HEPS

Machine Learning for Accelerators/High Energy Photon Source (HEPS)

Paul Chu

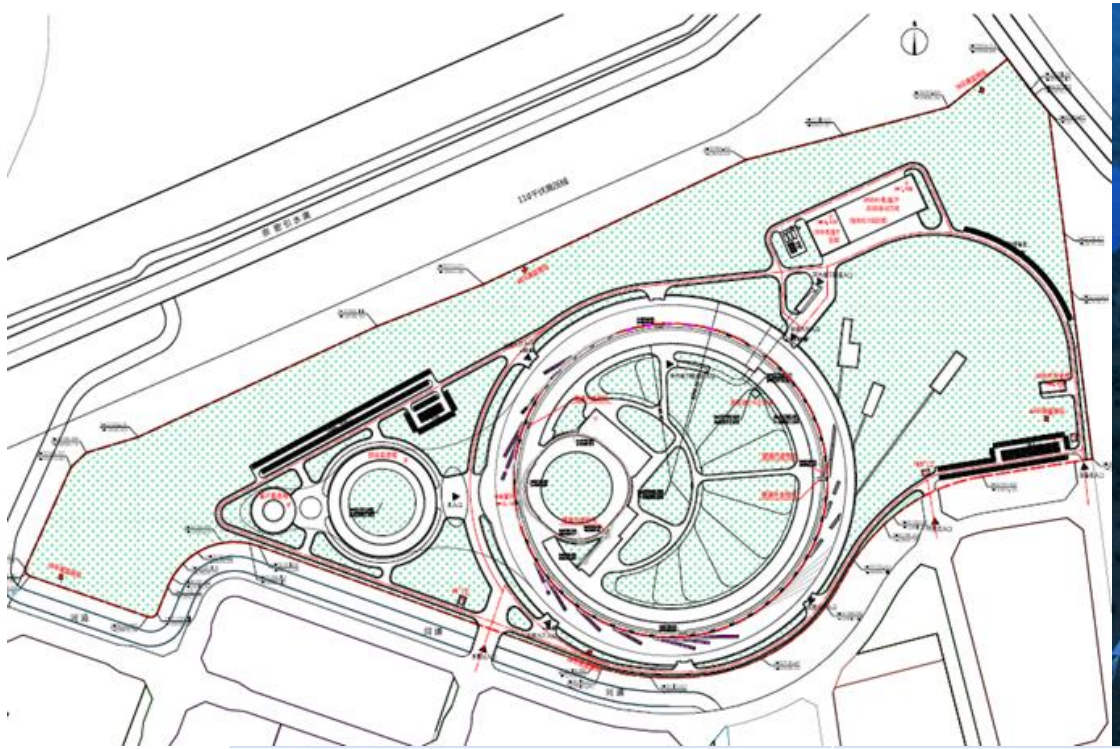
Institute of High Energy Physics, Chinese Academy of Sciences

Outline

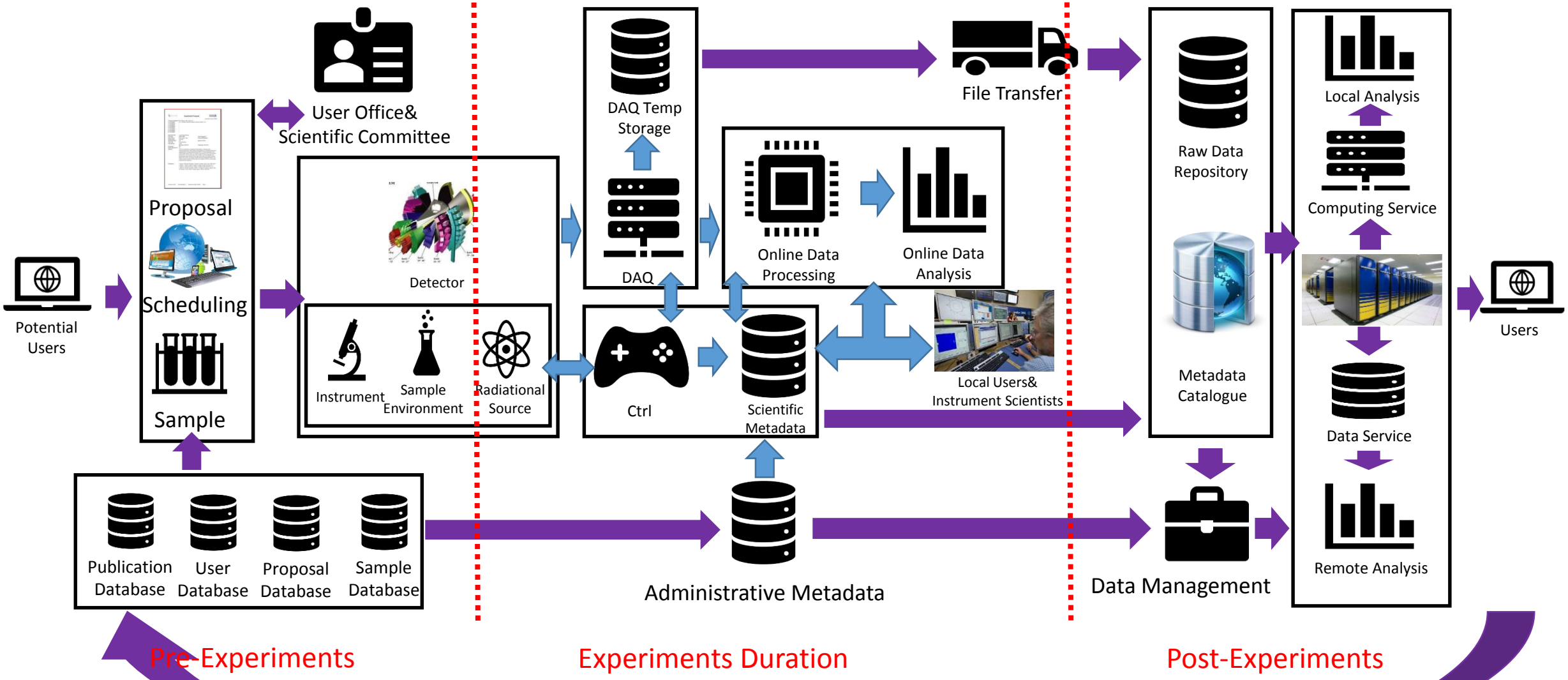
1. Introduction
2. Software Architecture for HEPS ML
3. Preparation for Machine Learning
4. ML Application Examples
5. Summary

Introduction

- HEPS– 4th generation synchrotron light source, 7BA-lattice
- Construction period – Jun. 2019 – Dec. 2025, ~US$700M

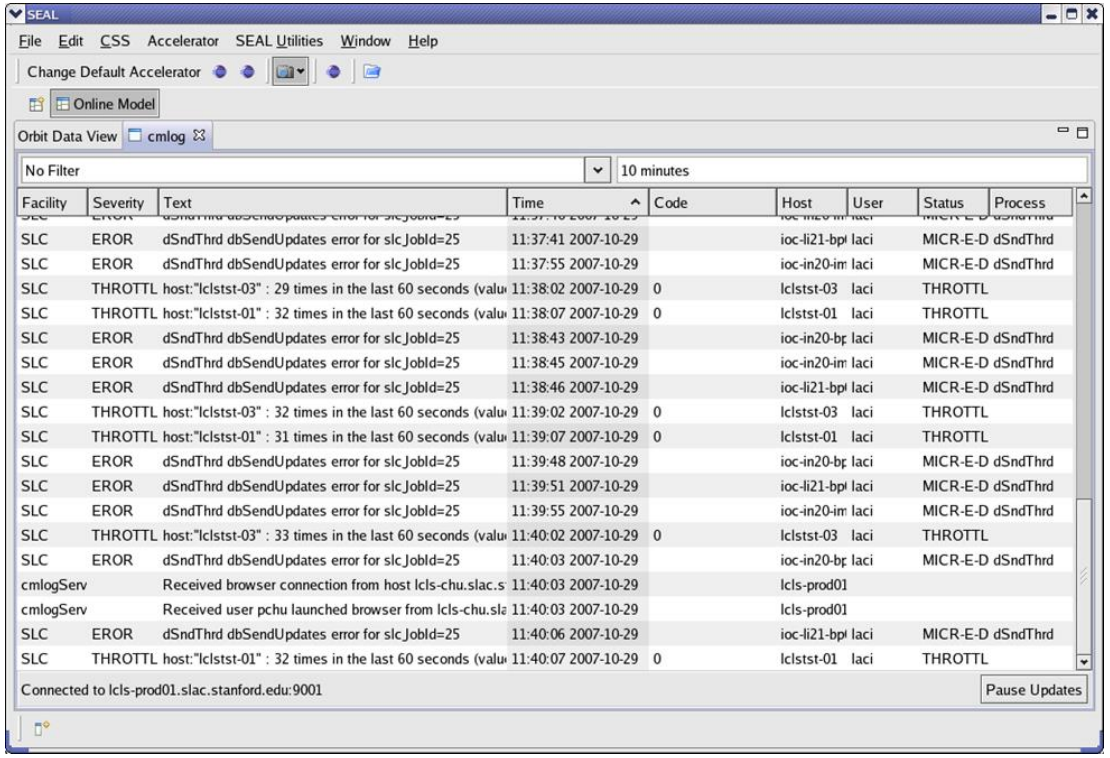


User Facility Software Architecture (courtesy F.Z. Qi)



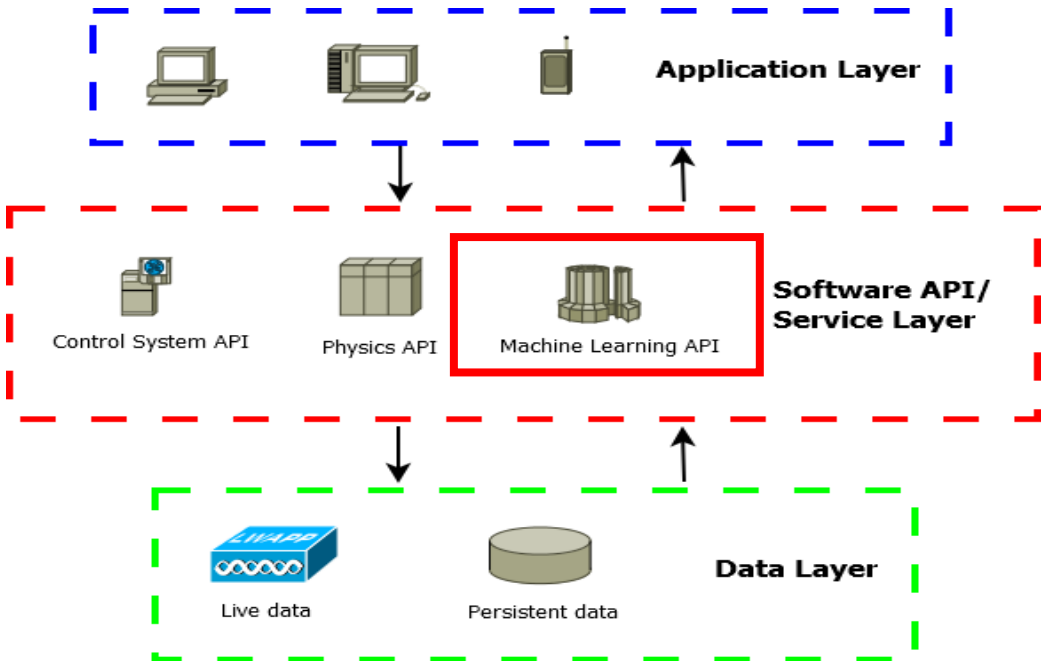
Central Message Logging System

- Recording every action occurred in the system
- CMLog data format
- CMLog database (noSQL)
- C/C++/Java/Python API
- CMLog client viewer
- Can serve for MPS postmortem analysis



HEPS

High-level Application Architecture



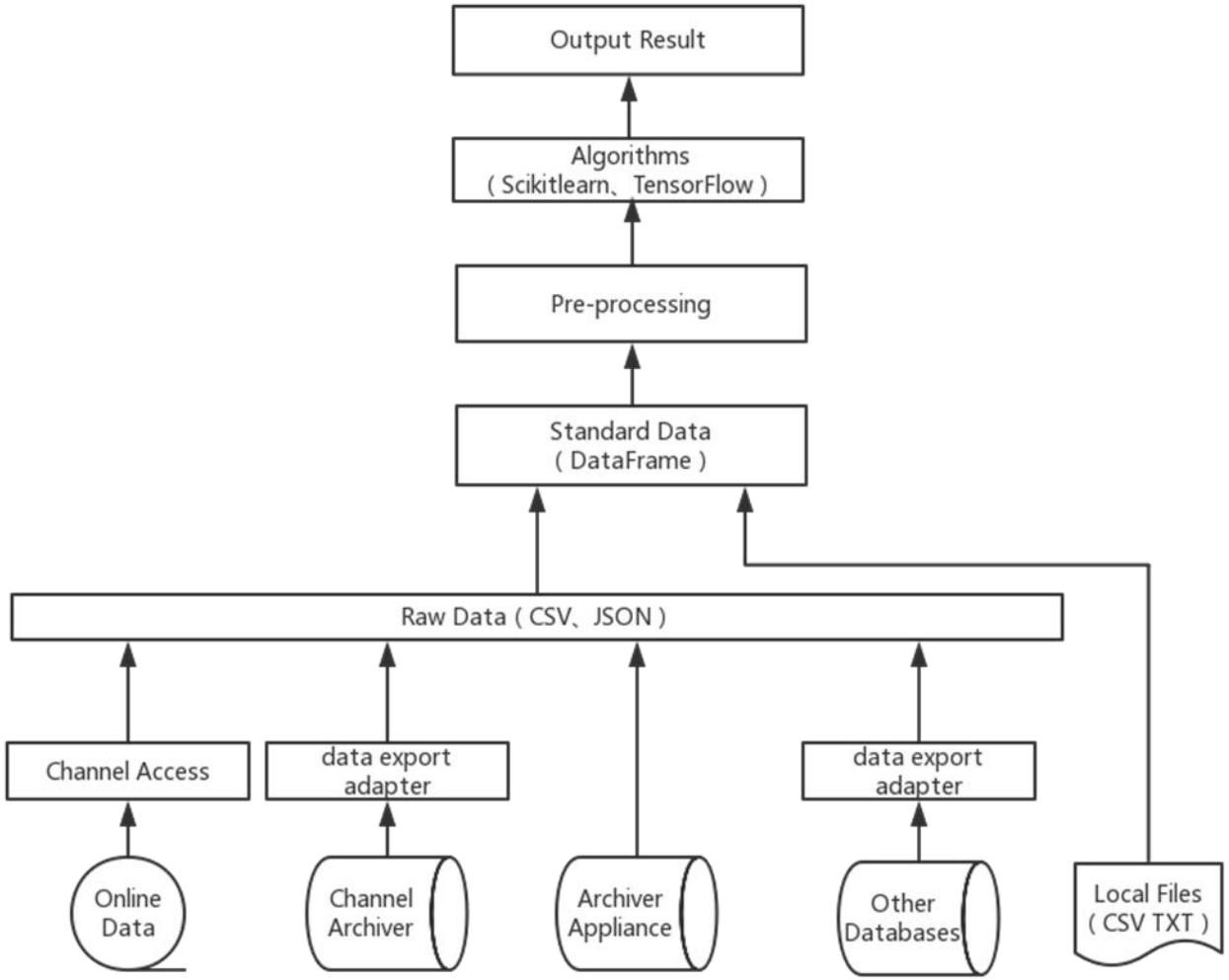
- Collecting as much data as possible
  - Central messaging logging
  - Operation logging
  - …
- Data Pre-processing
  - Clean up
  - Line up

- Programming applications with 3 categories of APIs
  - Software re-usability, cut development time
  - General-purpose, physics, and Machine Learning (ML)
- Standard data formats & popular algorithms

HEPS

Machine Learning Platform

- Getting data
- Pre-processing data
- Applying algorithms
- Displaying results
- Applying results/predictions



HEPS

ML Platform General Ideas

Machine Learning in Python

Scikit-learn/TensorFlow

- Simple and efficient tool for data mining & data analysis
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Machine Learning in MATLAB

MATLAB Machine Learning Toolbox

Data Preparation Example for Machine Learning

Data Collection

- Archived accelerator data
- Experiment data from Data Acquisition (DAQ) systems
- PVLogger (EPICS based) or time synchronized meta data
  - Define own EPICS PV list, logging period easily
  - Aligned timestamp PVs
  - MySQL DB
  - Periodic logging
  - On-demand logging

SVC_ID	SVC_NM
physics	region physics
NULL	NULL

SNAPSHOT_PER	SNAPSHOT_PER_NAM
0	no logging
1	1 sec
60	1 min
900	15 min
3600	1 hr

SNAPSHOT_RETENT	SNAPSHOT_RETENT_NM
0	forever
86400	1 day
2592000	30 days
7776000	90 days
15552000	180 days
31536000	1 yr

```
import sys
import math

from java.util import *

from xal.service.pvlogger import RemoteLoggingCenter

rL = RemoteLoggingCenter()

# prepare comments for the snapshot
time = Date()
comments = time.toString()

# take a snapshot
snapshotId = rL.takeAndPublishSnapshot( "Region physics", comments);
```

Data Handling

Data Sources

- EPICS live data
- TXT/Excel Files
- EPICS Channel Archiver
- EPISC Archiver Appliance
- Other data sources (e.g. PVLogger)

Output Data Format

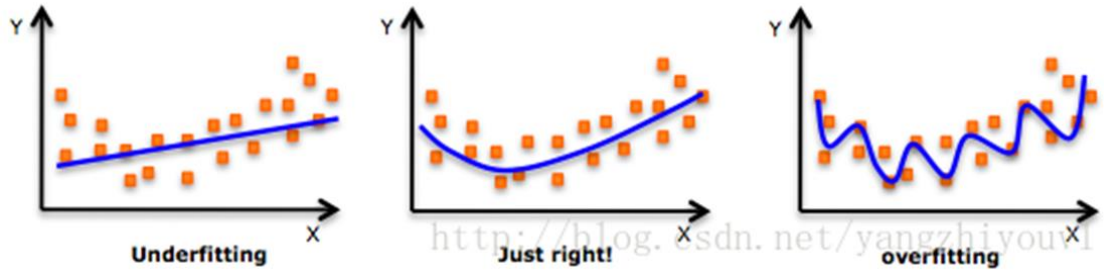
- Pandas DataFrame
- TXT/Excel Files
- Other format: HDFS

Code Snippet

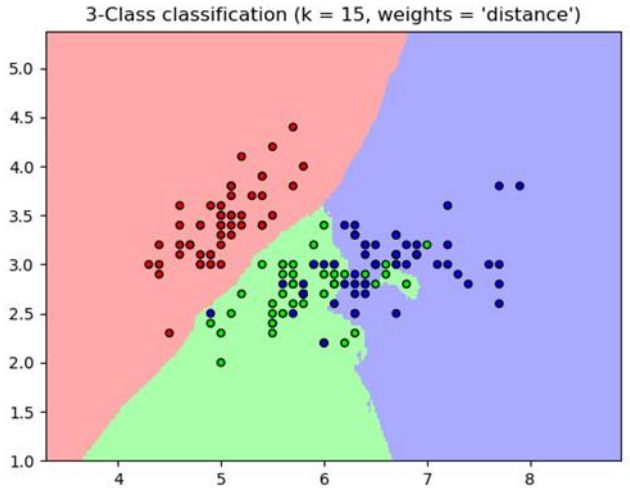
```
pvnames=['BIBPM:R1OBPM02:XPOS','BIBPM:R1OBPM03:XPOS','BIBPM:R1OBPM04:XPOS]
#also can load pvnames from files
engine=LoadData.getKey(server_addr,pvnames)
data=LoadData.getFormatChanArch(server_addr,engine,pvnames,start_time='11/30/2018
14:15:00',end_time='11/30/2018 14:16:00',merge_type='outer',interpolation_type='linear',
fillna_type=None,how=0)
```

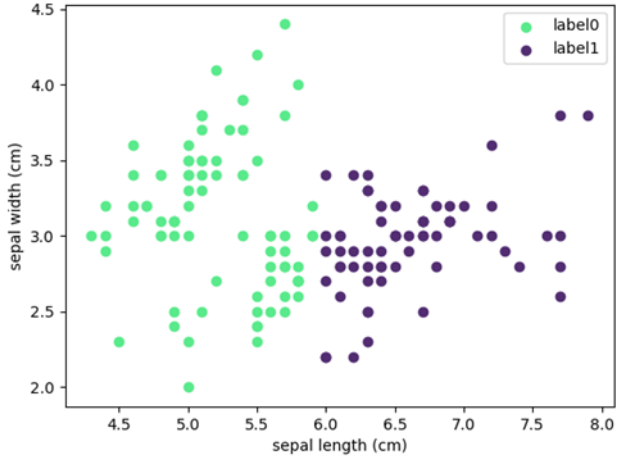
Algorithms

- Regression
  - Linear Regression
  - Bayesian Linear Regression
  - Polynomial Regression
- Decision Tree
- K-Nearest Neighbors
- Clustering
  - K-Means
  - DBSCAN
- Multi-layer Perceptron (MLP)
- …



http://blog.csdn.net/yangzhiyou11

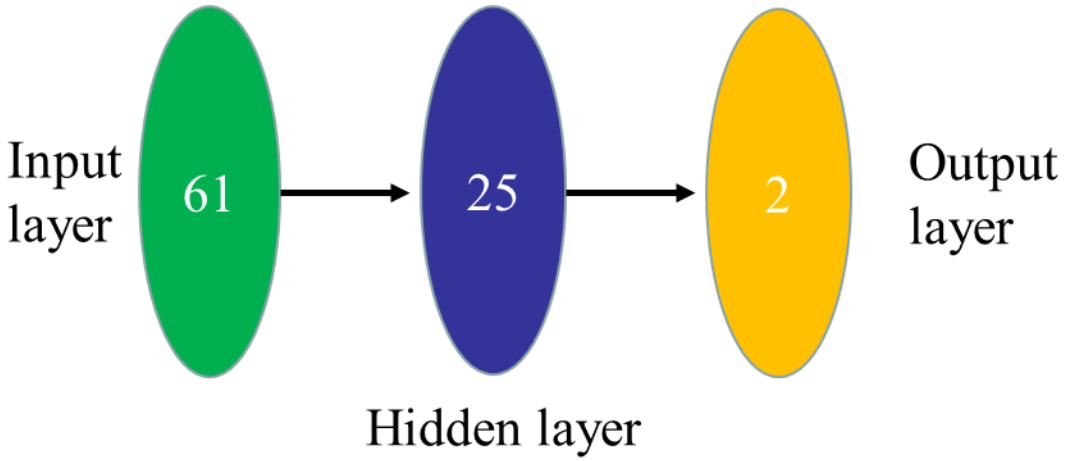


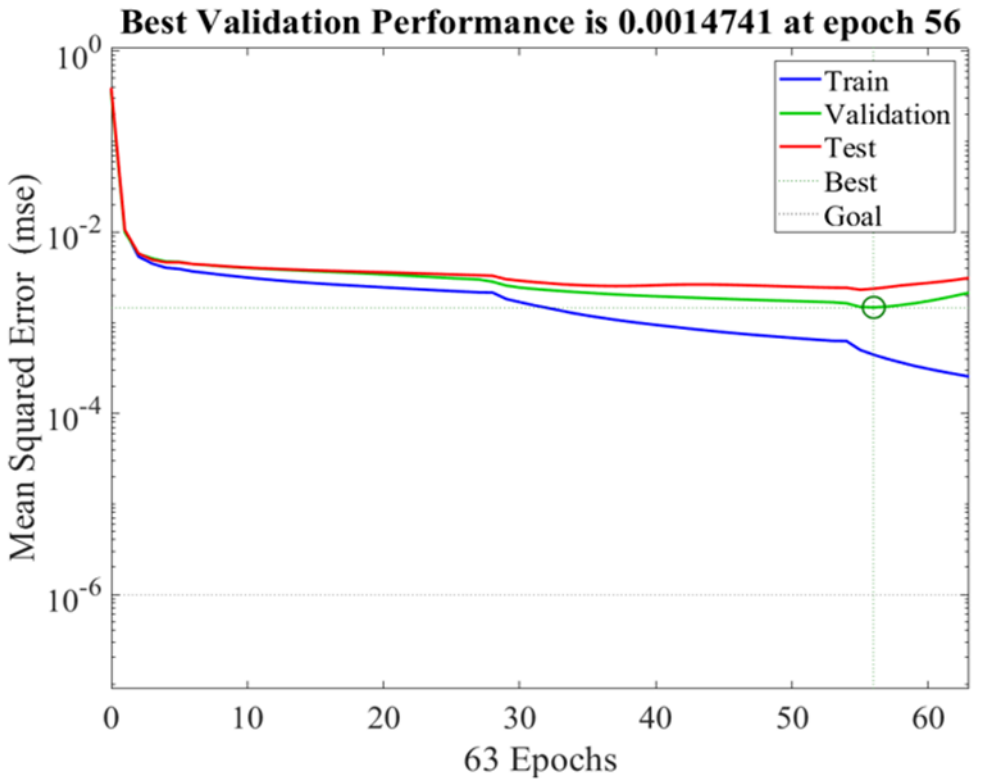


HEPS

ML for HEPS Lattice Design [1]

- Using DNN for HEPS lattice design
  - Highly nonlinear model
  - Applying HEPS optimized lattice data
  - Optimize brightness (BN) and dynamic aperture (DA)



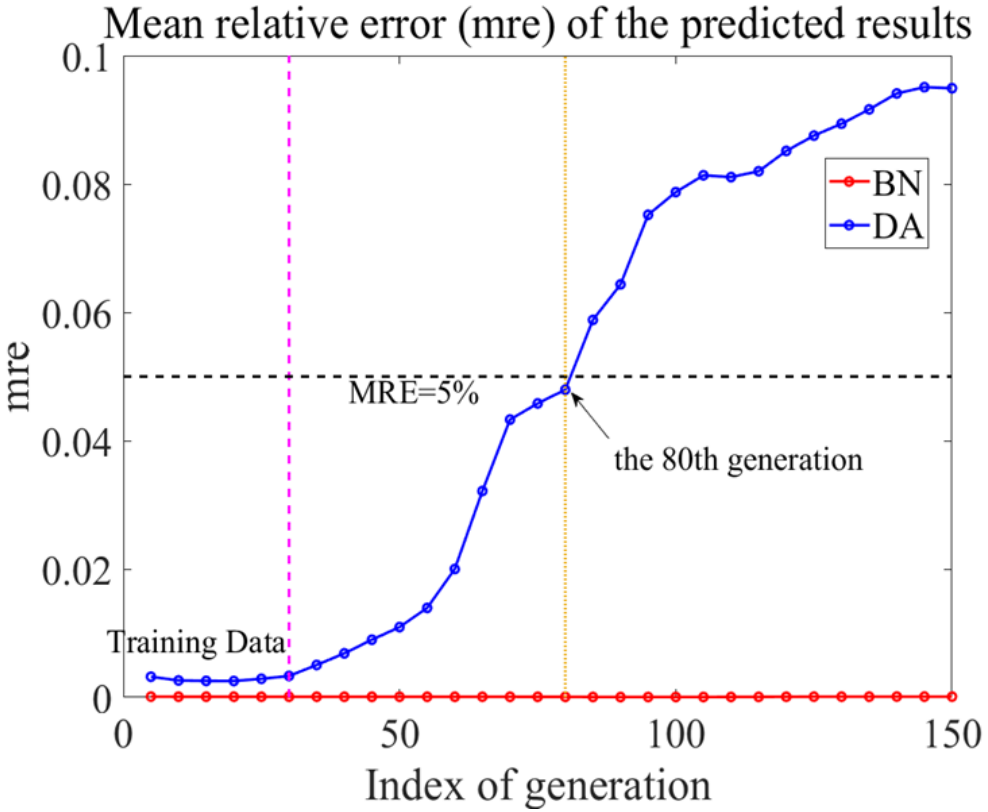


ML for HEPS Lattice Design [2]

- The DNN method can predict the BN close to 100% and DA over 95% at 80th generation

	Single thread (s)	62-thread parallel computing (s)
DNN	0.3944	0.0092
Particle Tracking	78020	1414.2
Improvement	O(5)	O(5)

DNN is 5 Orders better efficiency than Particle Tracking!



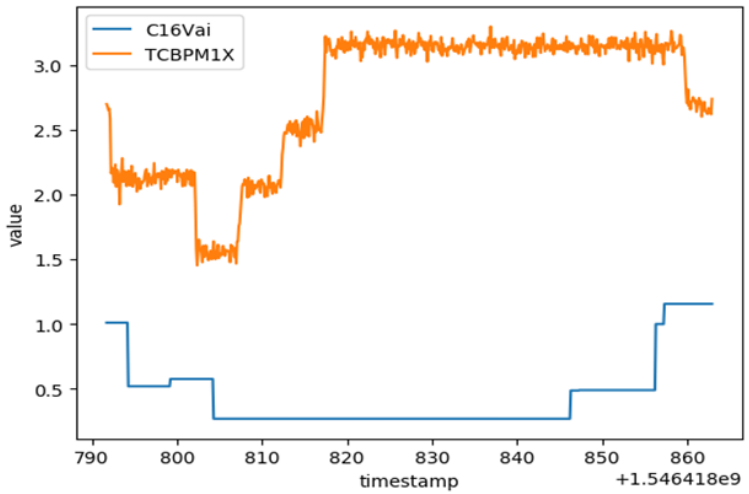
Machine Learning at Work

A test for BEPC-II timestamp correction

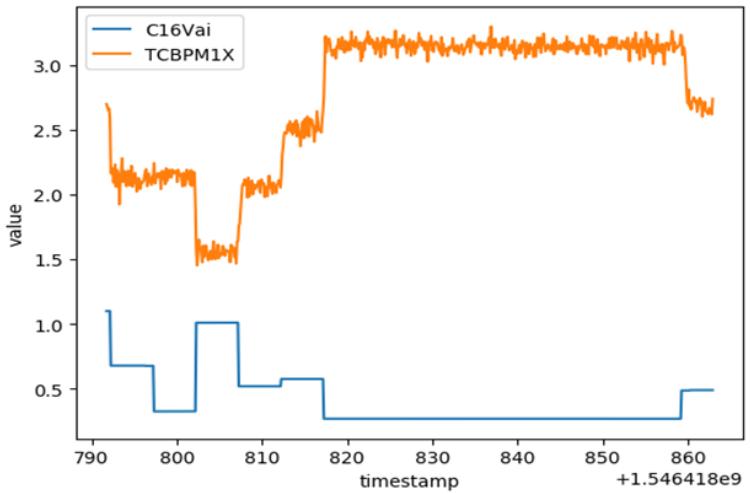
- Correlation function $R = [1 - \int \zeta (h(f_1(t + dt)) - f_2(t))]$
- Objective function $\arg\max R(dt)$

f1(t)&f2(t) : The relation between 'value' and 'timestamp' of two systems(such as correctors with BPM).
h(): Projection of one group of value to another.
ζ(): Integral coefficient. (Remove interference and noise. Keep normalization)

Input data



return

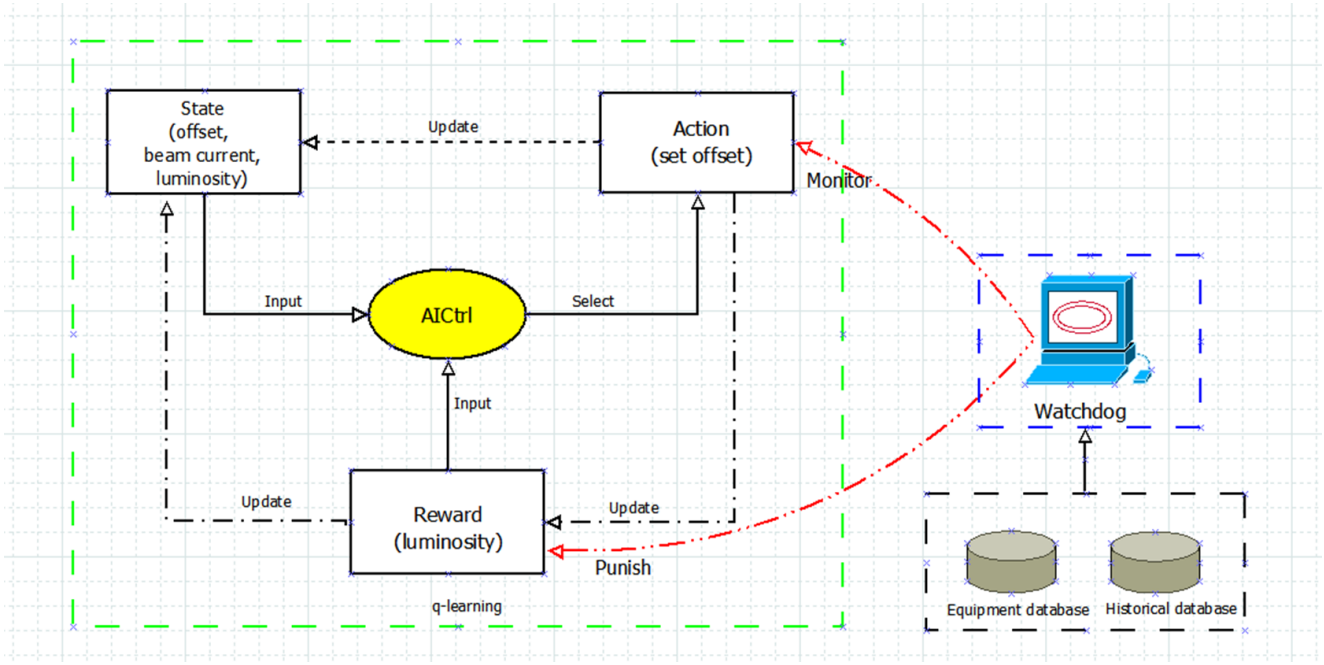


Accelerator Intelligent Control System [1]

AICtrl

- Could be applied to many conditions, such as beam loss reduction, luminosity optimization and so on
- Based on deep learning and reinforcement learning, it is better to find good condition with historical data then merely manual tuning

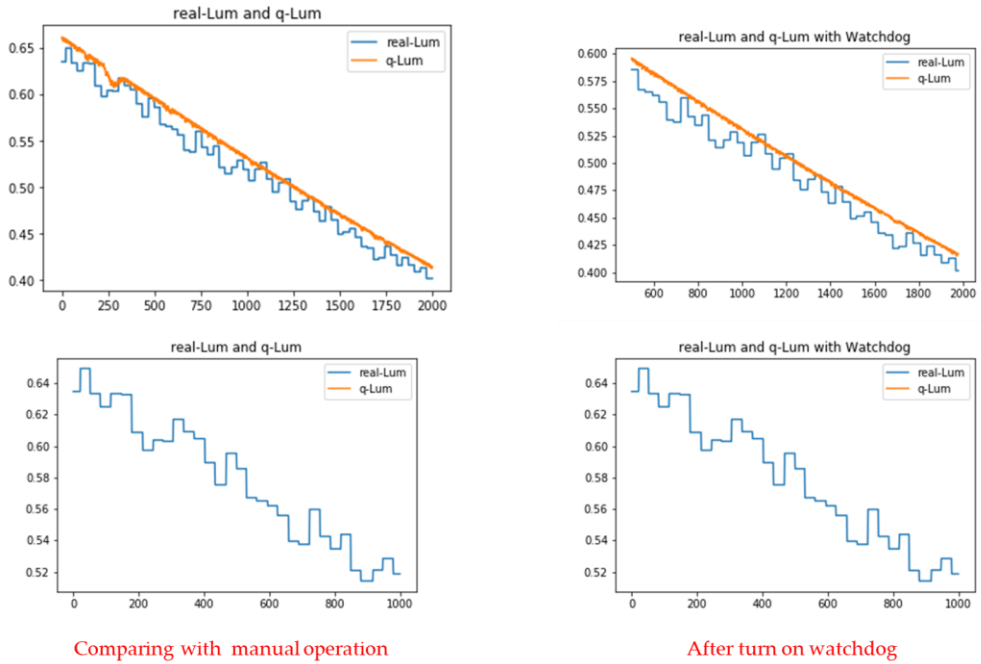
Application on luminosity optimization



HEPS

Accelerator Intelligent Control System [2]

- Application on luminosity optimization by adjusting vertical beam position offset
  - Comparing with manual tuning, AICtrl can make luminosity always in a good state
  - Strange behavior could be avoided by simply turning on the watchdog



Comparing with manual operation

After turn on watchdog

HEPS

Next Steps

- An Application “Template”
- Setting up virtual accelerator environment for ML App tests
- Setting up Hadoop environment
- Setting up GPU computing
- Getting more operation data

Summary

- HEPS Control/Data Systems designed for ML
- Overall consideration, modularized implementation
- Machine Learning for Accelerator Platform prototyped
- Many application ideas are emerging
- Collaborations are welcome

Thanks for your attention!

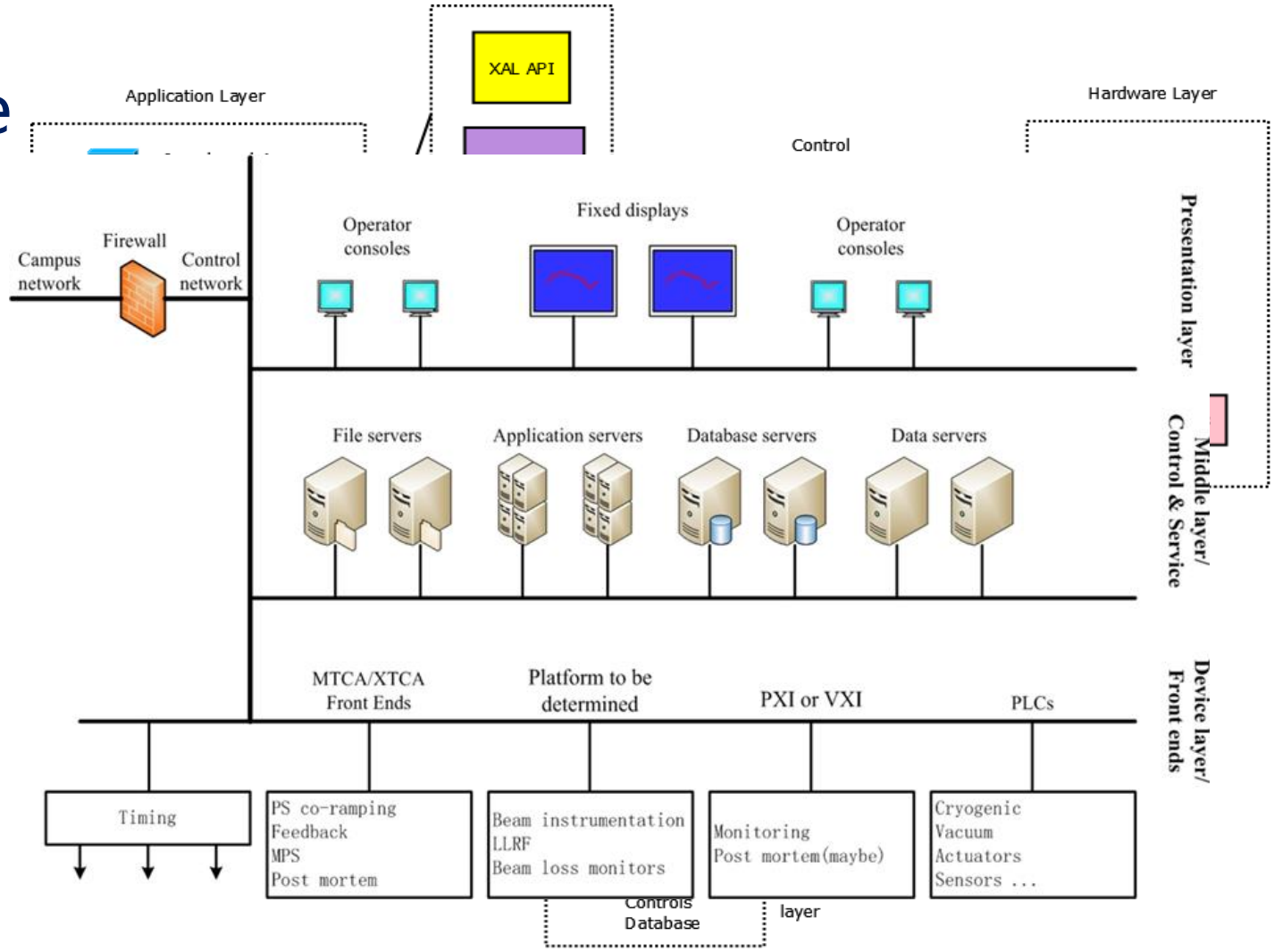
Machine Learning Platform Source Code Repository

- https://github.com/NicoleQiao/MLPlatform

Backup Slides

Controls and Online/Offline Software

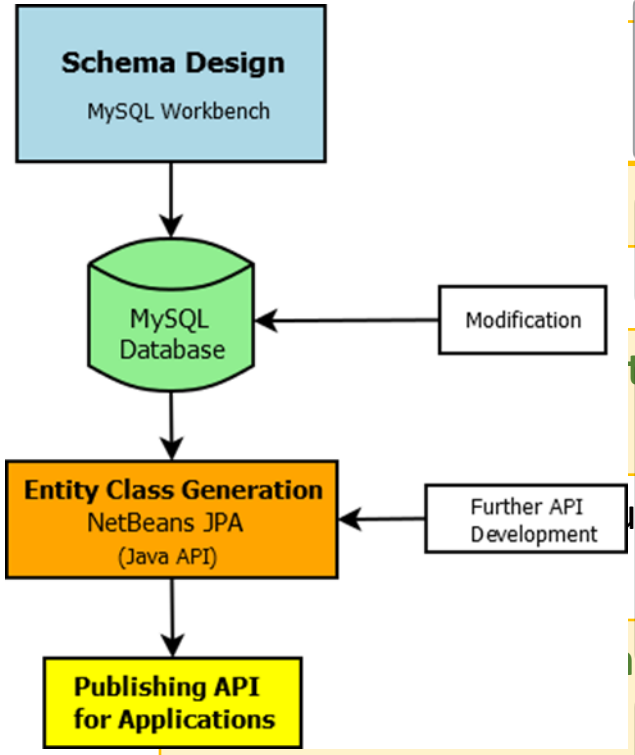
- Databases: online/offline
- Control systems
- Physics Modeling
- Services
- API
  - commonly used modules
- Online Applications
- Offline applications
  - Lattice Design w/ ML

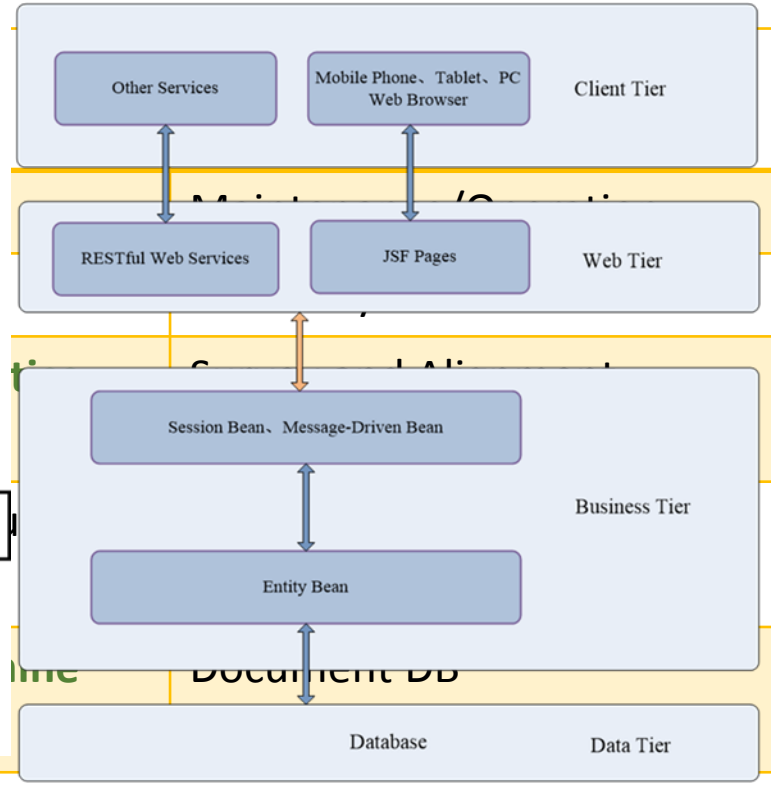


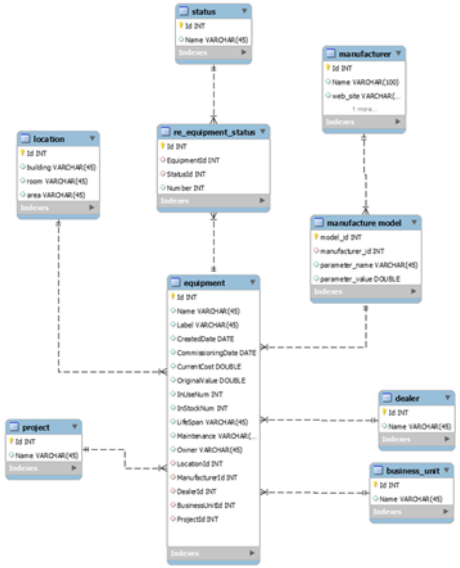
HEPS

Databases

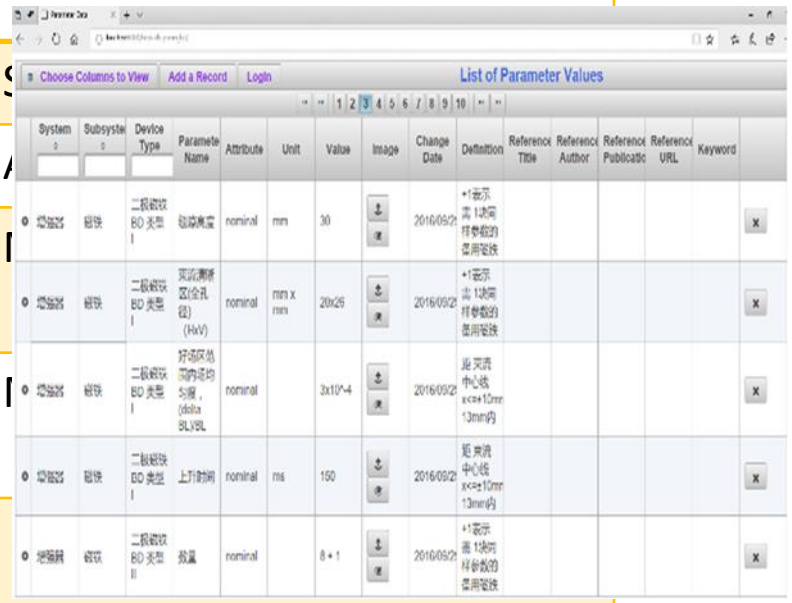
- 17 database domains in plan
- MySQL or Microsoft SharePoint/SQL







Cable

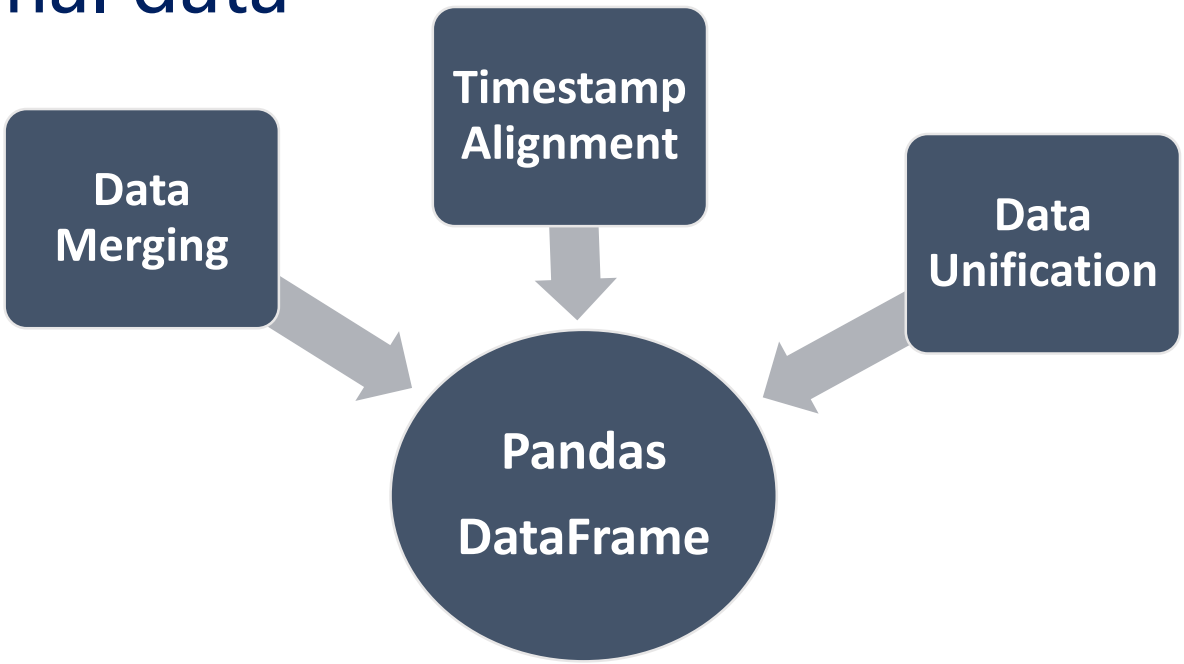


HEPS

Data Export & Unification [1]

Raw data characteristics

- Large amount of PVs as model features
- Different PV has different acquisition period
- Handling null or abnormal data



Data Export & Unification [2]

For temporal data from archiver

PV Timestamp alignment

3 data line-up types:

1. Outer -> smallest time period -> data addition
2. Inner -> biggest time period -> data deletion
3. Defined time period -> data addition & deletion

- Standardization
- Normalization
- Discretization (quantization or binning)
- Encoding categorical features

Data Pre-processing [1]

Data quality check

Handling empty, abnormal, inconsistent data

- **Padding**

Not changing over time

Bad machine status

- **Interpolation**

Linear, nearest, polynomial, cubic, spline……

- **Neural network**

Predict uncertain data through NN algorithm based on known data

Data Pre-processing [2]

- **Data feature analysis**
  - distribution analysis
  - comparative analysis
  - periodic analysis
  - contribution analysis
  - correlation analysis
- **Diagnostic functions**
  - common statistical indicators
  - histograms
  - scatter matrix diagrams
  - correlation tables & associated heat maps
  - box plots

Where Can ML Be for Accelerators

- Basically, everywhere…
- Best fit: non-linear issues with sufficient and good quality data
- Entire accelerator life cycle needs ML
  - Accelerator design
  - Accelerator/beamline controls
  - Beam Tuning – online and offline optimization
  - Operation optimization – productivity,
  - Machine reliability – maintenance,
  - Human resources: who performs the best at work

ML Application Ideas [1]

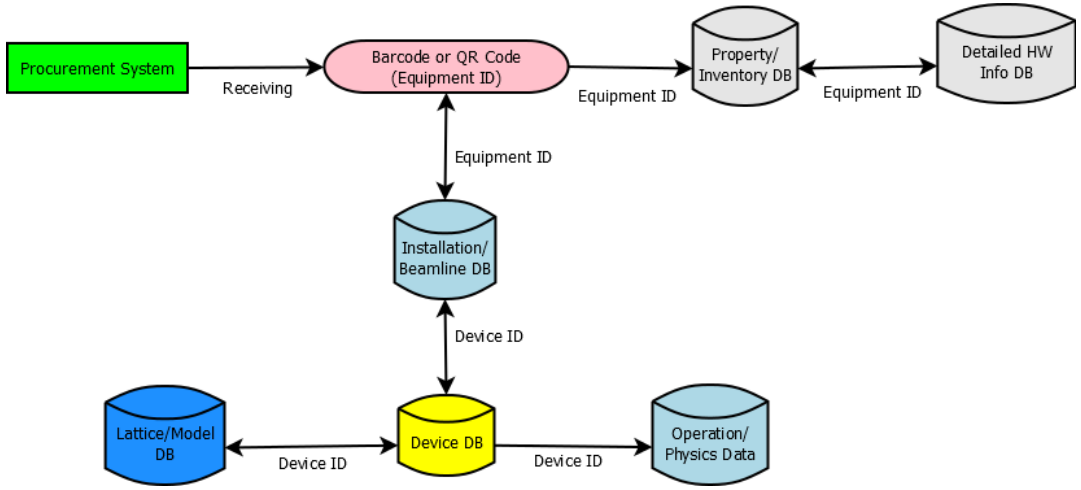
- Facility Operation Optimization
  - Use regression algorithms to improve the performance of key accelerator systems, such as high frequency cavity, superconducting system, water cooling system, etc.
- Beam Physics or experiment optimization
  - Apply the machine learning platform to beam physics optimization progress, to realize the automatic optimization of DA, emittance, current intensity, etc.
  - Data driven data acquisition
- Equipment maintenance
  - To avoid unexpected failures by analyzing equipment running data

ML Application Ideas [2]

- Big Data applications
  - Installing various sensors to collect as many data as possible
  - Data mining, data correlation, interdisciplinary data analysis
- Light source data center for data sharing
- Accelerator data archive center
  - Domestic and international data archive for various accelerator related data
- …

Project Control (Other Data)

- Procurement, Equipment
- Issue Tracking System, Maintenance, Operation Logbook…
- Work Breakdown Structure (WBS) for project management
  - Cost and schedule control/monitoring
- SharePoint based document system
  - Project Web site
  - Work flow control
  - Document and data sharing



HEPS