

模型引擎平台设计与实践

携程数据智能部 李媚

目录

CONTENTS

PART 01 项目背景

PART 02 整体架构

PART 03 技术方案

PART 04 进展与计划

PART

01

背景

Background

1.1 模型引擎的提出



实时部分



数据服务组

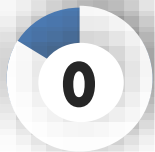


模型引擎

1.2 现有上线流程



1.3 现有服务流程



需求调研

算法提交说明书和模型文件，开发确认可行性



Online特征准备

T+1数据：酒店信息、用户画像、模型参数...
实时数据：用户点击、房价房态...



SOA服务开发

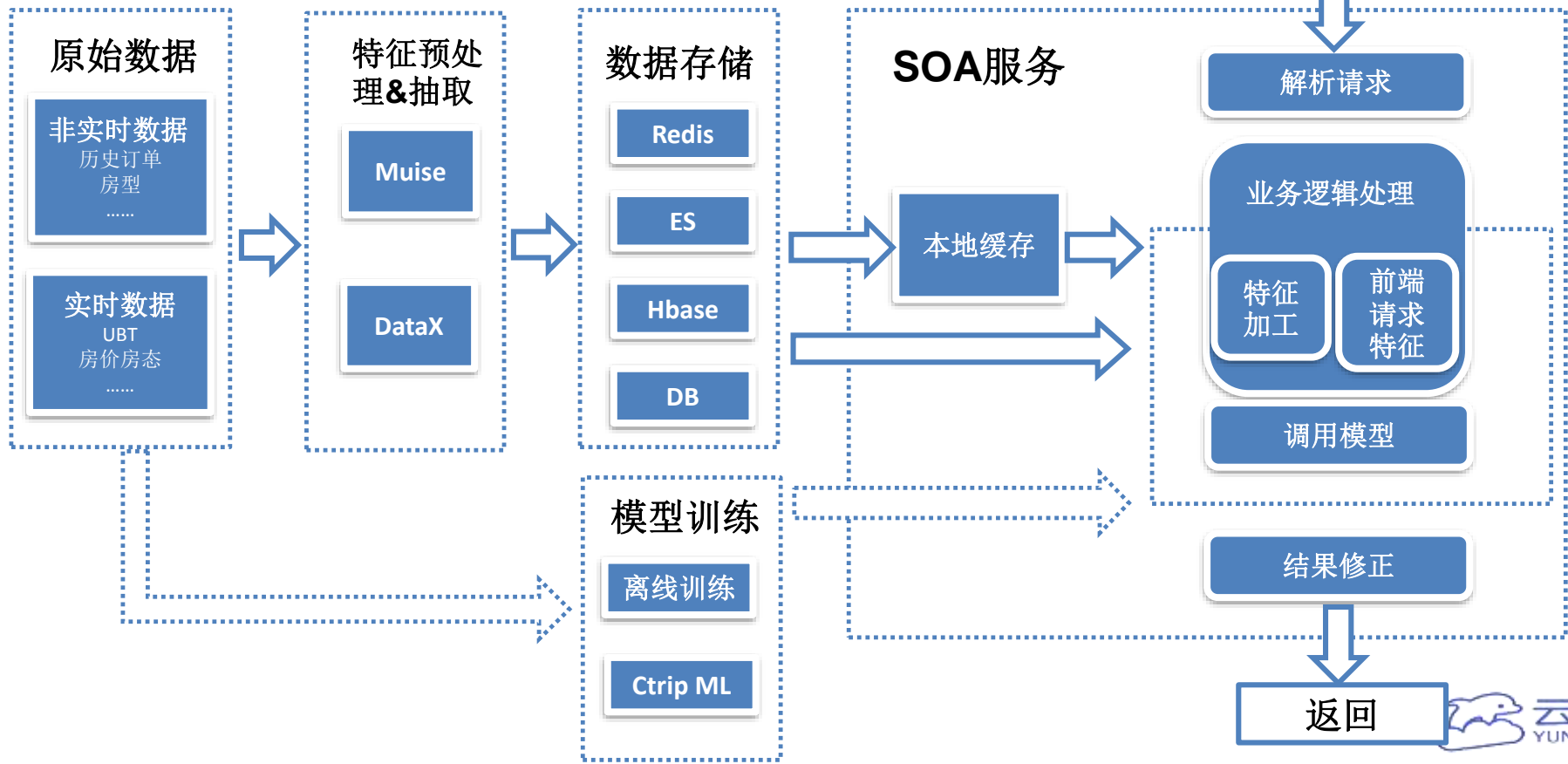
加载Online数据
实现业务逻辑
调用模型



测试验证

测试环境验证
与前端联调
生产环境抽样校验

1.4 现有服务架构



1.6 存在问题

SOA服务过重

- 业务逻辑、特征加载、模型调用耦合严重，迭代周期长
- 开发人员水平不一，项目质量层次不齐

特征管理缺失

- 各项目重复特征准备工作
- 实现特征时对复用考虑得不够，导致无法/不敢复用
- 模型和特征的依赖关系缺乏统一管理

模型训练孤立

- 训练数据需科学家自行准备
- 各团队训练成果缺乏共享途径

PART
02

整体架构

Framework

2.1 设计目标



服务于产品经理、数据科学家、工程团队、测试团队



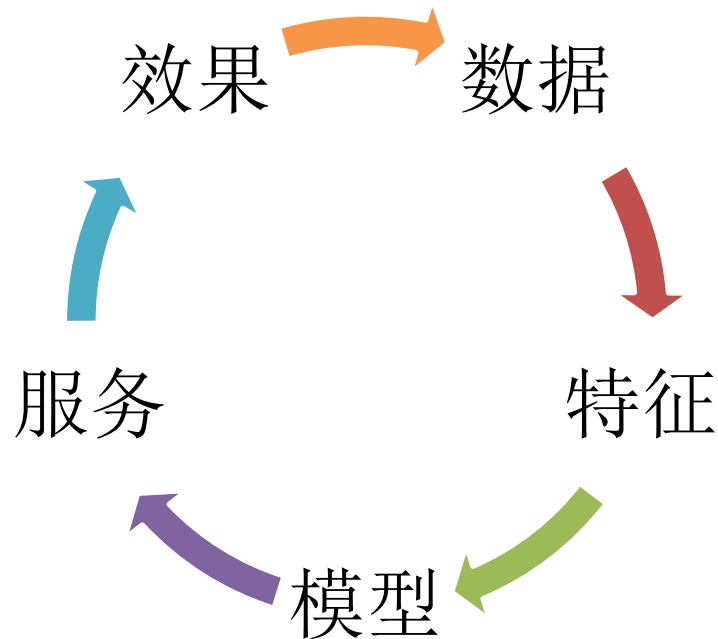
提供模型从诞生、上线到迭代的完整管理



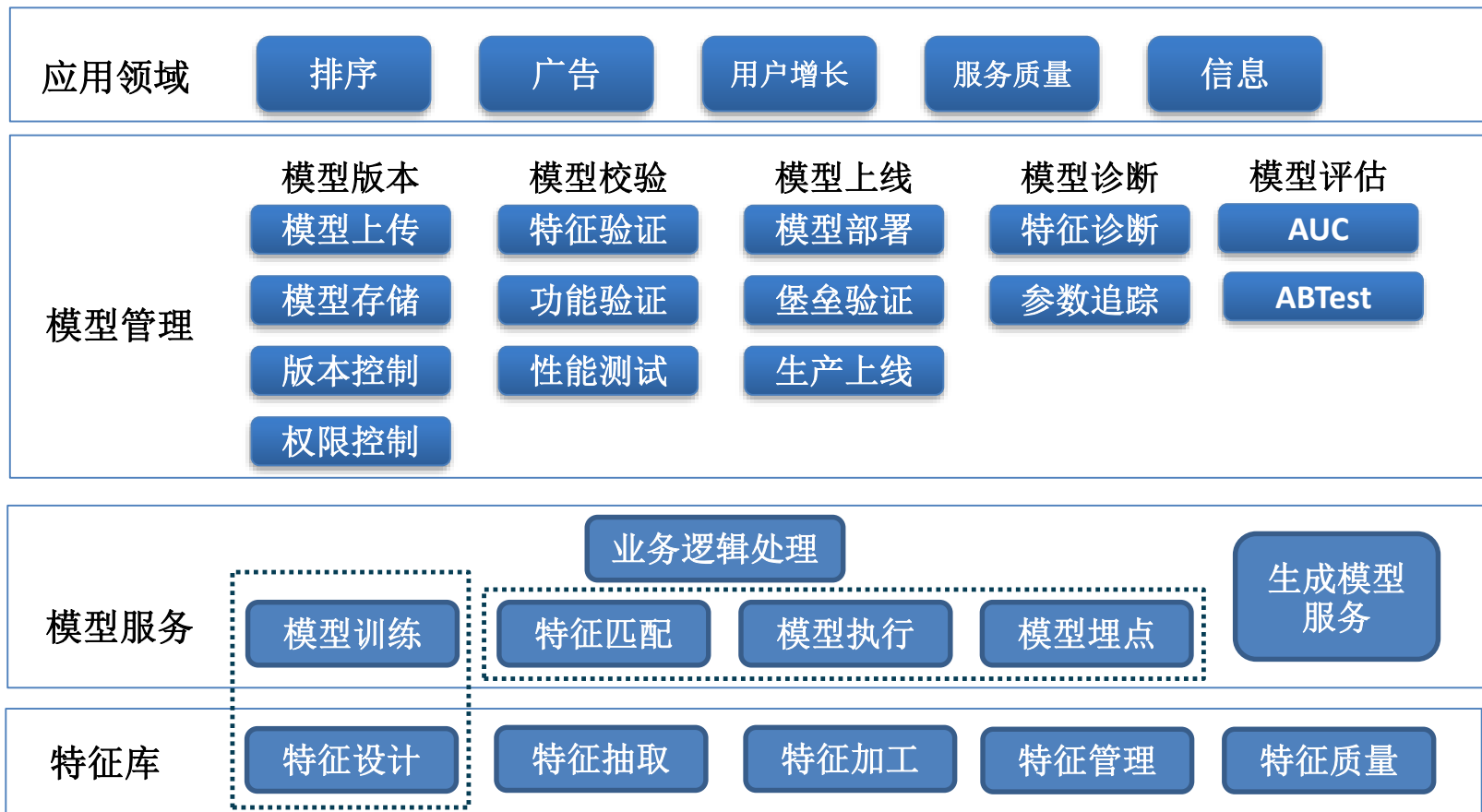
接受秒级延迟的实时数据，毫秒级地返回预测结果



广泛地应用于携程各类业务场景

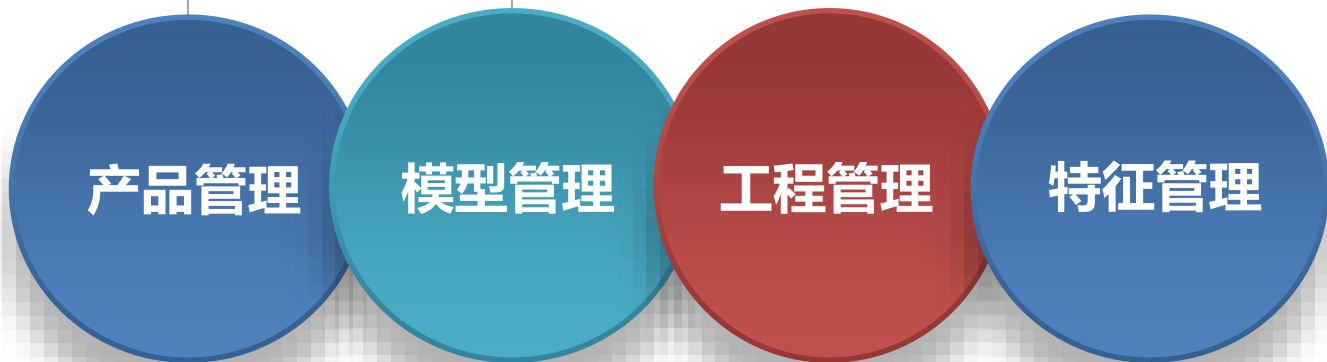


2.2 逻辑架构



2.3 主要功能

- 创建场景
- 添加AB实验
- 查看全景图、各模块状态
- 上传模型（堡垒\生产）
- 关联特征、修改特征
- 验证模型
- 更替模型



- 配置化加载特征
- 模型文件热更替
- 特征在线诊断
- 服务指标统一监控

- 添加新特征
(实时/T+1)
- 特征质量监控

2.4 预期效果

特征服务层

1

- 特征复用**70%**以上
- 特征开发时间缩短至**1/3**
- 提供实时特征供模型训练

模型计算层

2

- 自助生成模型服务，减少开发时间
- 支持多种类型的模型，如 `xgboost`、`pmml`、`python`
- 依靠平台实现模型热替换，迭代周期缩短至**1/3**

监控和质量

3

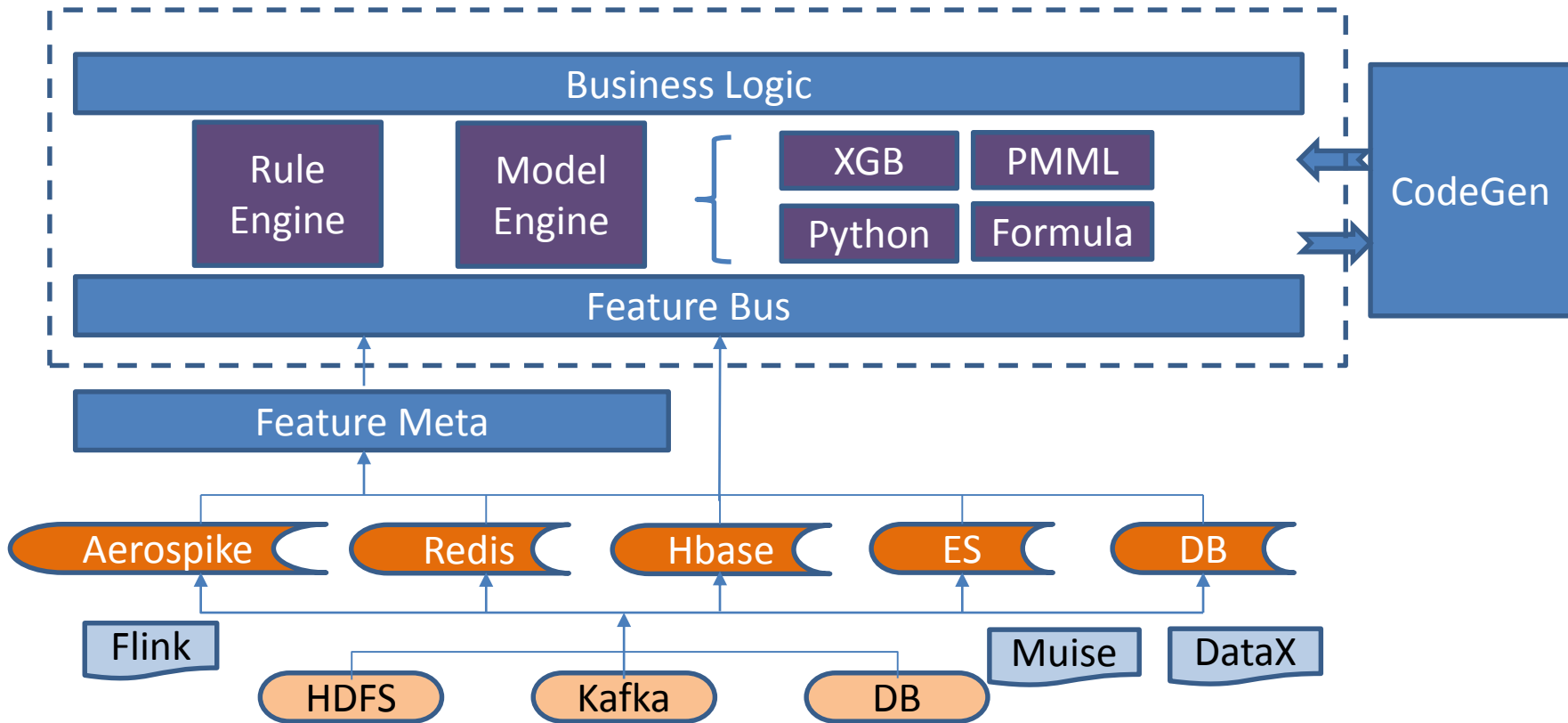
- 支持模型灰度上线
- 特征的各项指标监控，如缺失率、新鲜度、活跃度
- 模型服务的各项指标监控，如响应时长、调用频次、调用链路

PART
03

技术方案

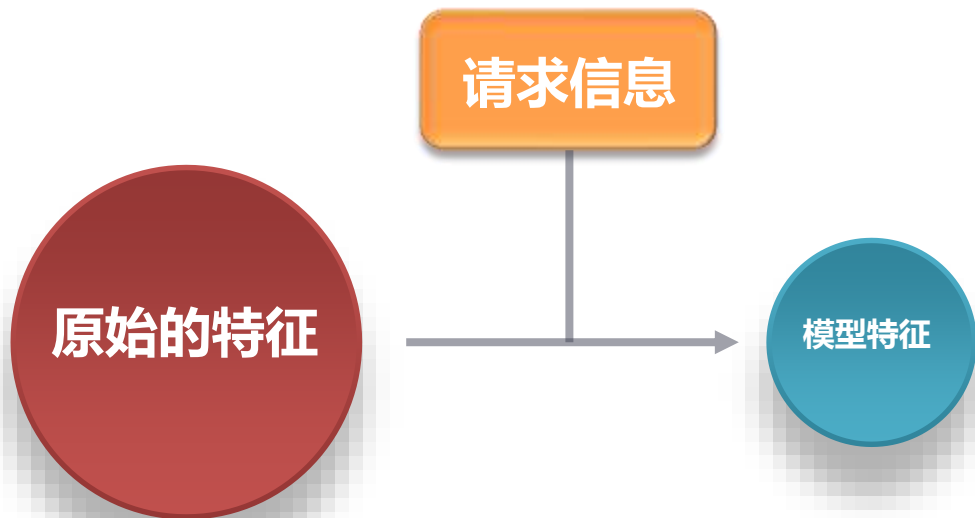
Technical Solution

3.1 技术架构



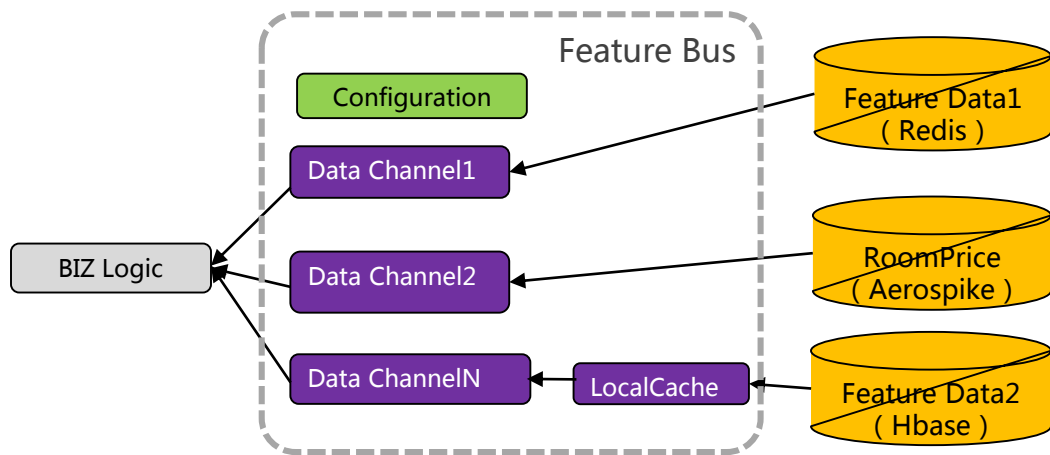
3.2 主要模块——Feature Meta

- 在线特征的元数据管理
- 描述模型与特征的关联关系



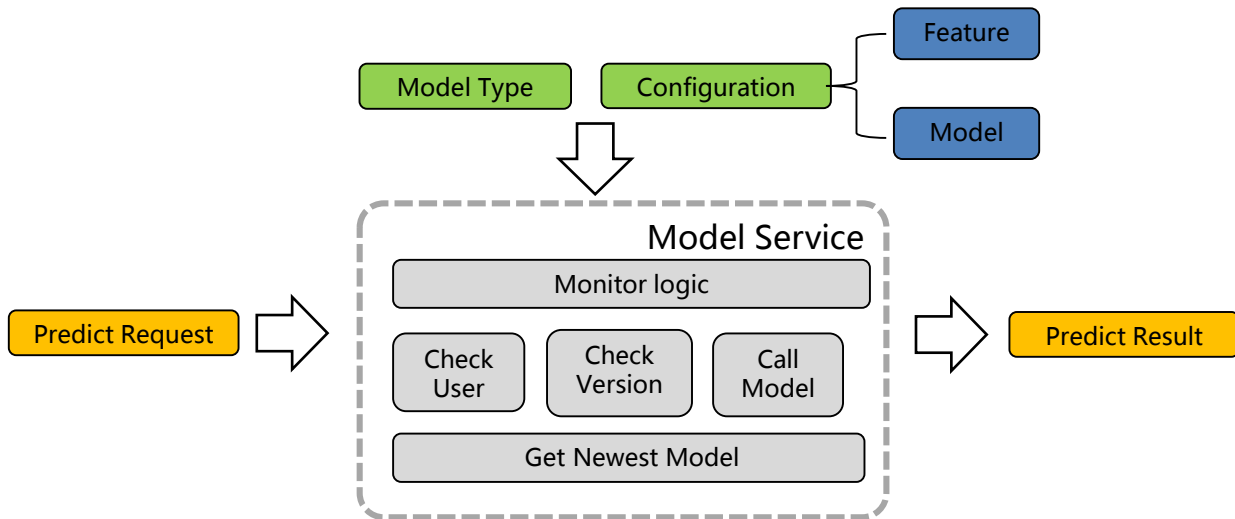
3.3 主要模块——Feature Bus

- 特征加载中间件
- 自动构建本地缓存
- 统一开发人员访问Feature Data的风格 (KV)



3.4 主要模块——Model Service

- Model Service : 支持主流机器学习算法
- Rule Engine : 支持规则、公式等广义模型



PART

04

进展与计划

PLAN

4.1 进展与计划

进展

计划

特征服务层

特征自助添加
特征关联模型

丰富特征库
完善特征监控、追踪机制

模型计算层

支持AB实验
支持模型灰度验证
支持多模型融合、模型特征在线诊断

支持更多类型的模型
支持常用业务逻辑处理的代码配置生成
模型训练功能，支持实时模型训练

模型管理层

打通模型上传、自助验证、审核上线、
更新迭代流程
酒店排序类项目七个，图像类项目两个

丰富模型服务监控
完善模型质量在线评估

谢谢

Q&A