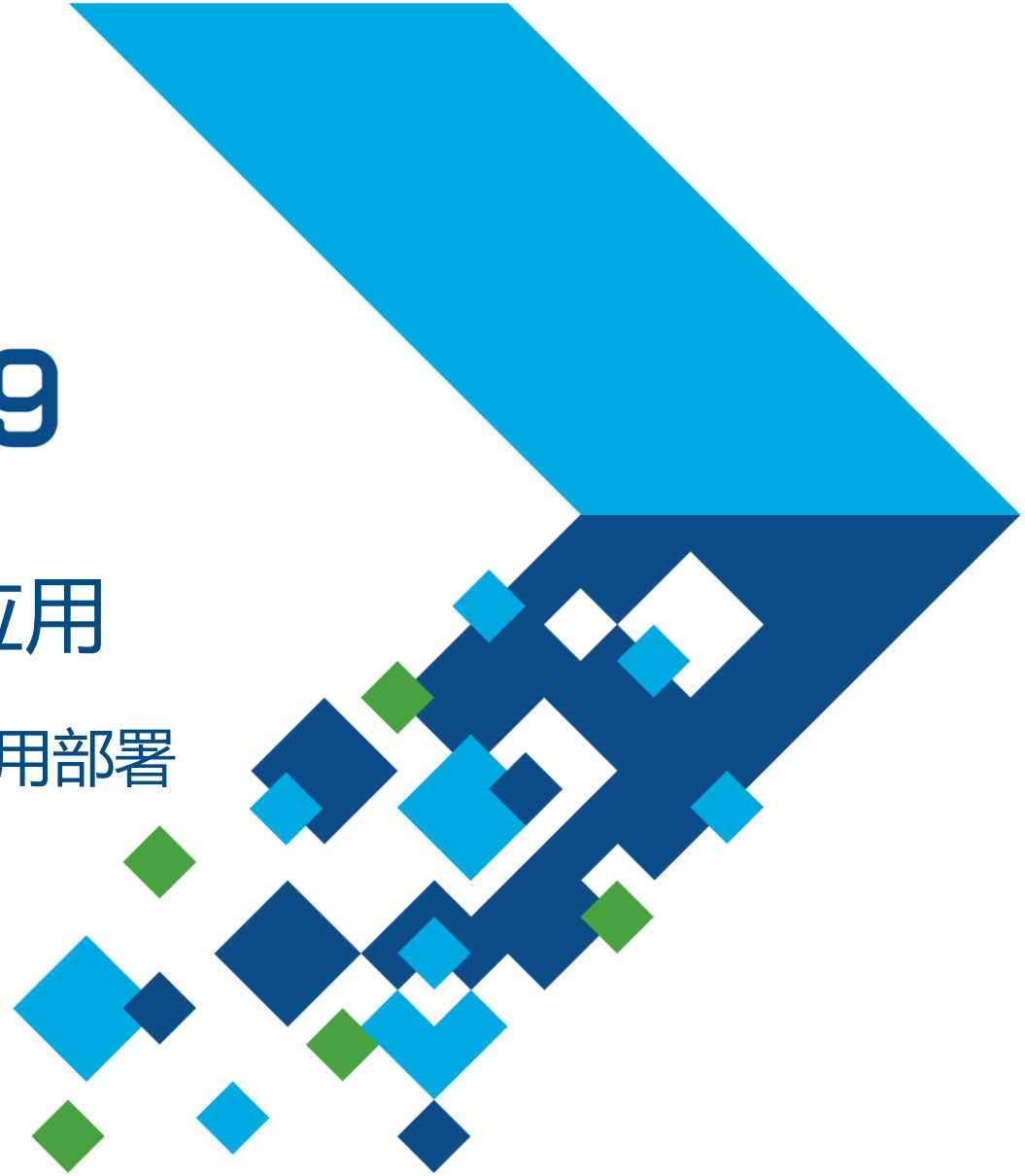
MATLAB EXPO 2019

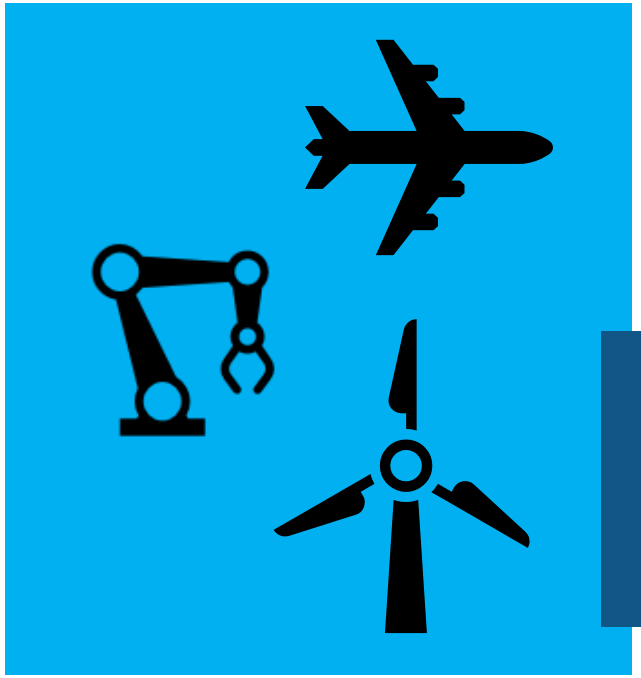
为近实时的制造业决策部署A.I.应用

-- 基于MATLAB的企业级人工智能应用部署

马文辉
MathWorks中国



大规模流式数据处理的需求



喷气发动机: ~800TB 每天
涡轮机:　　~ 2 TB 每天

预测性维护
(Predictive Maintenance)

提高运行效率

减少计划外宕机时间

More applications require near real-time analytics

医疗设备

患者安全

更好的治疗效果

车联网

安全、维护

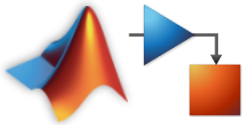
先进的驾驶性能

汽车: ~25 GB 每小时

案例导引: 开发并实施机器学习模型，以预测工业泵的故障

算法工程师

开发MATLAB和Simulink模型



系统架构师

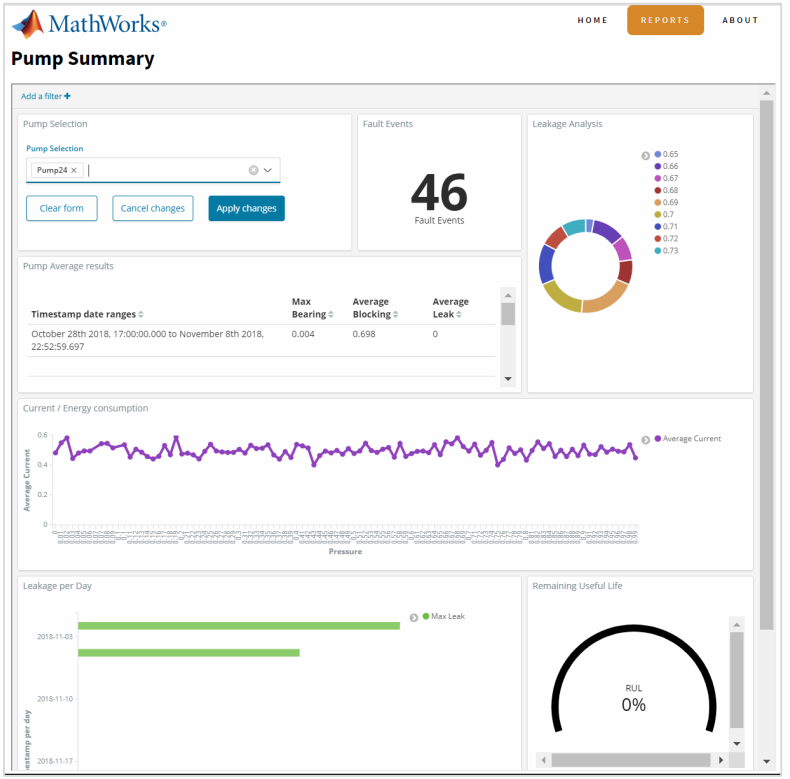
系统设计和模型部署环境设计（基于云）

Azure

运营人员（使用者）

根据模型输出做出运营决策

kibana



当前系统要求运营人员手动监控异常的指标。异常或故障的发现或预测需要专业知识的支持。


5" 12,000 PSI
5" 12,000 PSI
BAKER HUGHES

案例说明: 开发端到端的预测性维护系统

运营人员

1. 监控每个泵的流量，压力和电流，以便了解其运行状态
2. 当故障参数超出预定范围时警报，以便立即采取纠正措施
3. 持续估算每个泵的剩余使用寿命（RUL），以便安排维护或设备更换

A.I.系统部署面对的挑战

算法工程师

我们没有大量的故障数据，而且我们工厂为这个项目产生真正的故障成本太高

解决方案：使用基于泵的物理仿真模型生成仿真数据

A.I.系统部署面对的挑战

系统架构师

我们没有足够的硬件预算，但需要在投入生产环境之前进行测试并查看结果

解决方案：利用云资源配置和部署解决方案

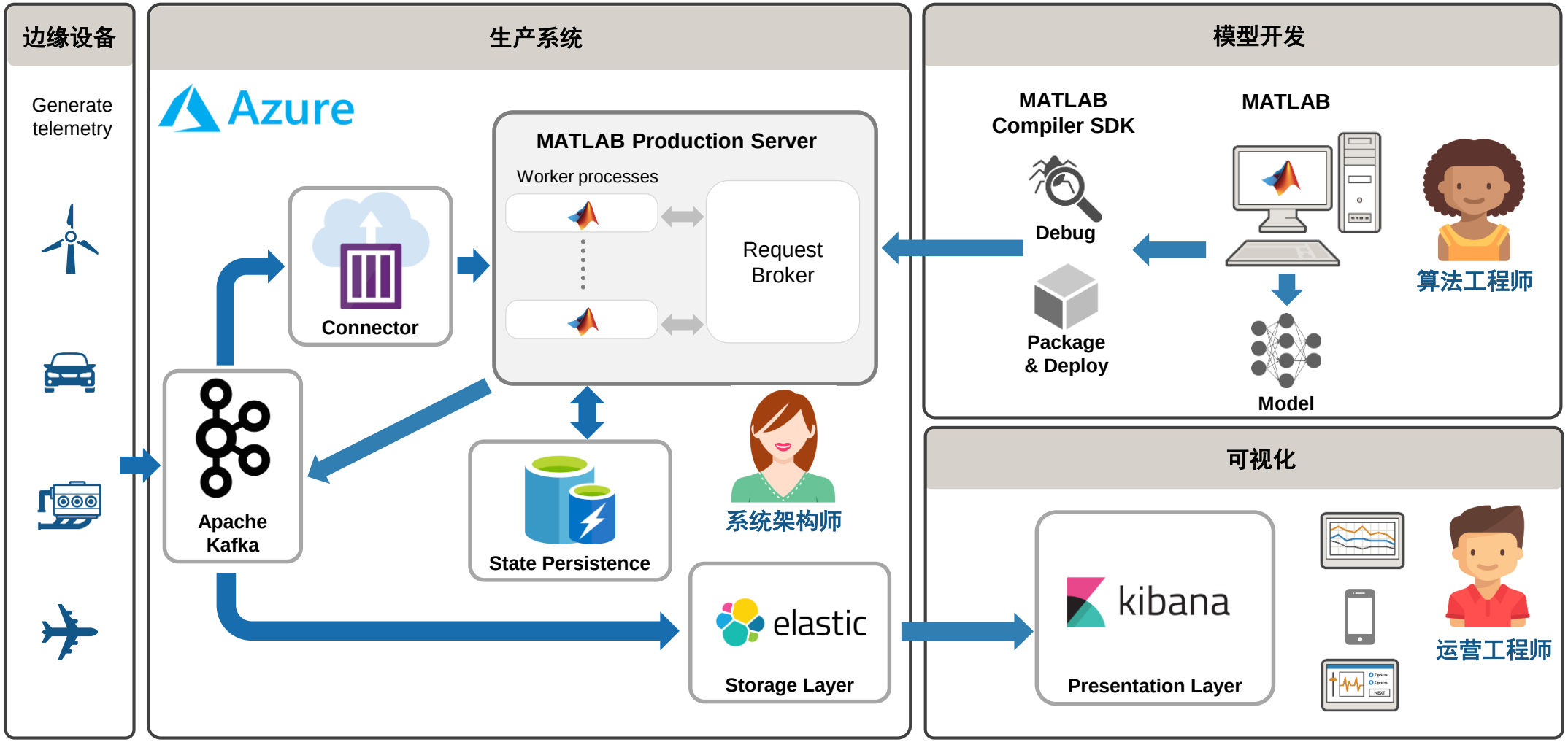
A.I.系统部署面对的挑战

算法工程师

需要跨团队的多学科问题的软件，以及与IT的集成

解决方案：使用MATLAB并与OSS企业系统集成

Azure（微软云）上的预测性维护体系结构



数据建模

算法工程师

1 数据访问

Files

Databases

Sensors

2 数据预处理

Working with Messy Data

Data Reduction/ Transformation

Feature Extraction

3 模型开发

Model Creation e.g. Machine Learning

Parameter Optimization

Model Validation

4 模型集成

Desktop Apps

Enterprise Scale Systems

kafka

AWS Kinesis

OSIsoft.

Embedded Devices and Hardware

5 结果可视化

3rd party dashboards

Spotfire® TIBCO Software

+ableau

Qlik

Power BI

Web apps

算法工程师

需求分析

来自运营人员的需求

- 可以实现故障预测与诊断
  - “Blocking”
  - “Leaking”
  - “Bearing”
  - Combination of above
- 可以预测设备的剩余寿命

来自系统架构师的需求

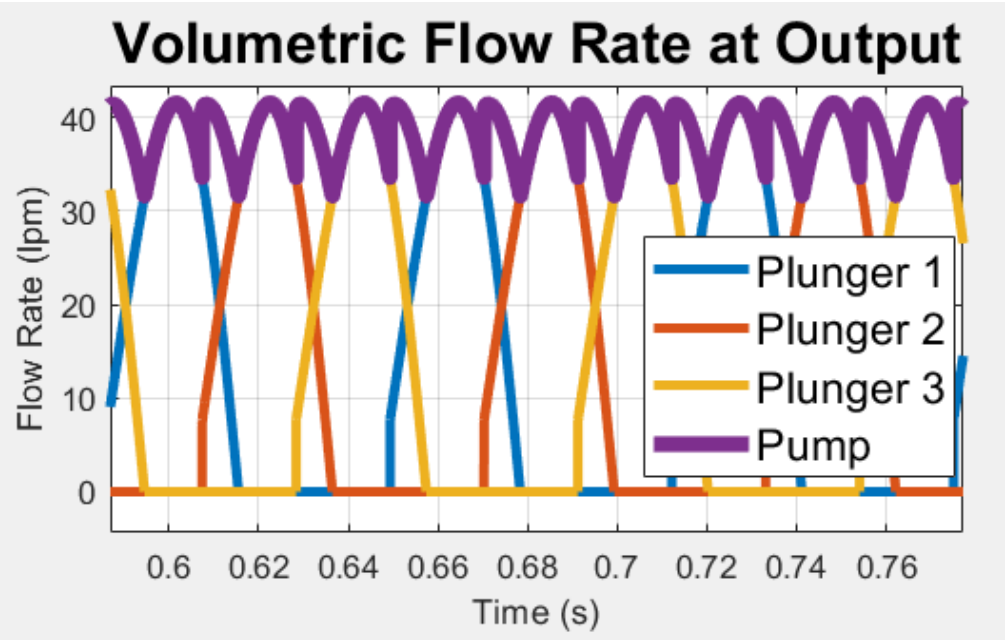
- 定义数据流的大小
- 定义结果和中间数据的数据类型和格式
- 可扩展性

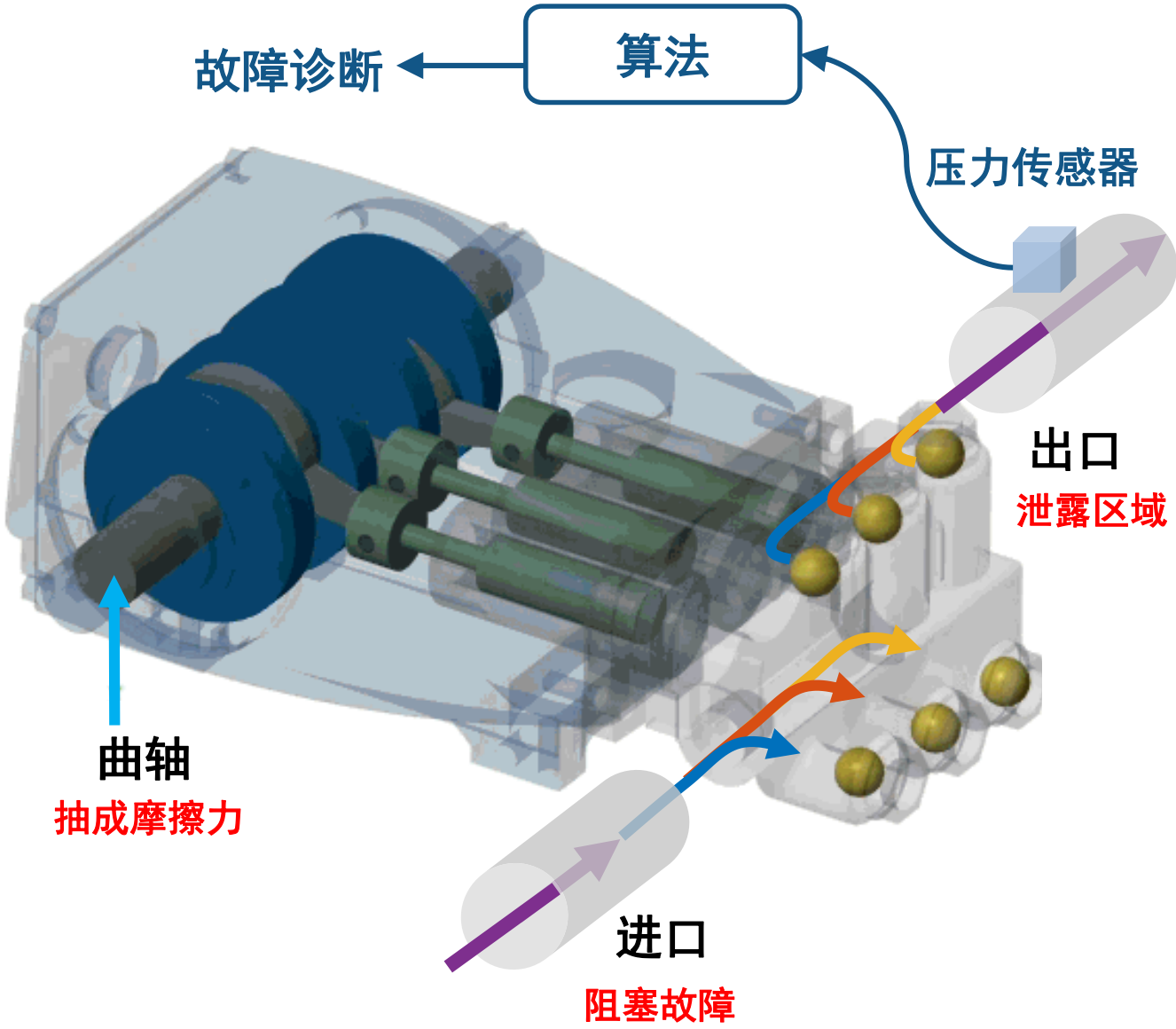

5" 12,000 PSI
5" 12,000 PSI
BAKER HUGHES

算法工程师

三缸泵（ Triplex Pump ）的仿真模型

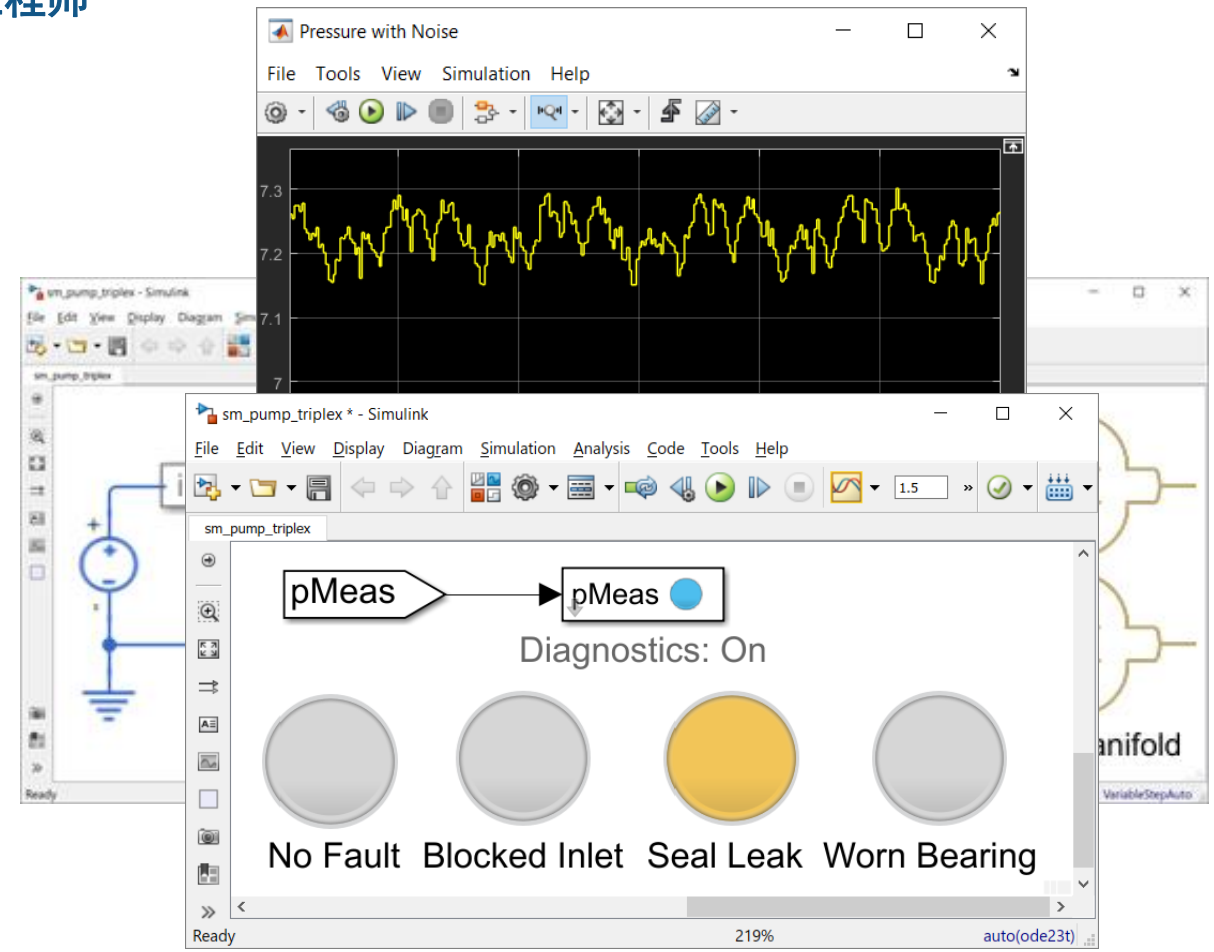
- 曲轴驱动三个柱塞
  - 每个120度异相
  - 一个腔室总是放电
  - 三种故障类型



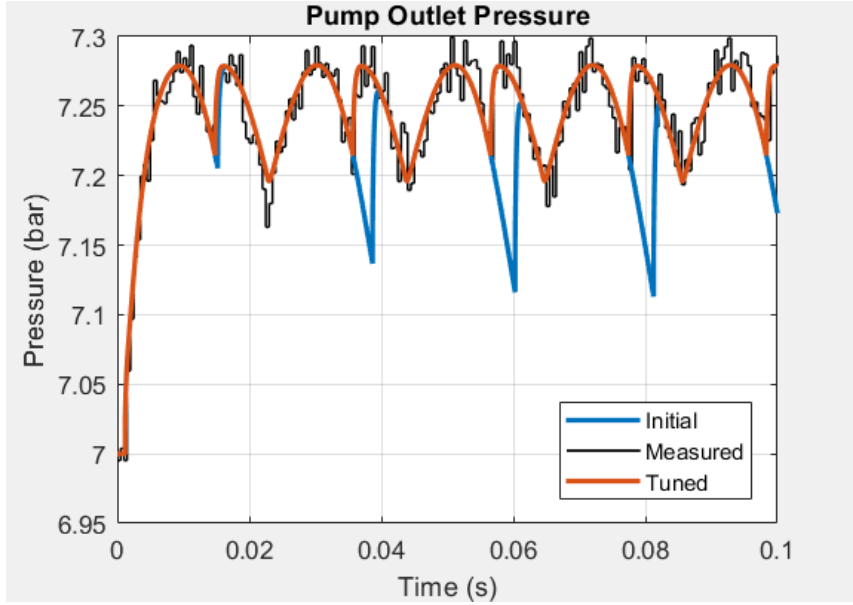


使用传感器数据识别故障

算法工程师



故障仿真



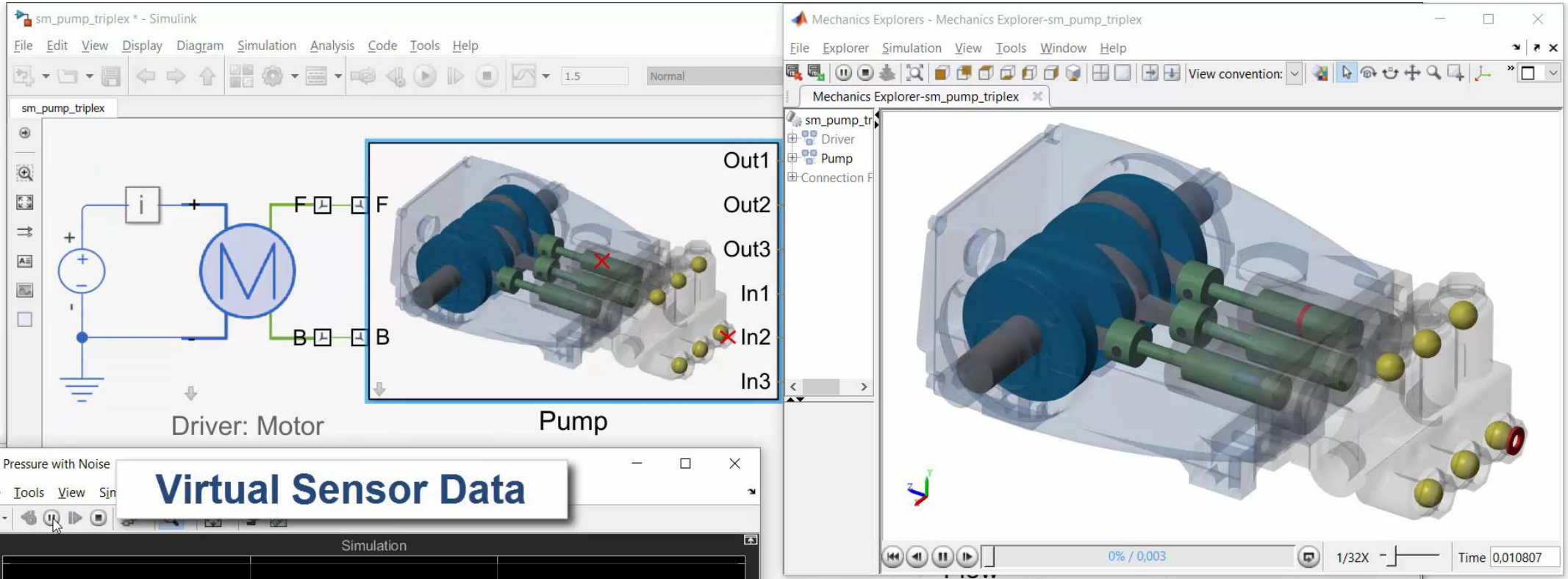
传感器数据

1

数据访问

构建数字孪生并生成数据

算法工程师

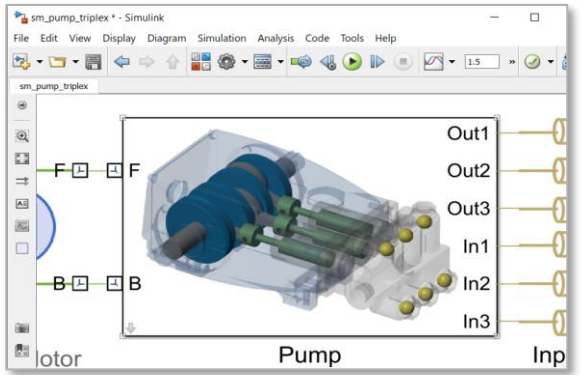


1

数据访问

算法工程师

带有故障的仿真数据生成



泄露区域 = [1e-9 0.036]

抽成摩擦力 = [0 6e-4]

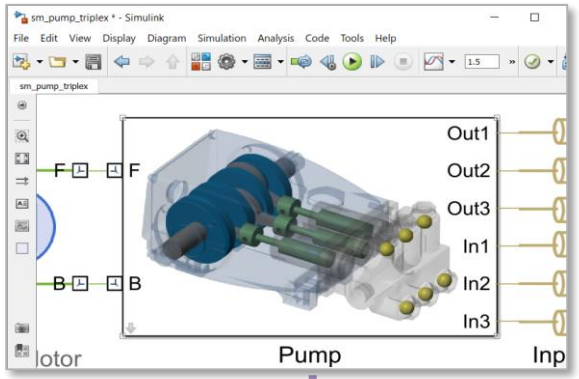
阻塞故障 = [0.5 0.8]

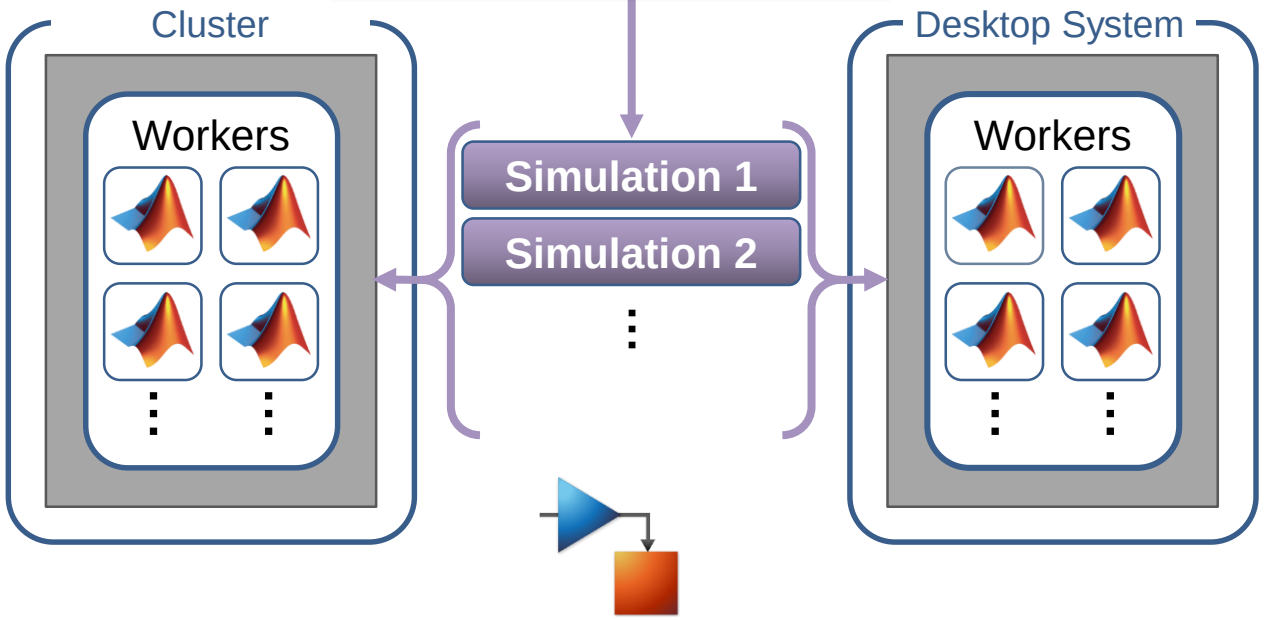
1

数据访问

带有故障的仿真数据生成

算法工程师





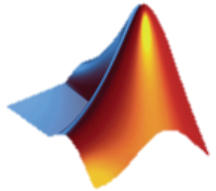
运行并行仿真

Access Data

```
ens = simulationEnsembleDatastore(location)
```

```
ens =
  simulationEnsembleDatastore with properties:

          DataVariables: [25×1 string]
    IndependentVariables: [0×0 string]
      ConditionVariables: [0×0 string]
       SelectedVariables: [25×1 string]
                ReadSize: 1
              NumMembers: 702
          LastMemberRead: [0×0 string]
                   Files: [702×1 string]
```



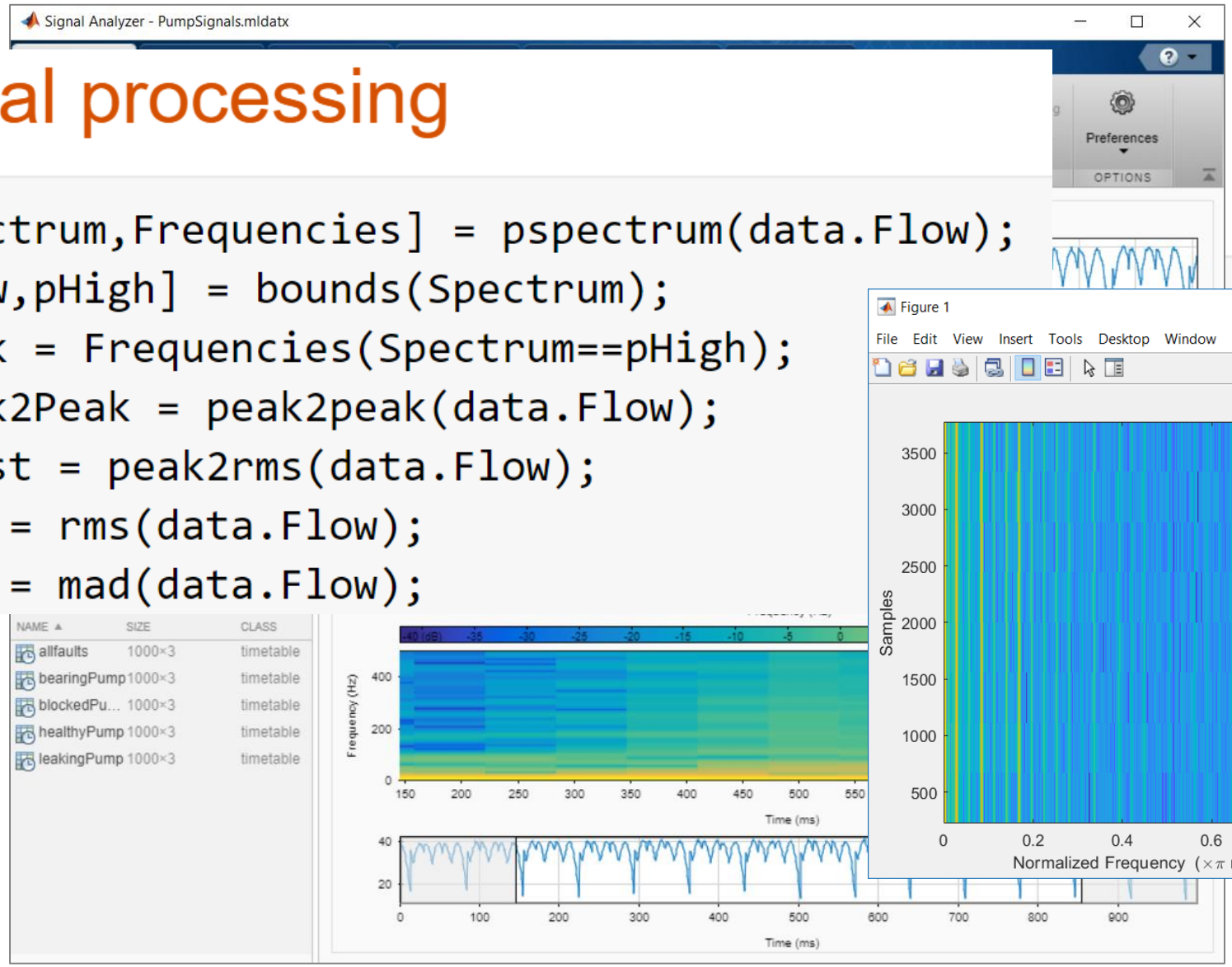
存储数据在HDFS

信号数据（时间序列）

算法工程师

Signal processing

```
[Spectrum,Frequencies] = pspectrum(data.Flow);
[pLow,pHigh] = bounds(Spectrum);
fPeak = Frequencies(Spectrum==pHigh);
qPeak2Peak = peak2peak(data.Flow);
qCrest = peak2rms(data.Flow);
qRMS = rms(data.Flow);
qMAD = mad(data.Flow);
```



3

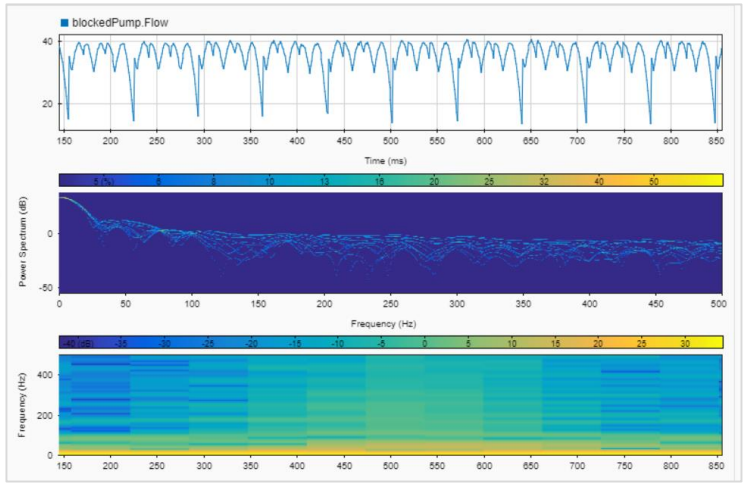
数据建模

利用MATLAB开发预测模型

算法工程师

	Time	1 LeakFault	2 BlockingFault	3 BearingFault	4 FaultType
1	0 sec	2.8472	-0.1477	1.8000	All
2	0.001 sec	-0.1498	-0.4207	1.3103	Bearing & Blocking
3	0.002 sec	0.6511	1.6521	-0.5557	Leak
4	0.003 sec	0.1469	-0.2775	1.0074	All
5	0.004 sec	-0.6480	0.7065	-0.8878	Blocking
6	0.005 sec	-0.8165	-0.5434	-0.3079	Blocking
7	0.006 sec	-1.0061	1.2083	0.0661	Bearing
8	0.007 sec	1.0125	-1.9098	-0.7027	Leak & Blocking

已标注数据

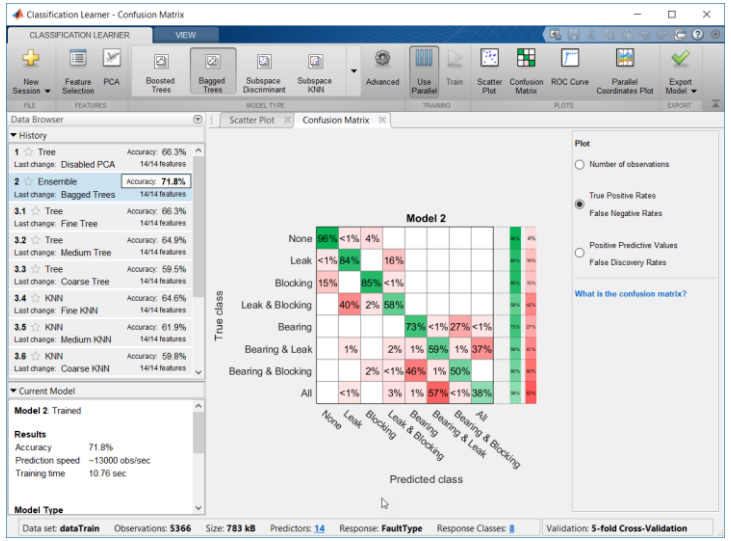


信号数据
（时间序列）

大数据处理

```
tt = tall(ds);
tt = preprocessData(tt);
model = TreeBagger(50,tt,'Event');
```

```
Evaluating tall expression using the Spark Cluster:
- Pass 1 of 2: Completed in 11 sec
- Pass 2 of 2: Completed in 2.3333 min
Evaluation completed in 2.6167 min
```



模型训练

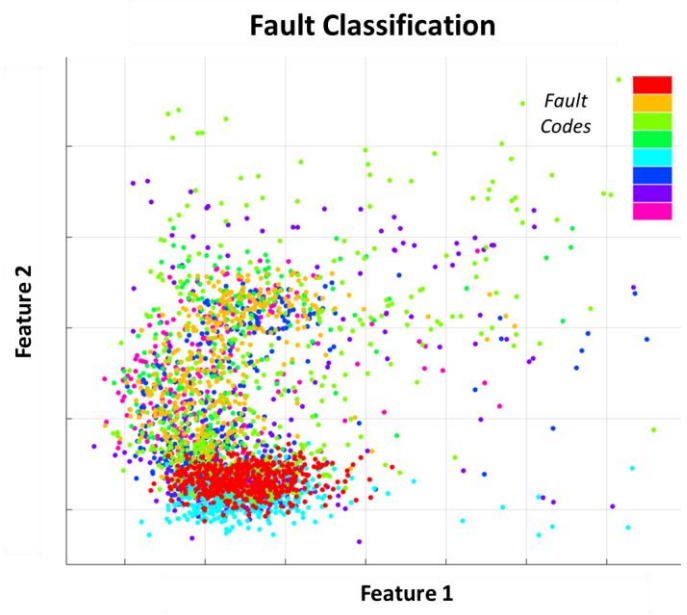
模型验证

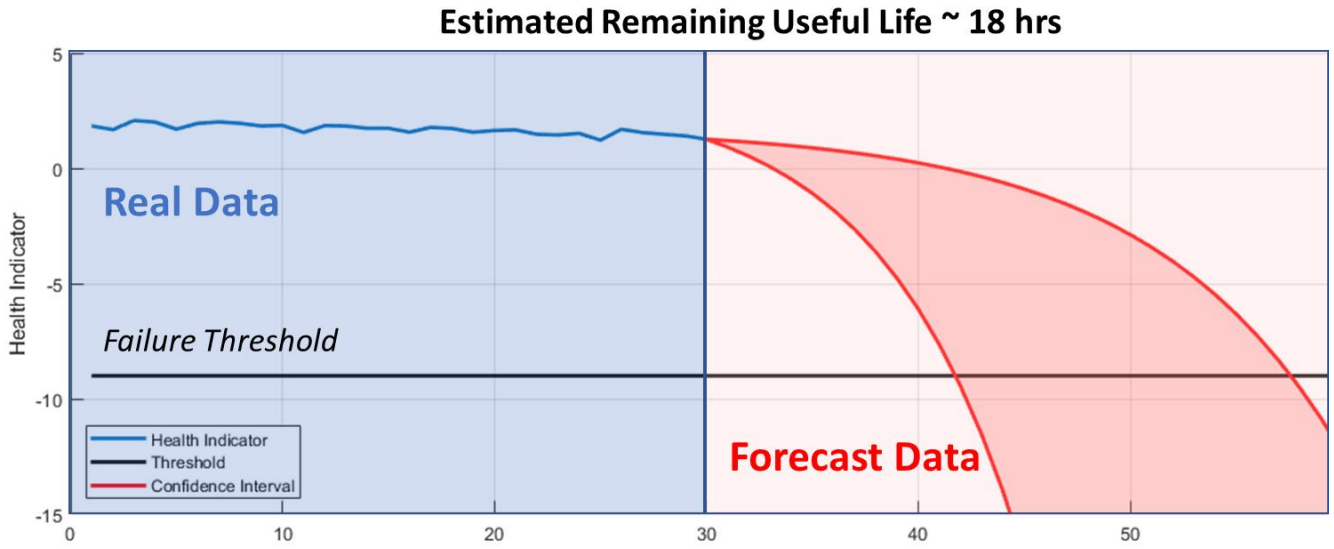
3

数据建模

利用MATLAB开发预测模型

算法工程师





故障识别
(分类)

剩余生命预测
(回归)

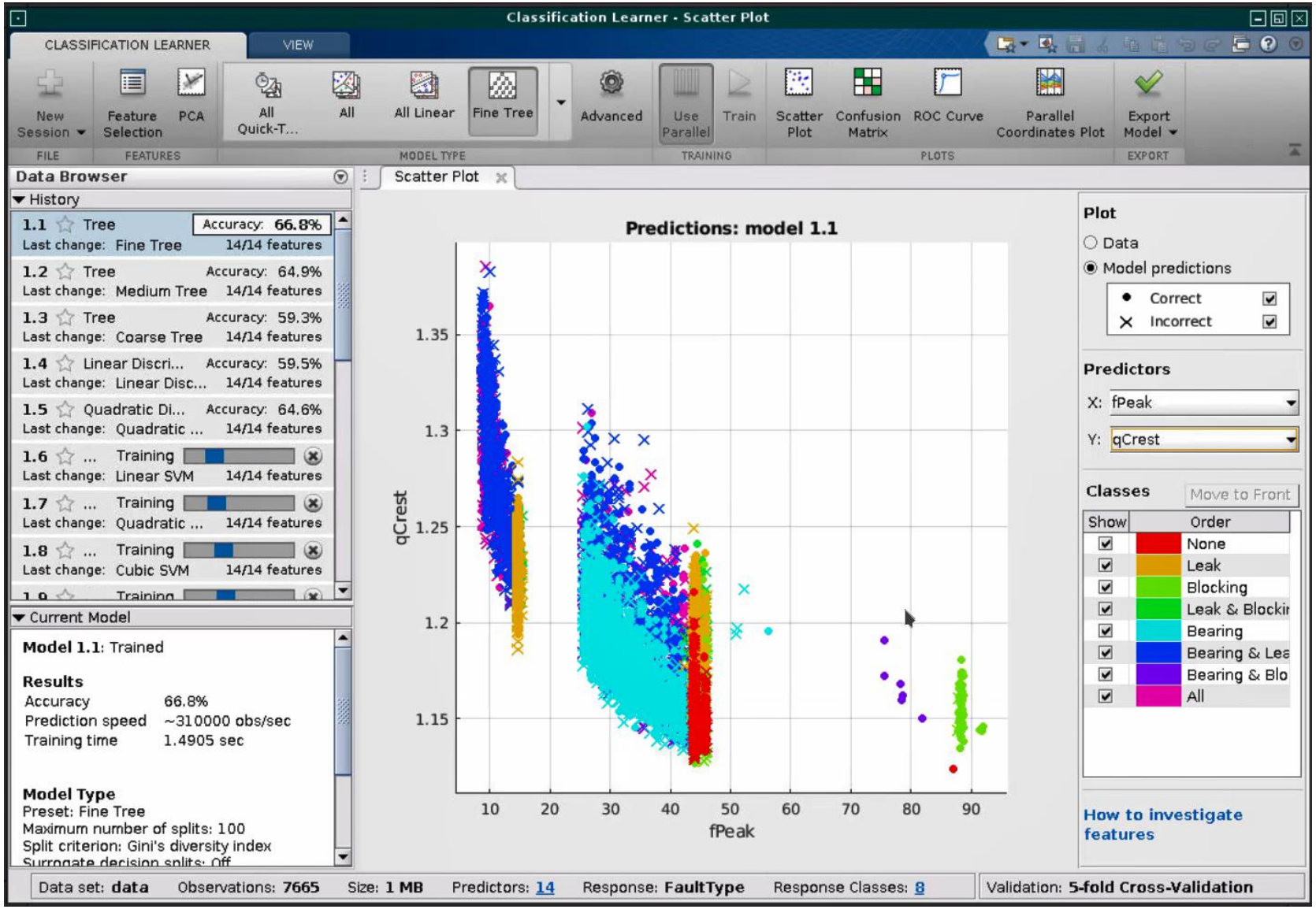
运营人员

3

数据建模

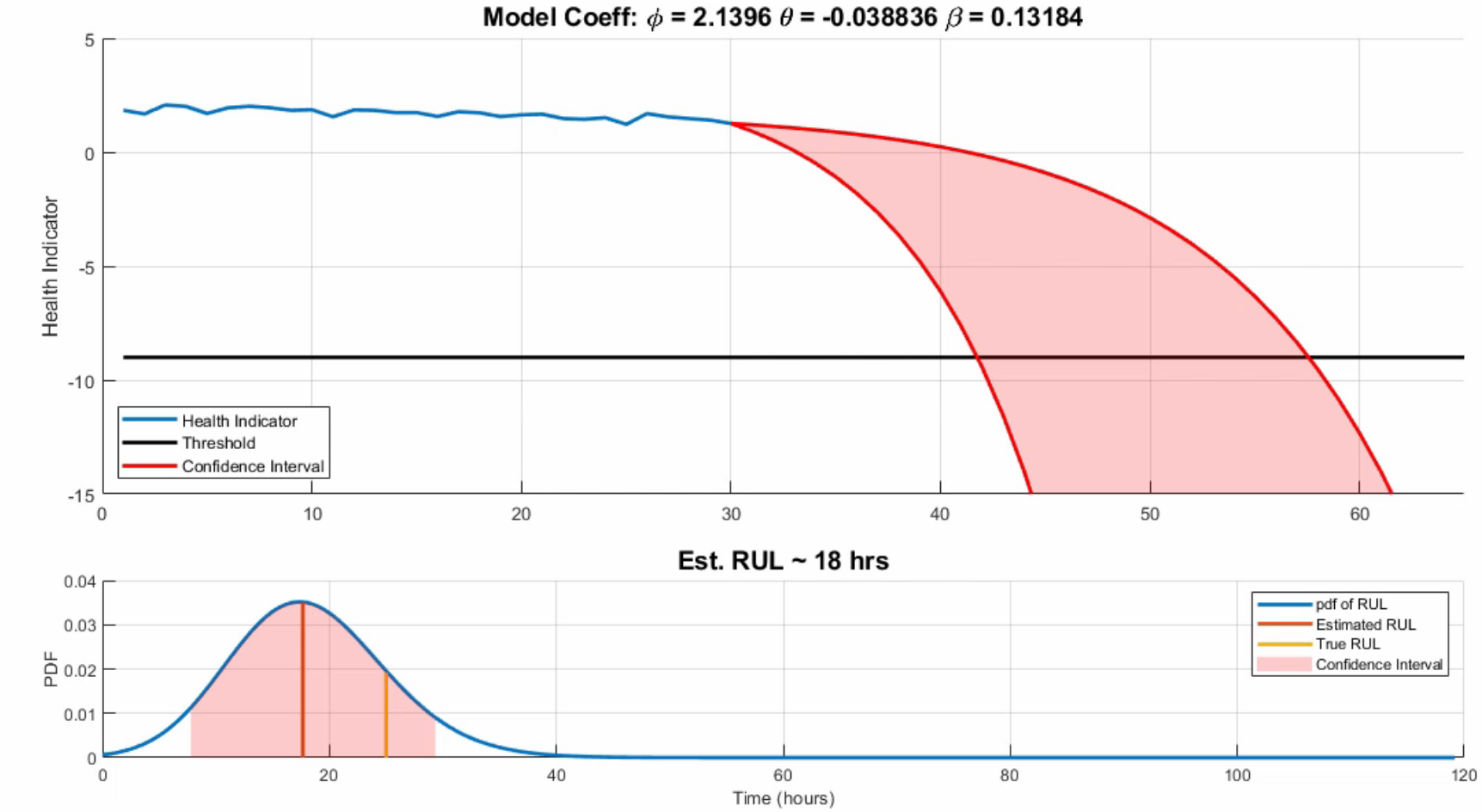
开发机器学习模型

算法工程师



算法工程师

评估剩余寿命



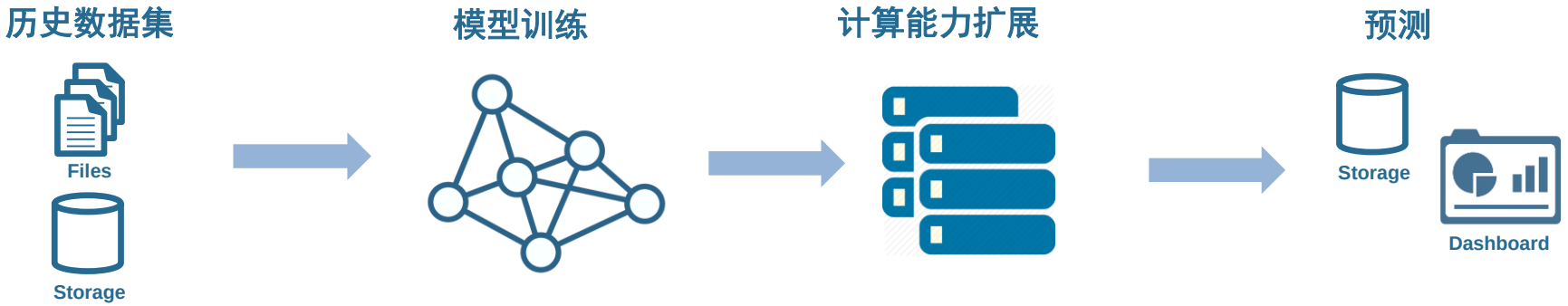
$$S(t) = \phi + \theta(t)\, e^{(\beta(t)t + \epsilon(t) - \frac{\sigma}{2})}$$

4 生产系统集成

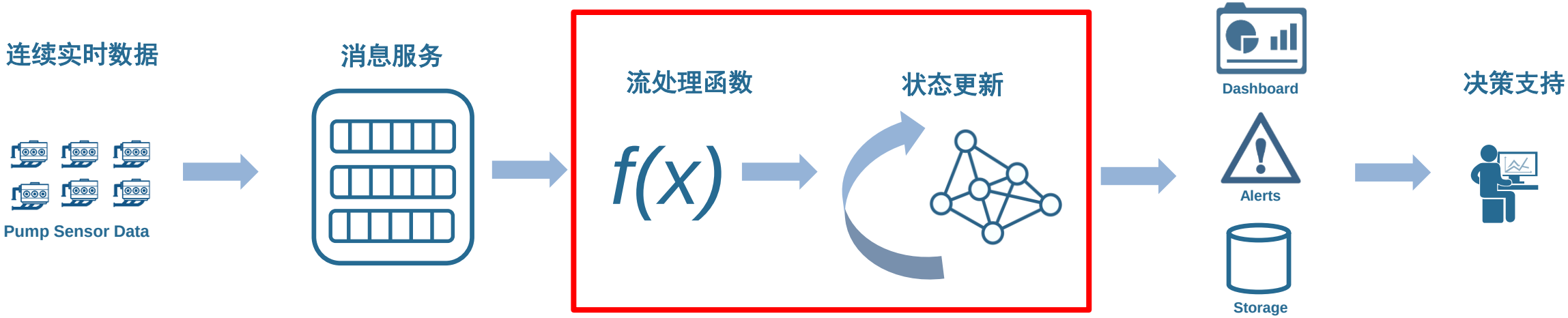
算法工程师

开发流式应用

- ***批处理:*** 模型训练与测试（历史数据集）



- ***流式处理:*** 模型应用（近实时传感器数据）



4 企业系统集成

算法工程师

开发流式应用

Streaming Function

```
function new_state = streamingFunction(data,old_state)
```

Preprocess signals

```
[data,features] = preprocessData(data);
```

Predict faults

```
[Leak,Blocking,Bearing] = predictFaultValues(features);
FaultType = predictFault(features);
[RUL,Model] = predictUpdateRUL(data.Timestamp,data.Flow,500);
```

Update state

```
new_state = updateState(data,old_state);
```

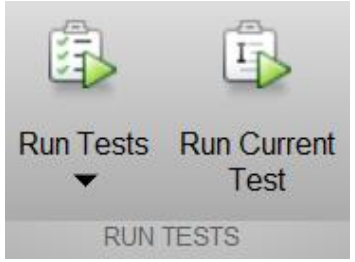
Write results

```
writeResults(Leak,Blocking,Bearing,FaultType,RUL,Model)
end
```

4 企业系统集成

算法工程师

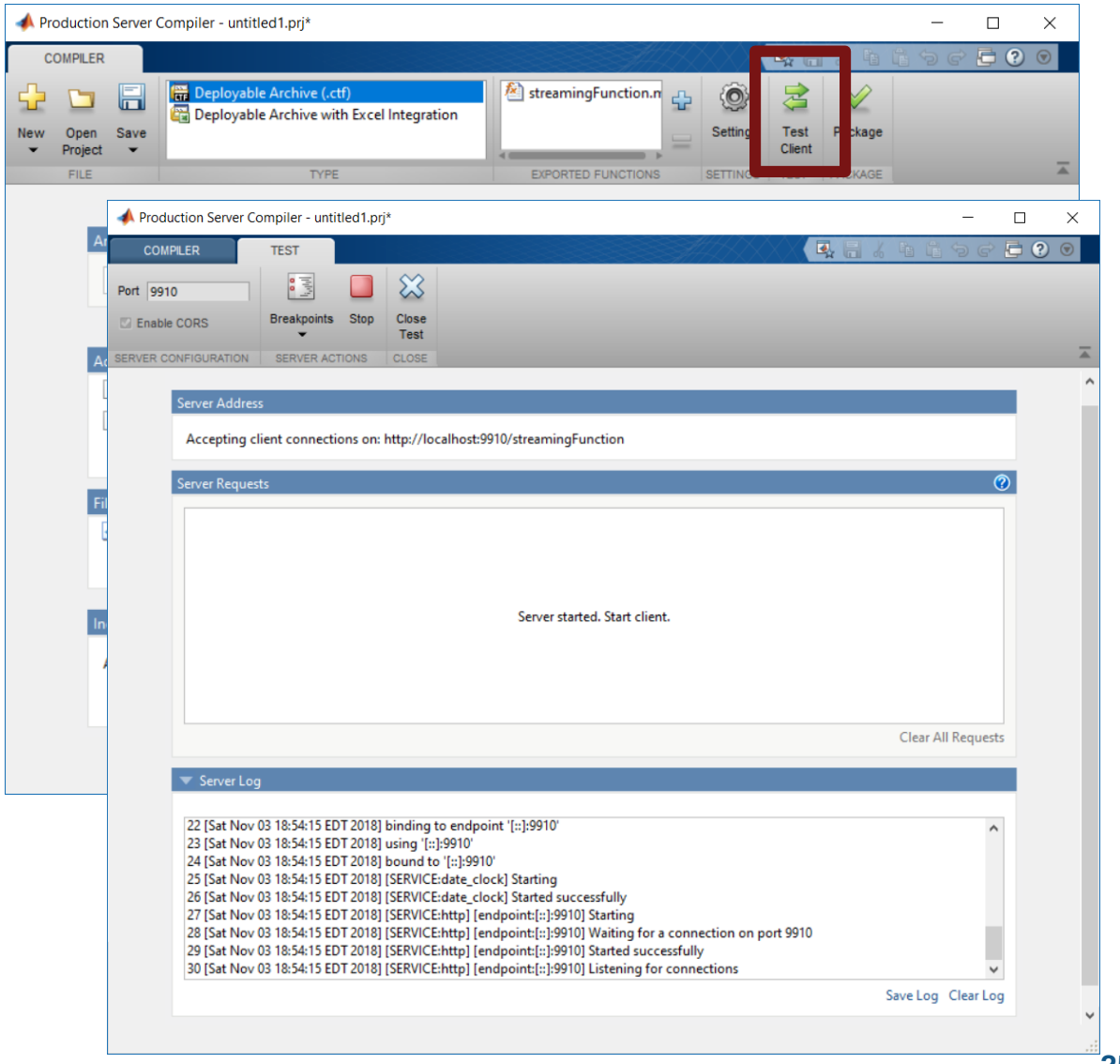
测试流式处理功能




```
results = runtests('predictFaults_tests')
```

```
Running predictFaults_tests
....
Done predictFaults_tests
__________
results =
  1×4 TestResult array with properties:

    Name
    Passed
    Failed
    Incomplete
    Duration
    Details
Totals:
   4 Passed, 0 Failed, 0 Incomplete.
   0.01614 seconds testing time.
```

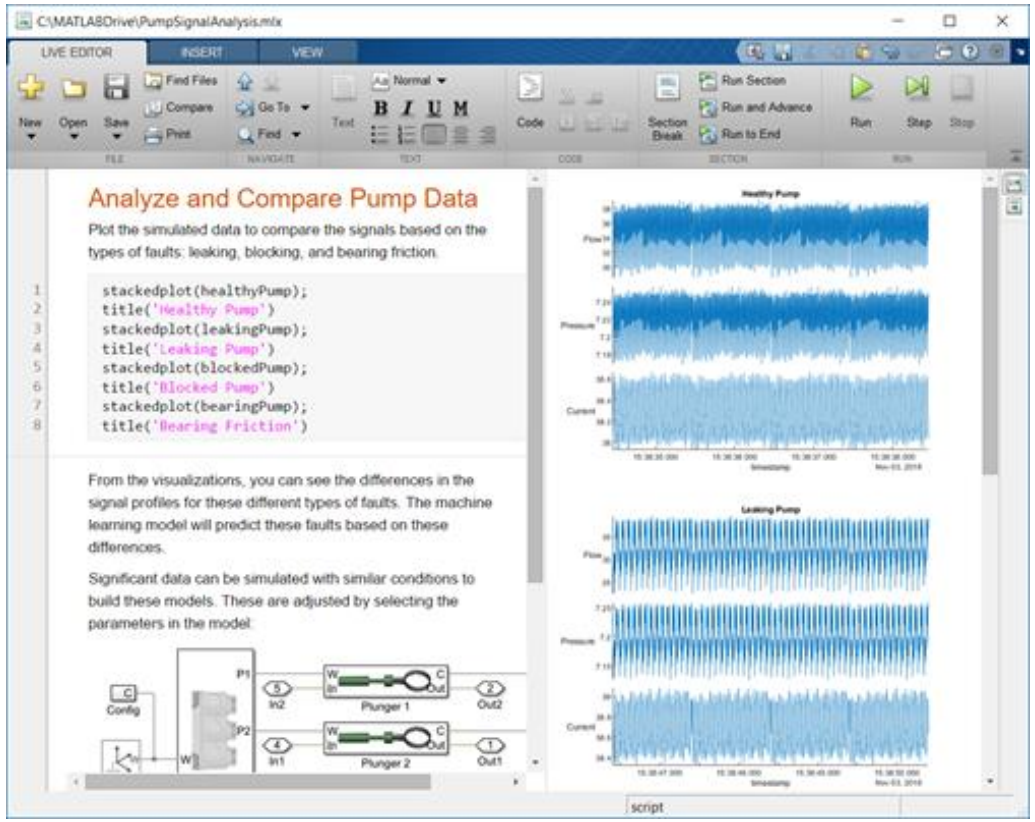


4 企业系统集成

与团队成员分享

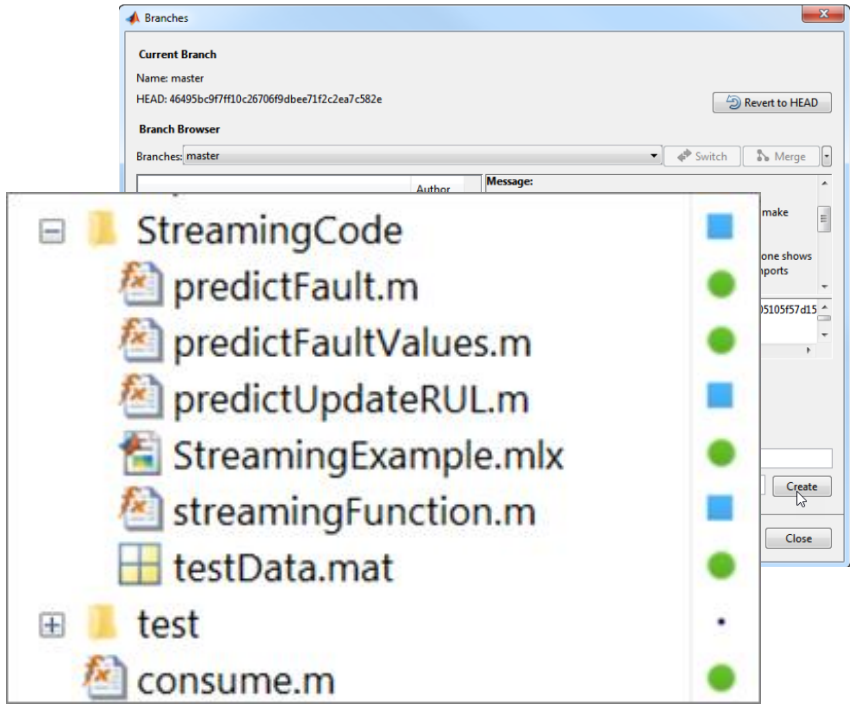
算法工程师

运行人员查看运行过程



.pdf, html, LaTeX

与系统架构师共享代码



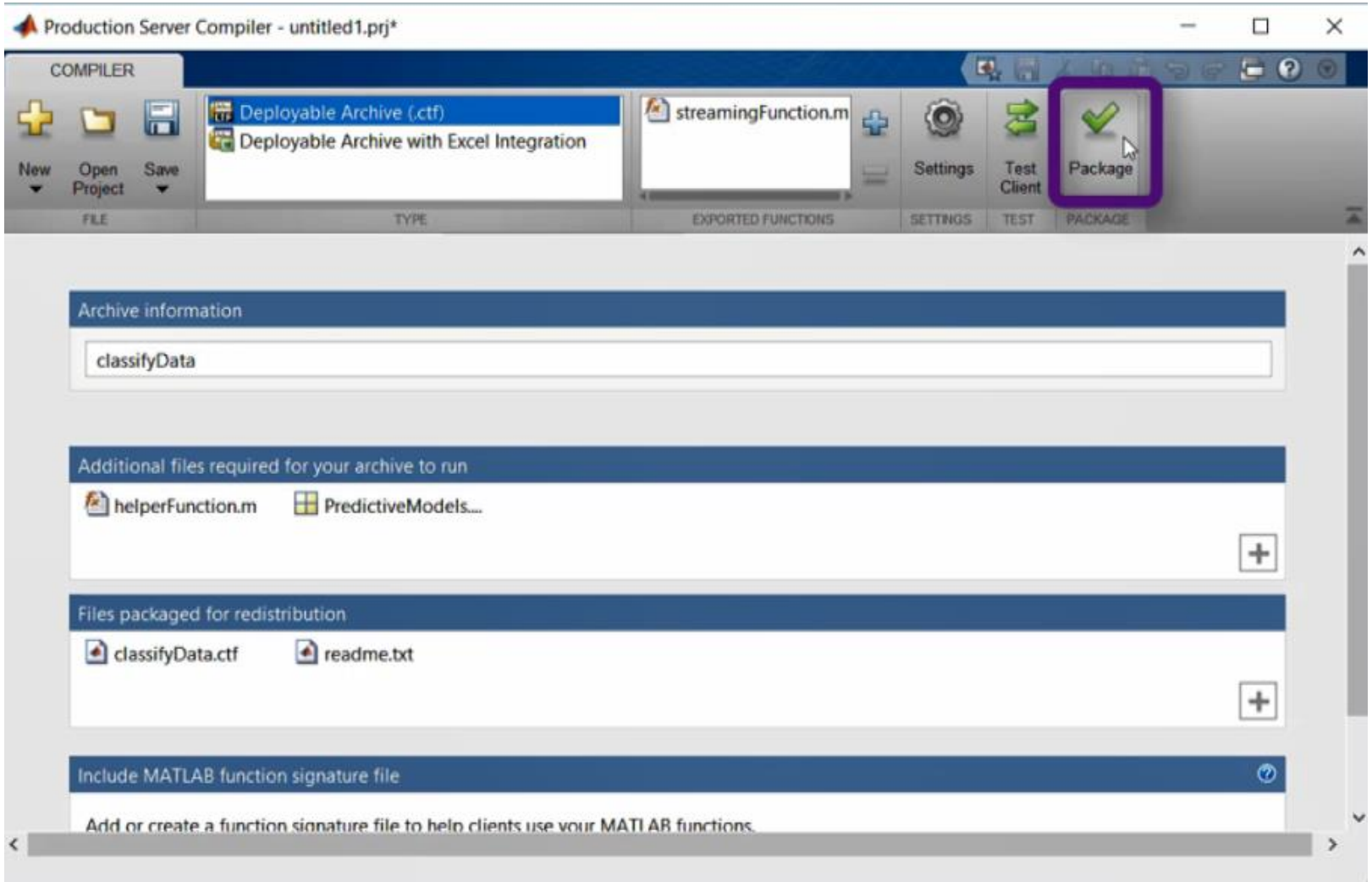
Source Control

4

企业系统集成

功能打包

算法工程师



系统架构师

4 企业系统集成

系统需求分析

- 算法工程师需求
  - 每毫秒，每个泵产生一组流量、压力和电流数据
  - 模型的数据窗口期是1秒
  - 初始是1到10台设备，但很快会扩展到100台

算法工程师

- 运营人员需求
  - 参数异常告警
  - 泵体剩余使用寿命的评估

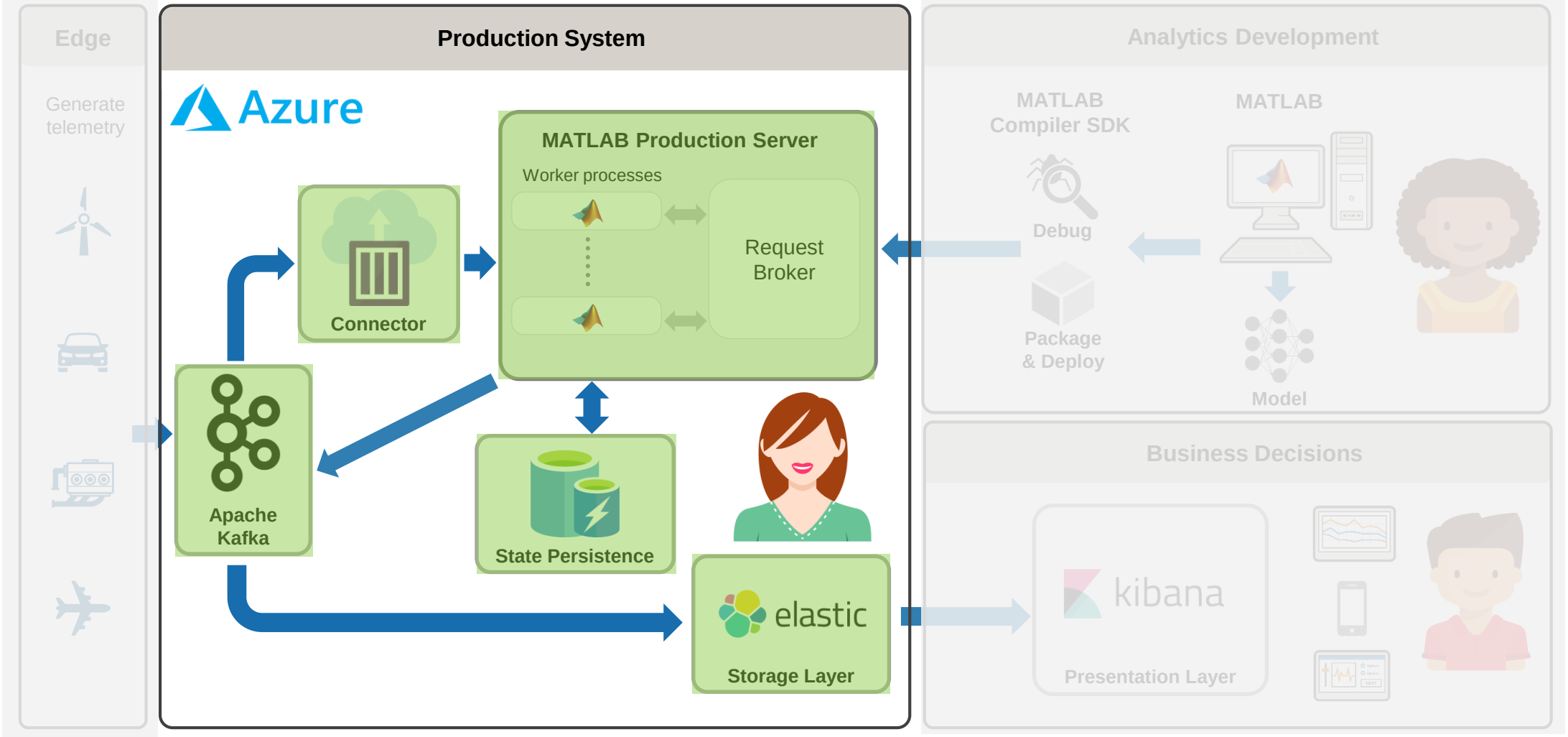
运营人员

4

企业系统集成

企业系统集成

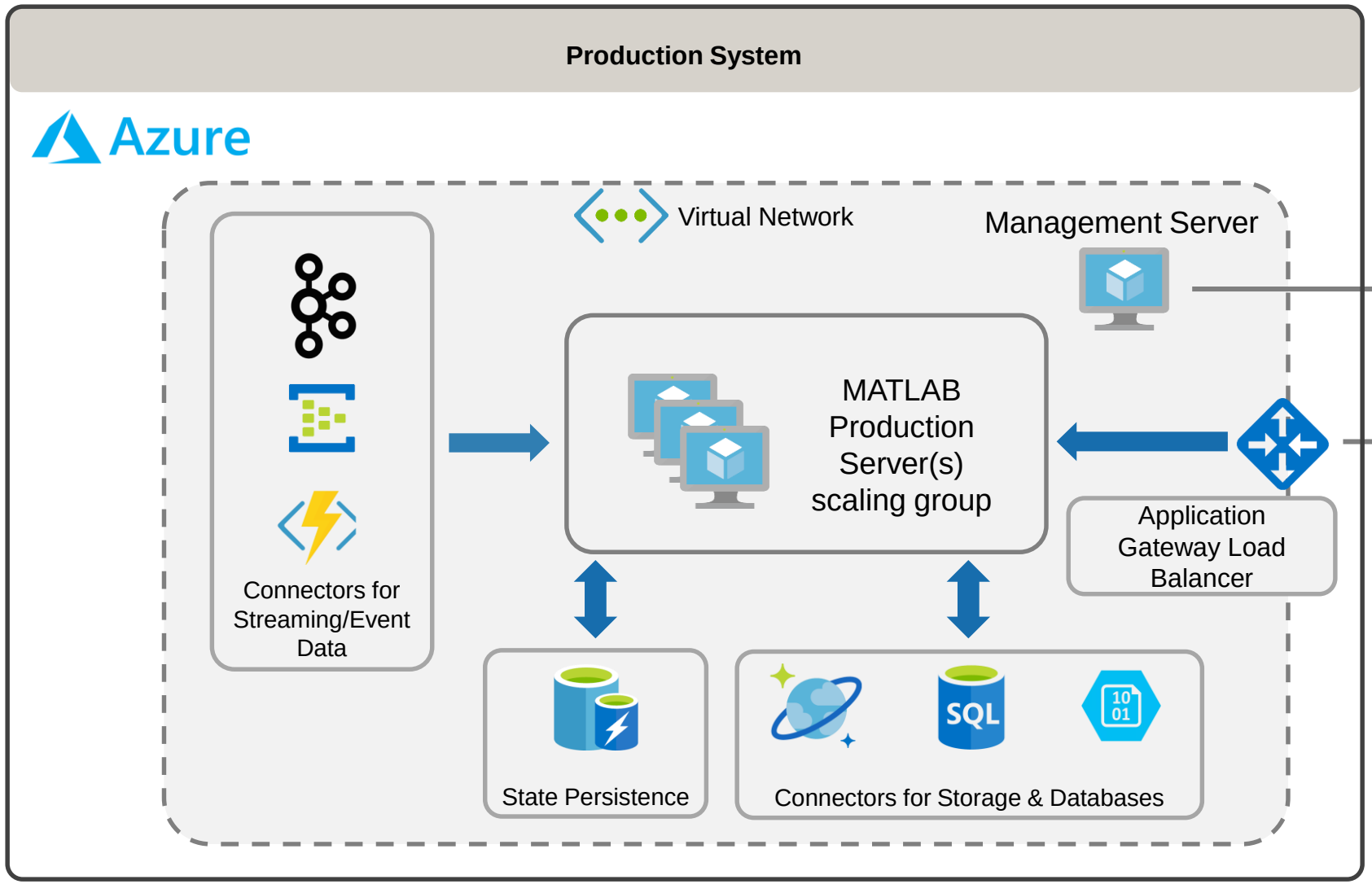
系统架构师

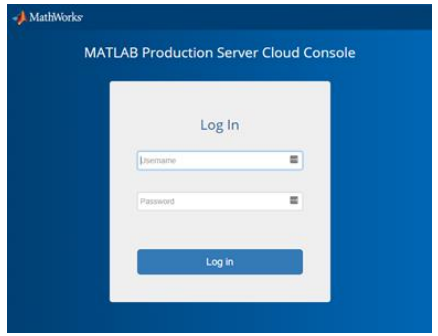


系统架构师

4 企业系统集成

基于Azure（微软云）的MATLAB Production Server





管理界面

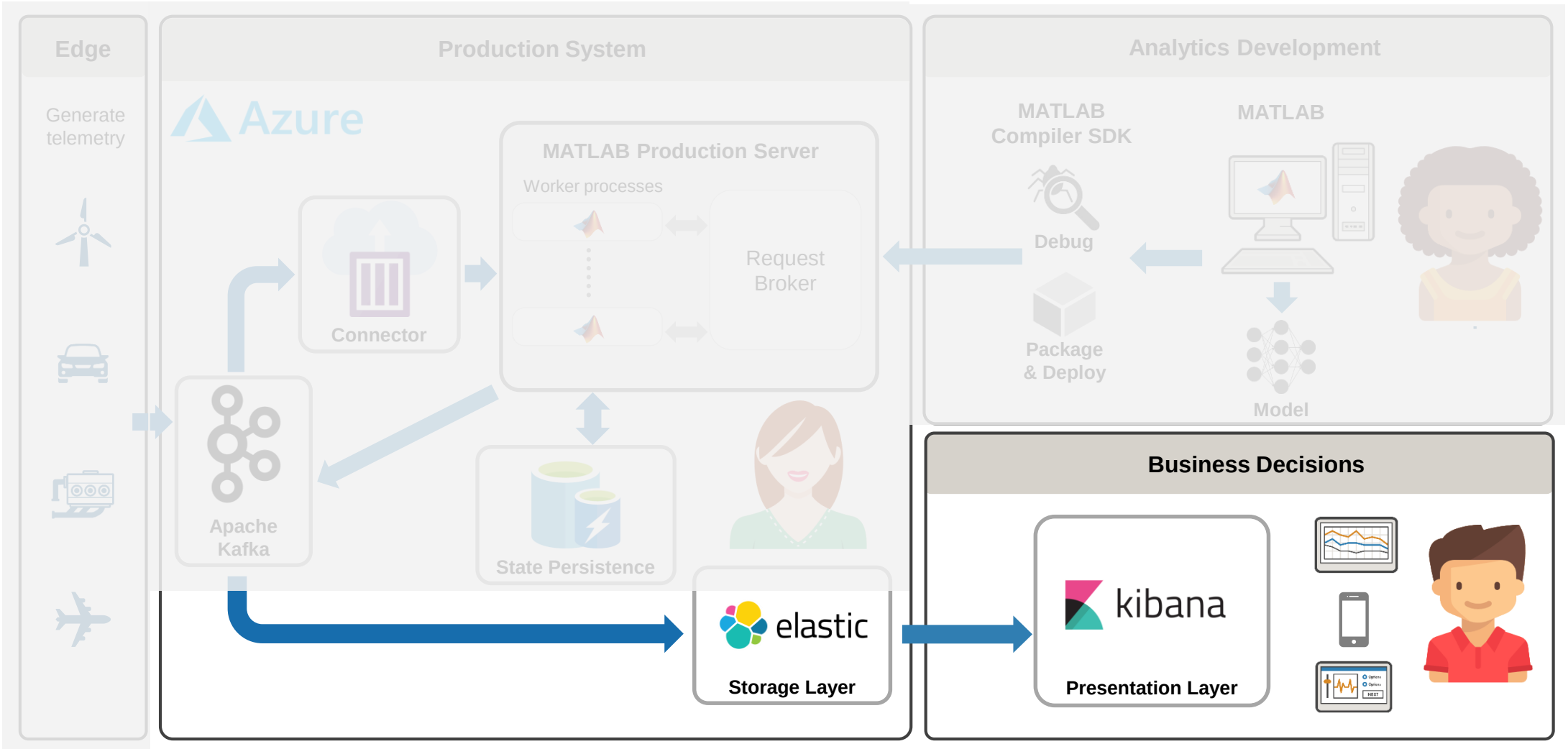
企业应用

4

企业系统集成

应用结果可视化

系统架构师



运营人员

5

结果可使唤

以可视化的方式显示模型预测结果