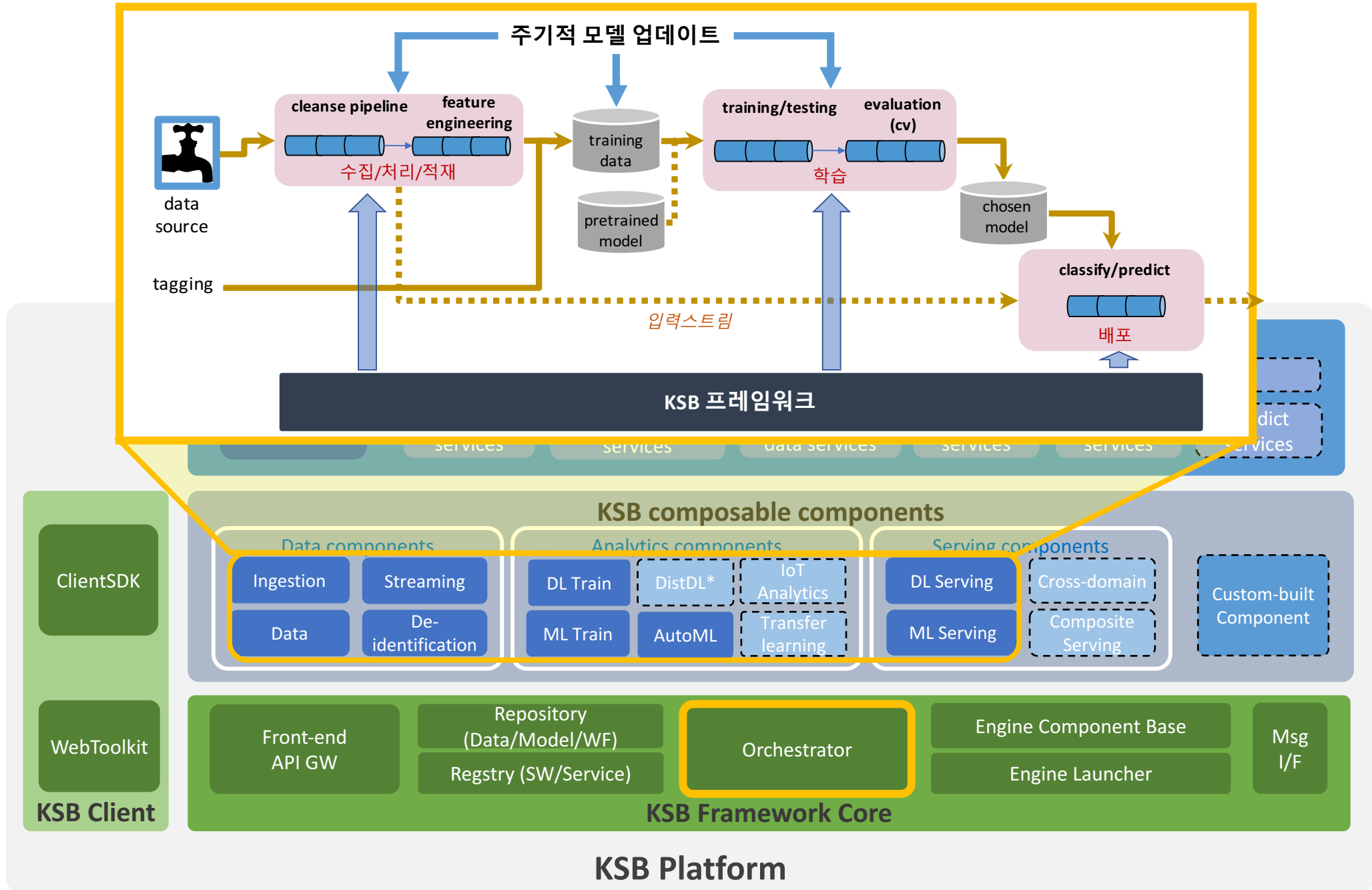




딥러닝 모델 스트리밍 서빙



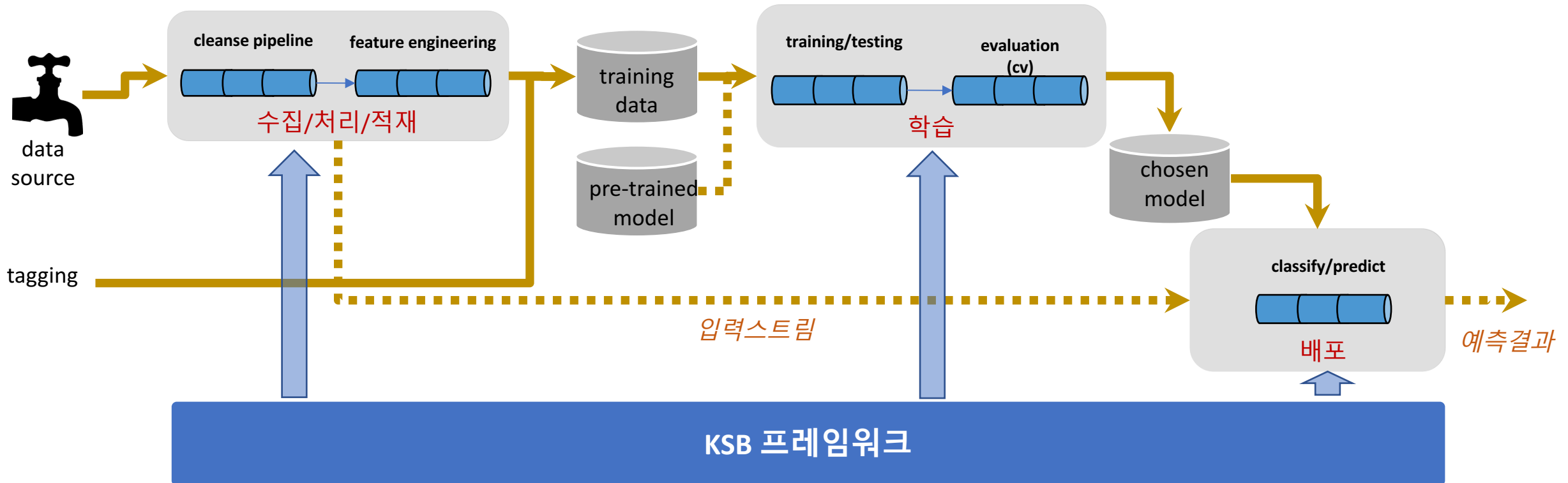




KSB Framework Key Feature

- End-to-end IoT Intelligence
 - From real-time streaming to streaming predictive serving
 - Predictive serving API to domain application serving API
- Easy of Use
 - DIY Workflow Editor
 - Develop and Deploy in One Place
- Extensibility
 - Accumulating SW component
 - Building Predictive Api using ML Model

End-to-end Machine Learning Workflow



Easy of use : DIY Workflow Editor

The screenshot displays the ETRI DIY Workflow Editor interface. The top bar includes a toolbar with icons for file operations and a 'Workflow Editor' title. The left sidebar shows a 'Components' panel with categories: Reader (5), Writer (5), Controller (1), Runner (3), and Operator (7). The main workspace contains a workflow diagram with three 'BatchDummyEngine' components. The top-left component has a 'PySparkRu...' runner. The top-right component has a '#2' runner and a 'SimpleSpark..' runner. The bottom component has a 'Runner' runner. Connections are labeled 'FileWriter', 'FileReader', 'TableReader', and 'TableWriter'. The right sidebar shows 'Workflow Instances' with a list of instances like 'aggClustering_20180105' and 'aggClustering_201806'. The bottom sidebar shows 'Properties' with fields for Name, Description, and isBatch. The bottom status bar shows a sequence of operators: 'Auto Spar..' → 'DecisionTr..' → 'LinearReg...'.

Workflow Name: Noname

Components

- BatchDummyEngine

Reader (5)

- FileReader
- HttpServerReader
- MongodbReader
- PhoenixReader
- Table Reader

Writer (5)

- FileWriter
- MongodbWriter
- PhoenixWrite
- StdoutWriter
- TableWriter

Controller (1)

- SparkSessioStreamContr..

Runner (3)

- PySparkRundener
- SimpleSparkRunner
- TensorflowRenner

Operator (7)

- AutoSparkMLOperator
- DecisionTreeClassifier
- DecisionTreeRegressor
- LinearRegressionRegressor
- NaiveBayesClassifier
- RandomForestClassifier
- RandomForestRegressor

BatchDummyEn...

- Reader
- PySparkRu...
- Controller
- Operator
- FileWriter

BatchDummyEn...

- #2
- SimpleSpark..
- SparkSession.
- #3
- FileWriter

BatchDummyEn...

- Reader
- Runner
- Controller
- Operator
- FileWriter

Workflow Instances

- aggClustering_20180105
 - 1:StreamtobatchEnging
- aggClustering_201806
 - 1:DataEngine
- aggClustering_201806
 - 1:StreamtobatchEnging
- aggClustering_201806
 - 1:StreamtobatchEnging
- aggClustering_201806
 - 1:StreamtobatchEnging
- aggClustering_201806
 - 1:DataEngine
- aggClustering_201806
 - 1:DataEngine

Properties

Workflow Properties

Name

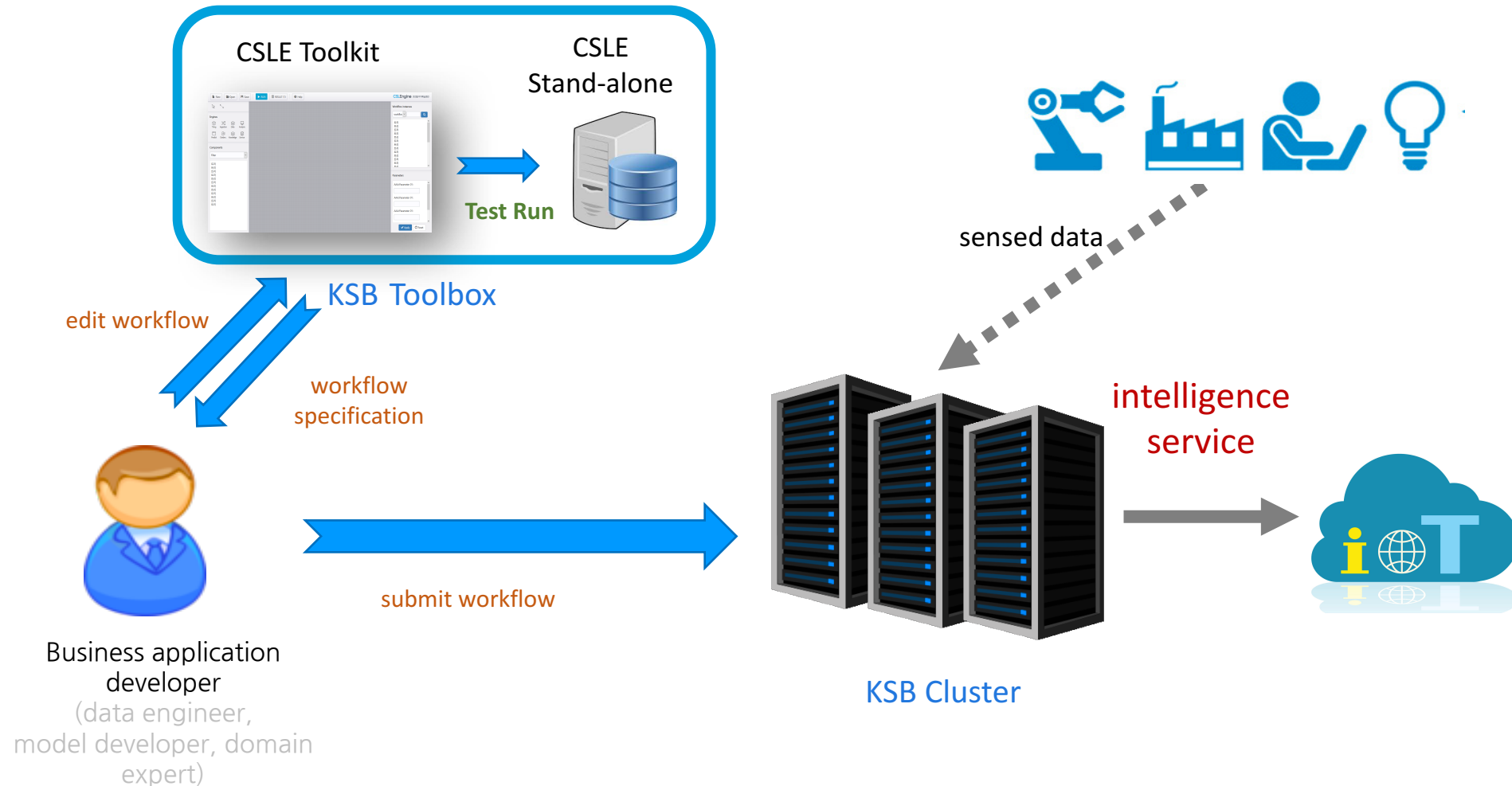
Description

isBatch True

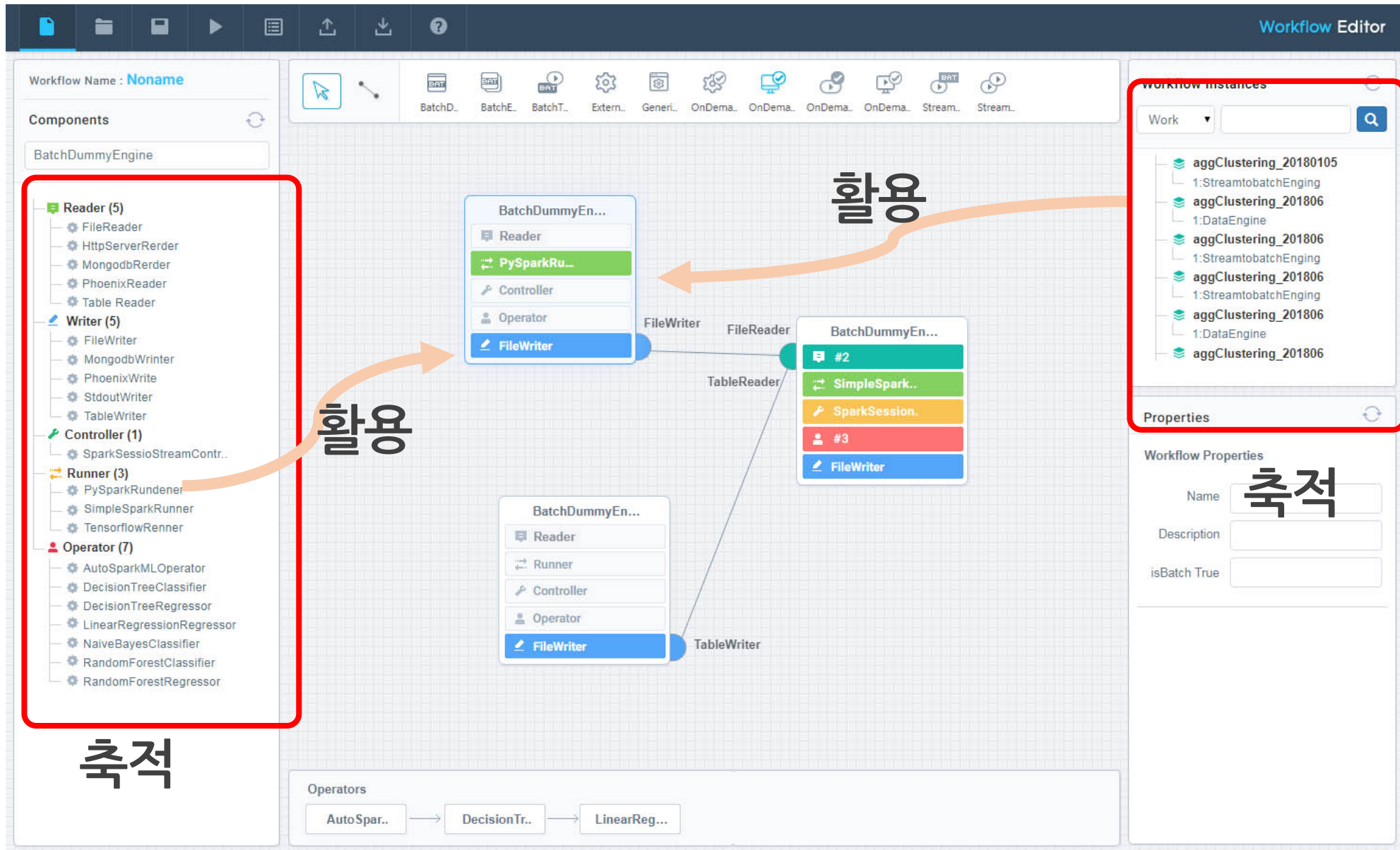
Operators

Auto Spar.. → DecisionTr.. → LinearReg...

Easy of Use: Develop and Deploy in One Place

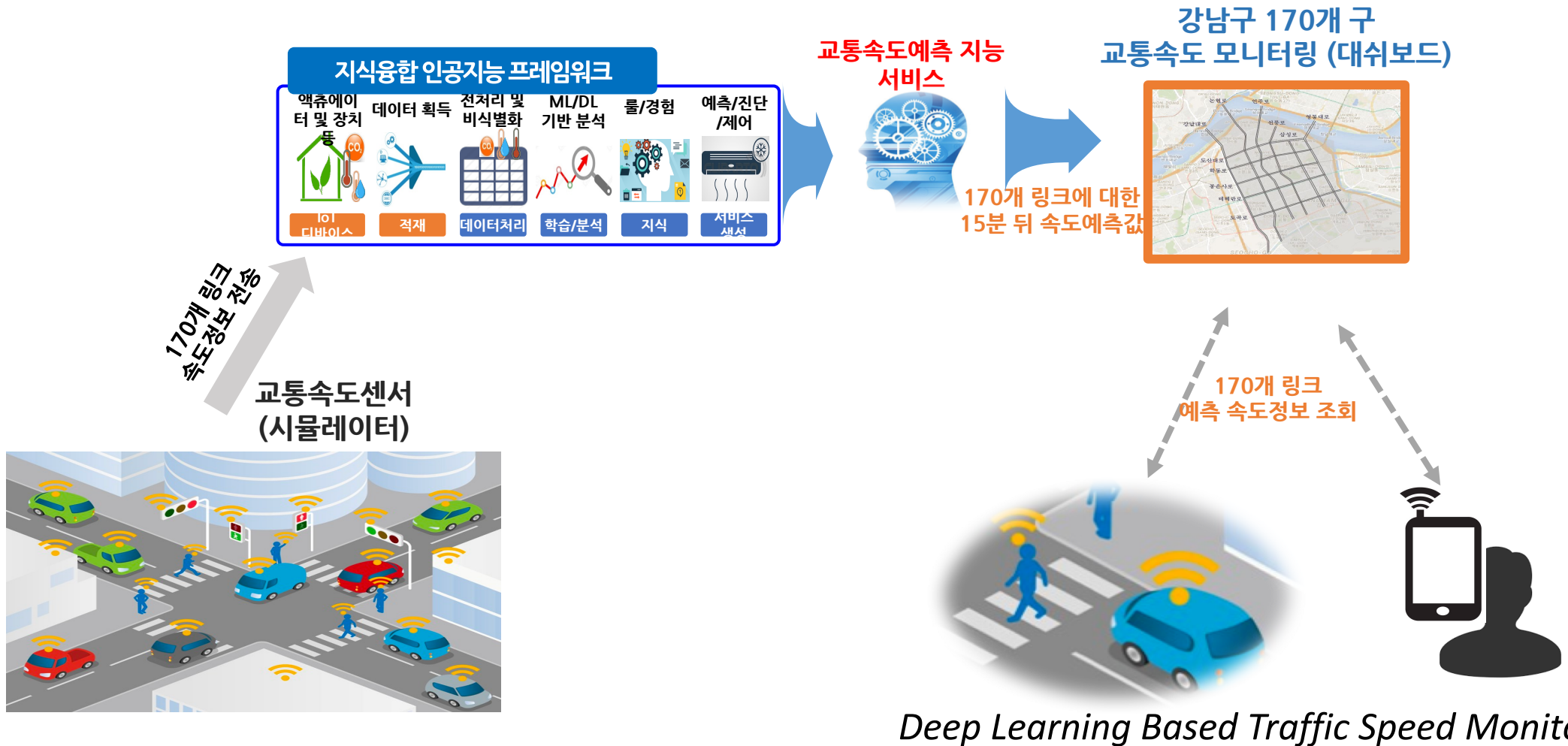


Extensibility: Accumulating SW component



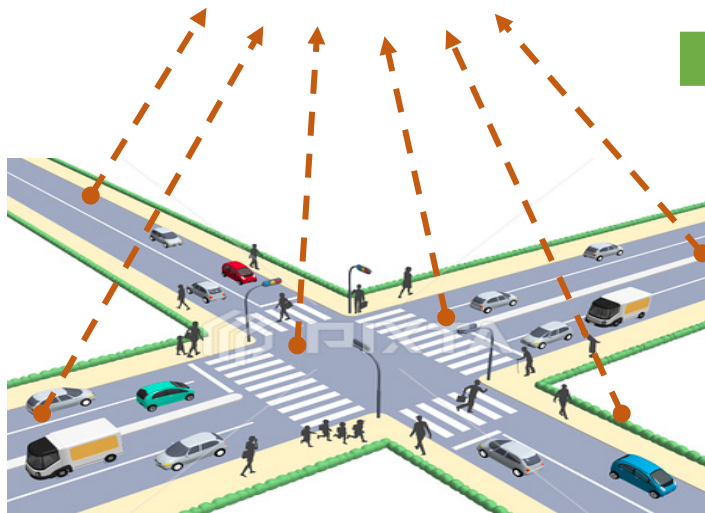
IV. Usecases

U1. Real-time streaming analytics with continuous model update



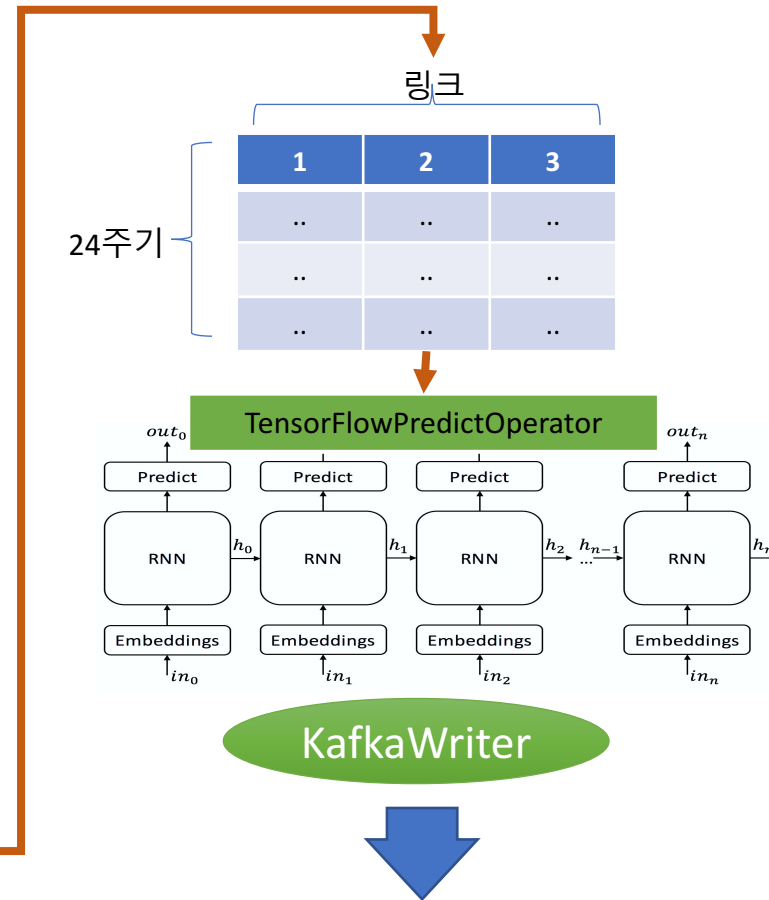
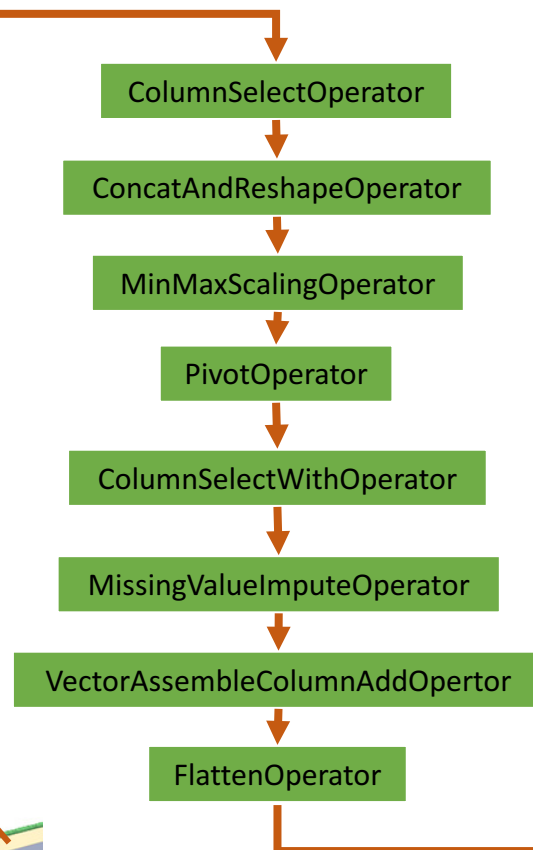
시간	ID	속도
..
..
..

httpReader



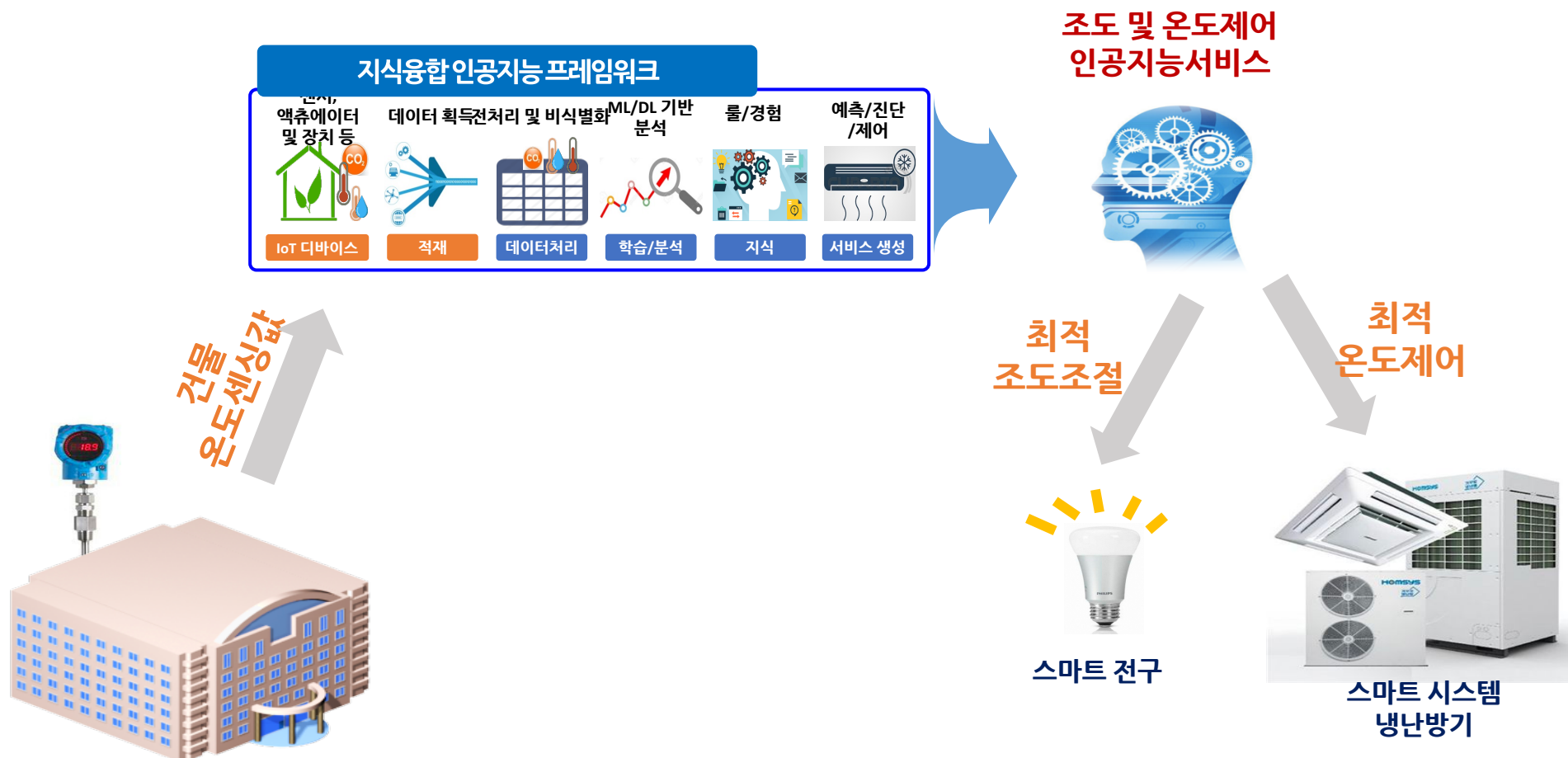
pixtastock.com - 21291829

1,440개 교통정보 수집 센서

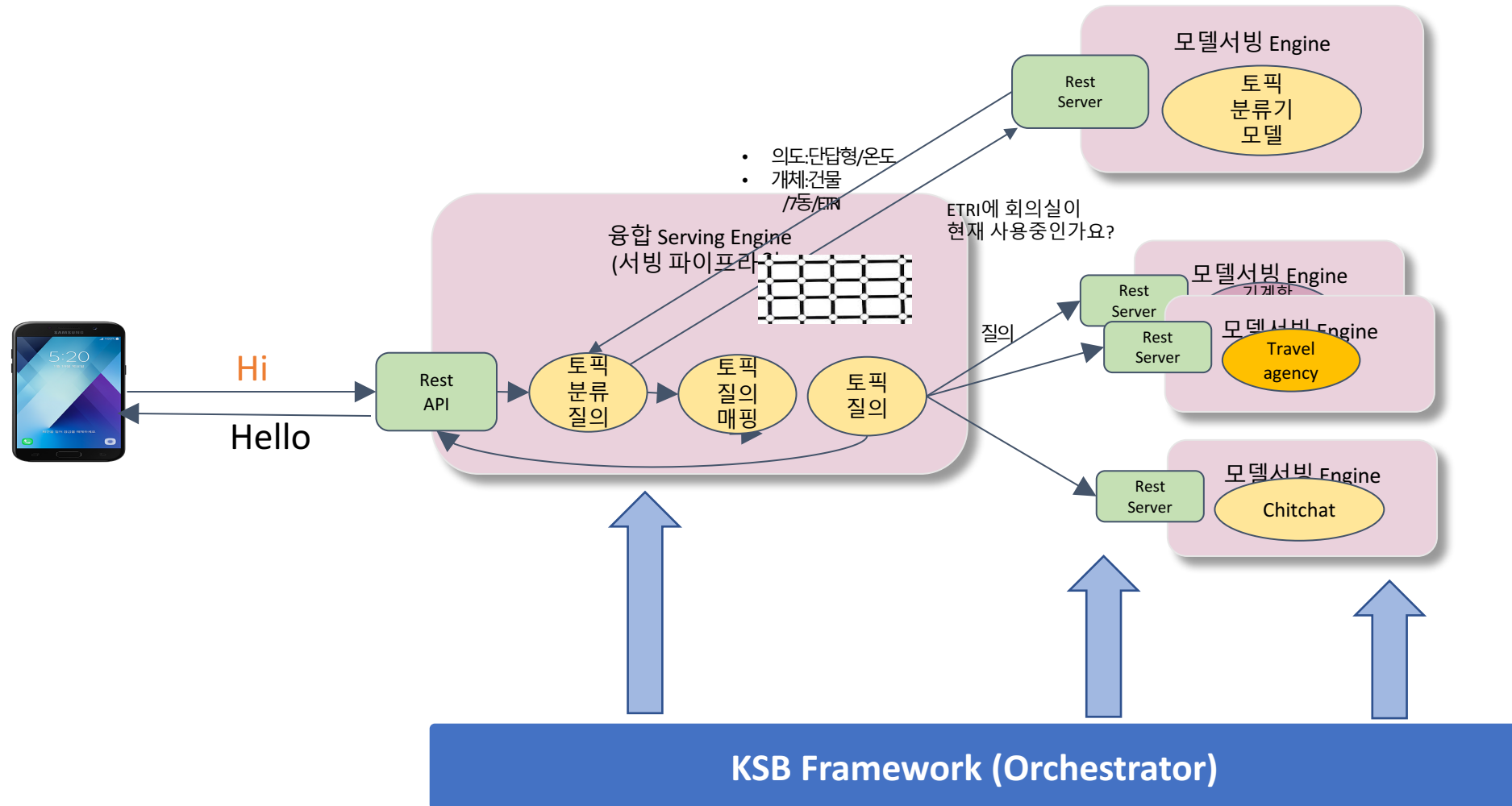


예측/모니터링

U2. Energy efficient building control (Prototype)

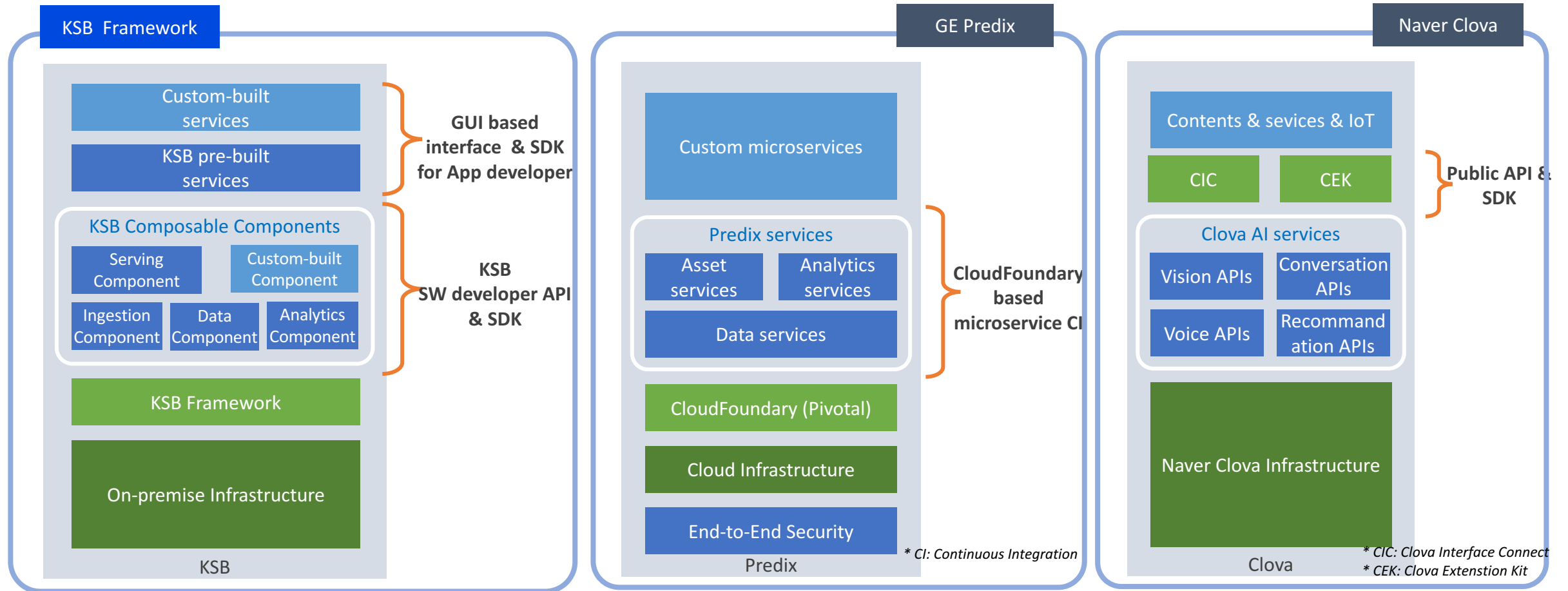


U3. Chatbot (prototype)



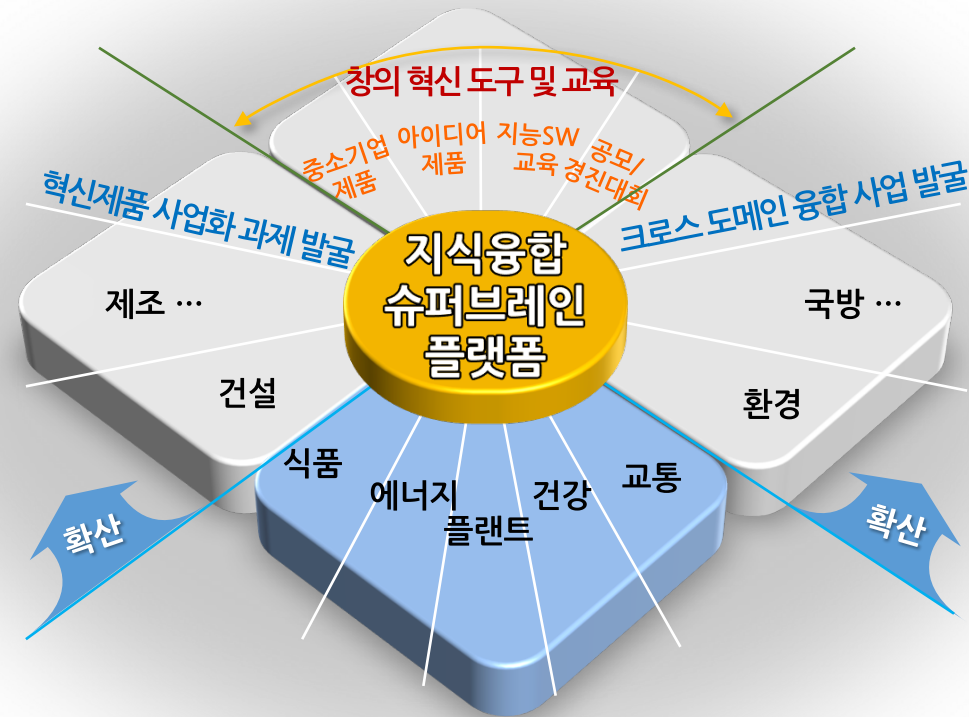
V. Conclusion

KSB vs. AI Platforms (Predix/Clova)



Conclusion

KSB VISION is to create Intelligence information industry ecosystem and to spread technology



공공분야 융합 선도 서비스 구축

Conclusion

기업의 경험 지속적으로 축적하고,
공유 및 협력을 위한 장과,
기술의 장벽을 넘어 끊임없이 진화할 수 있는 기반기술 제공

Thank you !