



Road to Enterprise Architecture for Big Data Applications

Mixing Apache Spark with singletons, wrapping, facade

Magneti Marelli, ICT Innovation
London, United Kingdom

#SAISEnt4

Company Overview



Magneti Marelli is an international company committed to the design and production of hi-tech systems and components for the automotive sector.

AUTOMOTIVE LIGHTING

(Headlamp, Rearlamp, Lighting and Body Electronics)

POWERTRAIN

(Gasoline and Diesel engine control, Electric Motor, Inverter and Transmission)

ELECTRONICS

(Instrument Clusters, Infotainment & Telematics)

SUSPENSION SYSTEMS AND SHOCK ABSORBERS

(Suspension Systems, Shock Absorbers and Dynamic Systems)

EXHAUST SYSTEMS

(Manifolds, Catalytic converter, Diesel Particulate Filter and Mufflers)

PLASTIC COMPONENTS AND MODULES

(Bumper, Dashboard, Central Console, Pedals, Hand Brake Levers and Fuel System)

AFTERMARKET PARTS & SERVICES

(Mechanical, Body Work, Electrics and Electronic and Consumables)

MOTORSPORT

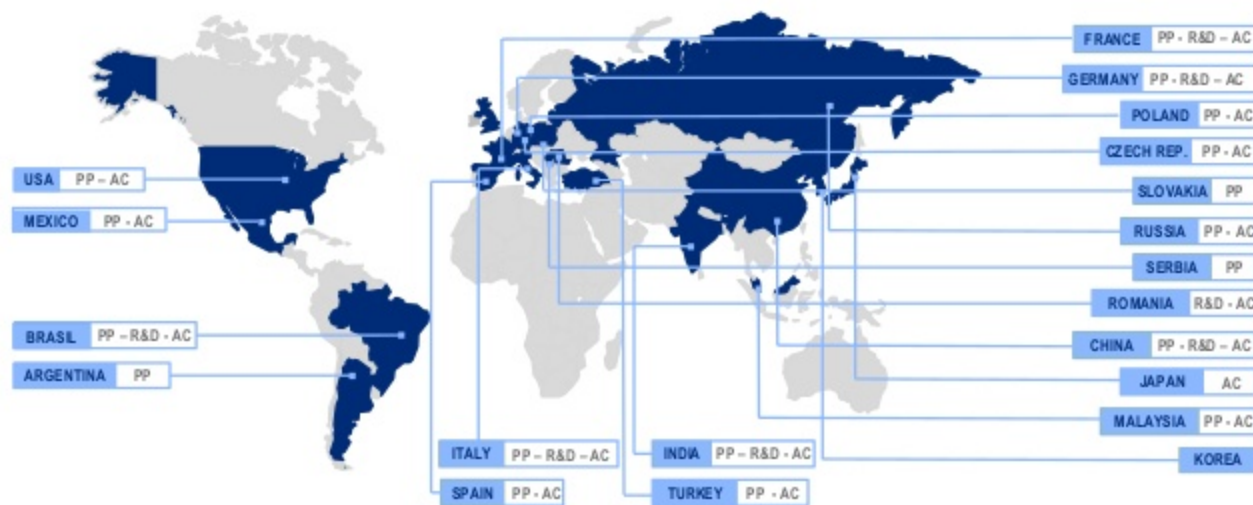
(Injection Systems, Electronic Control Units, Hybrid Systems, Telemetry Systems, Electric Actuators)



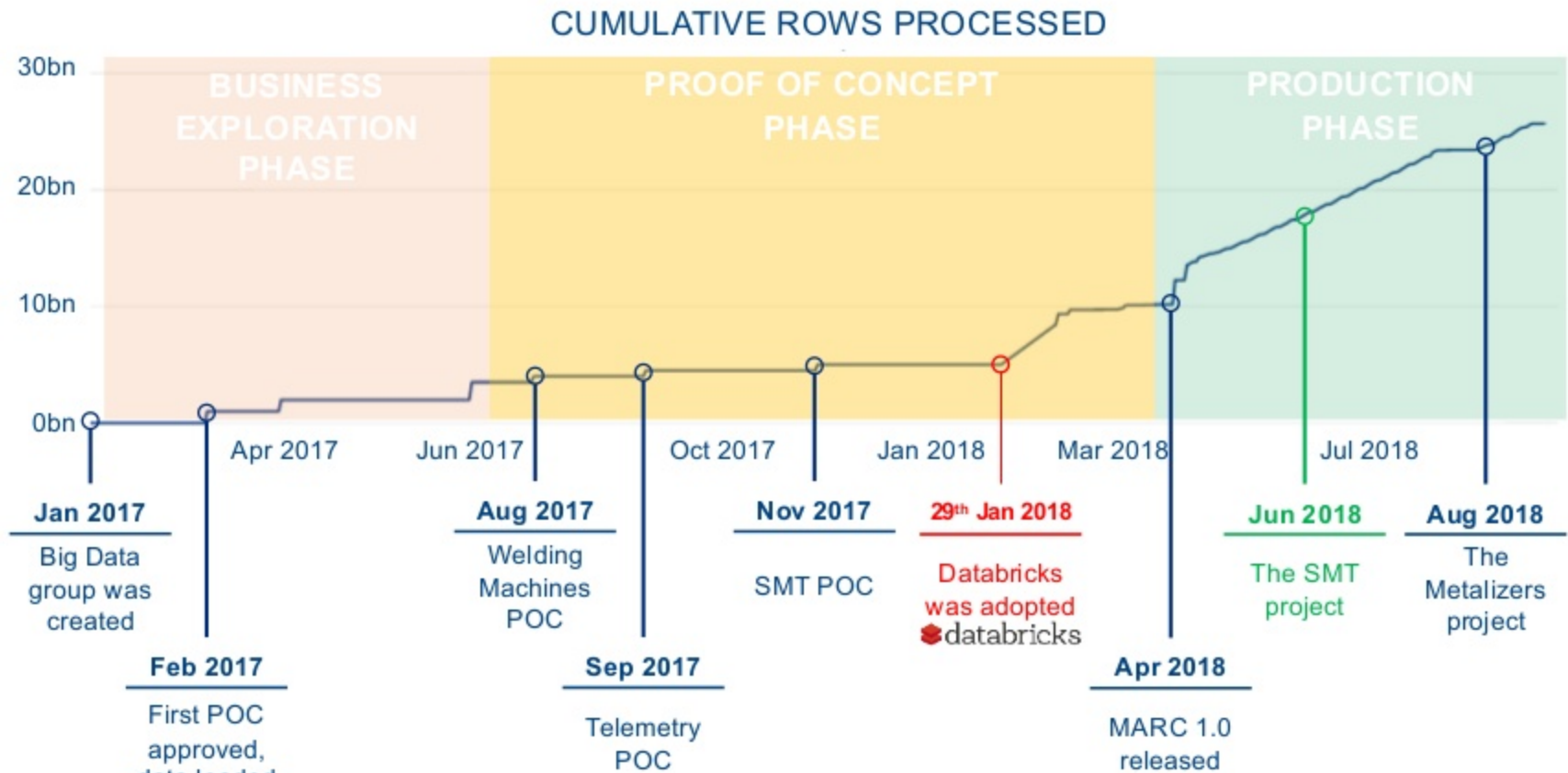
Magneti Marelli Worldwide Footprint



PP: Production Plant R&D: R&D Center AC: Application Center



Big Data storyline



The Surface-Mount Technology (SMT) project

Pre Production & Assembly Line



Surface-Mount Technology



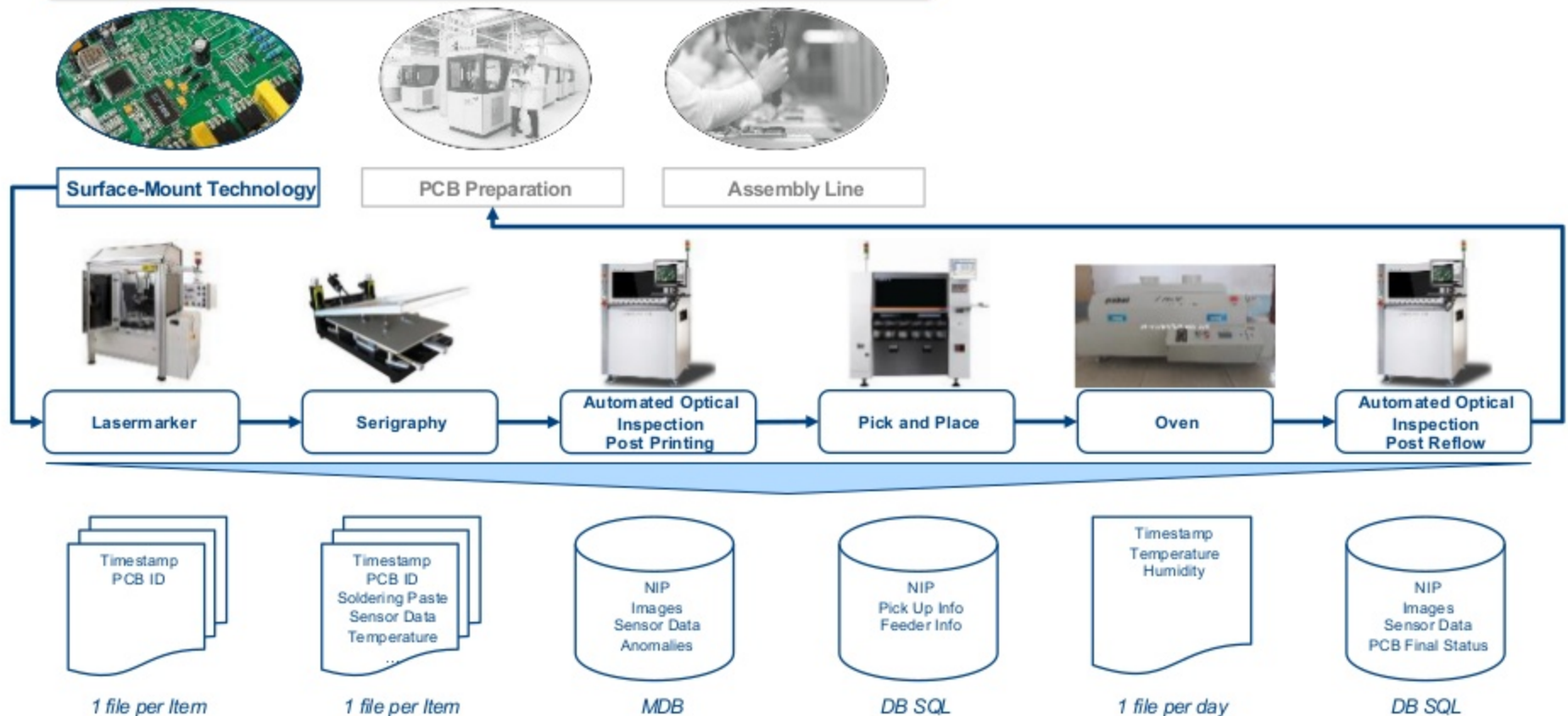
PCB Preparation



Assembly Line

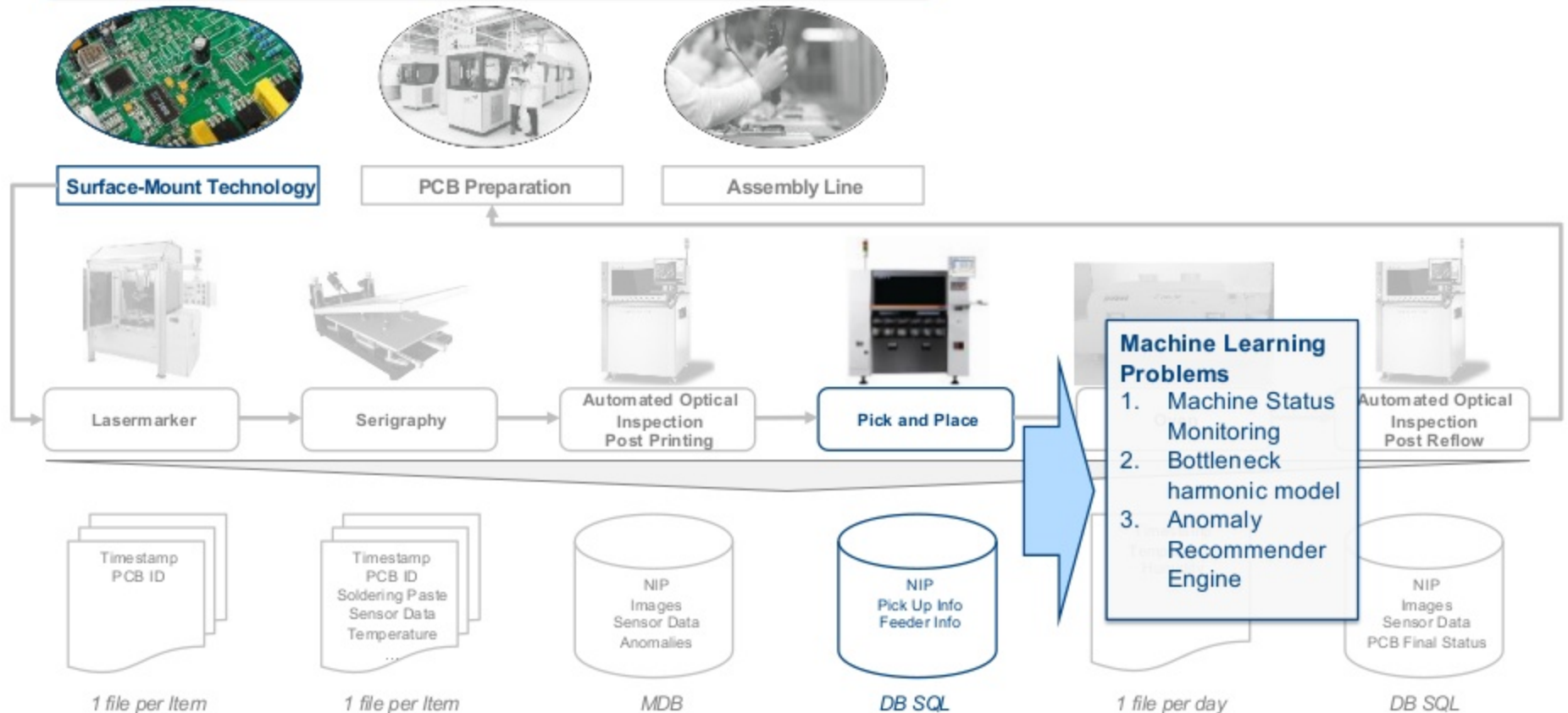
The Surface-Mount Technology (SMT) project

Pre Production & Assembly Line



The Surface-Mount Technology (SMT) project

Pre Production & Assembly Line



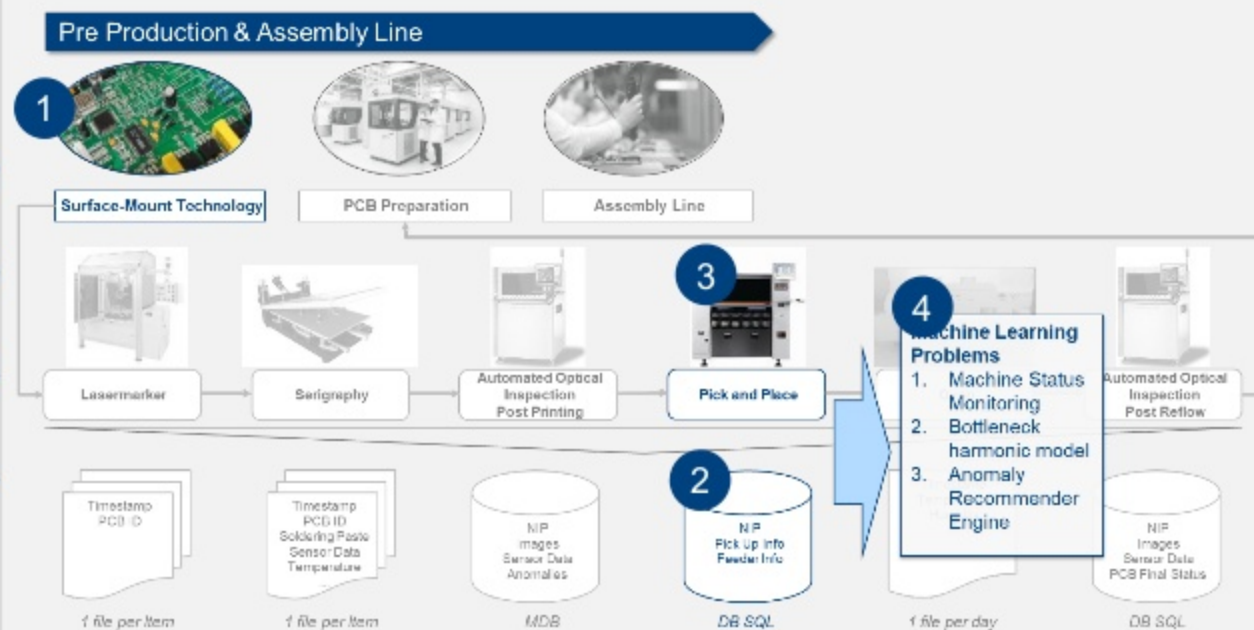
A Dream Project

1 Production process is well known

2 Data source is clearly defined

3 Need is raised by plant people

4 Algorithmic challenges are clear



Becoming a Nightmare

1 Production process is well known

2 Data source is clearly defined

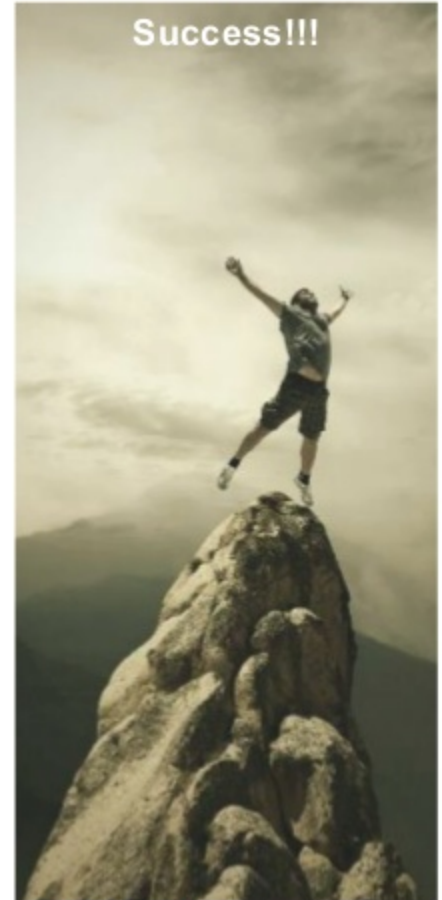
3 Need is raised by plant people

4 Algorithmic challenges are clear



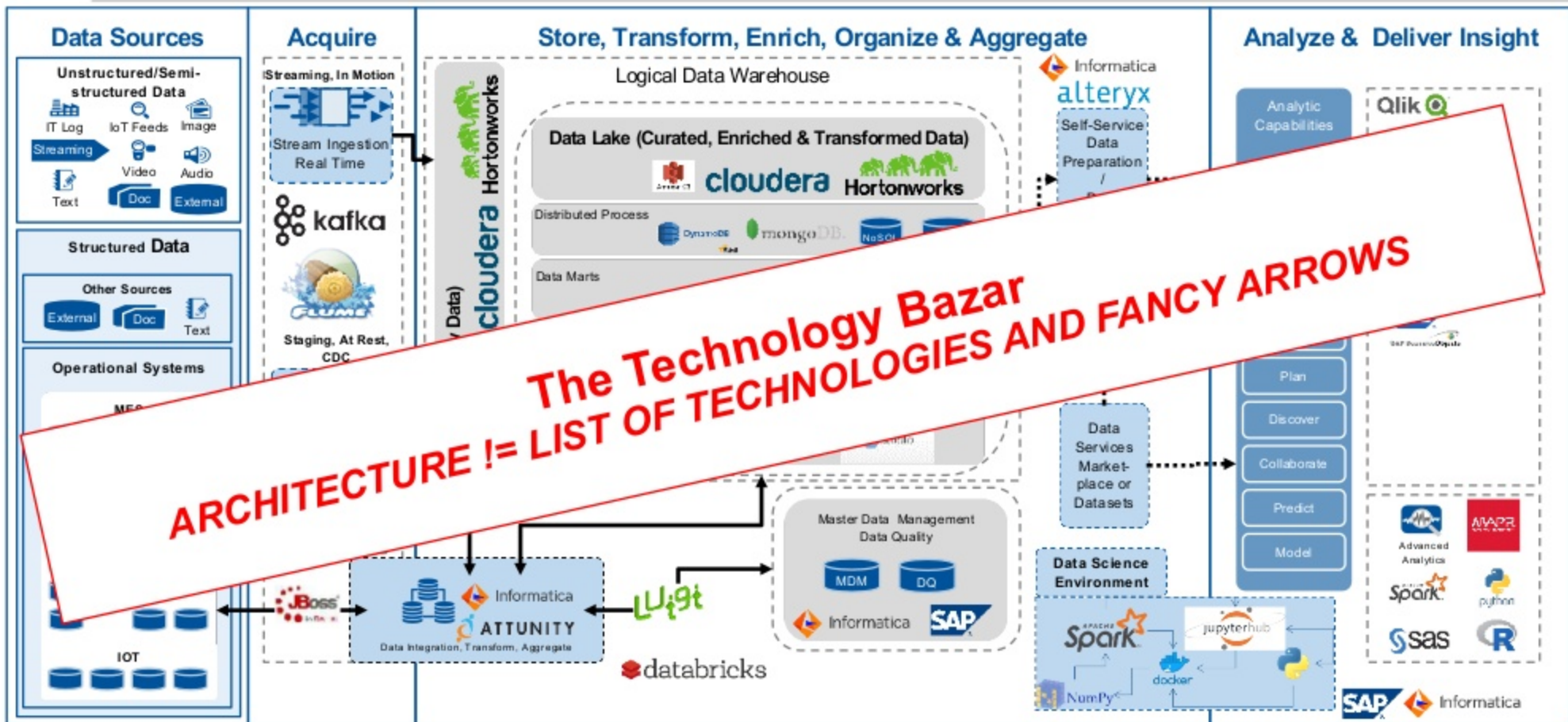
- So... Where is the data?
- How can I read/access the data?
- How can I be supported by my data scientist colleagues?
- How can I attach a Spark cluster to my Jupyter notebook?
- Who is going to port the notebook to production?
- What do you mean with production?

Success!!!

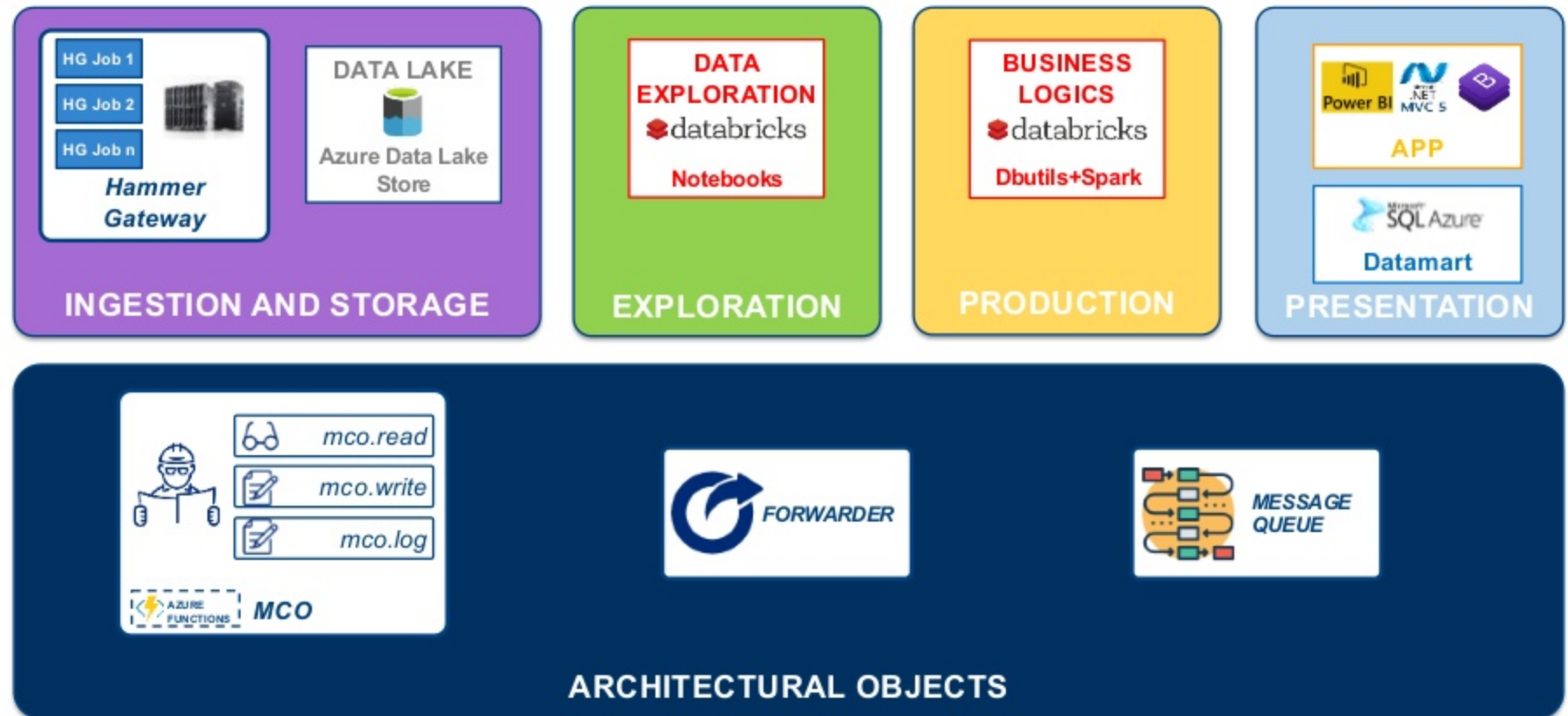


#SAISEnt4

Enterprise Architecture

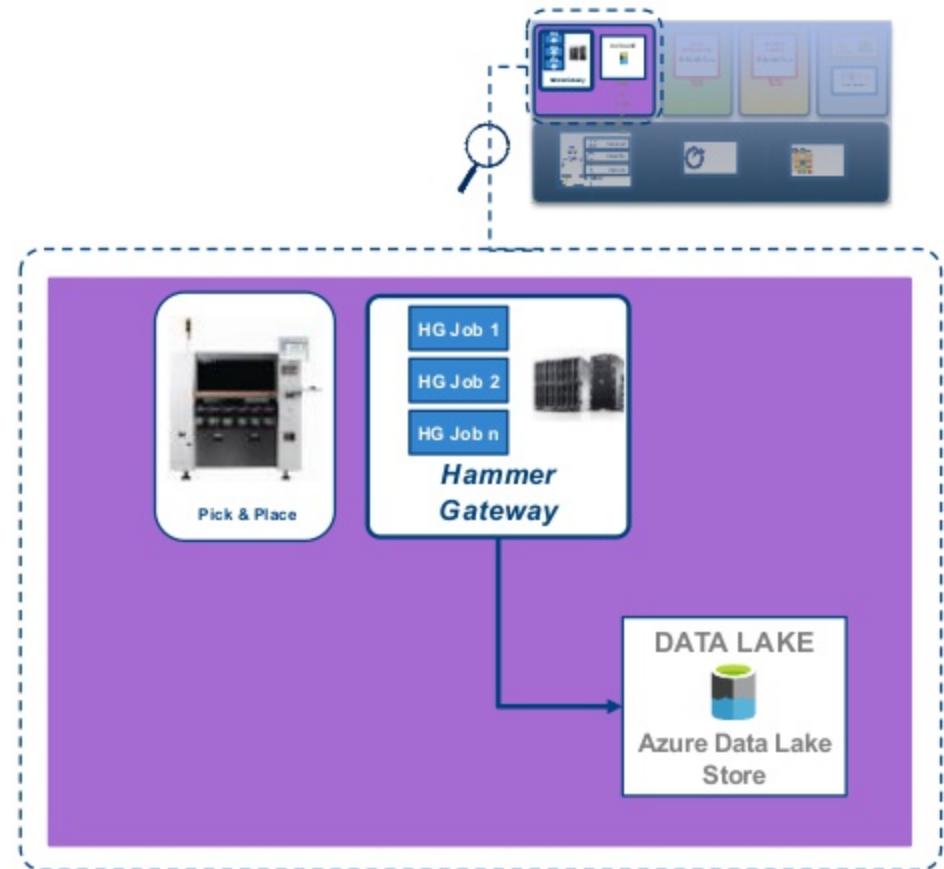


Keep it simple, stupid



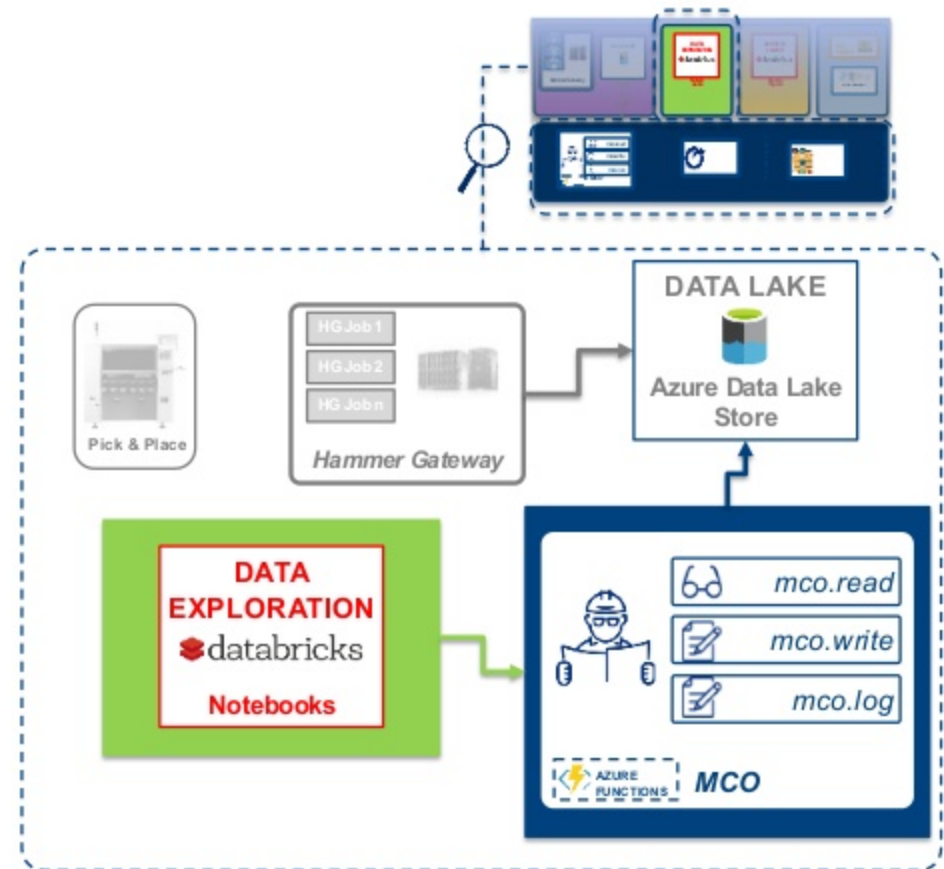
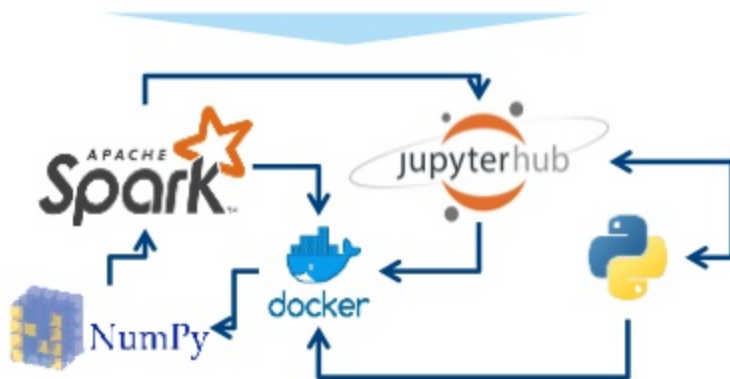
Enterprise Architecture – Data, where are you?

- Where are the csv?
- What do you mean with bcp out?
- How can I get a copy in the cloud?
- How can I update data on a regular basis?



Enterprise Architecture – The Jupyter case

- How could I work together with other data scientists?
- How can I deal with computation spikes?
- How can I attach an Apache Spark cluster to my Jupyter?
- Damn, Java Heap Memory Exception: what do you mean?



#SAISEnt4 |

One singleton to rule them all



mco

Pattern: Singleton

Use: bring tokens and technical access to notebook

Benefit:

- enhancing security
- enabling access control
- reducing vendor lock-in effect

mco.read

Pattern: Wrapping

Use: take data from data lake knowing only data names

Benefit:

- no one will need to know where data are or how data are stored
- incremental read capability out-of-the-box
- reducing time to port code in production
- reduced reading time
- reduce the vendor lock-in effect (to propagate a new HDFS PAAS vendor on all services is a matter of hours)

mco.write

Pattern: Wrapping

Use: save data everywhere

Benefit:

- no one will need to know where data must be put or how
- avoid dangerous behaviours such as writing on a SQL with a transformation action (connection pool, my beloved friend...)

mco.log

Pattern: Wrapping

Use: bring developer grade logging capability

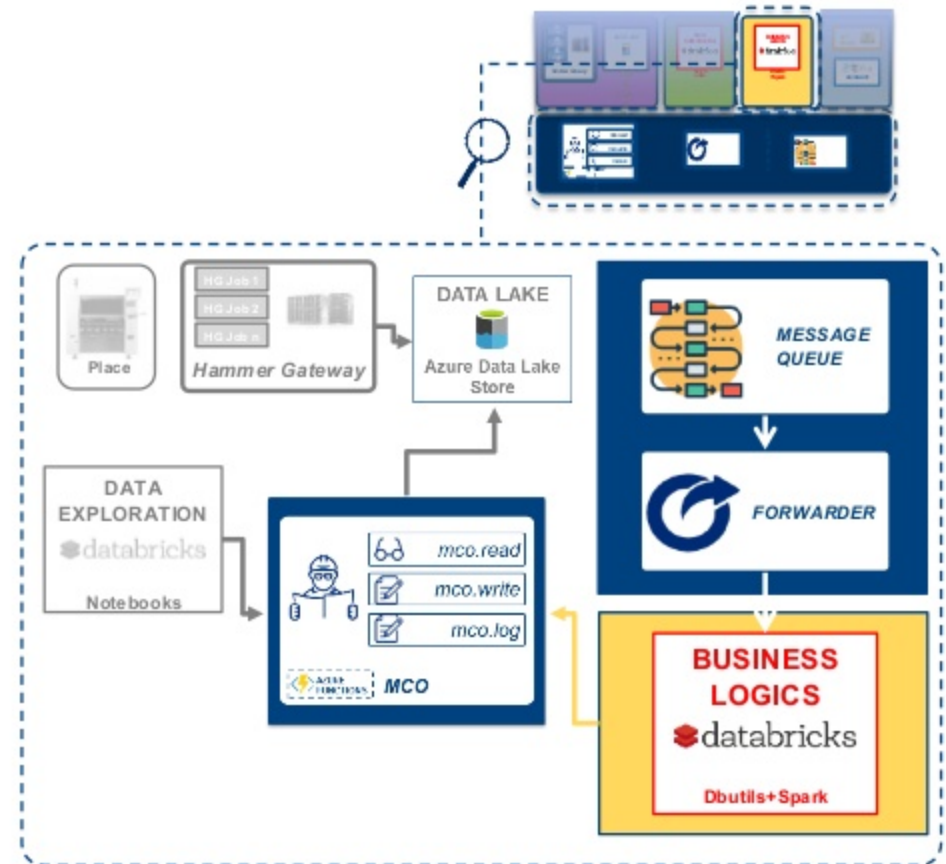
Benefit:

- reducing debug time
- enabling process audits

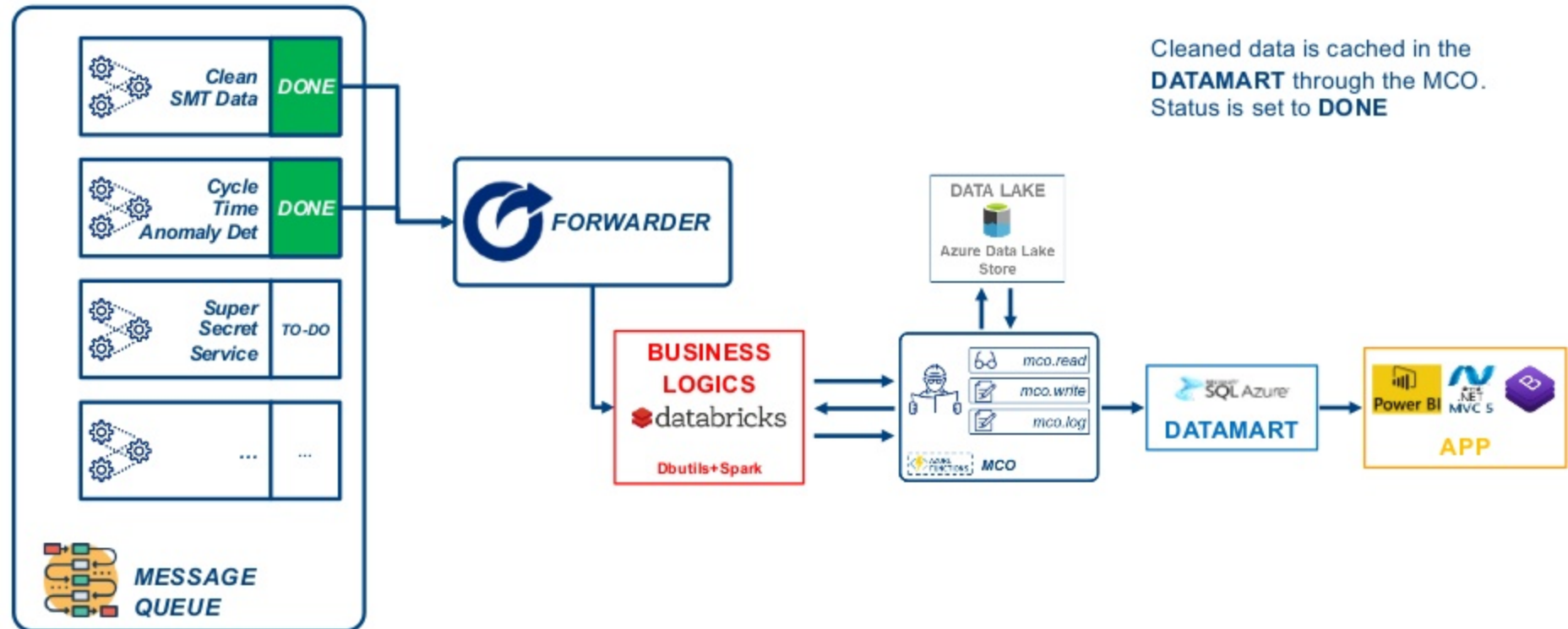
Enterprise Architecture – The model is ready!



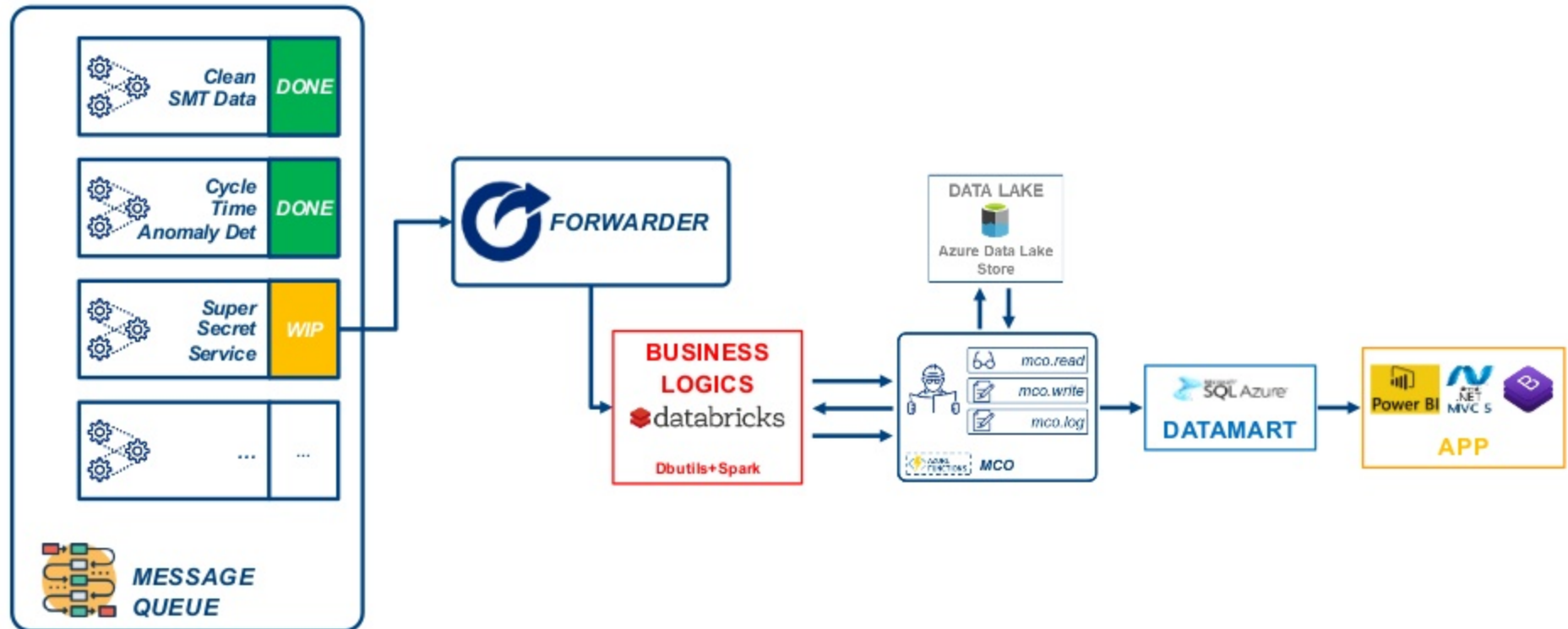
- Is the code production ready?
- Who is going to port the notebook to production?
- Developer algorithm is wrong: it produces different numbers...
- Ok, I got it! I'll need a crontab... but where?



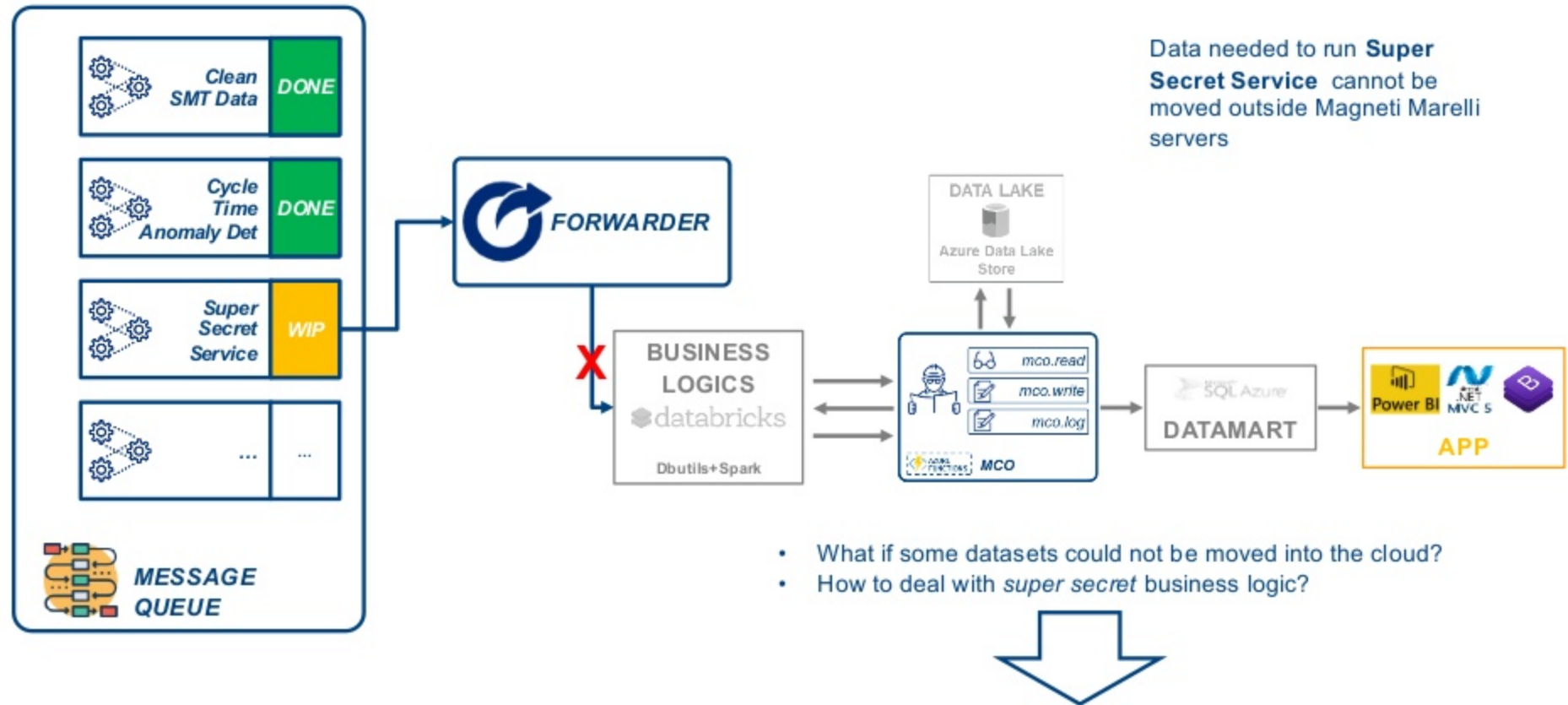
A For-what? *What the hell?*



A For-what? *What the hell?*

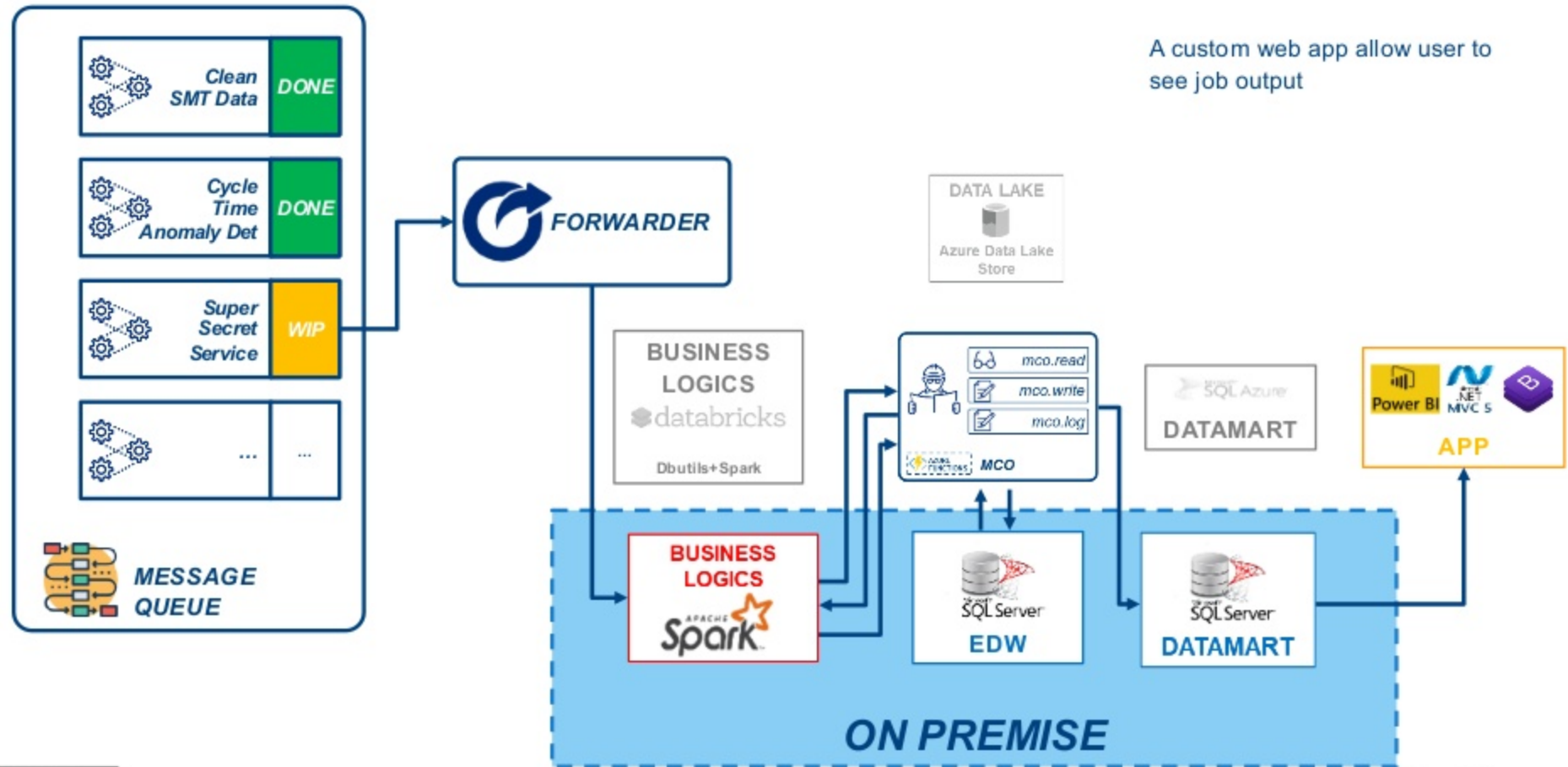


A For-what? *What the hell?*



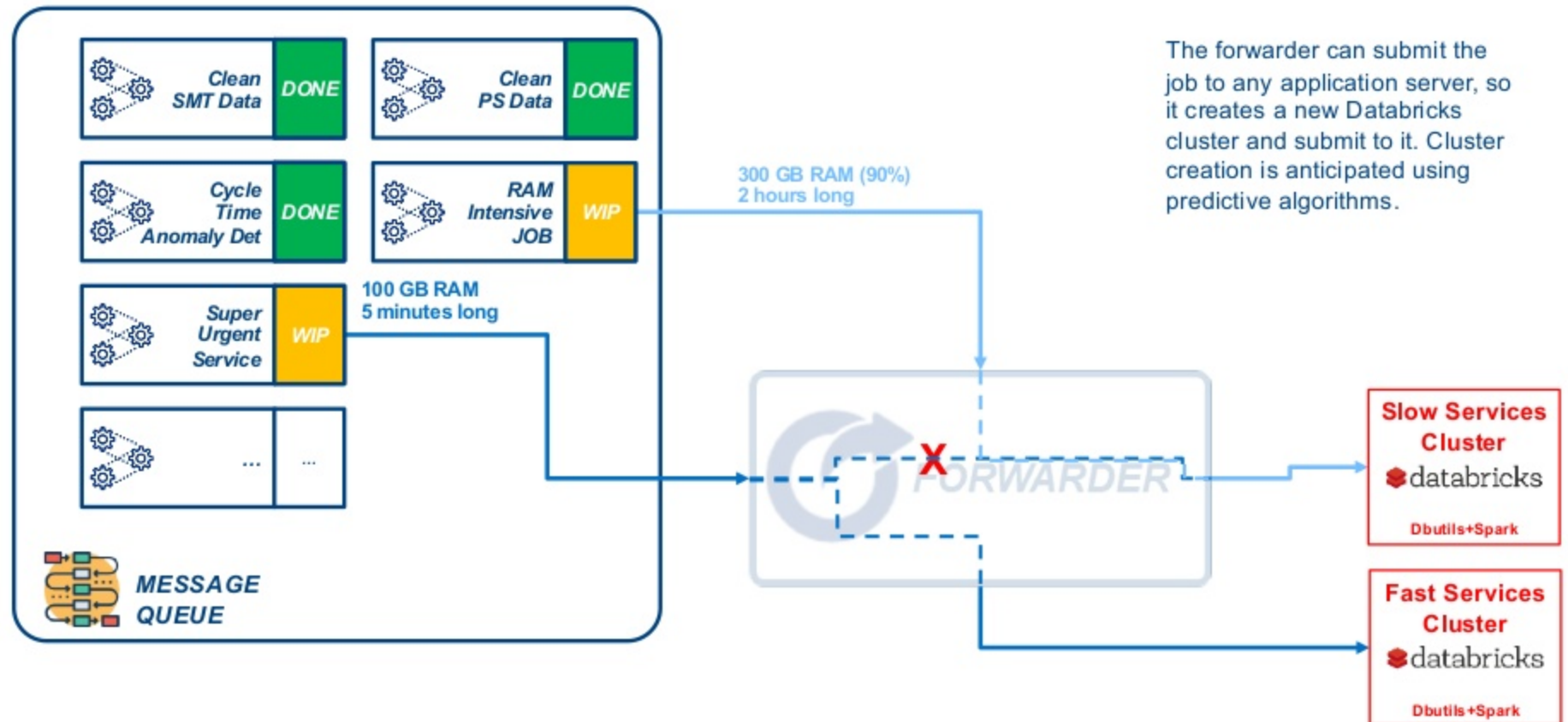
FORWARDER as the main component for cloud hybridization!

A For-what? *What the hell?*



A custom web app allow user to see job output

A For-what? What the hell? – Predictive balancing



Enterprise Architecture – Don't mind about nerd stuff



Data Scientists' presentation concerns:

- How do I write a web page?
- Do I need to bootstrap?
- MV-what? I thought Spring was just a season!
- Single sign-on? What do you mean?

coursera

Explore ▾

Q web development

Courses and Specializations



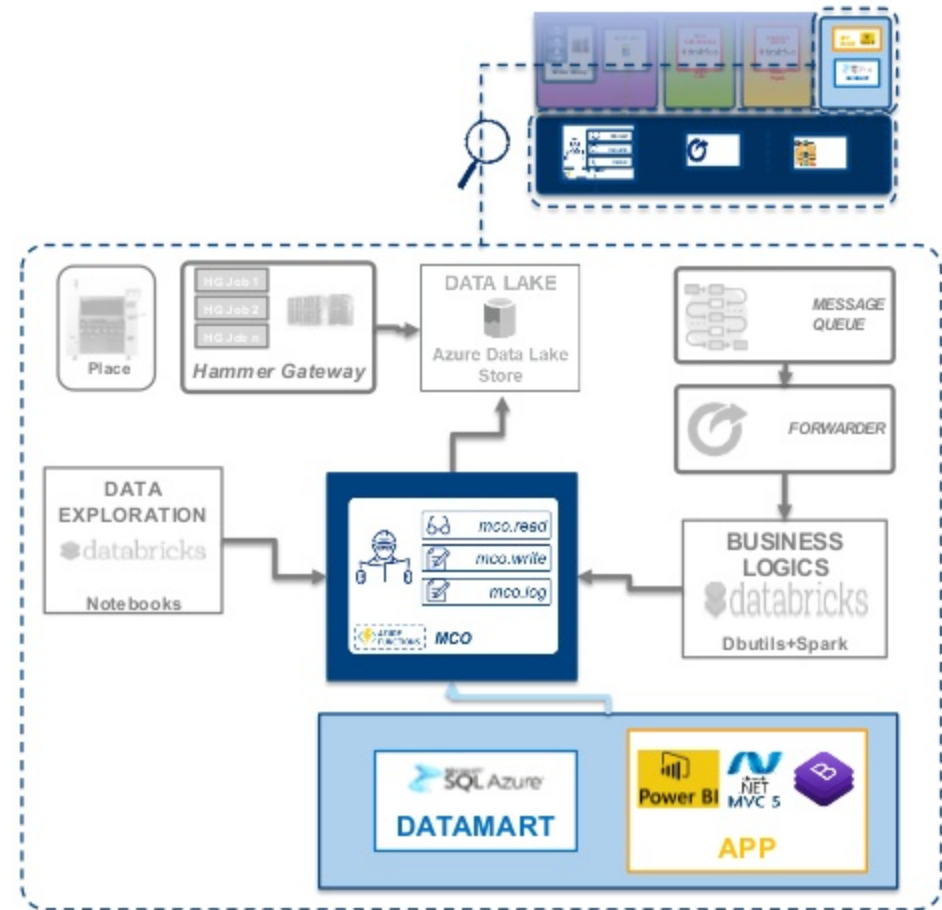
Ruby on Rails Web Development
Specialization · Johns Hopkins University



Web Design for Everybody (Basics of Web Development and Coding)
Specialization · University of Michigan



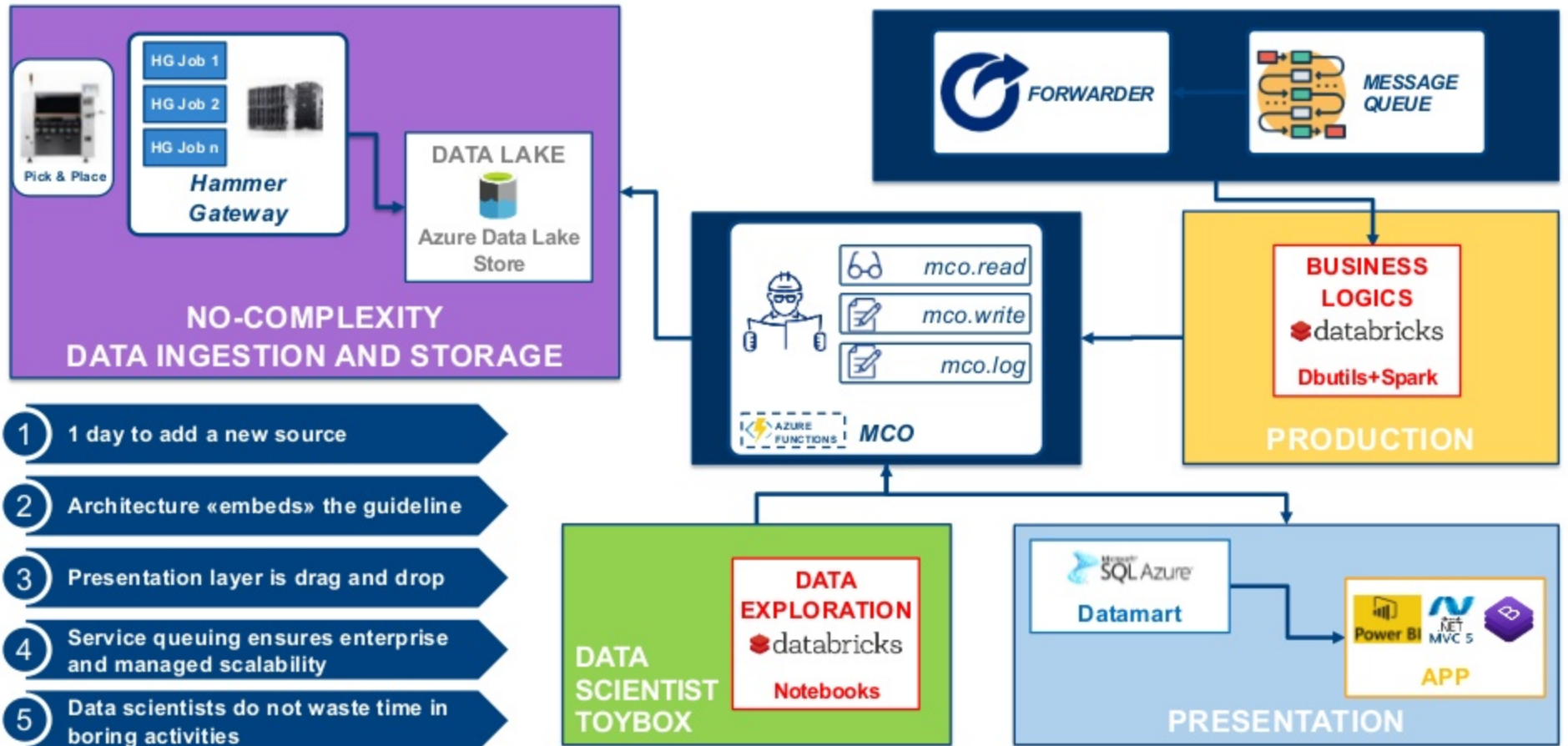
Introduction to Web Development
Course · University of California, Davis



#SAISEnt4

21

Enterprise Architecture (“...and in the darkness bind them”)



- 1 1 day to add a new source
- 2 Architecture «embeds» the guideline
- 3 Presentation layer is drag and drop
- 4 Service queuing ensures enterprise and managed scalability
- 5 Data scientists do not waste time in boring activities

"I have done the deed. Did you hear a noise?"

- 1 Production process is well known
- 2 Data source is clearly defined
- 3 Need is raised by plant people
- 4 Algorithmic challenges are clear



"The Guide says there is an art to flying", said Ford, "or rather a knack. The knack lies in learning how to throw yourself at the ground and miss."

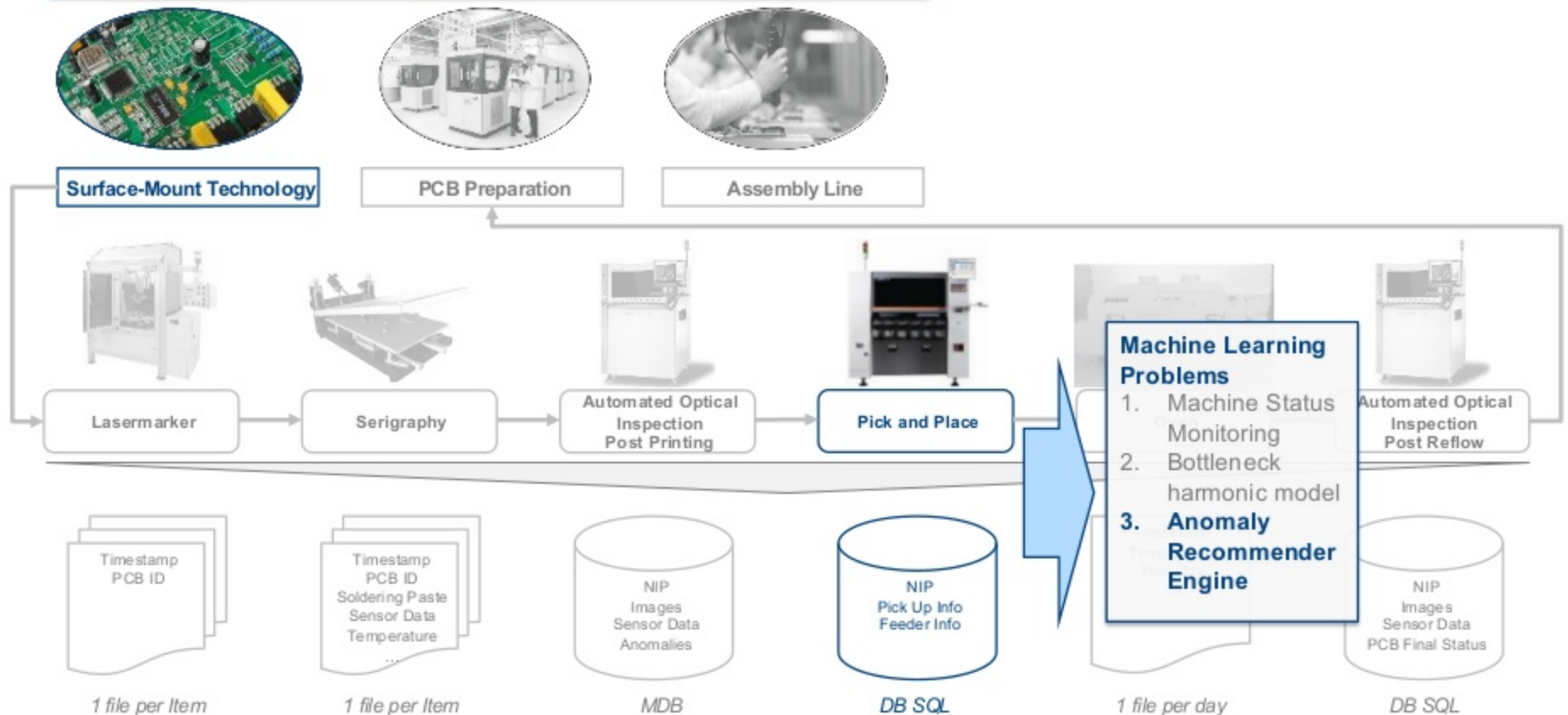
Success!!!



#SAISEnt4

The Surface-Mount Technology (SMT) project

Pre Production & Assembly Line



Anomaly Recommender Engine



Description

A **summary dashboard** shows the health of each part of the line. A drill down with details is available.

Use Case

Support maintenance team to **prioritize standard** and extraordinary maintenance activities.

Benefit

Reducing machine stoppage losses per year per line.
Down time reduction.

Much ado about nothing... ?

Pre Production & Assembly Line



Surface-Mount Technology



PCB Preparation



Assembly Line

Break-even point reached after
8 months

Cost per line reduced by 90%
after the first one

Return On Investment: 12X in 3
years

- **Databricks** and Microsoft **PowerBI** allow a very cost effective first project

- **Hammer Gateway** allows cost effective **ingestion**
- **MCO** enabled Data Scientists to convert notebooks in services with a very very low effort
- **Microsoft Azure** and **Databricks** ensure *endless* scalability

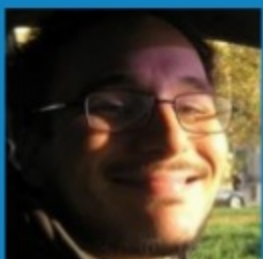
People behind



THE TEAM



**Florindo
PALLADINO**



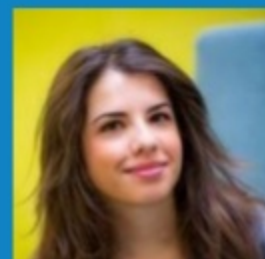
**Andrea
CONDORELLI**



**Giovanni
FAZZI**



**Alessandro
SICOLI**



**Manuela
DETOMASO**

SUPPORTERS



**Heinrich-Gerhard
SCHUERING**



**Dario
CASTELLO**