



Hurence

*IoT: un allégement nécessaire
des technologies Big Data*

Laurence.Hubert@hurence.com





Quatre pôles d'activités

Conseil

Cadrages d'initiatives Big Data
Sizing et installation d'infrastructures Big Data
Développements Big Data du POC à la mise en production



Formations

Plus de 25 modules de formations Big Data
Hadoop, bases NoSQL, moteurs de recherche, développements sur stacks diverses temps réel ou non, machine learning, deep learning etc.



~ 50% des acteurs du CAC nous font confiance

Support

Support d'infrastructures Hadoop, Smack

Monitoring, configuration, patches, upgrades



Produits

Suite LogIsland

Traitement temps réel d'événements sur des grosses volumétries (open source basé sur Spark / Kafka)



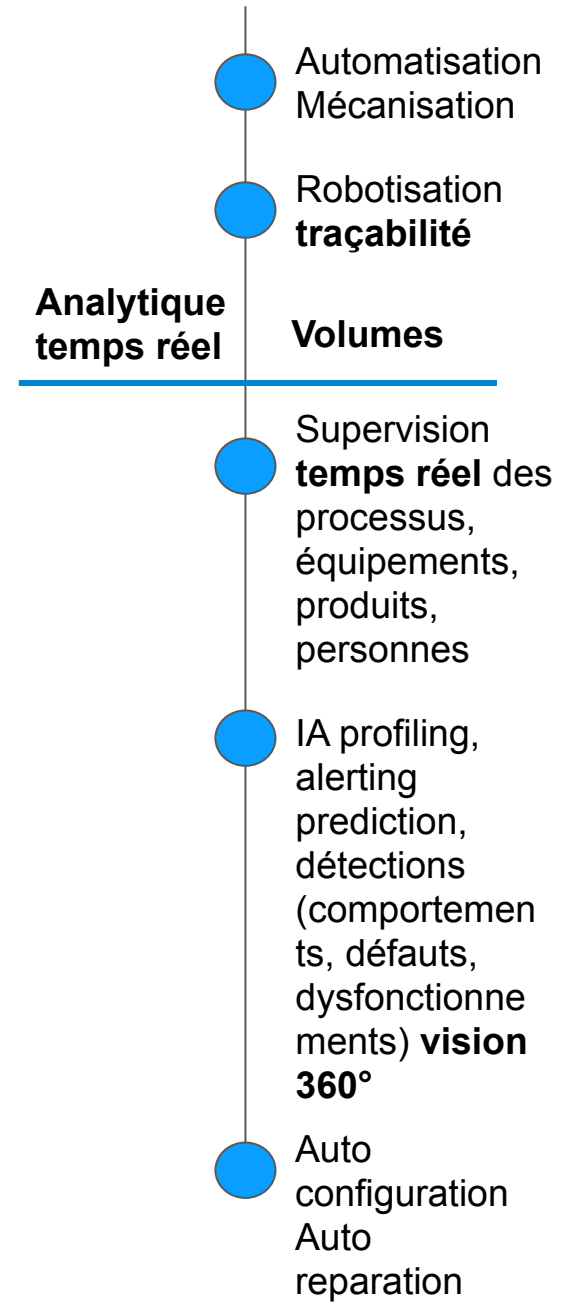
Equipe de senior lead dev anciens de Xerox, IBM, HP, Sun Microsystems, Oracle, Qwant etc.



La technologie induit une révolution culturelle

L'internet, la 4G et bientôt la 5G et leurs applications ont changé la relation aux **hommes**, à la **distance**, au **temps** et aux **choses**:

- Internet relie l'ensemble de l'humanité sur le même réseau.
- La mobilité : tout le monde peut se connecter de partout et réaliser des travaux à distance avec de nouveaux modes d'organisation et d'entreprises intégrées.
- Le temps réel abolit les barrières temporelles
- L'IA donne de la vitesse à l'identification de problèmes
- l'IOT permet de faire entrer le digital dans la matérialité physique de nos vies. Nos "choses" (équipements, produits) sont connectées et on peut interagir avec.



Autonomie des "choses"



Volumétries

=

Data Lake et ses applications



Data Lakes

From tiny ones to XXL ones



Un super-calculateur Low Cost...



- Des technologies matérielles plutôt « low-cost » (commodité)
- Des technologies logicielles plutôt « low-cost » (open source)
- Une facilité à rajouter de la puissance par simple ajout de machines
- **Scalabilité horizontale** versus scalabilité verticale
- Une facilité à dé-commissionner du matériel défaillant ou obsolète
- Une redondance des serveurs clés (les serveurs sont en **High Availability**)
- **Aucun SPOF** si on configure bien son cluster (théorie)
- **Zéro “downtime” (interruption de service)**



Un super-calculateur Low Cost...



- Des technologies matérielles plutôt « low-cost » (commodité)
- Du **stockage “attaché”**
- Des technologies logicielles plutôt « low-cost » (open source)
- Une facilité à rajouter de la capacité par simple ajout de machines (CPU et stockage)



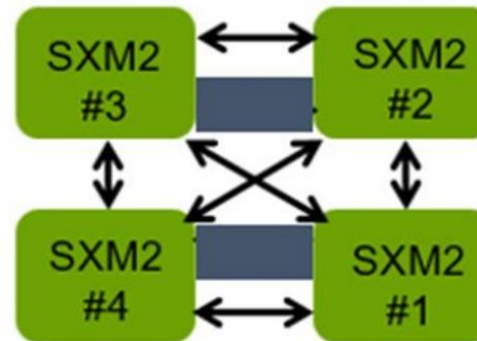
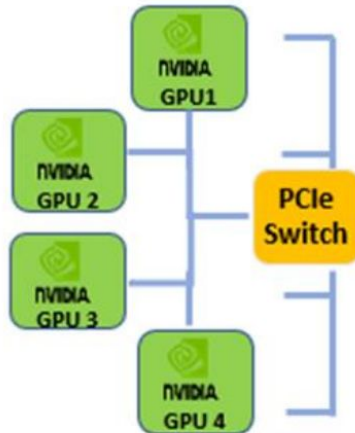
Traditionnellement pour les puristes...

Un Data Lake minimum c'est deux racks de serveurs physiques



Pour l'IA des machines supplémentaires...

Une petite infrastructure de machine learning à base de GPUs (machines dédiées avec 4 cartes NVIDIA NVLink 16 ou 32 Gigas) si possible dans le cadre d'un Data Lake Hadoop/Spark) existant ou dans le cloud (offre à l'usage)

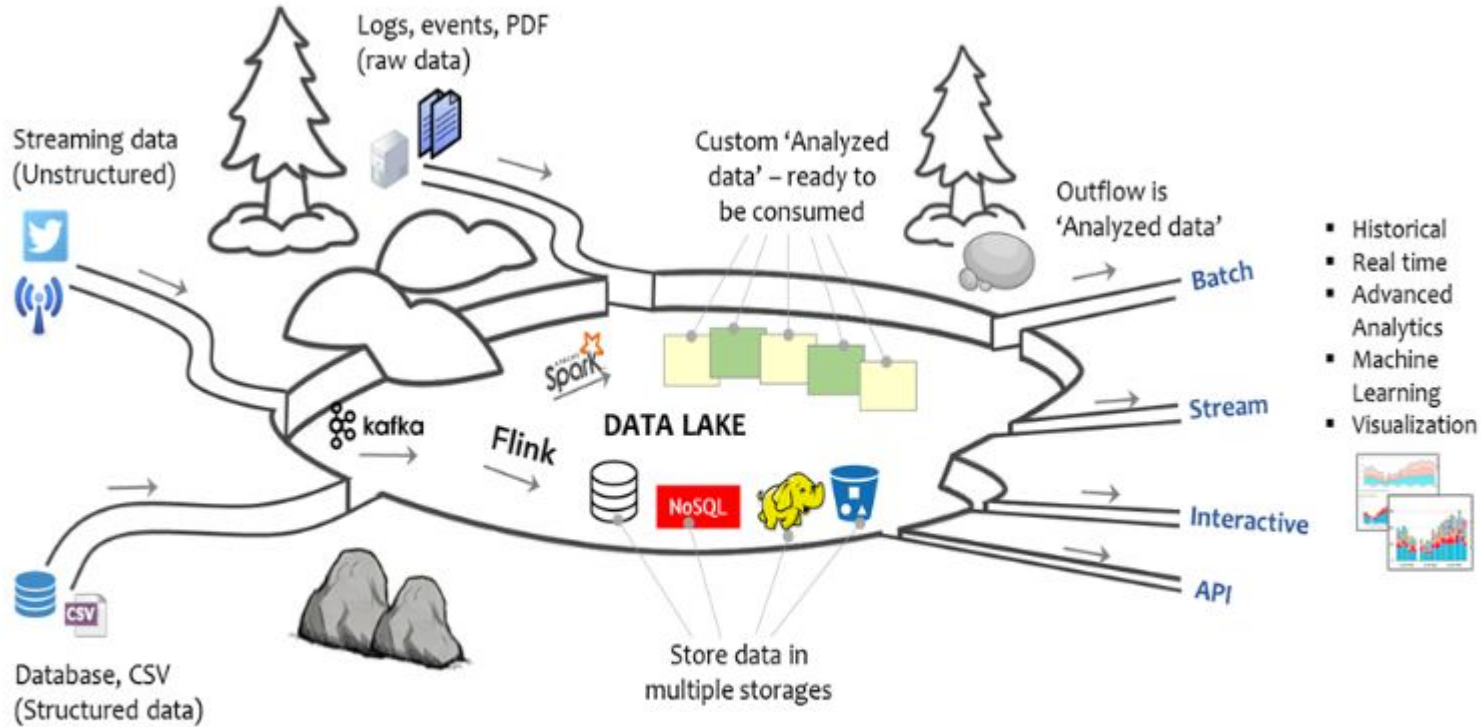


Sans GPUs l'entraînement des modèles deep learning prend trop de temps

Faire coopérer les processeurs GPUs est un accélérateur
=> NVLink en attendant les nouvelles génération de bus PCI



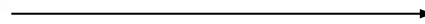
The Data Lake : a pretty “static” thing at the very beginning



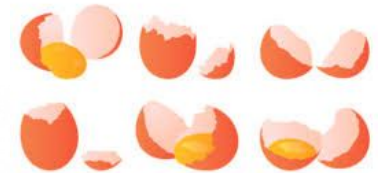
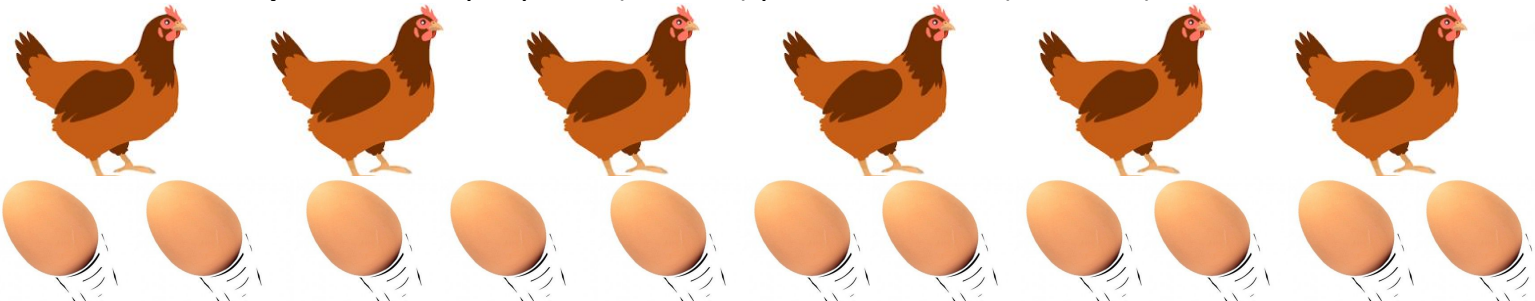
Data Lake Representation



Le Big Data expliqué: le célèbre Map Reduce ...



Map tasks : chaque poule (serveur) pond des oeufs (résultats), sa tâche "coeurs"



stockage distribué

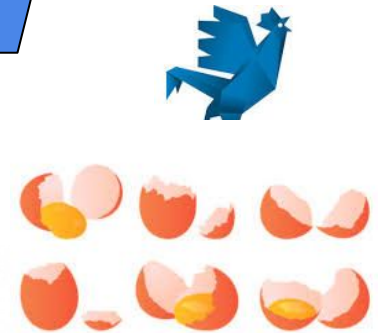
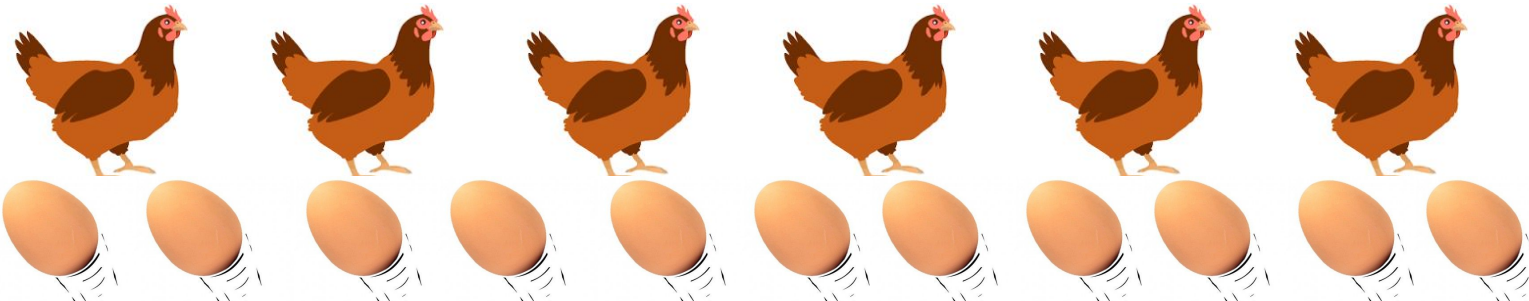
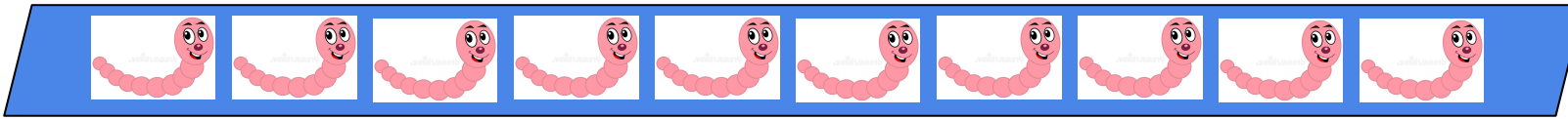




Le Big Data expliqué: le stream processing



Evénement: un ver de terre passe à portée de bec...

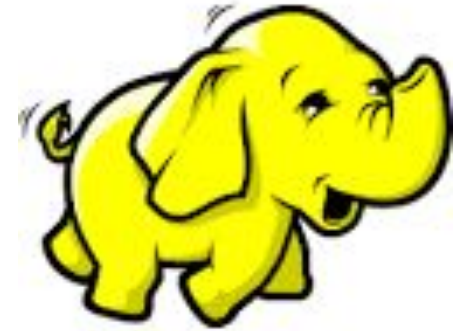


stockage distribué





Les technologies les plus employées



Hadoop

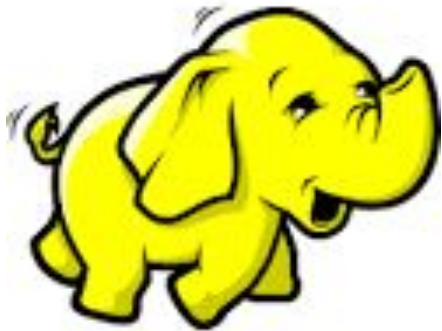
- HDFS pour le stockage distribué (S3 dans Amazon, ...)
- Yarn pour la gestion des ressources des machines
- Kafka un bus de messages sur des grosses volumétries
- Spark calculs de “type” Map Reduce (filtres, jointures, groupBy, etc.)
- Spark streaming pour le stream processing (ou Flink, Storm etc.)
- Des surcouches pour éviter de programmer (LogIsland)
- Des outils de gouvernance de données (Atlas, etc.)
- Des outils de sécurisation des données (Kerberos, Ranger)
- Des outils de monitoring
- Des outils de configuration

Dans le cloud on retrouve plus ou moins les mêmes... Spark étant une constante...



Allo Houston: on a un problème !

- Près des machines d'une usine on ne peut mettre qu'un petit serveur mais on veut les mêmes technologies
- Dans des bâtiments municipaux on ne va pas mettre deux racks mais on veut les mêmes technologies
- Hors toute la mécanique Big Data traditionnelle des data centers ne rentre pas... et c'est trop lourd à gérer...

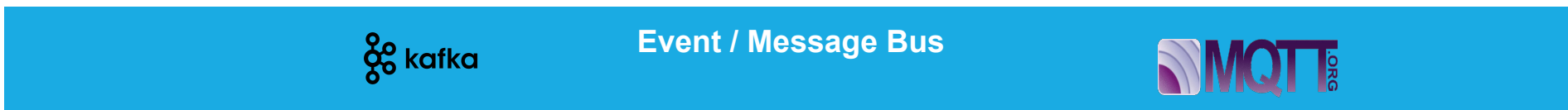
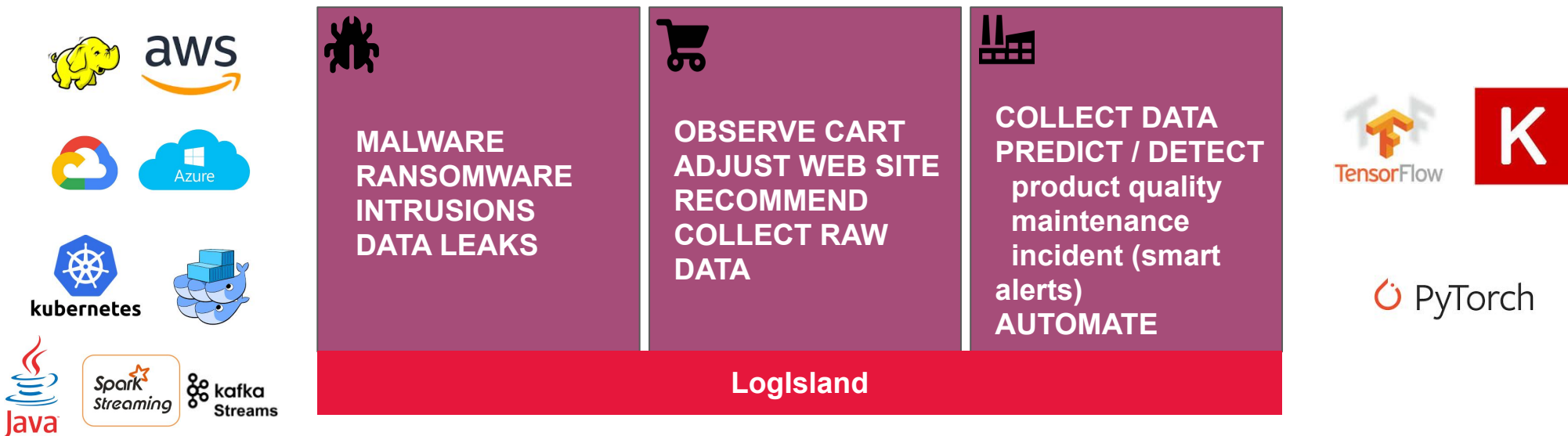




La supervision d'usines...
L'IloT, le début d'une vraie
révolution vers la machine qui
s'auto-diagnostique,
s'auto-reconfigure, s'auto-répare



Event Processing at Enterprise scale

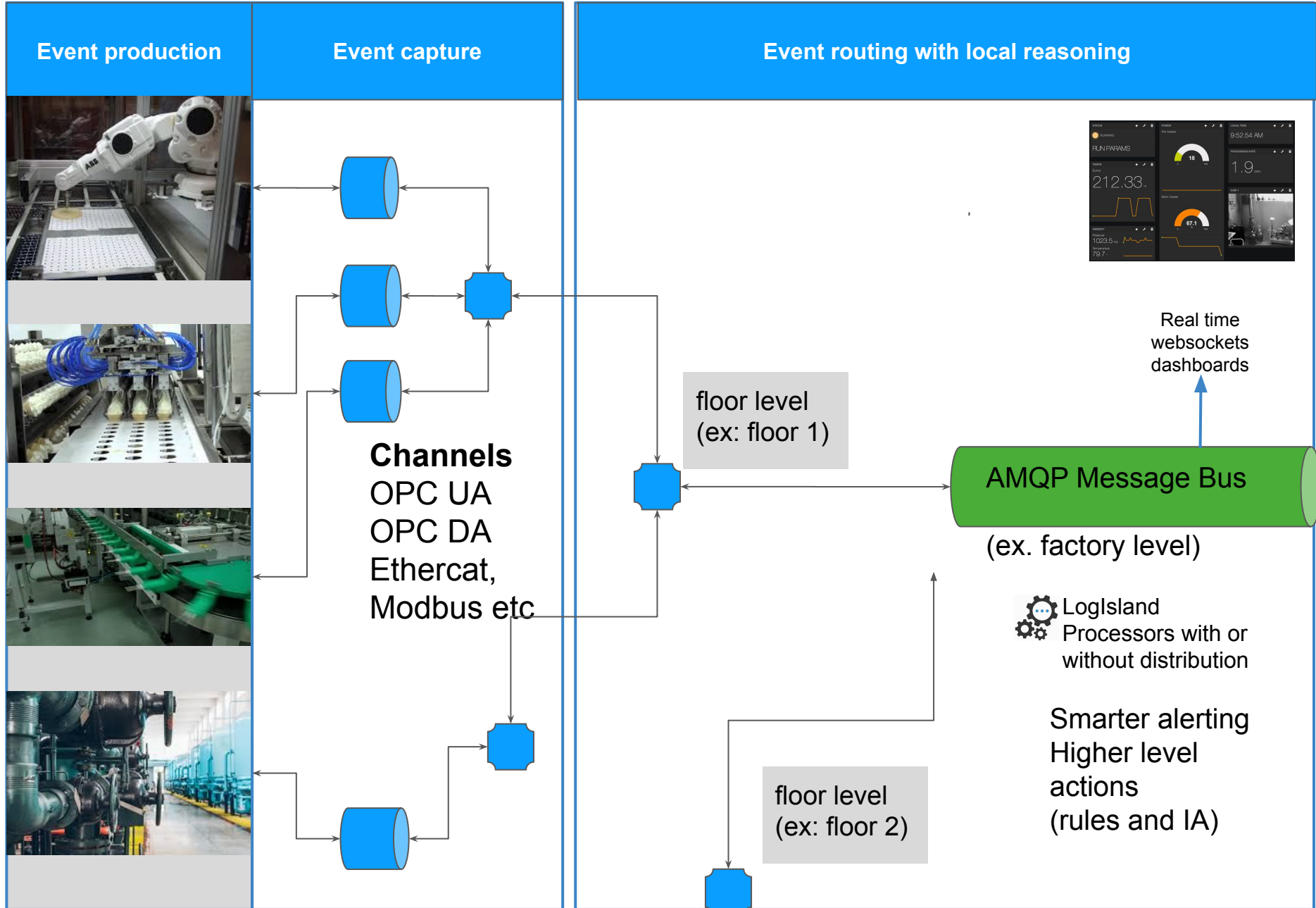




Le edge processing... ou la captation des événements in-situ



LogIsland industrie 4.0 : data flows (edge processing side)



Floor level

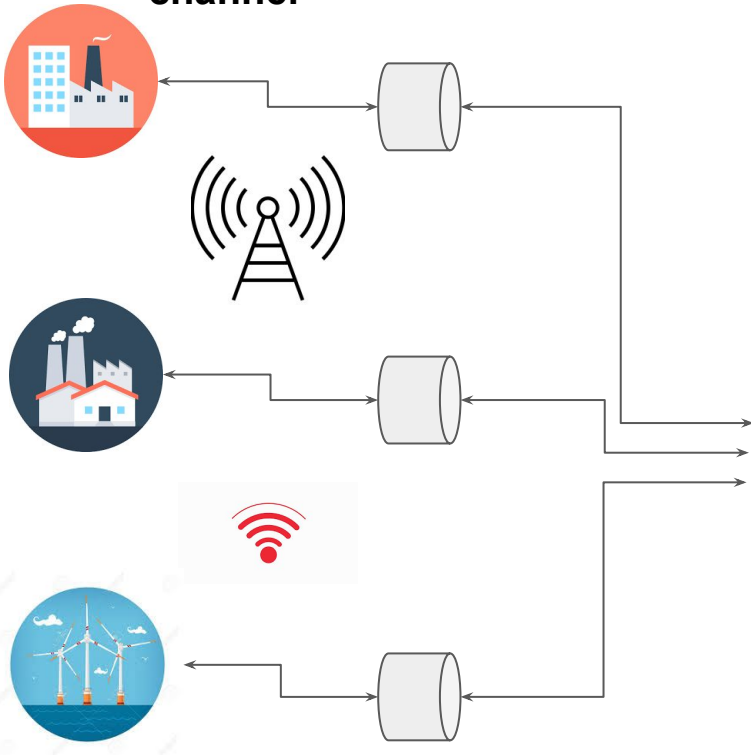
Factory Level



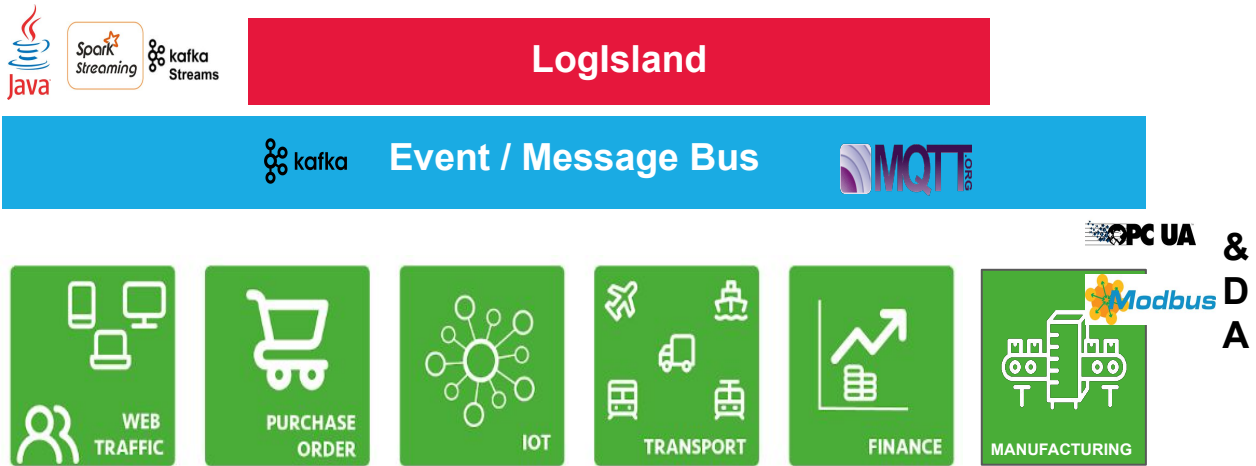


LogIsland industrie 4.0 : data flows (big data - data lake side)

LogIsland
AMQP to
Kafka
channel



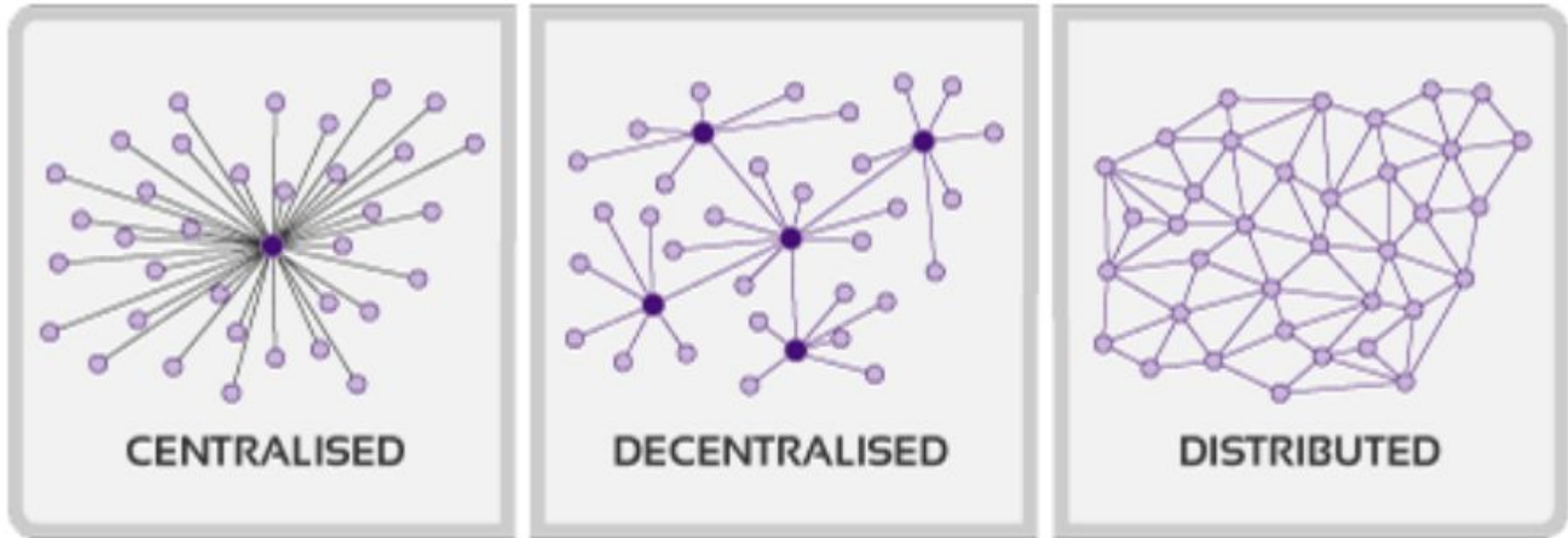
product quality (defect detection)
 predictive maintenance
 anomaly detection
 supervision / smart alerting
 machine auto reconfiguration
 tracking



Data Center or Cloud Level (data lake)



Le pouvoir au peuple ! A bas la centralisation !



- Les décisions ne peuvent pas être laissées au Data Lake tel un pouvoir “dirigiste”
- On veut avoir une décentralisation de certaines décisions, notamment les décisions qui demandent une réactivité !
- Donc une partie des technologies utilisées pour l’analyse et la décision doivent être portées au plus près des machines...

=> notion de EDGE BOX intelligente!



Cerveau reptilien côté edge: le cas usine



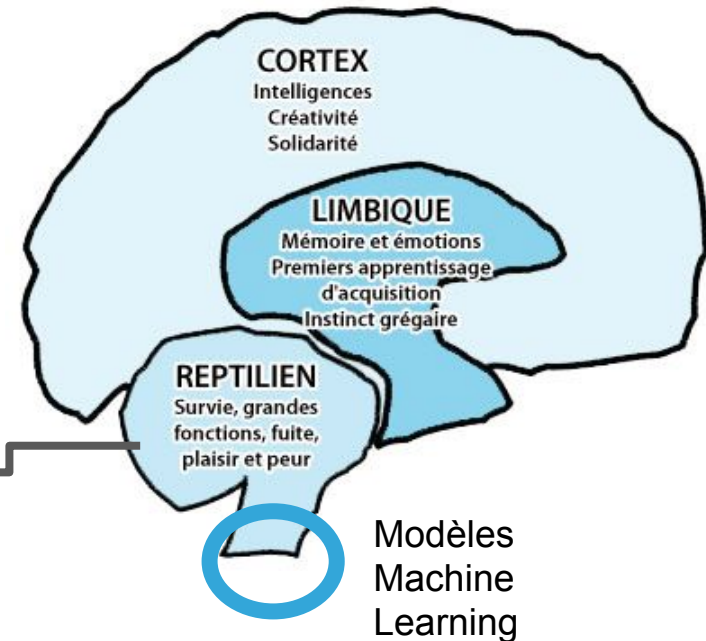
Une machine d'emboutissage qui coûte des millions

Le cerveau reptilien ne peut être dans un data center distant
A terme il sera dans les machines... mais là on doit le concevoir **exogène**

Automates programmables



Cette partie doit marcher sans internet !





Cerveau reptilien côté edge: le cas usine



Une machine d'emboutissage qui coûte des millions

Le cerveau reptilien ne peut être dans un data center distant
A terme il sera dans les machines... mais là on doit le concevoir **exogène**

Automates programmables



channels et bus

Cette partie doit marcher sans internet !



Data Lake

REPTILIEN
Survie, grandes fonctions, fuite, plaisir et peur

Modèles Machine Learning



Cerveau reptilien côté edge: le cas de l'usine CEVESO

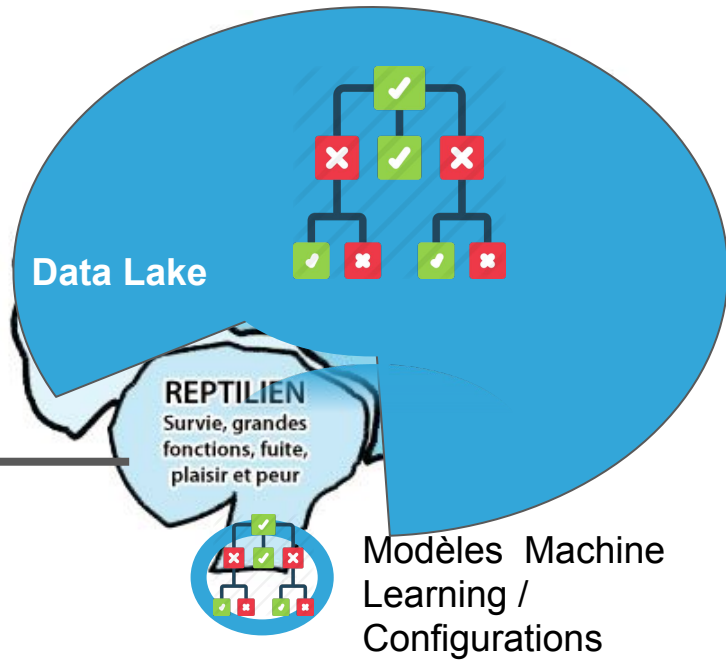


Edge boxes : avec des capteurs et connectées à l'internet/intranet



Cette partie doit marcher sans internet !

Instructions individualisées pour des écoles, des particuliers, des hopitaux





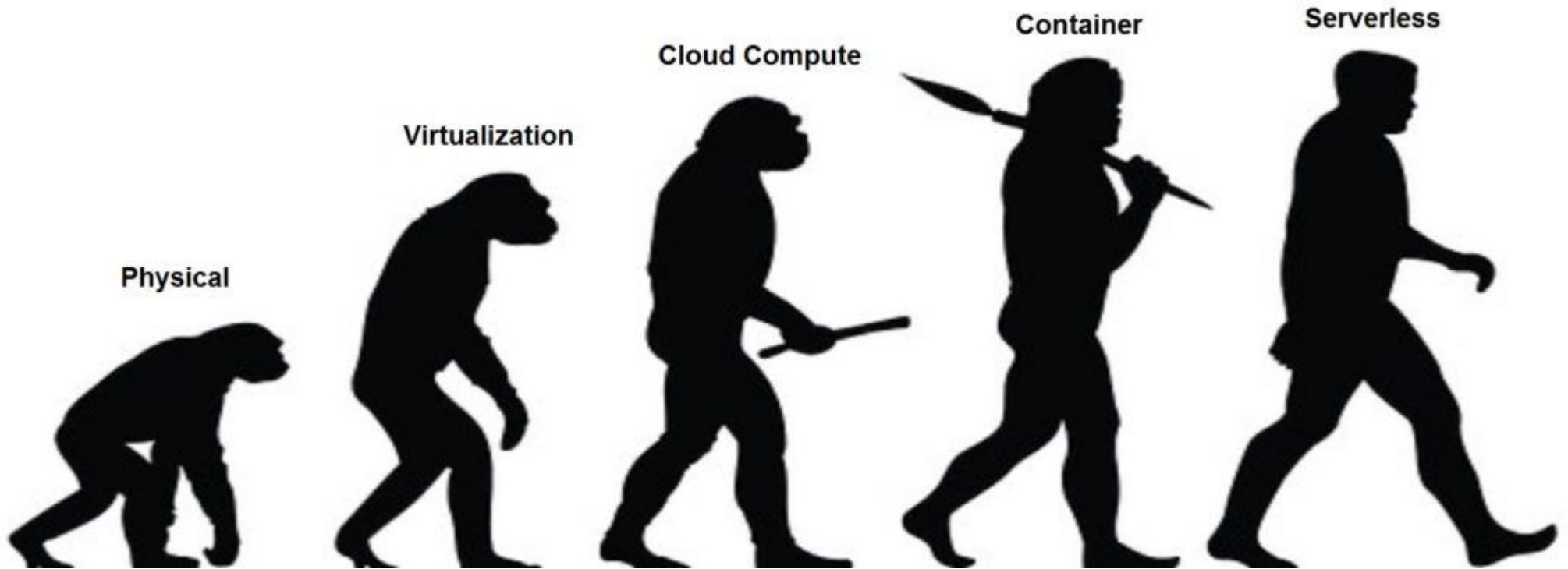
Besoin d'intelligence aussi sur le edge mais...

- On est sur des **petites infrastructures** (des petits serveurs)
- La edge box c'est donc comme votre box internet.
- Elle va avoir un peu d'intelligence mais c'est sur le data lake qu'on construit ses modèles de machine learning, ses configurations de service et on les déploie dans la partie "cerveau reptilien"
- Elle a intrinsèquement du stockage (minimaliste), de la capacité de traiter des événements locaux, de prédire sur la base de ses événements locaux...
- Elle se met à jour de ses modèles et configuration sur événements





L'architecture **serverless** est idéale pour les edge box



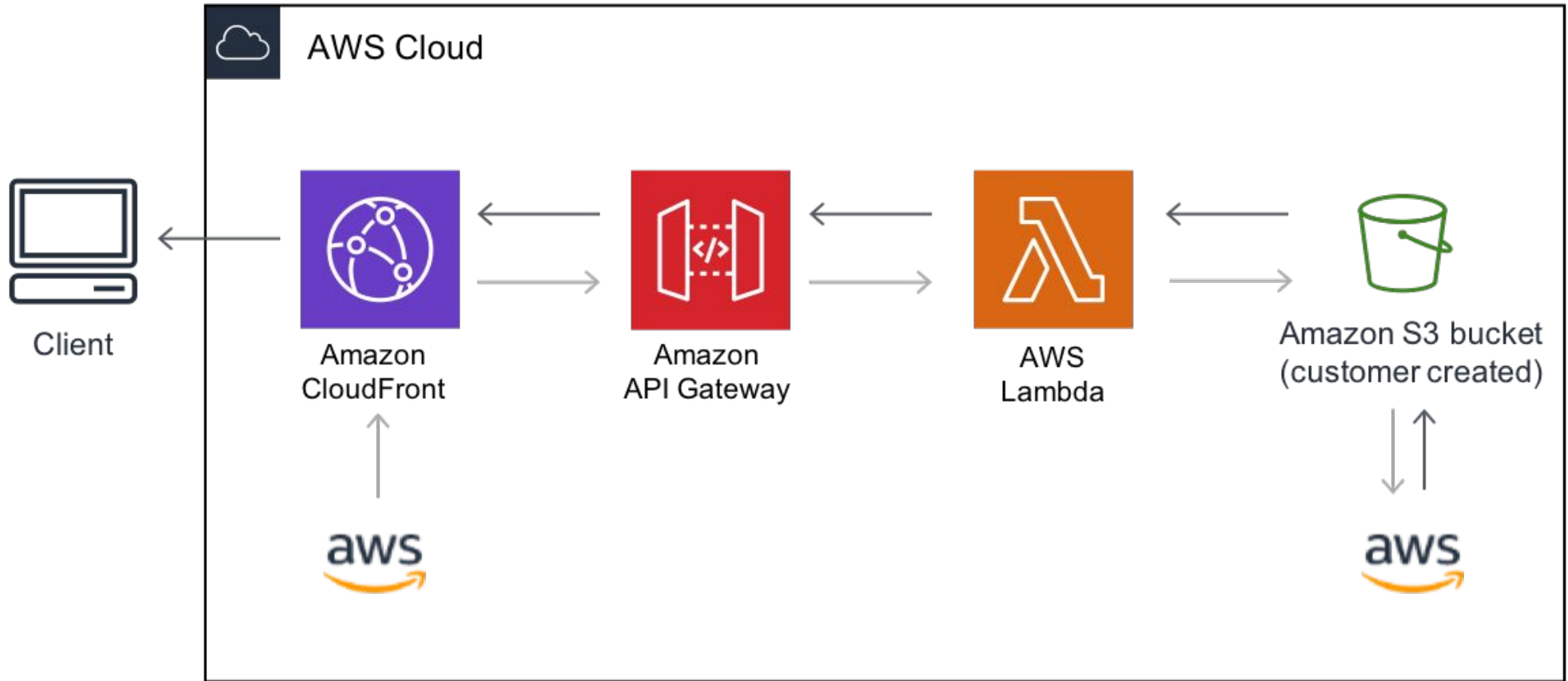
Architecture Serverless : généralement associée à l'idée de FAAS (Function As A Service) - des fonctions mises à disposition - dans des containers - qui s'exécutent sur événement et qui n'ont pas d'état (stateless).

Elles prennent des données en entrée (des événements dans un petit bus, des modèles), les traitent et se finissent (en émettant par exemple de nouveaux événements)

Il y a l'idée que l'on peut ajouter facilement de nouvelles fonctions ou mettre à jour facilement celle qui existent (comme pour votre box internet).



Amazon: les lambdas



Beaucoup d'alternatives existent pour des déploiements "on premise".



Event Processing at Edge Box side...

CONFIGURE
Pipelines &
Parameters

ALERT



PREDICT



ACTION



PLUG AI
Deep Learning

Simple
Functions
& IA models



LogIsland



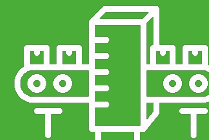
Event / Message Bus (Mini Bus)



IOT



TRANSPORT



MANUFACTURING



Supervision de la fabrication et Smart Alerting



Le challenge des fabs automatisées



Opérateur

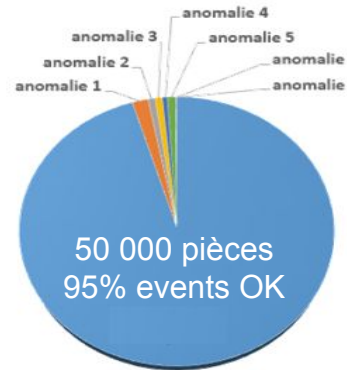
Conducteur de machines

Alimente les machines et gère les encours

Présence physique dans les ateliers de production

Utilise ses 5 sens pour analyser en continu la situation.

**Patrouilles en salle,
Détection temps réel sur le terrain
et diagnostic quasi instantané**



Ce qui était évident

Détection des blocages transitoires, des comportements anormaux

Devient compliqué

Comment alors garantir la productivité de l'unité de fabrication 7j/7 - 365j/an ?

Technicien

Expert domaine technique

Maintenance

Diagnostic Process

Résolution des problèmes et des situations de blocage



**Travail sur PC,
personnel statique**



Houston, j'ai un petit problème avec mon IA ...

31

Une IA ne sait pas généraliser, abstraire un processus complexe. Elle ne sait que restituer ce qu'elle a déjà vu (ce sur quoi on l'a entraînée). Donc on a du "coder" l'expertise des humains et n'utiliser que l'IA pour ce à quoi elle est bien: choisir les paramètres optimaux.



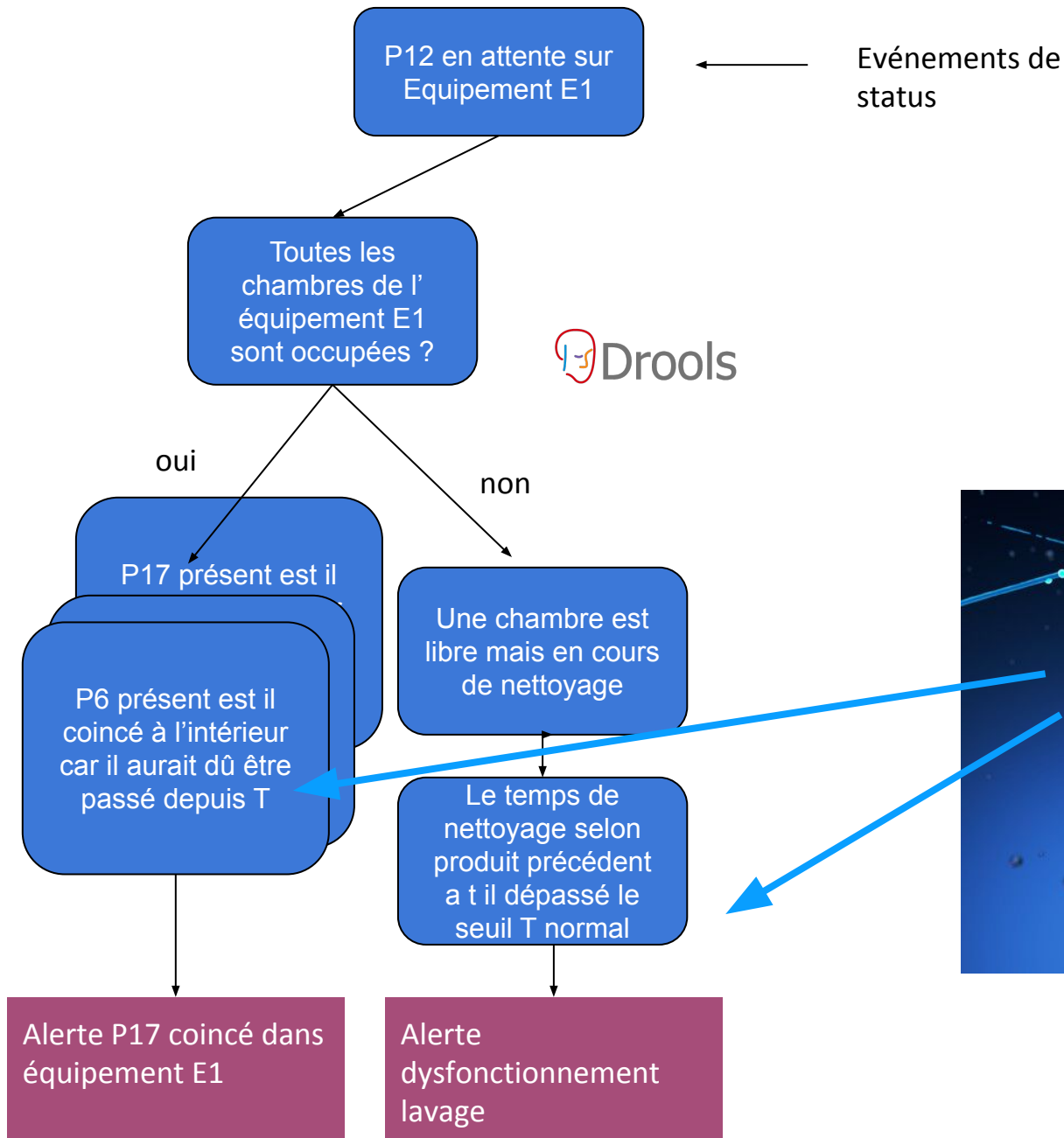
Quand est ce qu'on doit lever une alerte si un lot n'est pas passé dans une machine ?

- Cela dépend de ce que la machine a fait avant (elle peut être en train de se rincer), ou de ce qu'elle fait maintenant (certains dépôts ou abrasions prennent plus de temps selon le produit) etc...

Approche hybride: règles comme dans nos bons vieux systèmes experts et l'IA pour trouver les bons seuils de fonctionnement.



Approche hybride à la détection d'incidents



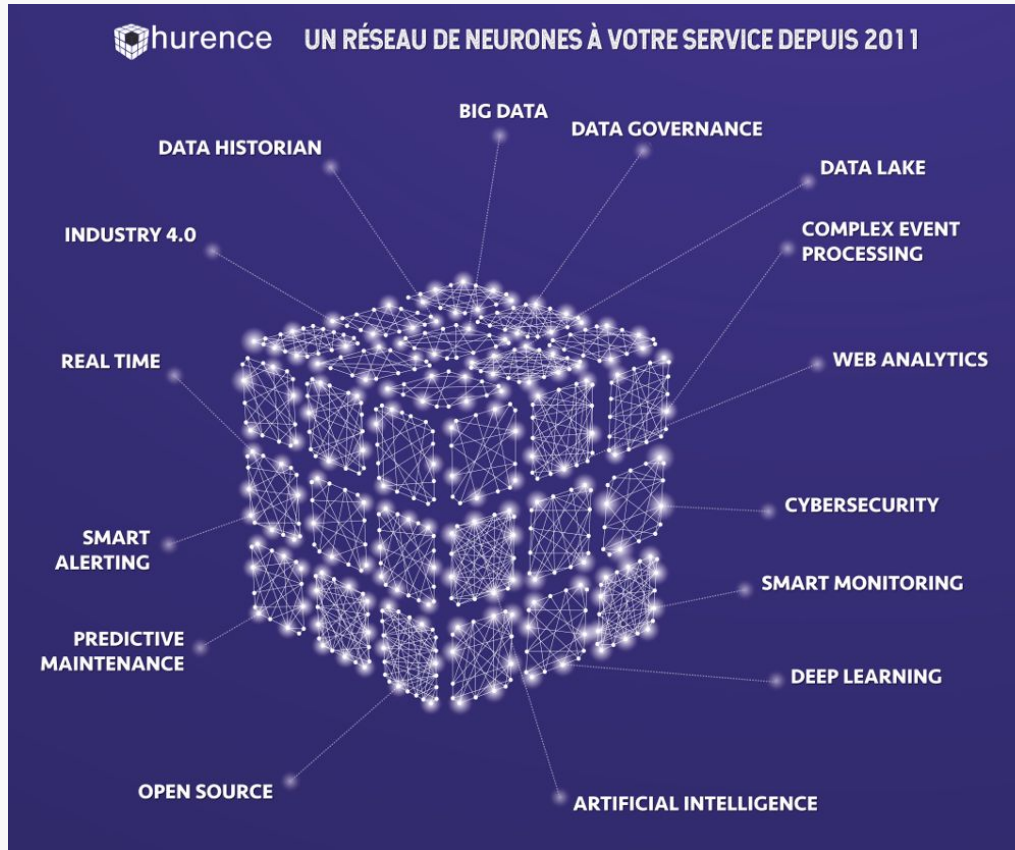


Conclusion

- On redécouvre de vieilles technologies de type systèmes experts que l'on remet au goût du jour (parallélisation, couplage IA).
Hybridation IA et autres techniques
- On ajoute cette intelligence “hybride” sur des EDGE BOX en déployant une portion “épurée” d'un système Big Data temps réel
- On adjoint des mécaniques de déploiement de services via des infras serverless (openFaas ou autres...)
- Prochaine étape : outils de supervisions de nos edge boxes à très grande échelle ;)



Merci



Laurence.Hubert@hurence.com

Mon neurone ou le réseau entier
sont à votre disposition