Dec 12, 2023

Bring AI/ML into Action: Informatica ModelServe

- Damerla Chinnari, Senior Solutions Architect, CSA

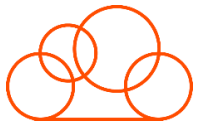
Informatica

Housekeeping Tips

- Today's Webinar is scheduled for 1 hour
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the Q&A option and we will respond at the end of the presentation
- The webinar is being recorded and will be available on our Success Portal - where you can download the slide deck for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the post-webinar survey and provide your feedback and suggestions for upcoming topics.

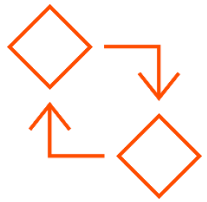
Informatica

Feature Rich Success Portal



Bootstrap trial and POC Customers

Enriched Customer Onboarding experience



Product Learning Paths and Weekly Expert Sessions

Informatica Concierge

Tailored training and content recommendations

Informatica

More Information

Success Portal

https://success.informatica.com

Communities & Support

https://network.informatica.com

Documentation

https://docs.informatica.com

University

https://www.informatica.com/in/services-and-training/informatica-university.html

Informatica

Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

Informatica

Agenda

- What is ML ?
- What is MLOps ?
- Benefits of MLOps
- Informatica's solution for MLOps
- What is InfaCore ?
- What is ModelServe ?
- Why ModelServe ?
- How to Operationalize Informatica ModelServe
- Demo
- Q&A

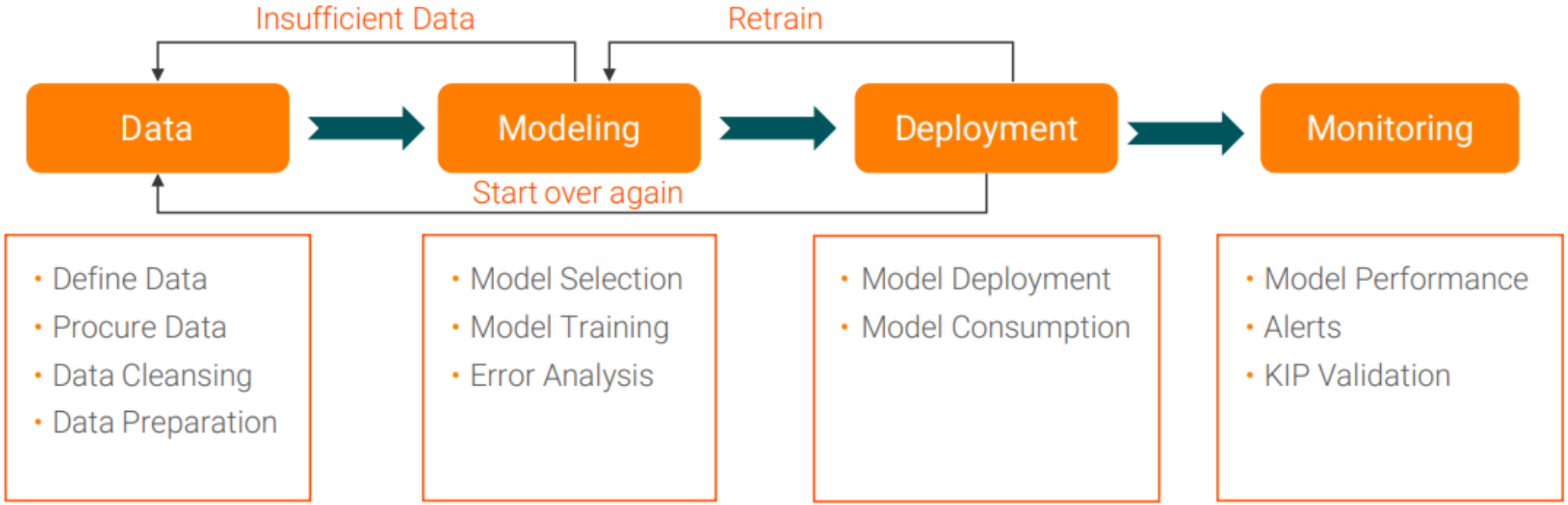
What is ML(Machine Learning)?

- ML is a subset of artificial intelligence (AI) in which an algorithm discovers patterns in data that include batch and streaming data.
- ML models become more intelligent and more accurate with exposure to new data. An example is how Google maps gives faster routes by adding the traffic conditions.
- ML allows businesses to be more innovative, efficient and sustainable.
- The reason many ML projects fail is because many proofs of concept never make it to production. The ML community has always focused on building ML models but not on building production-ready ML products
- This includes the roles required to automate and operate an ML system in a real-world setting.

What is MLOps ?

MLOps (Machine Learning Operationalization) enables the operationalization of the end-to-end pipeline that supports the continuous delivery and continuous integration of ML models in a production environment.

MLOps focuses on data model deployment, operationalization and execution.



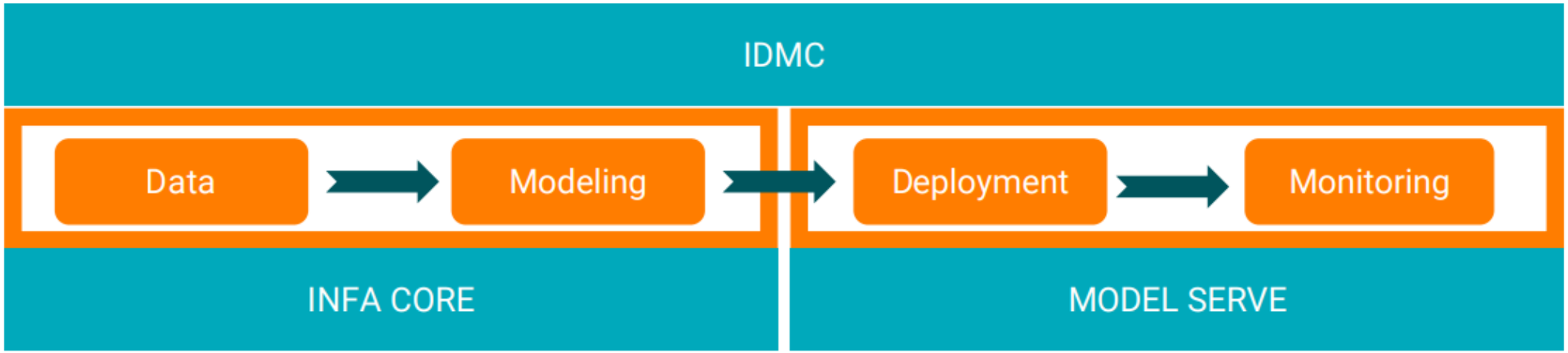
Benefits of MLOps

There are many benefits of implementing MLOps, including the ability to:

- Deliver business value for data science projects
- Improve the efficiency of the data science team
- Allow ML models to run more predictably with better results
- Help enterprises improve revenue and operational efficiency
- Speed up AI/ML initiatives with high performing models
- Advance ML for digital transformation

Informatica's solution for MLOps

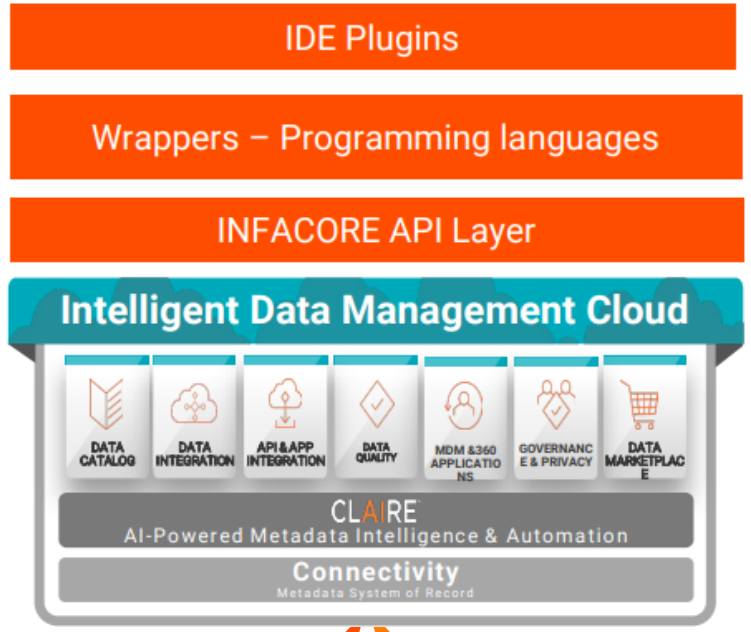
Informatica provides end-to-end MLOps solution with the help of InfaCore API platform and Model Serve, which is essential for high-performing AI solutions.



What is InfaCore?

An open , extensible ,embeddable framework for democratization of data management for all developers and all data driven applications.

Initial Features
Connectivity
Advance Transformations
CLAIRE Capabilities
Operationalization



Informatica

InfaCore Example for low code persona

Scenario

Hierarchy processing

Hand code/low code

```
from pyspark.sql import SparkSession
from pyspark.dbutils import DBUtils
from pyspark.sql.functions import col
from pyspark.sql.types import StructType

spark = SparkSession.builder.getOrCreate()

# declare dummy data to demonstrate how the collapse mechanism works
jsonStrings = ['{"car":{"color":"red", "model":"jagsar"},"name":"Jo","address":{"city":"Houston","state":"Texas", "zip":{"first":1234,"second":4321}}}']
otherPeopleRDD = spark.sparkContext.parallelize(jsonStrings)
df = spark.read.json(otherPeopleRDD)

# Recursively iterates over the schema, creating an array of arrays, whereby each item
# of the master array, is an array of column names
#
# For example, lets say there are three columns of which two are hierarchical and the following schema/structure
#   name
#   address
#     street
#     city
#     state
#     zip
#       first
#       second
#   details
#     age
#     gender
#
# the function will return the following array:
# [["name"],["address","street"],["address","town"],["details","age"],["details","gender"]]
def get_all_columns_from_schema(source_schema):
  branches = []
  def inner_get(schema, ancestor=None):
    if ancestor is None: ancestor = []
    for field in schema.fields:
      branch_path = ancestor+[field.name]
      if isinstance(field.dataType, StructType):
        inner_get(field.dataType, branch_path)
      else:
        branches.append(branch_path)

  inner_get(source_schema)

  return branches

# collapse_columns is passed the dataframe schema, which is then passes
# to get_all_columns_from_schema.  On return, it iterates through the array
# of columns in order to build up the select list that will be used
# to collapse the hierarchical columns into a single 2d structure
#
# for example, lets say _all_columns has the following array: [["name"],["address","street"]]
# after iterating through the array, the function response will be
# [col("name"), col("address.street").alias("address_street")]
def collapse_columns(source_schema, columnFilter=None):
  _columns_to_select = []
  if columnFilter is None: columnFilter = ""
  _all_columns = get_all_columns_from_schema(source_schema)
  for column_collection in _all_columns:
    if (len(columnFilter) > 0) & (column_collection[0] != columnFilter):
        continue

    if len(column_collection) > 1:
      _columns_to_select.append(col('.'.join(column_collection)).alias('_'.join(column_collection)))
    else:
      _columns_to_select.append(col(column_collection[0]))

  return _columns_to_select

# as above but for individual columns
def collapse_column(source_df, source_column):
    column_name = ""
    if isinstance(source_column, Column):
      column_name = source_column.name
    else:
      column_name = source_column

    return collapse_columns(source_df.schema, column_name)

# returns a dataframe that has been collapsed.  Input is the dataframe to be collapsed
def collapse_to_dataframe(source_df):
  return source_df.select(collapse_columns(source_df.schema))
```

Informatica Mapping

InfaCore API code

```
Infacore.parse_hierarchy(input, type_def_sch)
```

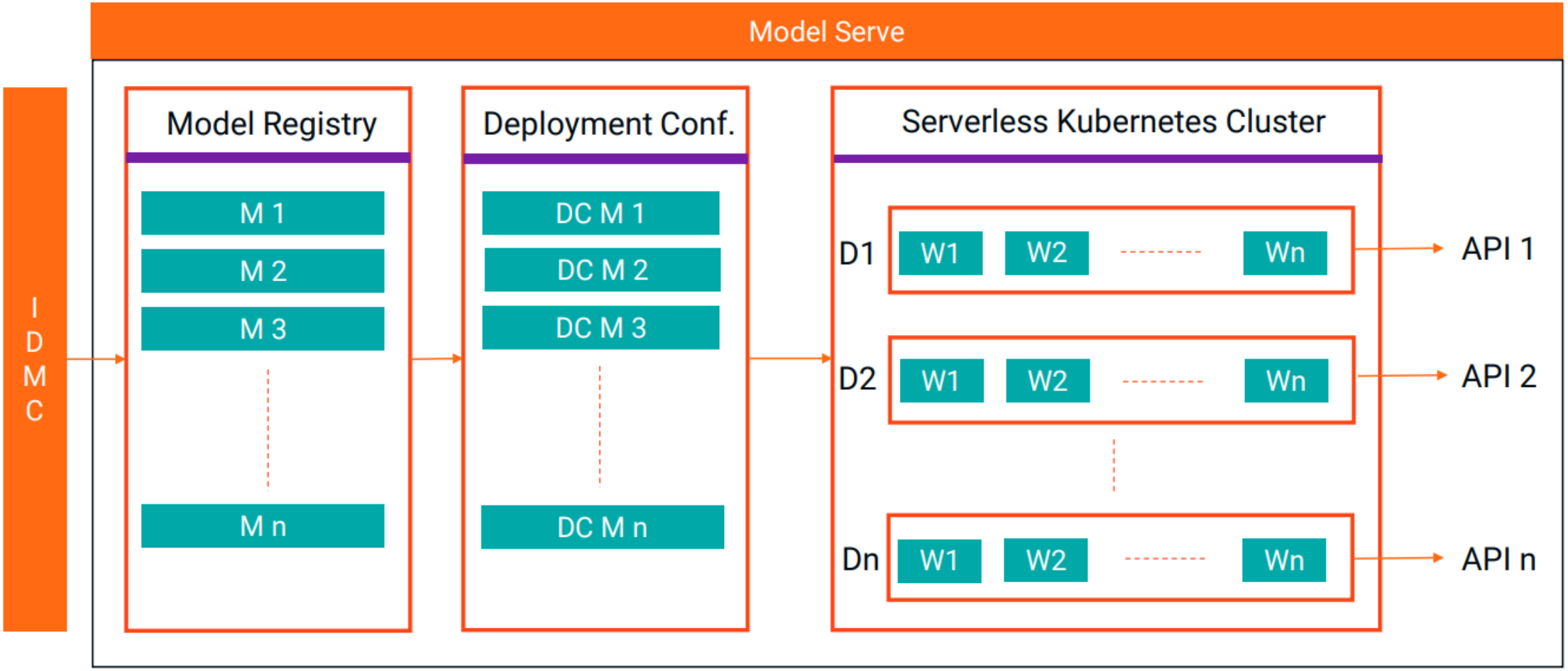
What is ModelServe?

- ModelServe is Centralized model registry & management within IDMC
- ModelServe is Informatica's cloud-native MLOps solution that simplifies ML model deployment
- ModelServe service empower teams of data scientists to create and govern high-quality machine learning models with MLOps.

ModelServe provides:

- Build nearly any AI/ML model in no time
- Deploy serverless AI/ML at scale
- Monitor, alert and consume

ModelServe Architecture



Informatica

Key Benefits of ModelServe

- Deploy and operationalize just about any AI/ML model at scale with simple, wizard-driven approach
- Provide flexibility in building AI/ML models in most frameworks and consume them in just about any application
- Accelerate AI/ML initiatives with high-quality, trusted and governed data
- Boost productivity of data science teams
- Enhance AI/ML model performance with timely delivery of trusted data using integrated DataOps

How to Operationalize Informatica ModelServe

1. Model Registry: Data scientists build their models using AI/ML frameworks like Python, TensorFlow, Spark ML, Keras, etc. and register the models seamlessly.

Register the model using just about any machine learning framework.

2. Model Deployment: Once registered, data scientists can deploy the AI/ML models in a serverless environment in minutes (versus days or weeks) without worrying about model provisioning infrastructure.

One-click serverless deployment.

3. Model Monitoring: With Informatica ModelServe, you can monitor the performance of the deployed model in a single pane of glass, identify inconsistencies and take the right action. Post monitoring, the model can be consumed in any application.

Monitor the deployed models for anomaly detection.

Demo

References

ModelServe

https://docs.informatica.com/data-science-and-machine-learning/model-serve/current-version/model-serve/preface.html

https://www.informatica.com/in/products/cloud-data-integration/informatica-modelserve.html

https://www.informatica.com/content/dam/informatica-com/en/collateral/data-sheet/informatica-modelserve_data-sheet_4469en.pdf

https://www.youtube.com/watch?v=mexJ4POMGMo

We would love to hear from you

- supportvideos@informatica.com
- https://twitter.com/INFAsupport

Q & A

Thank You

Informatica™