

AI/ML Framework Project Proposal

Samsung Open Source

28th July 2022

Samsung Electronics

AI/ML Framework Project Proposal

Project objectives:

1. Framework that can be used by Network Operators to create AI/ML based use-cases for RAN and easily deploy the use-cases as rApp or xApp.
2. Framework to realize AI/ML training required for various AI/ML use-cases.
3. Framework that enables Non-RT RIC and Near-RT RIC to host AI/ML capable xApps and rApps.

Project Scope :

AI workflow management functions

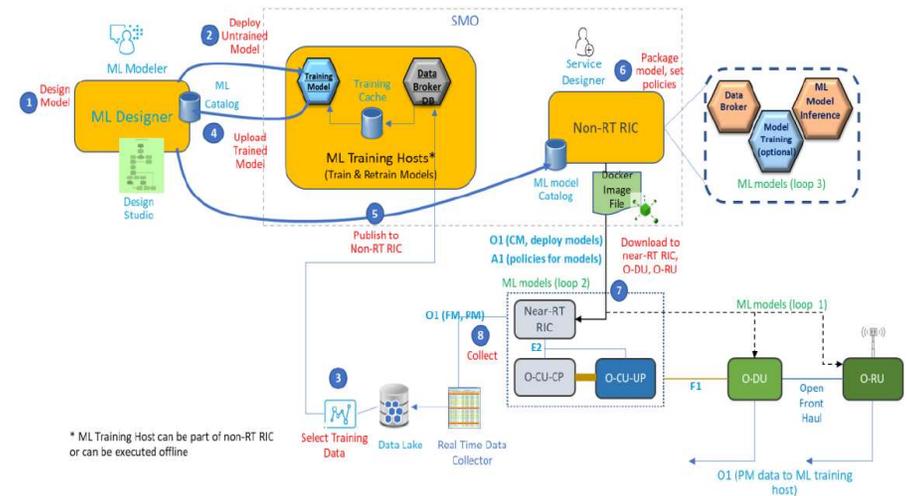
- Training job management
- AI/ML trained model management
- Service Management

AI training platform functions

- TMS (Training Management System)
- Data extraction/Data broker
- Feature store
- AI Model storage
- TMS Adapter

AI inference platform functions

- IMS (Inference Management System)
- IMS adapter



References:

1. O-RAN.WG2.AI/ML-v01.03
2. <https://www.kubeflow.org/docs/components/pipelines/introduction/>
3. <https://github.com/kserve/kserve>

Key Facts:

Facts	Info
PTL	Hoejoo Lee
Project ID	AI/ML Framework
Link to wiki space	<Need to request>
Jira Project Name	<Need to request>
Jira key	<Need to request>
Repository	<Need to request>

Resource Details (Samsung Contributor List):

- Hyunsung Cho
- Youhwan Seo
- Hunje Yeon
- Joseph Thaliath
- Sandeep Kumar Jaiswal

Open to contributors from any company to participate in this project

Project Name: AI/ML Framework

AI/ML Framework Subcomponents

1. AI management functions

- **Training job management:** Create/Edit/Delete usecases and Training Pipelines. Train/Retrain, Provide status of Training jobs.
- **AI/ML trained model management:** Model versioning and provide APIs to download models and metadata information

2. AI training platform functions

- **Use Kubeflow Pipelines (open source) for TMS (Training Management System):** Platform for building and deploying portable, scalable machine learning (ML) workflows based on Docker containers.
- **Data extraction/Data broker:** Module which retrieves required data for training from Data Lake
- **Feature store (open source):** DB to store features required for training
- **AI Model storage (open source):** DB to store trained models
- **TMS Adapter:** Module to interact with Kubeflow platform for training

3. AI inference platform functions

- **Use Kserve (open source, LF AI project) for IMS (Inference Management System):** Provides a Kubernetes Custom Resource Definition for serving machine learning (ML) models on arbitrary frameworks

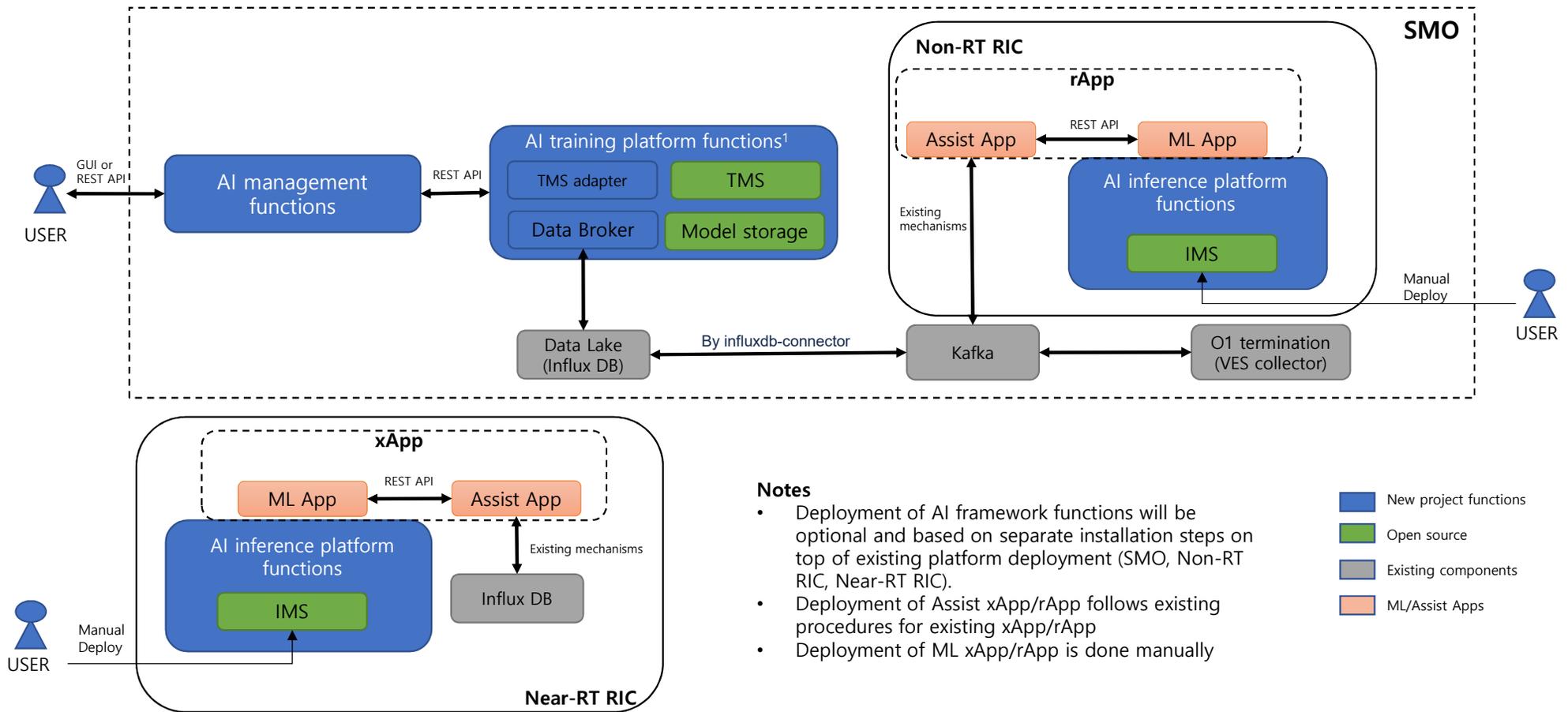
Impact on other O-RAN SC Projects

For G release, we are expecting no impact to other projects (maybe RICAPP project if we want to update or create new xapp). But in future, there may be impact expected in SMO, Non-RT RIC, Near-RT RIC

References:

1. <https://www.kubeflow.org/docs/components/pipelines/introduction/>
2. <https://github.com/kserve/kserve>

Architecture, Workflow – ‘G’ Release Scope



Notes

- Deployment of AI framework functions will be optional and based on separate installation steps on top of existing platform deployment (SMO, Non-RT RIC, Near-RT RIC).
- Deployment of Assist xApp/rApp follows existing procedures for existing xApp/rApp
- Deployment of ML xApp/rApp is done manually

¹Location of AI Training functions is not fixed and may vary according to usecase. One example is shown here

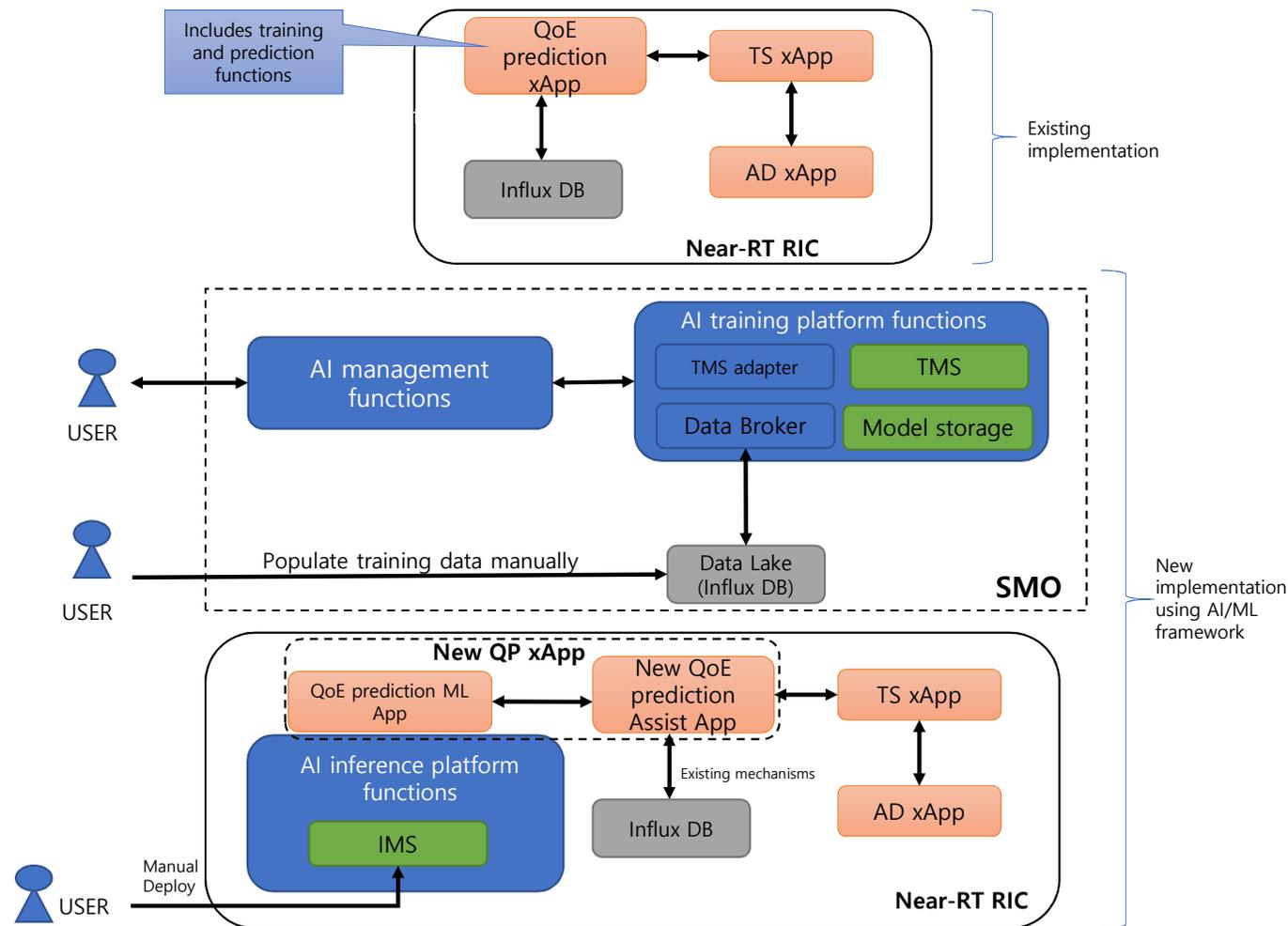
Example use case – ‘G’ Release Scope

Traffic steering use case

- New “QoE Prediction (QP) xApp with AI/ML Framework”
 - Make updates to remove the ML model training and model saving functionality within the xApp
 - Make updates to remove predictions performed within the xApp
 - Make updates to trigger API calls to perform predictions
- Anomaly Detection (AD) xApp
- Traffic Steering xApp

Notes

- Deployment of Assist xApp/rApp follows existing procedures for existing xApp/rApp
- Deployment of ML xApp/rApp is done manually
- Data required for training is populated in SMO Data Lake manually



Thank You