

An IIoT Platform for Agile Development and Deployment of Data-Driven Solutions in E&P*

Tailai Wen¹ and Xinwo Huang¹

Search and Discovery Article #42387 (2019)**

Posted July 15, 2019

*Adapted from oral presentation given at 2019 AAPG Annual Convention and Exhibition, San Antonio, Texas, May 19-22, 2019

**Datapages © 2019 Serial rights given by author. For all other rights contact author directly. DOI:10.1306/42387Wen2019

¹Arundo Analytics, Houston, Texas (tailai.wen@arundo.com)

Abstract

In the recent decade, upstream companies' investment in digital transformation and data-driven analytics increased rapidly. Machine learning is expected to bring enhanced insights into E&P workflows that are traditionally based on physics and engineering knowledge. However, the majority of data science projects did not go beyond research papers and few were put in constant operation in actual business processes. Data scientists in upstream companies spend most their time on their local computing environments, which are isolated from real-time data and real users. This gap obstructs continuous testing and improving of data science solutions in real-world scenarios.

Agile methodology is a widely-used product development workflow in Silicon Valley. It emphasizes the importance of continuous feedback from end users and fast cycles of product operationalization. We built an IIoT (Industrial Internet of Things) platform that aims at enabling agile development and deployment of data science solutions in E&P. The platform allows data scientists to operationalize machine learning solutions with minimal IT supports, and it accelerates the feedback-improvement iterations between end users and data scientists effectively.

Core components of this integrated platform includes:

- 1) Edge agent: a software installed on assets (e.g. equipments, sensors, IoT devices) that standardizes real-time data streams. It also hosts models of edge computing which are usually for smart processing of raw data.
- 2) Cloud-based model hub: the center of entire system that connects data, models, and applications. The hub automates enterprise-scale model management (e.g. resourcing, CI/CD, version management). Containerization technology plays a key role here.
- 3) Model deployment software: a tool that enables data scientists to publish models from local environment to model hub and/or to edge agent. Data scientists may operationalize a new solution such that it is ready to end users in just a few minutes.

4) End-user application: it includes flexible connector to common BI tools, a web-based dashboard builder, and an automated web-based interface renderer.

We tested the platform with our industrial partners in real-world E&P environments. We will highlight a couple of geology-related cases: one on automated well log interpretation, and the other on machine learning assisted seismic imaging.

ARUNDO

**An IIoT Platform for Agile Development and Deployment of
Data-Driven Solutions in E&P**

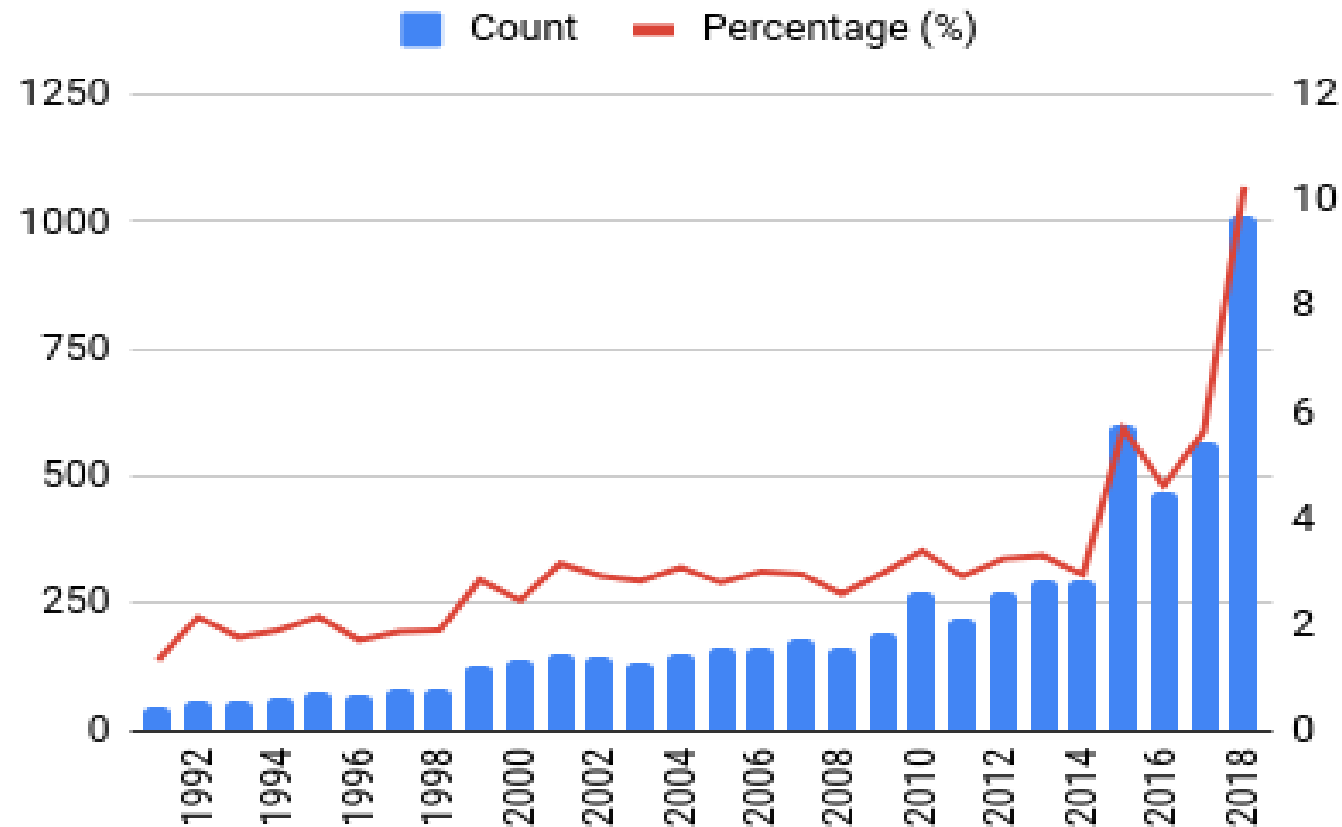
Dr. Tailai Wen*, Dr. Xinwo Huang

May 22, 2019

AAPG ACE @ San Antonio, TX

More Attention to Machine Learning / Artificial Intelligence in O&G

Count of papers on machine learning in OnePetro



What is Missing?

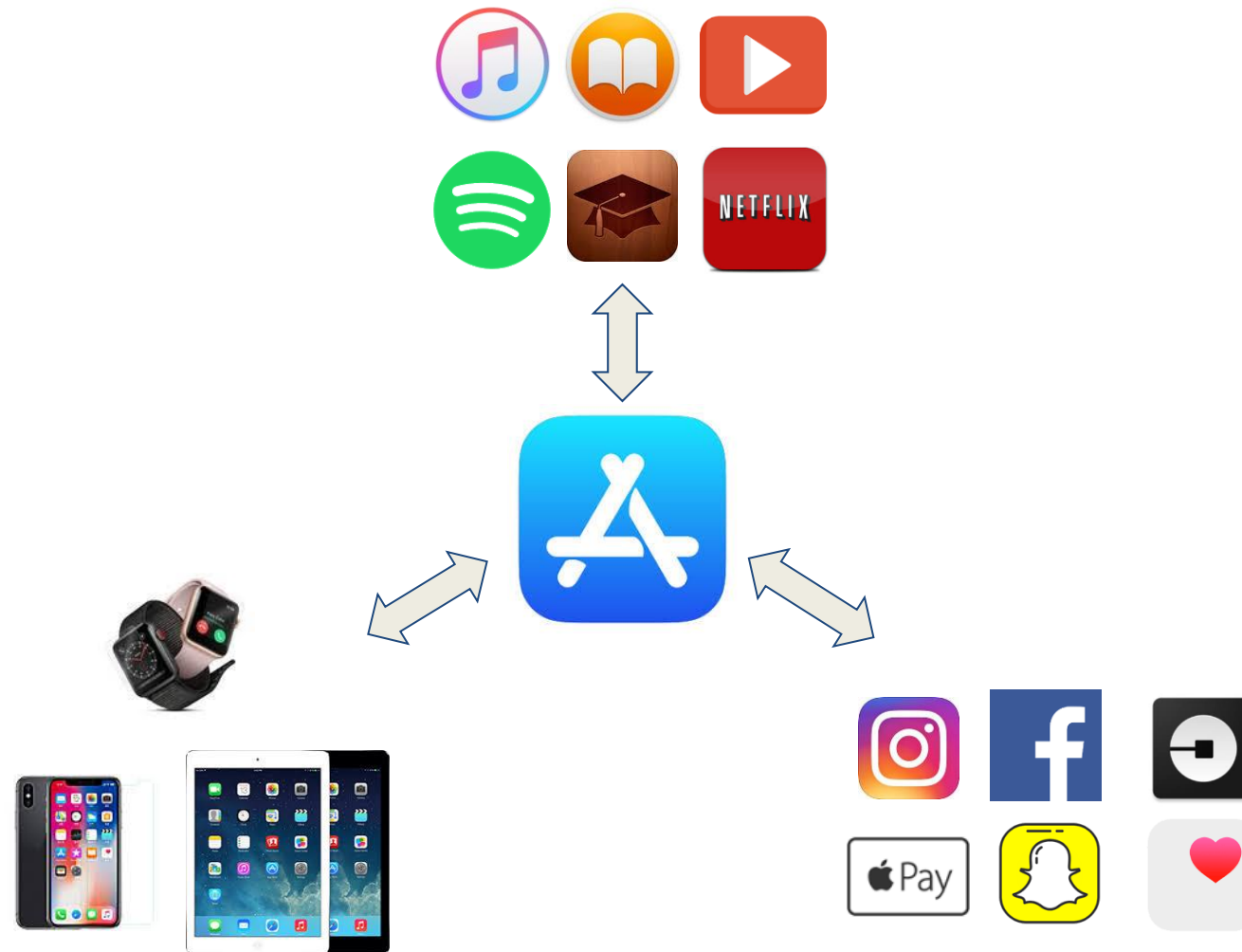
McKinsey Global Institute interview with 2,135 executives

“What is the biggest hurdle for you to achieve digital excellence?”



Source: McKinsey

Transformation of Ecosystem is Initialized by New Platform



Infrastructure vs. Platform vs. Software

Cloud Clients

Web browser, mobile app, thin client, terminal emulator, ...



Software as a Service (SaaS)

CRM, Email, virtual desktop, communication, games, ...

Platform as a Service (PaaS)

Execution runtime, database, web server, development tools, ...

Infrastructure as a Service (IaaS)

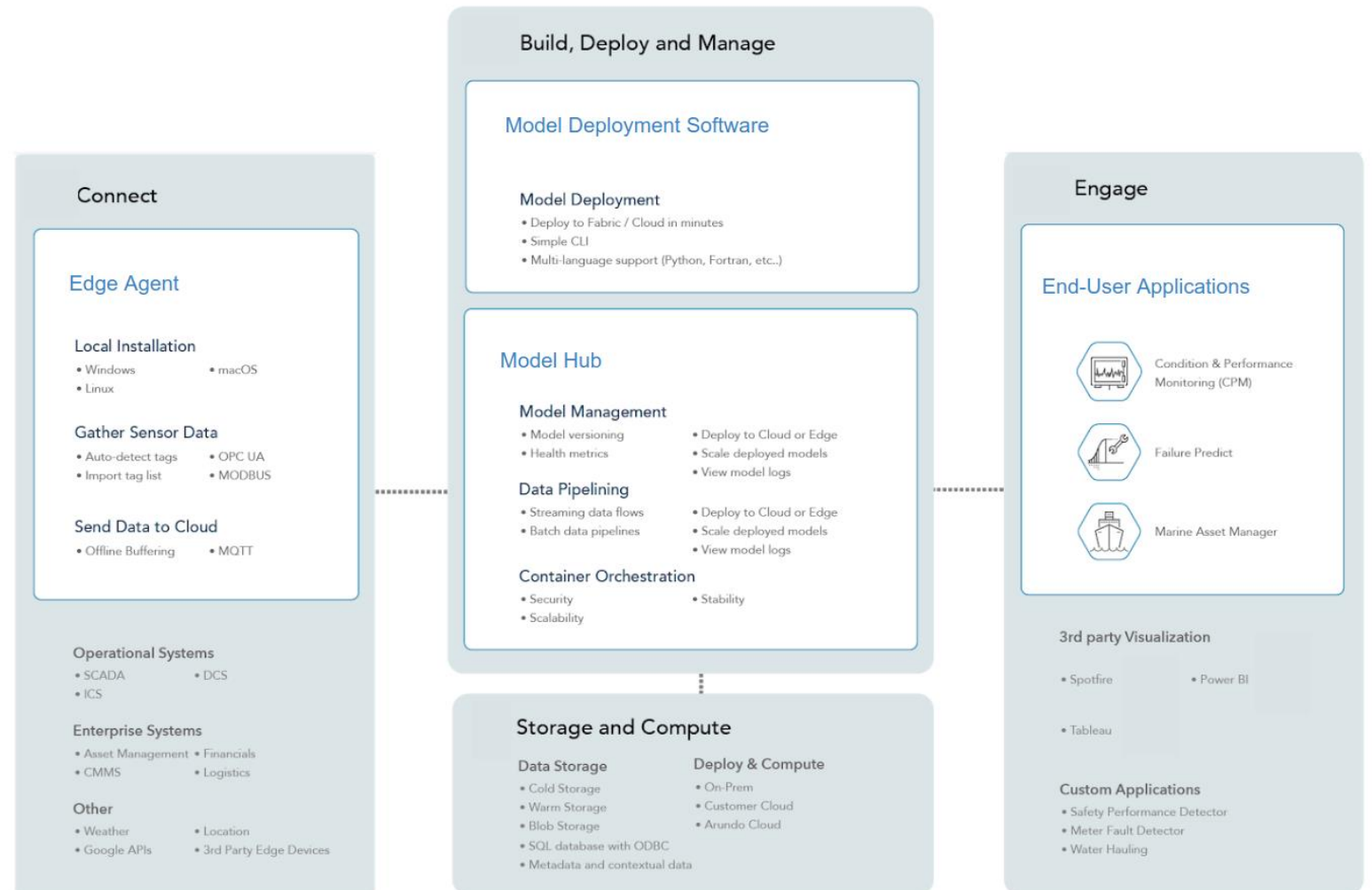
Virtual machines, servers, storage, load balancers, network, ...

Source: Wikipedia

An Integrated Platform for IIoT

Four key components:

- Cloud-based model management hub
- Edge agent software
- Model deployment software
- End-user interface



Edge Agent for Data Streaming



Industrial Connectors	Edge Agent	Microsoft IOT Hub
<ul style="list-style-type: none">→ Supports up to 10 hz frequency from today’s main control system interfaces→ Easily browse for available tags or upload large csv files with tag information→ Build a compute chain to modify or scale any raw data→ Timestamping can be handled by the control system or edge agent	<ul style="list-style-type: none">→ Buffer offline data for days (configurable) and push to the cloud when internet connectivity is established→ Report health of the system so you can monitor it remotely→ Build compute chains to add virtual sensors to your stream→ Use a simple viewer to check current streaming data	<ul style="list-style-type: none">→ Out of the box support for Microsoft’s IOT Hub→ Stream to your own IOT Hub or Arundo’s→ Cloud connection can be configured in less than a minute→ Ability to configure “at least once” delivery confirmation→ In poor bandwidth situations data sample rates can be changed

Model Deployment and Management

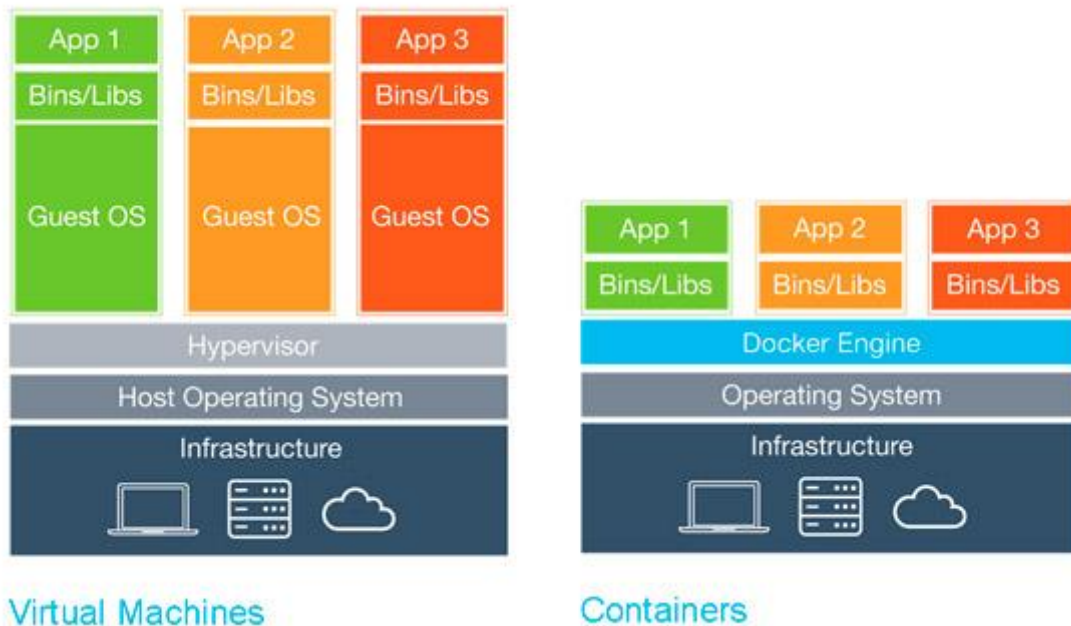
- **Model hub connects data, models, and applications**
- Each model is a SaaS deployed at a docker container
- Model deployment software is an interface for data scientists to publish their models as an API (application programming interface)
- Model hub is fully responsible for enterprise-level management work (resourcing, CI/CD, version controls, scheduling, performance monitoring, etc.)

Model Deployment and Management

- Model hub connects data, models, and applications
- **Each model is a SaaS deployed at a docker container**
- Model deployment software is an interface for data scientists to publish their models as an API (application programming interface)
- Model hub is fully responsible for enterprise-level management work (resourcing, CI/CD, version controls, scheduling, performance monitoring, etc.)

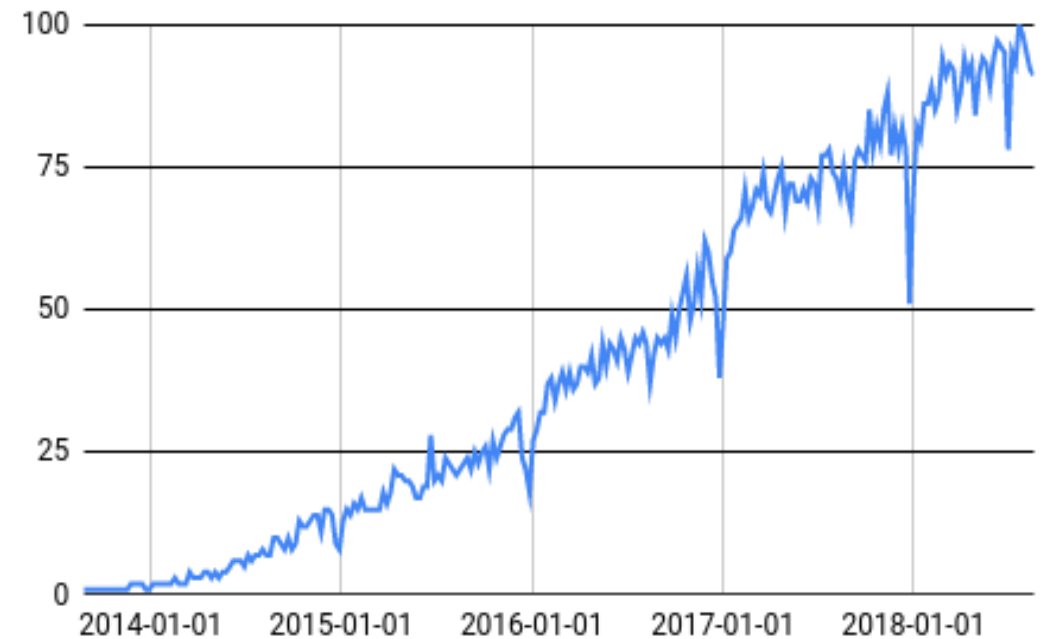
Containerization

- Operating-system-level virtualization technology that has emerged in recent year
- Reduced resource usage, faster, cheaper and more scalable
- Isolated environments, more secure



Source: Docker.com

Google searches of “docker container” (normalized)



Model Deployment and Management

- Model hub connects data, models, and applications
- Each model is a SaaS deployed at a docker container
- **Model deployment software is an interface for data scientists to publish their models as an API (application programming interface)**
- Model hub is fully responsible for enterprise-level management work (resourcing, CI/CD, version controls, scheduling, performance monitoring, etc.)

Model Deployment

- A Python library to wrap a machine learning model into a deployment-ready model

```
10
11 @endpoint()
12 @argument("demoSensor_01", type=float)
13 @argument("demoSensor_02", type=float)
14 @returns("status_flag", type=float)
15 def equipment_status_predictor(demoSensor_01, demoSensor_02):
16     '''
17     Input the reading of demoSensor_01 and demoSensor_02 to pr
18     '''
19
20     status_flag = mySVC.predict([[demoSensor_01, demoSensor_02
21
```

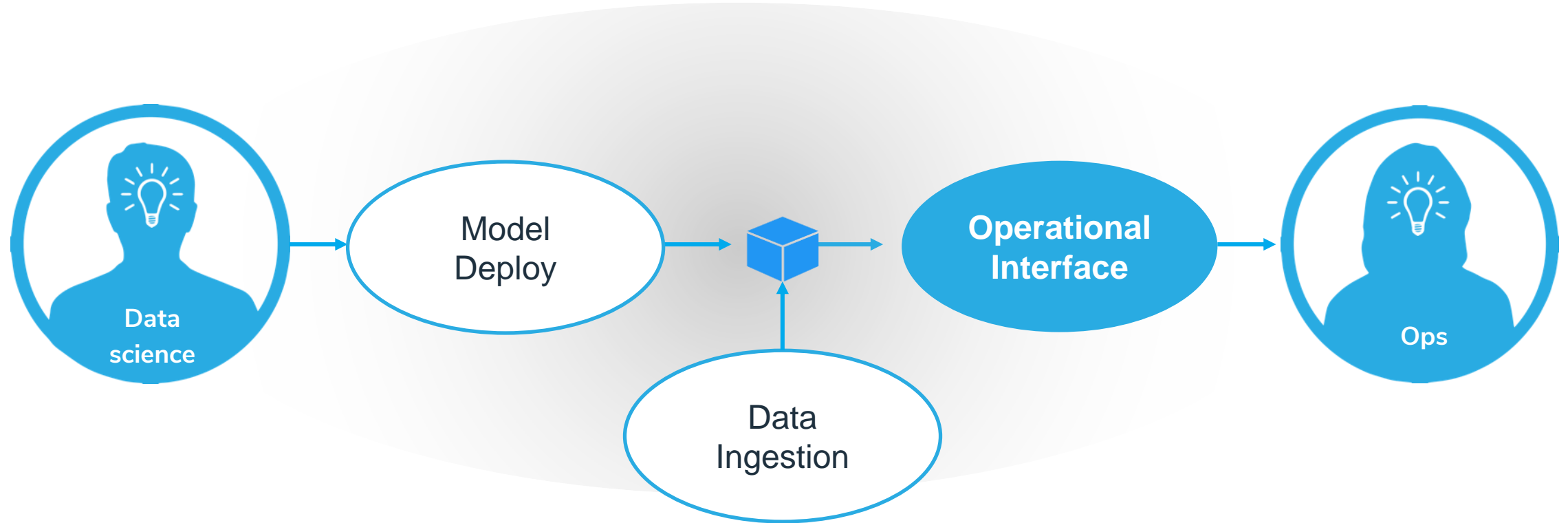
- A CLI (command line interface) to push models into model hub

```
(arundo) tailaiw@tailaiw-XPS-15-9560:~/work/arundo-demo/end-to-end/equipment_status_predictor$ composer publish
Validating model...
Model validated!
Bundling model...
Bundling app in folder /home/tailaiw/work/arundo-demo/end-to-end/equipment_status_predictor
Making zip file: equipment_status_predictor.zip
Model bundled!
Creating model doc...
Model doc created id = 813999fd-574d-a971-8d29-3a549d1bb4ee
Uploading model zip file...
Model upload succeeded
Build triggered
```

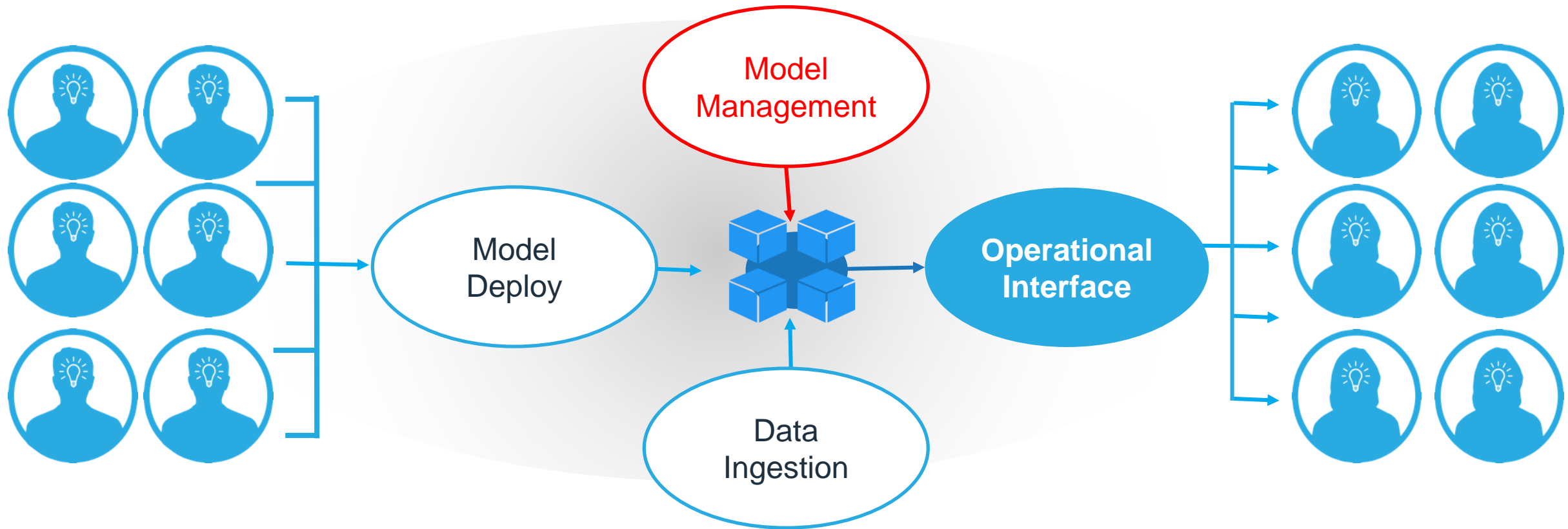
Model Deployment and Management

- Model hub connects data, models, and applications
- Each model is a SaaS deployed at a docker container
- Model deployment software is an interface for data scientists to publish their models as an API (application programming interface)
- **Model hub is fully responsible for enterprise-level management work (resourcing, CI/CD, version controls, scheduling, performance monitoring, etc.)**

Model Management in Individual Scale

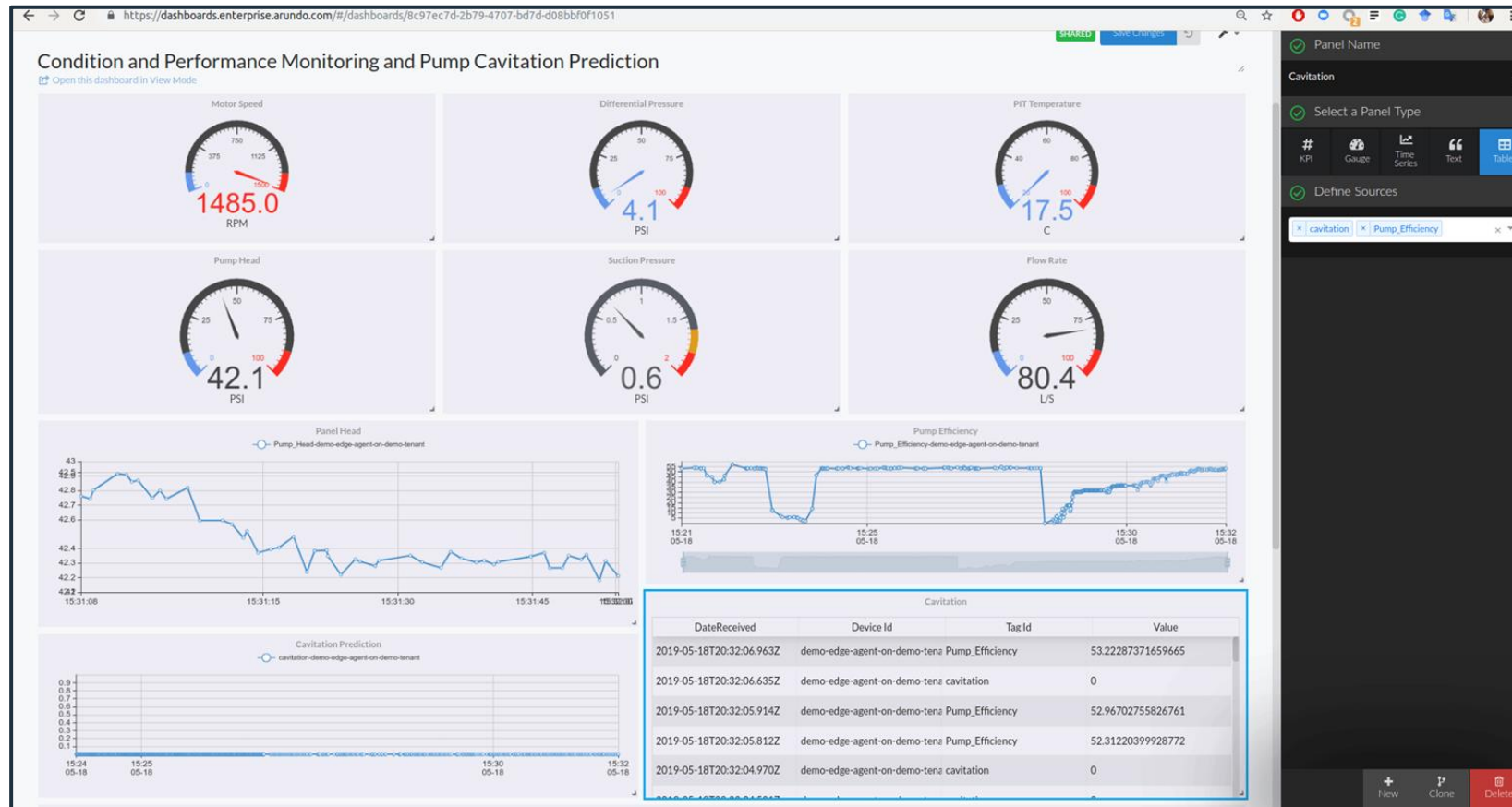


Model Management in Enterprise Scale



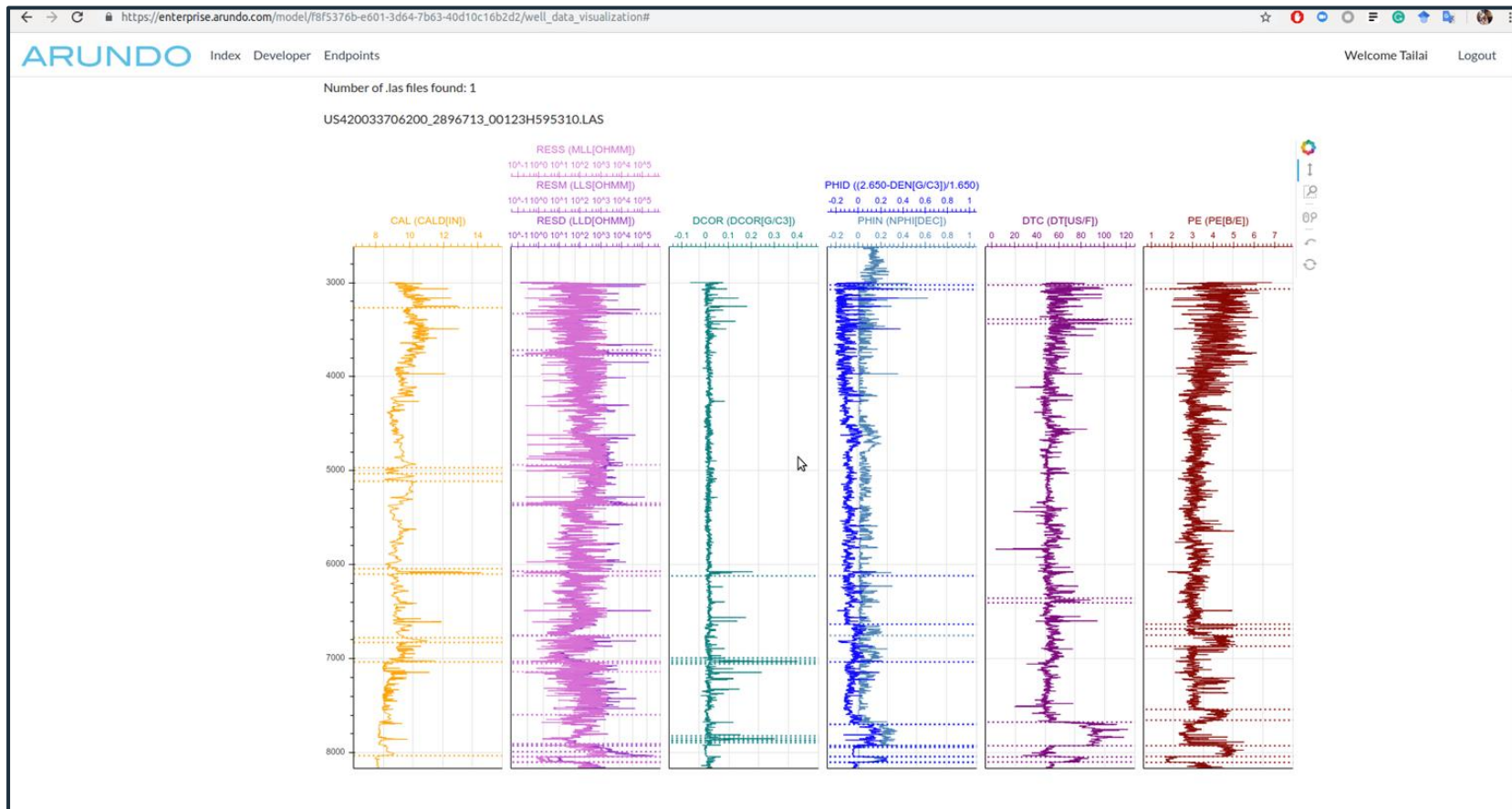
End-User Interface

- Flexibility to connect common BI tools
- Customized web-based dashboard
- Customized user interface



End-User Interface

- Flexibility to connect common BI tools
- Customized web-based dashboard
- Customized user interface



Why is This Important to Data Scientists in E&P Companies?

Common skill set of a data scientist in 2019

Includes:

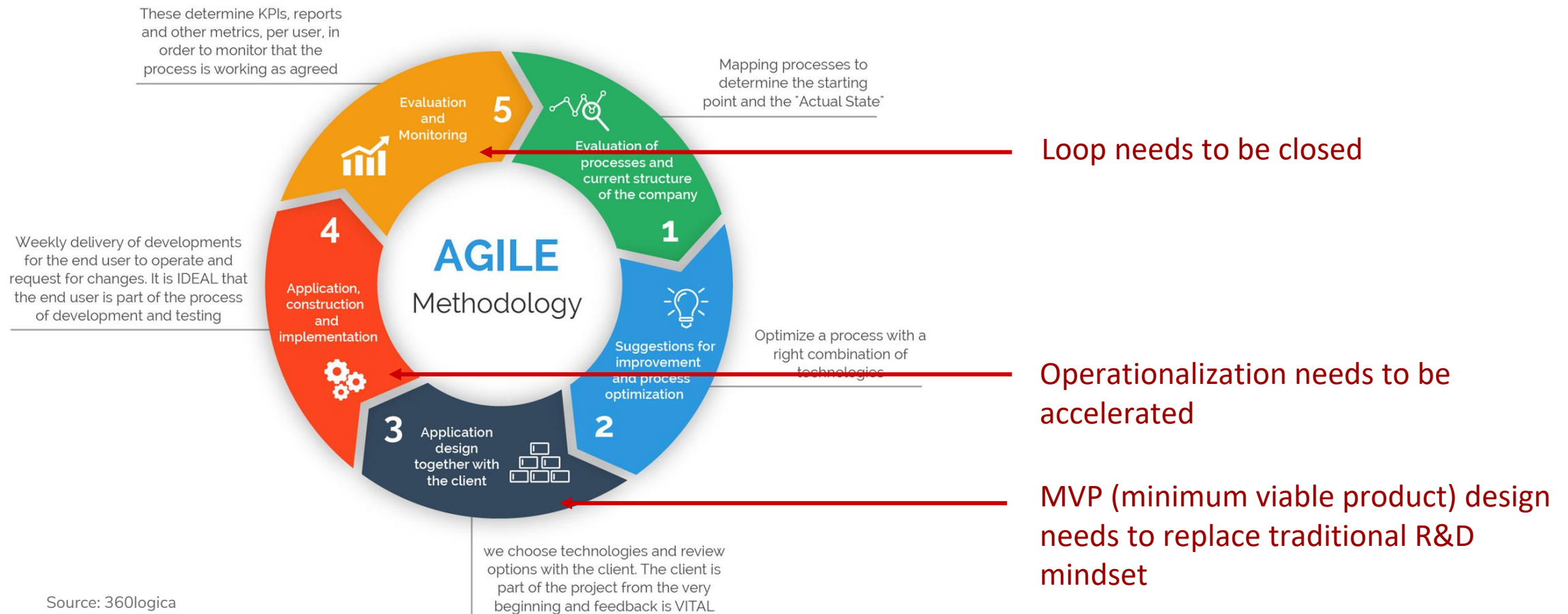
- Python, R, MATLAB, Julia, ...
- Jupyter notebook, PyCharm, ...
- Machine learning, deep learning, statistics, ...
- Scikit-learn, Keras, TensorFlow, PyTorch,

Does NOT include:

- Docker, Kubernetes
- OPC, Modbus, MQTT, Kafka, ...
- NodeJS, PHP, C#, GoLang, ...
- Authentication, network security
- Travis CI, Jenkins, Circle CI, ...
- ReactJS, AngularJS, VueJS, CSS, ...

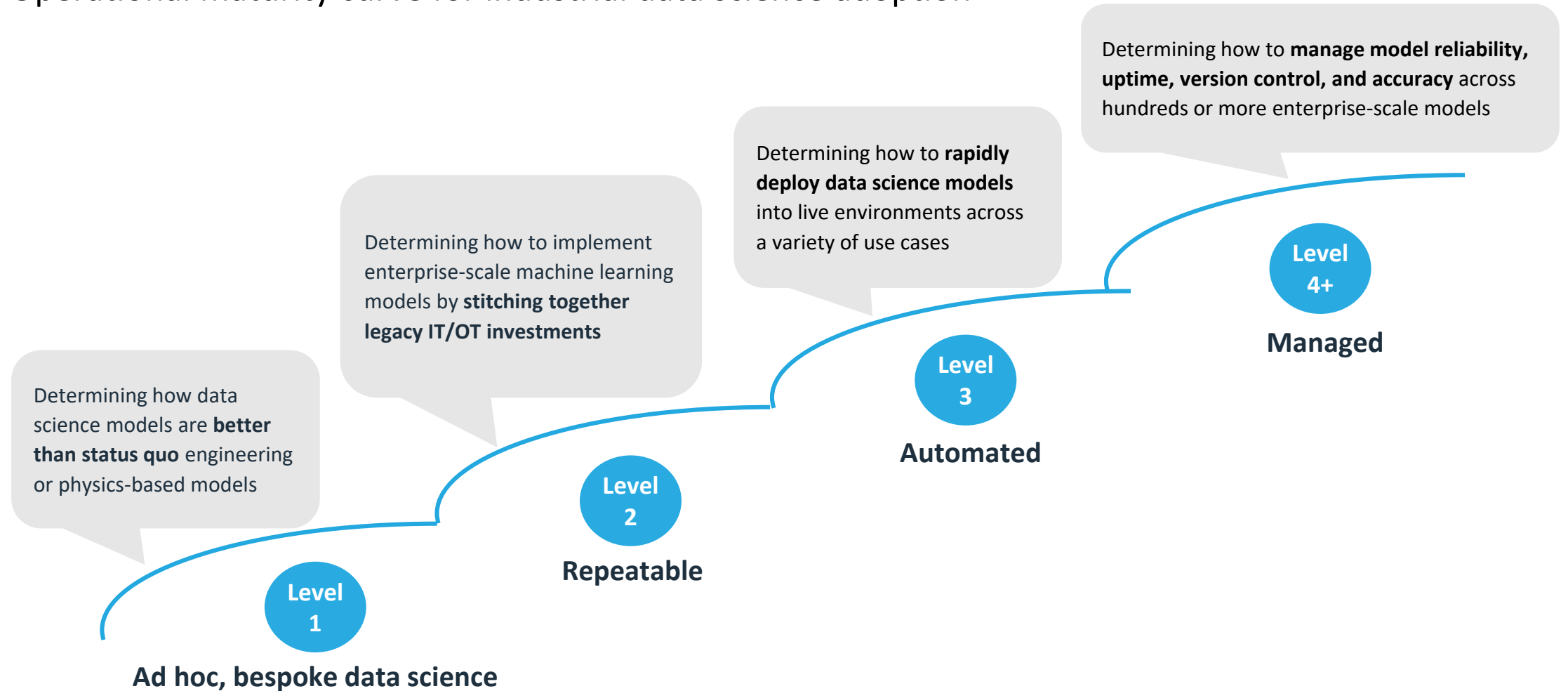
Why is This Important to E&P Companies?

Building an agile ecosystem of data-driven decision making



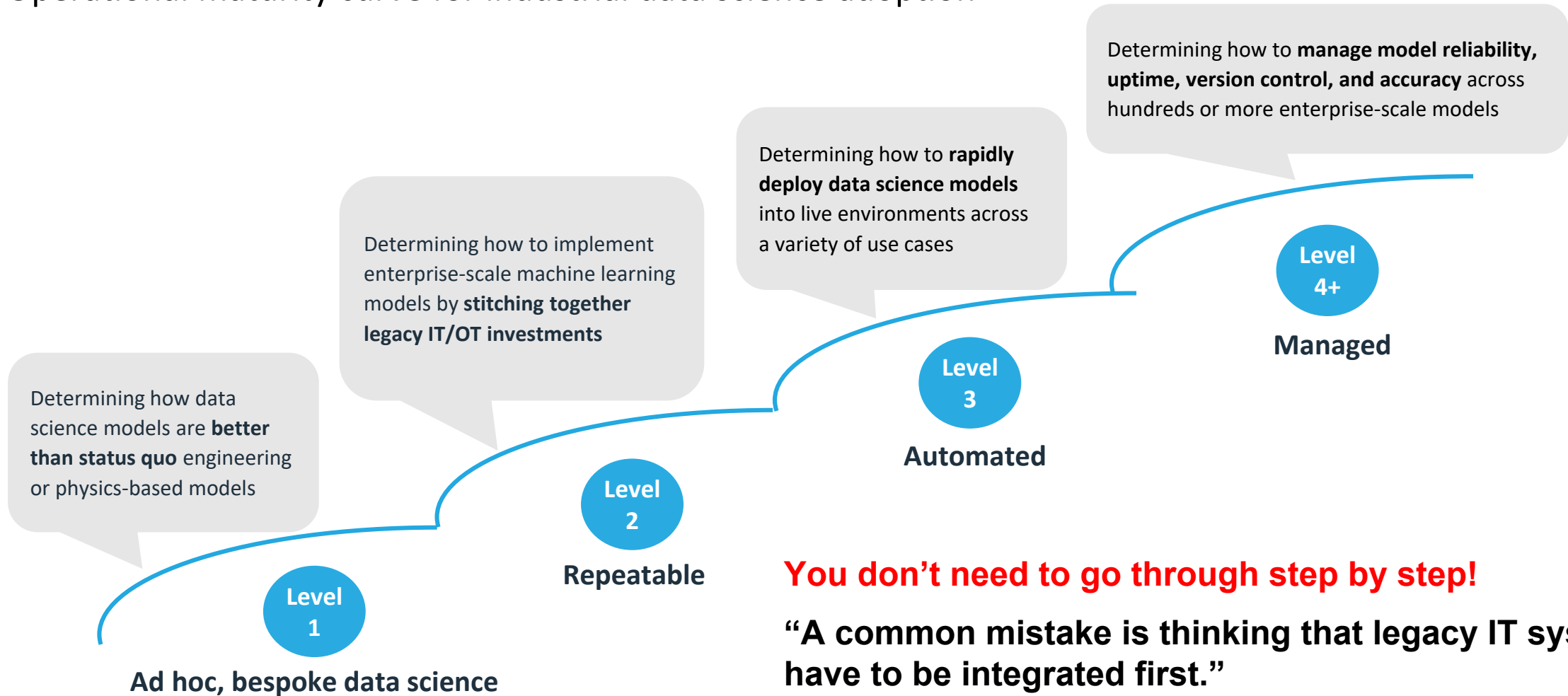
Why is This Important to E&P Companies?

Operational maturity curve for industrial data science adoption



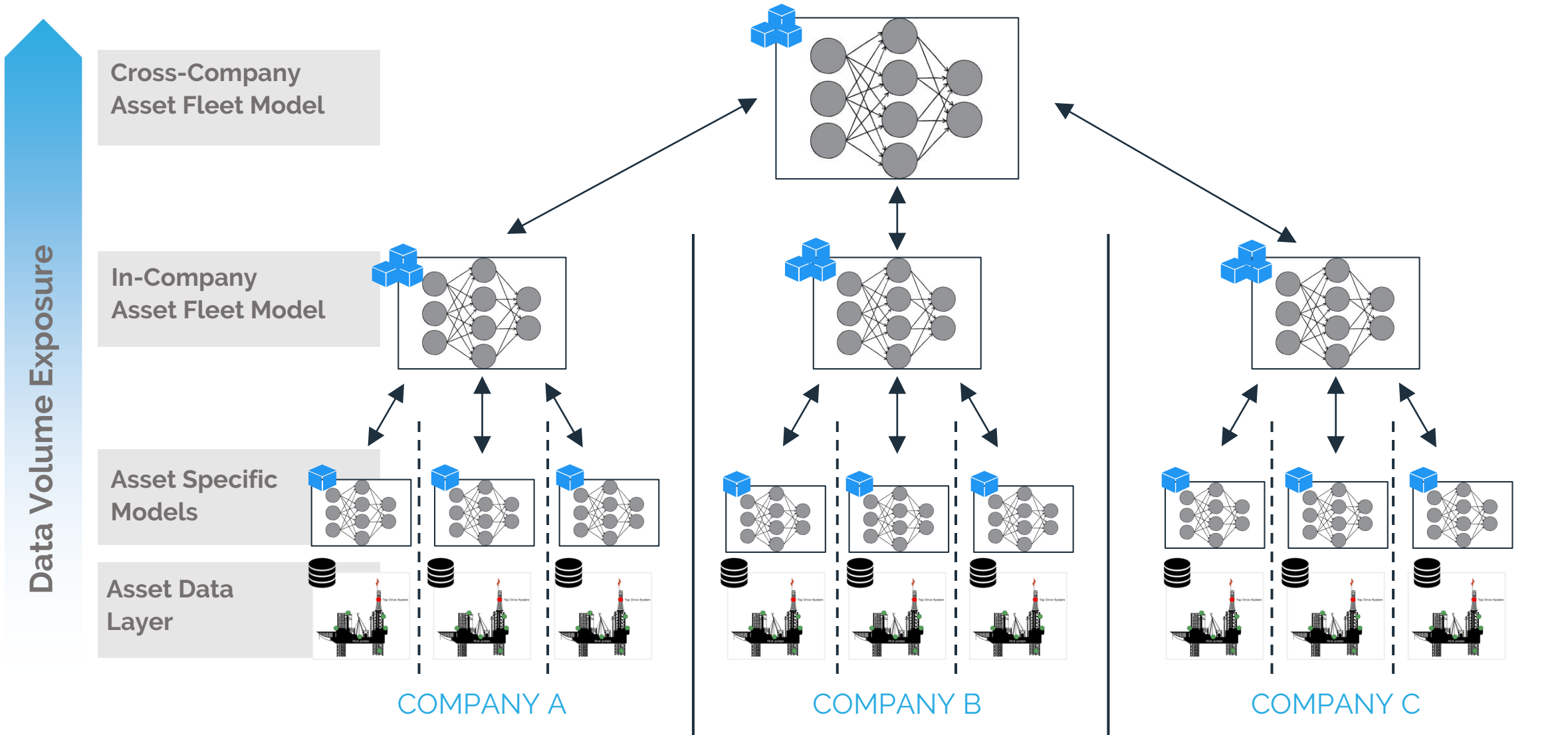
Why is This Important to E&P Companies?

Operational maturity curve for industrial data science adoption



---Ten Red Flags Signaling Your Analytics Program Will Fail by McKinsey

Why is This Important to E&P Sector?



ARUNDO

HOUSTON | OSLO | PALO ALTO