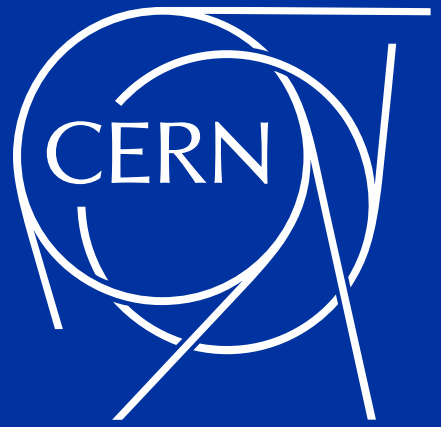


Requirements & Prototyping

Alexander Zoechbauer (CERN)

28.02.2022

Requirement updates

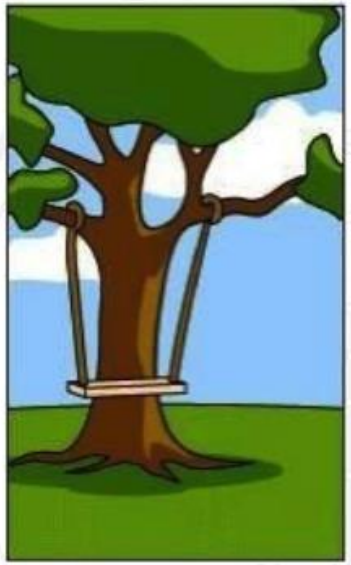
USE CASE	USER			MODEL						DATA			TRAINING			DEPLOYMENT	
Column1	Familarity with ML	Level of interaction	Task	pre-processing	input	output	Type	Size	data Size	update frequency	dimensionality	federated learning	HPO	retraining frequency	post-processing	Inference	Hardware
QCD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	prob our requirement	prob our requirement	TBD	TBD	TBD	TBD
Radio Astronomy	TBD	TBD	classify timeframes of radioastromical data into different levels of interest to find pulses of pulsars	convert binary data stream to 2D images (prob using fourier transform)	frequency-time heatmap (freq_res x time_res)	Classification outcome	CNN, RNNs, LSTMs, neural ODEs	TBD	O(10TB) @ 10MB/s	stream		prob our 2requirement	prob our requirement	TBD	TBD	curently where the data is, future at telescopes experiment	FPGA
Virgo	TBD	TBD	classify timeframes that have noise in gravitational wave signal. Then veto those events and in the future denoise them	1D binary stream is Fourier transformed in (freq_res x time_res) heatmaps	frequency-time heatmap (freq_res x time_res)	Classification outcome (veto/ denoise)	GAN	TBD	50 channels = 2 MB/s	stream		prob our 2requirement	prob our requirement	1 day	TBD	At INFN computing facilities, close to the incoming data	TBD
Detector Sim	High	Low-level	generate images with 3DGAN to simulate energy depositions for fast simulation instead of slow classical MC simulations	trajectory rollout and scaling with power	incoming particle energy and angle and a latent vector of random variables ((254+3+1) x 1)	image of energy depositions (res x res x depth)	3DGAN	5M parameters	O(10 GB)	currently never, maybe in future		prob our 3requirement	prob our requirement	currently never, maybe in future	currenty none	currently where .root files are, future maybe closer to the detector	CPUs
Fire, Storms	TBD	TBD	1. The generation of synthetic Fire Weather Index (FWI) maps 2.Classification if extreme storm is present, then locate center of the storm	Cropping, regridding, patches generation (sub-grids), data augmentation, feature selection & scaling	1. satellite time series with precipitation, temp, humidity (4 x res x res x time) 2. satellite time series with pressure, wind, temp, etc (4 x res x res x time).	1. Fire Weather Index (res x res x time) 2. Storm center (1 x 1 x time)	CNNs, GANs, CGNN/GNNs	TBD	2TB	1. monthly 2.weekly		prob our 4requirement	prob our requirement	TBD	TBD	TBD	CPUs
Floods and droughts	TBD	TBD	predict droughts in alpine regions	data fusion of 6-7 grid data, remove part of res data (cloud removal)	openEO raster- datacubes: satellite time series temperature, pressure, etc (fused_dim+ x res x res x time) and openEO vector datacubes	classification outcome (warning for specific region) or generative (map with drought likelyhood)	TBD	TBD	2PB	daily		prob our 4requirement	prob our requirement	TBD	TBD	HPC center where data is stored	GPUs
Climate Change Impact	TBD	TBD	use historical data to make projections about the future floods/droughts/storms/fires	use tool icclim to convert 3 daily variables (temperature, precipitation, wind) to monthly climate indices. (3x res x res x 365) -> (50x res x res x 12)	satellite time series (50x res x res x 12)	delta in the climate for the area (e.g. more/less droughts than in the past)	Unsupervised ML (prob not true)	TBD	O(100 GB)	very rarely, negligible		prob our 4requirement	prob our requirement	5-10 years, negligible	TBD	TBD	TBD

Requirement collection

A communication problem

Need

What the customer wanted

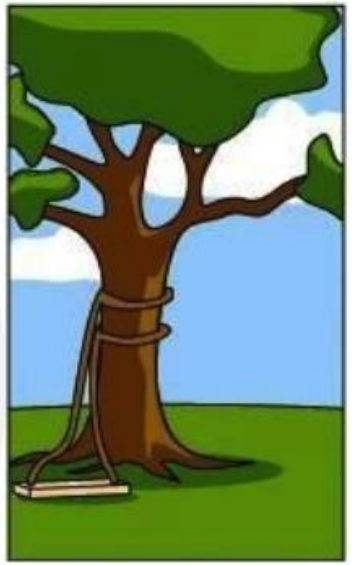


Analysis

What the analyst understood

Design

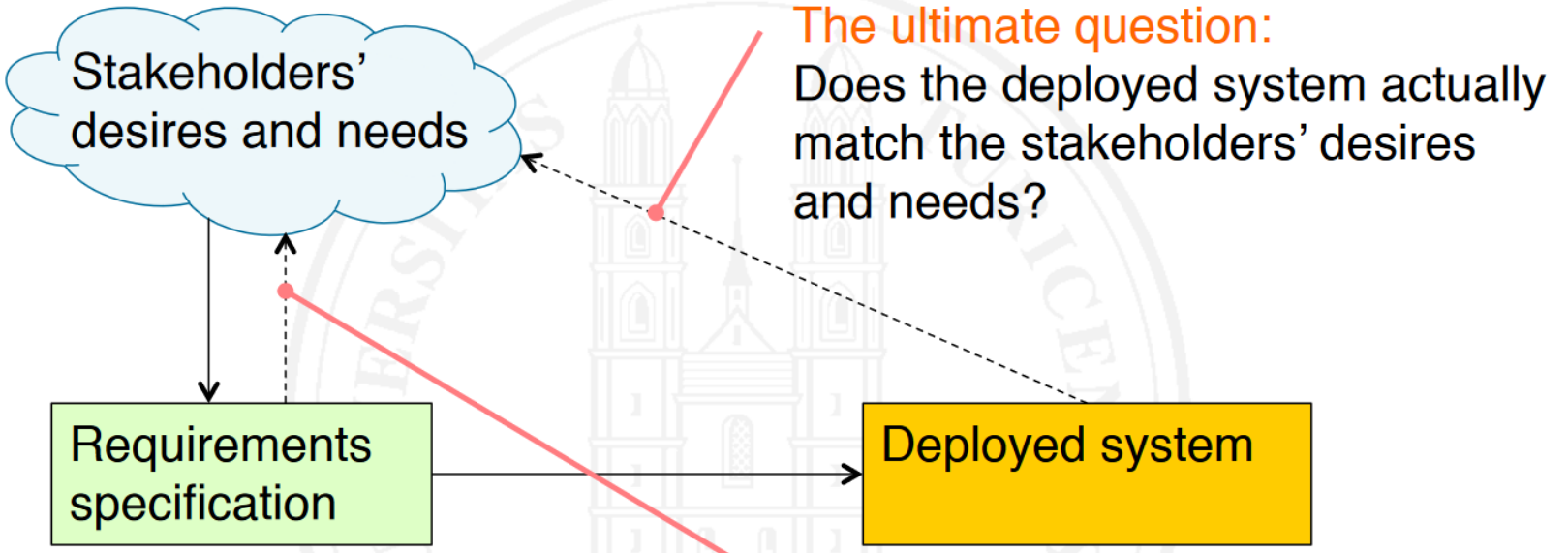
What the architect designed



Deployed System

What the programmers implemented

Requirement collection

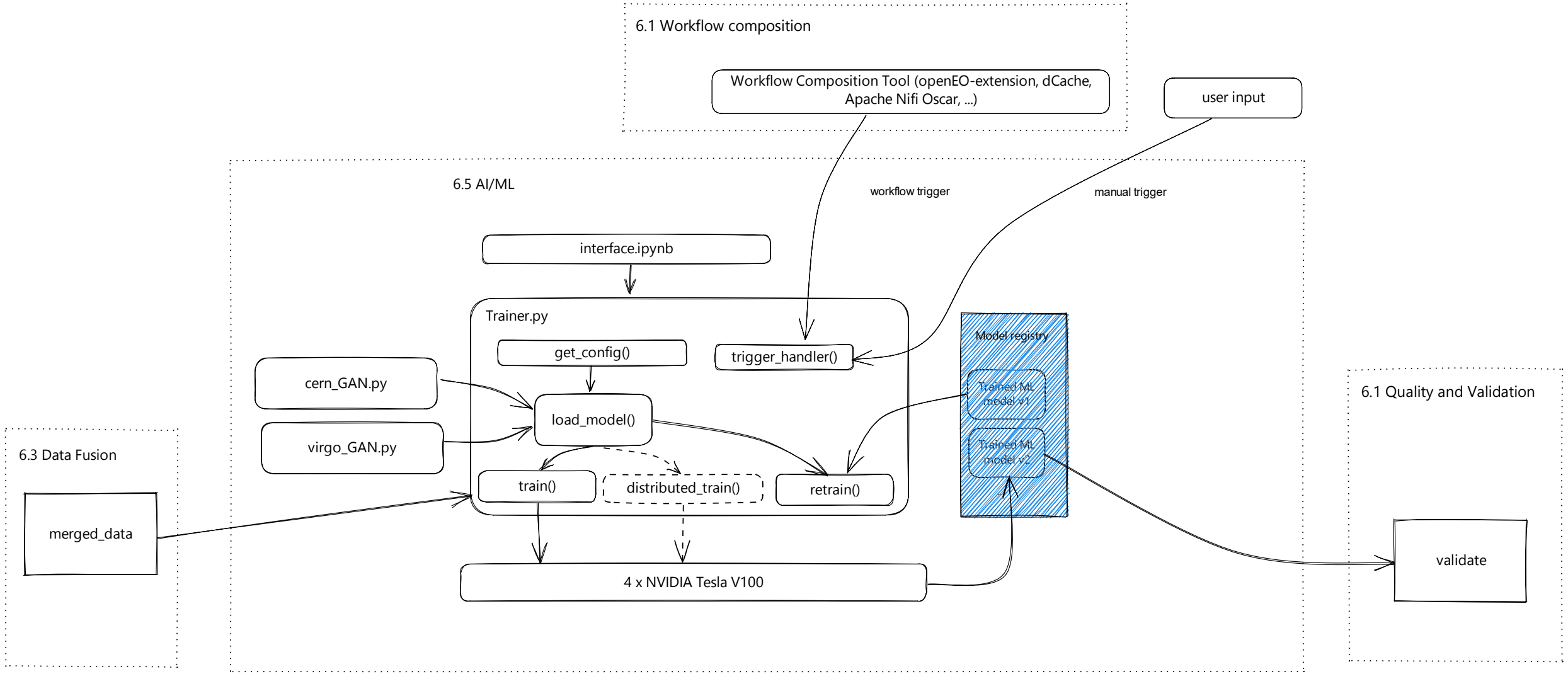


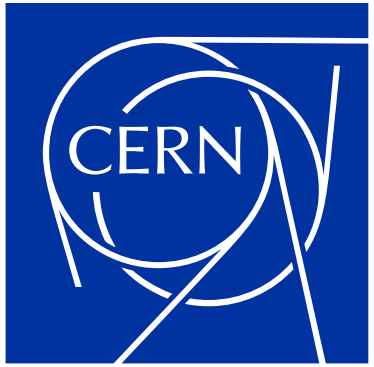
The ultimate question:
Does the deployed system actually match the stakeholders' desires and needs?

The risk-reduction question:
Do the documented requirements match the stakeholders' desires and needs?

Every requirement needs to be validated

Prototype layout





home.cern