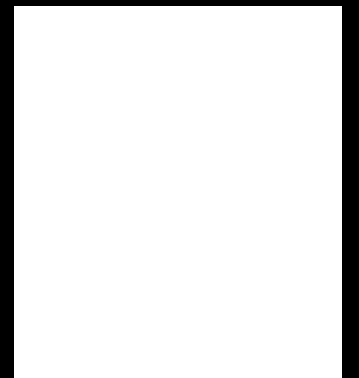


# Academia ML Iberia Summit 2020



Ricardo Paiva  
Pierre-Yves Duquesnoy

Sales Engineers



# Objetivos

- InterSystems IRIS como plataforma de Datos HTAP
  - Acceso desde entornos ML a datos en IRIS
  - Extracción Online, Transformación, Normalización, Proyecciones para ML
- Despliegue de Algoritmos ML en InterSystems IRIS
  - Servicios REST, SOAP
  - Motor de Interoperabilidad
  - Ejecución de PMML
  - Ejecución de Python y R
- ML Fácil desde SQL: IntegratedML
  - Generación automatizada de algoritmos ML con varios motores (InterSystems, H2O)



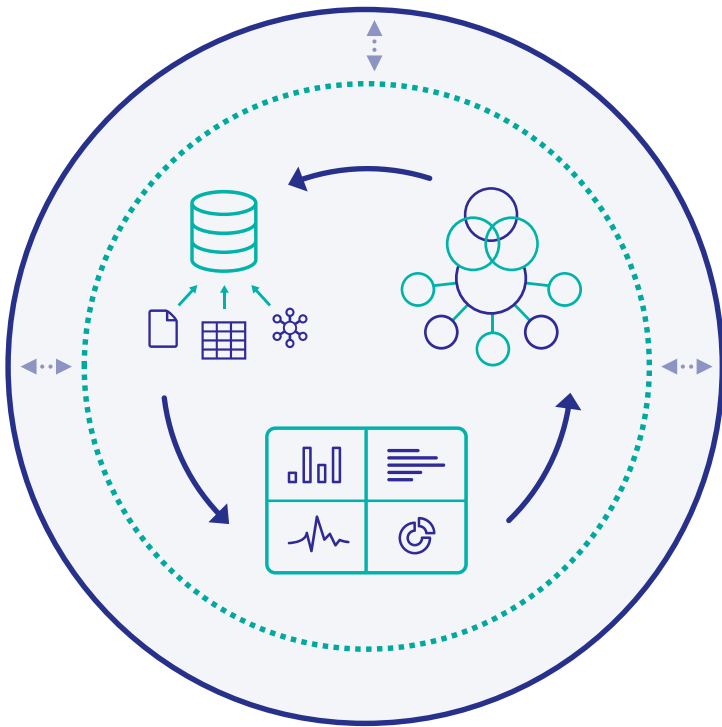
# InterSystems IRIS como plataforma de Datos HTAP

---

Ejercicios 1 y 2

# InterSystems IRIS Data Platform

# A Unified Platform for Data Intensive Applications



Multi - Model	<ul style="list-style-type: none"><li>• Relational</li><li>• Document</li></ul>	<ul style="list-style-type: none"><li>• Multidimensional</li><li>• Object</li></ul>
Multi - Workload	<ul style="list-style-type: none"><li>• HTAP with Isolated Workload Processing</li><li>• Unique Approach to Big Data Analytics</li><li>• ACID Transactions + Real-Time Indexing</li></ul>	
Open Analytics	<ul style="list-style-type: none"><li>• Analytic SQL</li><li>• Business Intelligence</li><li>• Natural Language Processing</li></ul>	<ul style="list-style-type: none"><li>• Predictive Model Runtime</li><li>• Connector Architecture</li></ul>
Interoperability	<ul style="list-style-type: none"><li>• Languages and Tools</li><li>• Data Integration</li><li>• Application Integration</li></ul>	<ul style="list-style-type: none"><li>• Composite Business Process Orchestration</li></ul>

Flexible Cloud | Hybrid | On Premises Deployment Options



# Incorporando ML a sus procesos de negocio

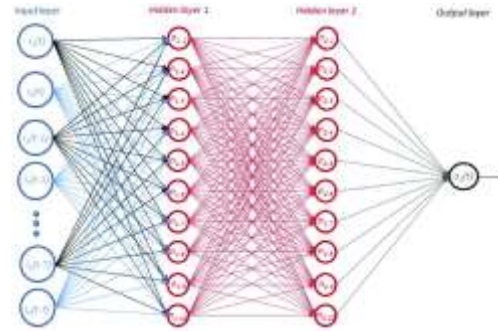
## DataOps

Capturar, normalizar and preparar datos de una manera eficaz



## Construcción de Modelos

Interoperar con herramientas especializadas de ML



## Despliegue

Desplegar y operar su modelo en aplicaciones de negocio en tiempo real





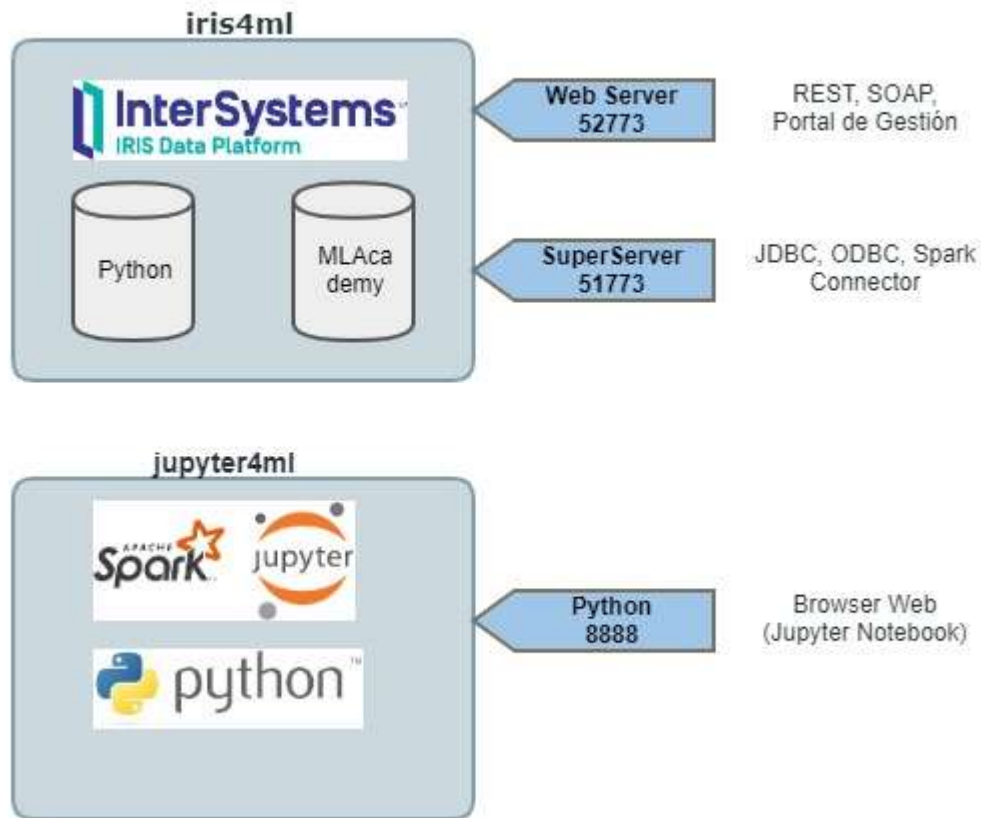
# Escenario ML

## Estudio con Información demográfica de Citas medicas

- 100 00 citas en Brazil, fuente: Kaggle, cargado en una tabla IRIS SqlUser.Appointments
- Variables disponibles
  - Edad, Genero, Día de Reserva, Día Citado, Hipertensión, Diabetes,...Barrio de origen
- Variable a predecir
  - NoShow (True/False): si el paciente no se presenta a la cita



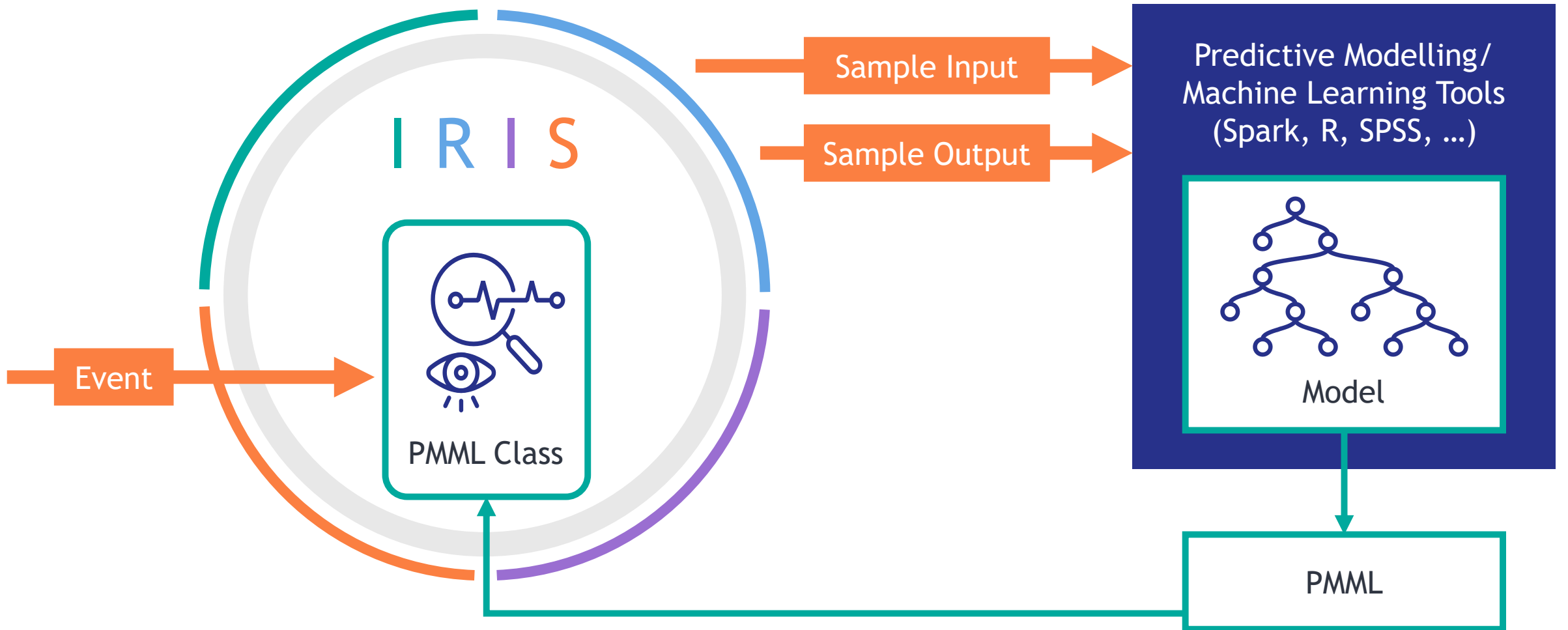
# Arquitectura del Entorno



- 2 contenedores docker
  - Iris4ml
    - 2 bases de datos: MLACADEMY y PYTHON
    - Puerto Web: 52773
    - Puerto SuperServer: 51773
  - Jupyter4ml
    - Jupyter en puerto 8888
    - Python 3 & pySpark
    - librerías ML
- 1 volumen “./shared”
  - Compartido con el host
  - Compartido entre contenedores



# Desarrollo y Despliegue de ML vía PMML



# Ejercicios 1 , 2 , 3

1. Arrancar los contenedores
2. Abrir Jupyter Notebook : `MLAcademy_IRIS_Pred_Ausencia_LogReg_PMML.iypnb`
  - Ejecutar para entrenar un modelo “Logistic Regression” sobre los datos
3. Generar y Exportar el PMML del algoritmo
  - Opcional:
    - Conexión Spark con el adaptador nativo
    - Otro Notebook `MLAcademy_XGBoost_Pred_Ausencia.ipynb`: Compara 2 Algoritmos de ML



# Despliegue de Algoritmos ML en InterSystems IRIS

---

# Despliegue de Algoritmos

- Beneficios de InterSystems IRIS para Despliegue
  - Capa de Interoperabilidad
    - Capa de ETL (Extracción de varias fuentes, Transformación de Datos, Normalización)
    - Servicios (REST, SOAP...)
    - APIs publicadas vía el InterSystems API Manager (IAM)
    - Reacción a Eventos
    - Trazabilidad de todas las llamadas
    - Motor de workflow para Interacción humana



# Despliegue de Algoritmos

- Ejecución del Modelo
  - PMML
  - Llamadas directas a Python o R desde el Gateway
  - Acceso directo a los datos Transaccionales o Analíticos almacenados en BBDD
    - Mismo formato para Train/Test y Operaciones
- ML Robotizado
  - Producción de Integración que re-entrena un modelo cuando sus resultados pasan debajo de un umbral y lo despliega
  - Ejemplos en: ML Toolkit
    - <https://openexchange.intersystems.com/package/PythonGateway>
    - <https://openexchange.intersystems.com/package/Python-Gateway-Samples>



# Ejercicio 4: Despliegue PMML

- Servicio REST con algoritmo pre-cargado
  - Predicción de NoShow según Identificador de fila de datos
    - Solo hacer falta referenciar los datos “transaccionales”
    - IRIS extrae y expande la columnas a pasar al algoritmo (Gender\_F, Gender\_M)
  - Predicción con datos nuevos
    - Se pasan todas la características necesarias al algorithmo
- Servicio de Importación de PMML
  - Importa y despliega nuevo algoritmo



# Ejercicio 5: Llamar a Python desde IRIS

- Uso de la capa de Interoperabilidad
  - Llamadas a Python o R
  - Trazas de Mensajes



# IntegratedML

---



# IntegratedML

- Objetivo: Un acceso simplificado a ML desde una API SQL
- Características
  - Gestiona la preparación de Dato
  - Prueba varios algoritmos
  - Escoge el mejor Algoritmo, lo entrena y lo optimiza
  - Desde SQL
  - Motores de ML intercambiables ( InterSystems, H2O,...)



# Sintaxis IntegratedML

**Create Model** `MySurvival` Predicting (Survived) from  
`titanic.Passenger`

**Train Model** `MySurvival` from `Titanic.Passenger`

Select ID,Name, **Predict** (`MySurvival`) from `Titanic.Passenger` where..

Select \* from `%ML.TrainingRun`

- Detalles de training en columna **Log**



The power behind what matters.



**Gracias**

