Operationalize Machine Learning for Real Impact

Zongjie Diao (Z)
Director of Product Strategy and Management, Cisco

AI Initiatives Seen Across All Industries

Finance

Fraud Detection

Cryptocurrencies

Algorithmic Trading

Healthcare

Cancer Cell Detection

Drug Discovery

Medical Research

Media and Entertainment

Video Captioning

Content Based Search

Nl, Vr and Ar

Security and Defense

Face Recognition

Crowd Analytics

Cyber Security

Retail

Theft Detection

Auto Checkout

Targeted Marketing

Manufacturing

Reduce Product Defects

Increase Production Speed

Shorten Downtime

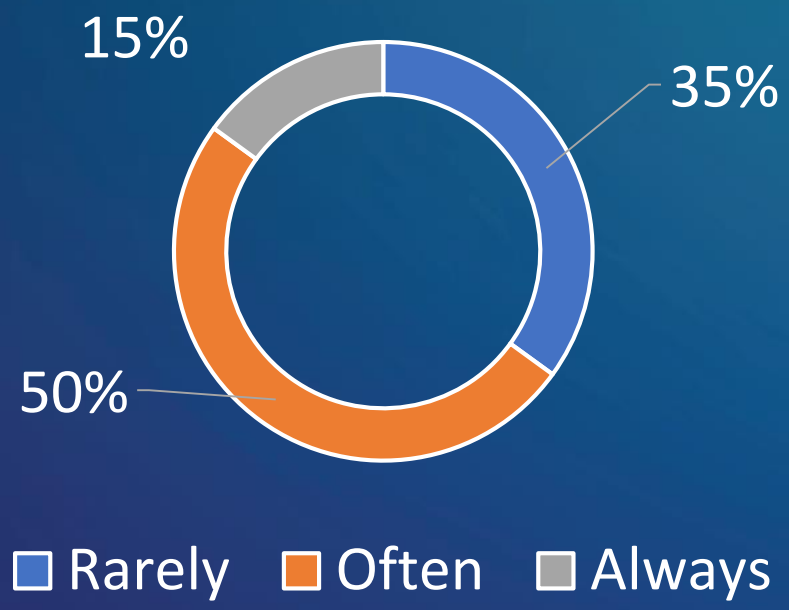
Machine Learning is

useless… and a waste of money

So said many companies who have tried…..

Low deployment

How often are trained model deployed?



Long deployment

Expected AI deployment timeline

How long it really takes

3-5X

Start planning

Still planning

Piloting

Restarting

Finally launch

Source: .iianalytics.com/2019-analytics-predictions-and-priorities

75% of early AI projects will underwhelm due to operational oversights

Source: forrester.com/predictions (2018)

Which process(es) below is (are) the most critical component(s) of ML deployment?

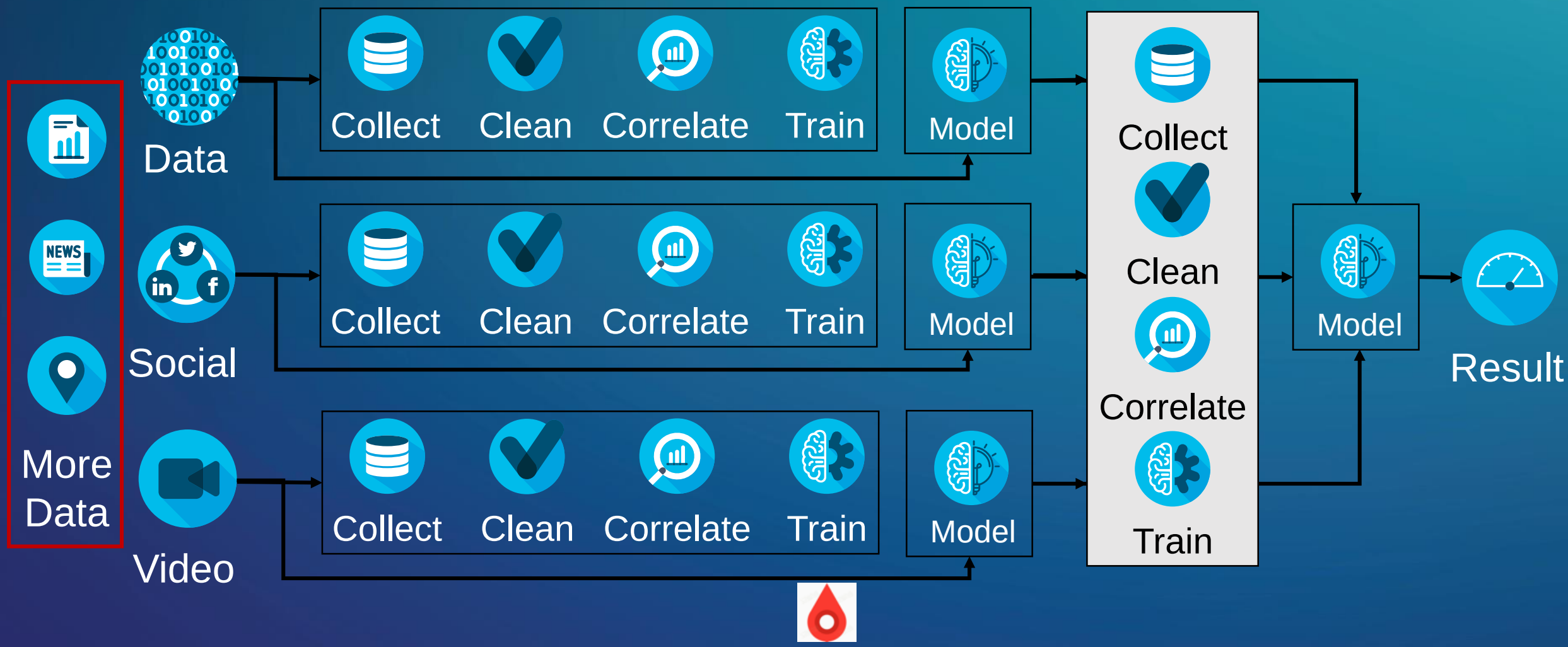
A. Data collection

B. Data preparation

C. Training and evaluation

D. Inference and deployment

Data Pipeline for Multiple Data Sources



Who are the key stakeholders to ensure a successful ML deployment at scale? (select multiple choices)

A. Data scientists

B. Data engineers

C. CXO &Line of business

D. IT

Infrastructure Implications of AI Pipelines

- Data Fusion
- ETL Pipelines
- Data Visualization
- Feature Identification

- Weekly Retraining
- Model Management

- Model Ranking and Validation

- Production Deployment
- Establishing Feedback Loop

IO Intensive | Compute Intensive | Latency Sensitive

Framework and SW instruments need to be supported

- Data Fusion
- ETL Pipelines
- Data Visualization
- Feature Identification

- Weekly Retraining
- Model Management

- Model Ranking and Validation

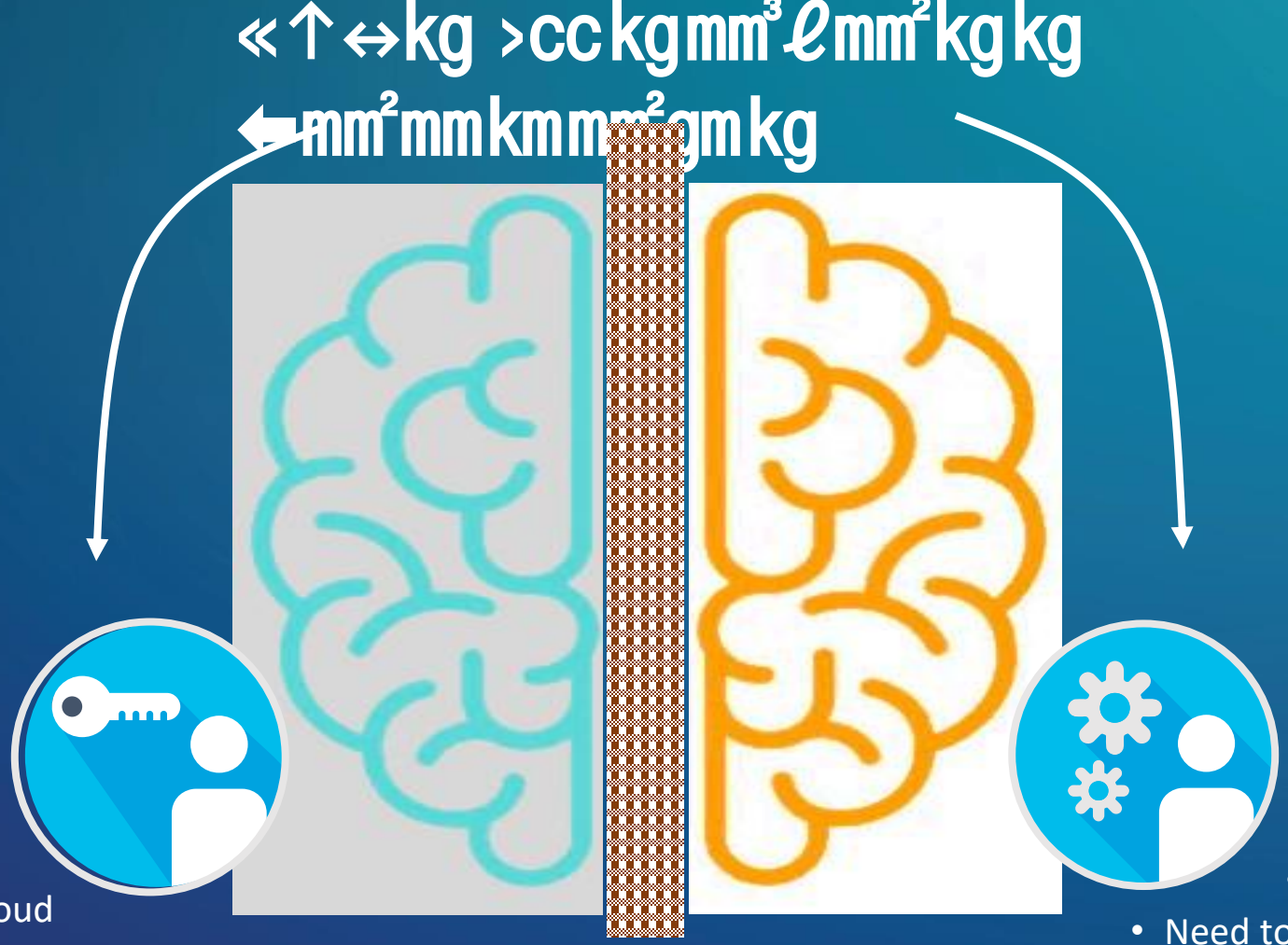
- Production Deployment
- Establishing Feedback Loop

Data Infrastructure

ML/DL Framework / Infrastructure

Inferencing & Ingestion End Point

Major challenge gaps exist between key stakeholders in realizing ML impact



- Start with workstation or cloud
- Hard to scale for production
- Want cloud like experience on-prem

- Lack of AI/ML expertise
- Need new infra architecture
- Need to solve silos and manageability
- Need enterprise readiness and security

How much percent of Machine Learning models are developed on Cloud vs. on-premise?

A. 90% vs. 10%

B. 70% vs. 30%

C. 40% vs. 60%

D. 10% vs. 90%

The benefits of both – it's a hybrid world when rubber meets the road

~60%

On-Prem

Data gravity and integration

Performance

TCO

Governance

Remodel, retraining

At scale production

~40%

Cloud

Fast deployment

Test-dev

Simplicity

APIs

One time

Source: Gartner, Market Guide for Machine Learning Compute Infrastructures, Sep 2108, ID: G00362287; Figure 3

Full Data Life Cycle Support

Validated design with eco system

Natural extension of data center

Cisco AI/ML Computing Platforms & Eco-System Partners

Compute

Inference

UCS & HX
1U Servers

CISCO INTERSIGHT

Test/Dev ML

UCS & HX
2U Servers

2 x PCIe V100
4-6 x PCIe T4*

Accelerate Learning

UCS & HX
C480 ML

8 x SXM2 V100, NVLink, 24 Drives, 6 x NVMe3 TB DRAM, Choice of OS,

© 2018 served. Cisco Confidential

ISV Partners

Kubeflow

NVIDIA

DataRobot

ANACONDA

Google Cloud

Cisco CCP

FASTDATA

H2O.ai

RED HAT OPENSHIFT

sas

SAP

awiros

and growing...

Storage, CI & Big Data

FlexPod

FlashStack

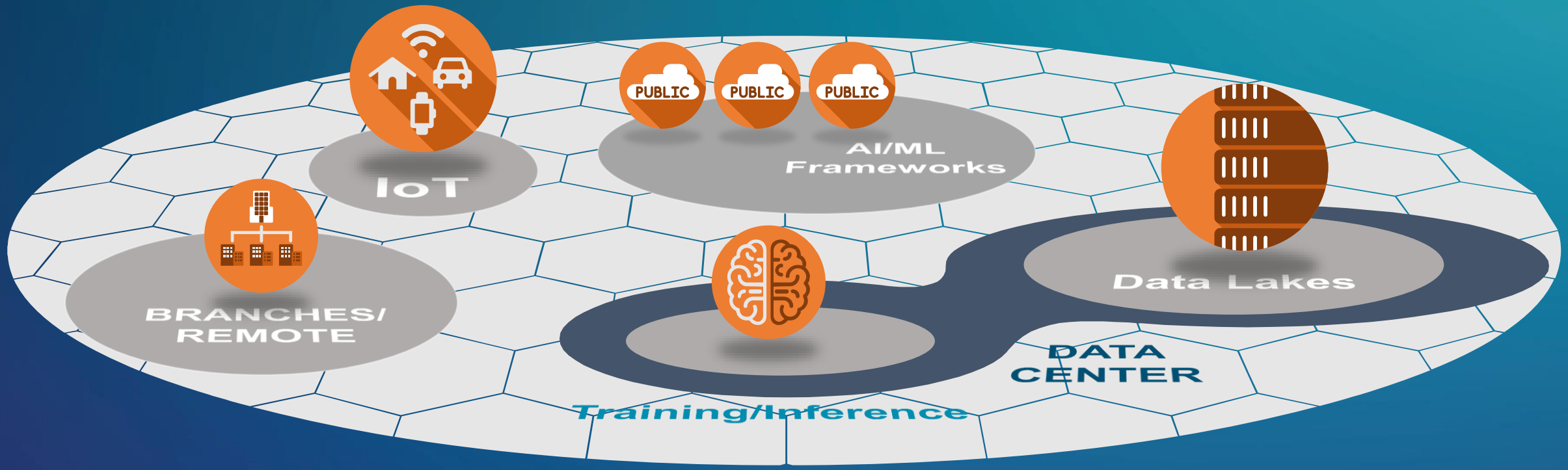
HORTONWORKS
POWERING THE FUTURE OF DATA

cloudera

SwiftStack

and growing...

Cisco AI/ML, A Holistic Approach



Accelerated Computing for Inference at the Edge

ML/AI Stack Partnerships

Accelerated Computing for ML/AI in the Core

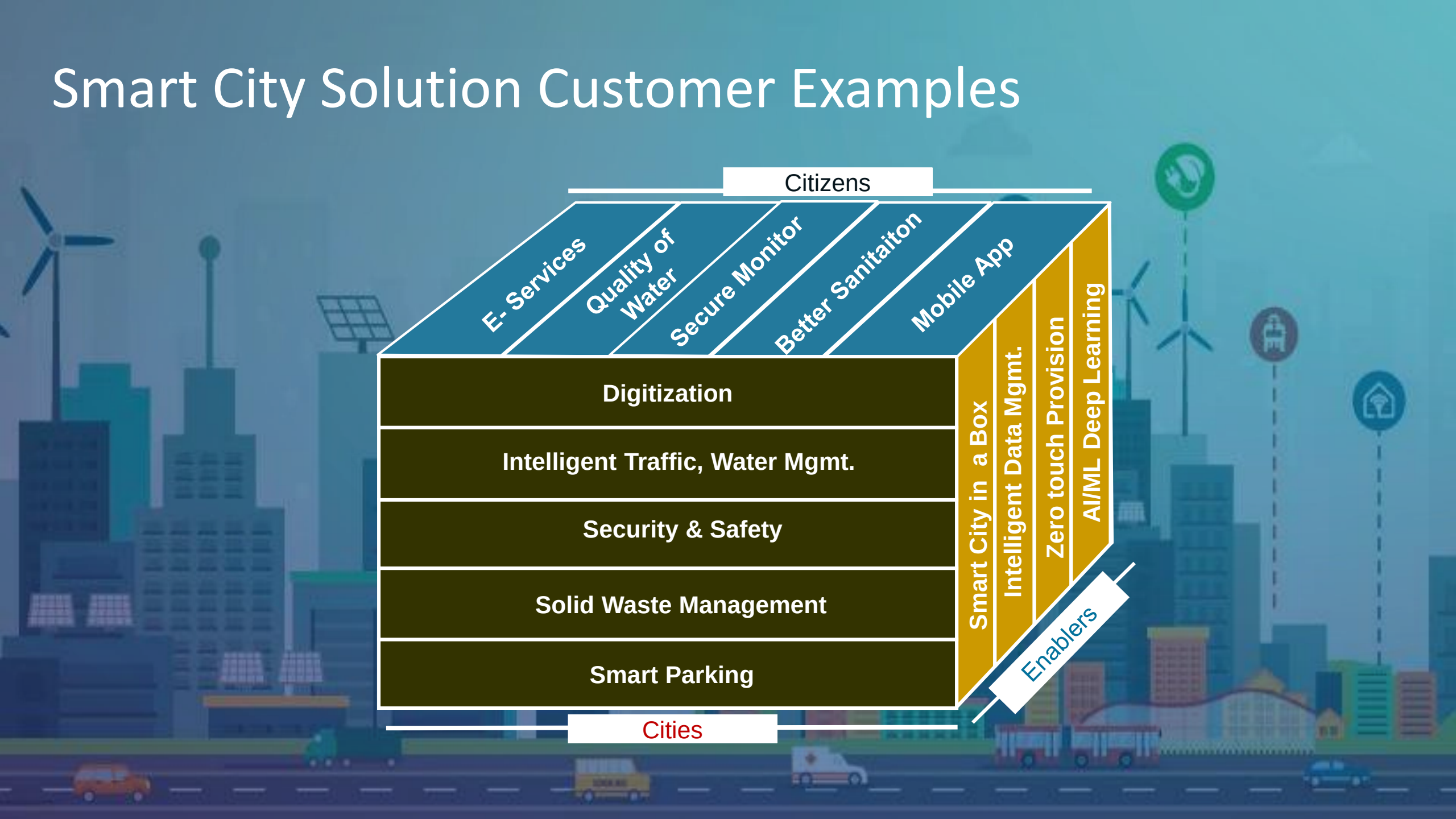
Big Data/Analytics and ERP

Cisco Servers (UCS and HyperFlex)
Cisco Multicloud Portfolio

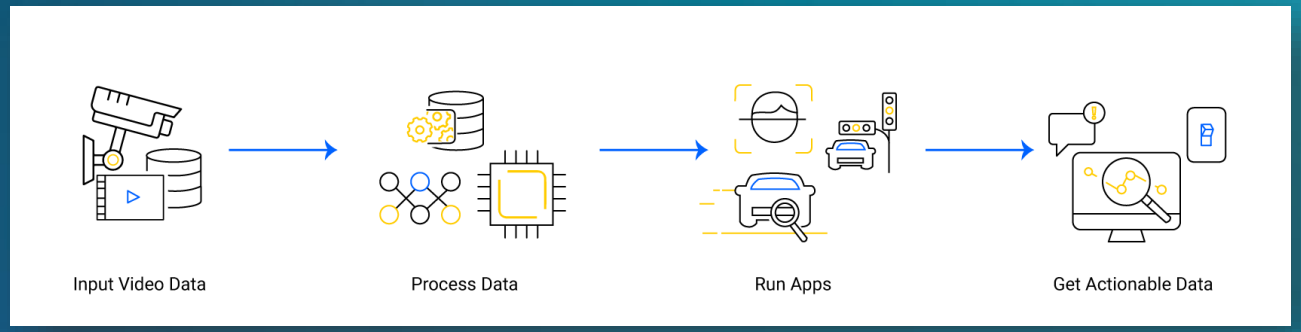
CISCO INTERSIGHT

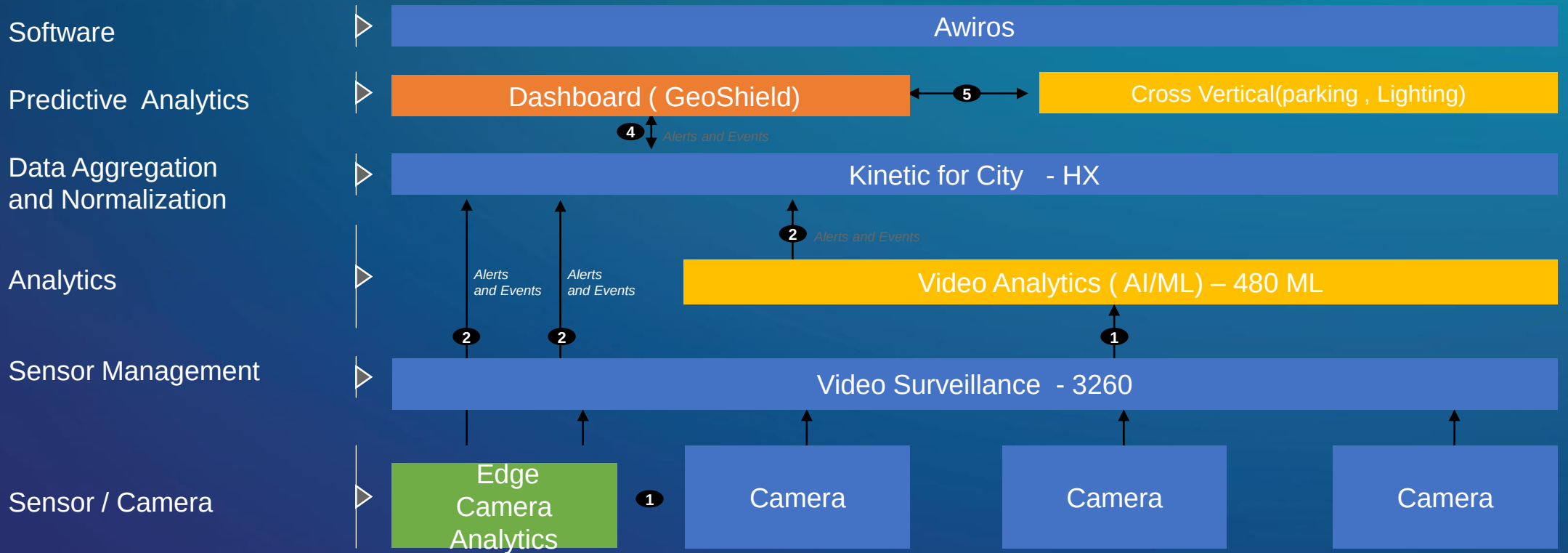
Cisco IoT
Cisco Network

Smart City Solution Customer Examples



Cisco Security Architecture for Smart Cities





Thank you!

Operationalizing ML is the key to success!

Cisco has solutions and is ready to help!

Join us to make the world a better place!

Email:
zdiao@cisco.com

https://www.cisco.com/c/en/us/solutions/data-center/artificial-intelligence-machine-learning/index.html