DATA & CLOUD

Expertise

Fort de plus de 20 ans d'expériences en ***architecture technique***, avec pour domaines d'expertise le ***Cloud***, la ***Data Science*** et les architectures ***Big Data***, DATA & CLOUD Expertise vous apporte son savoir-faire dans le cadre de ses missions, et notamment :

Expertise Cloud
Move2Cloud, Optimisation des coûts, Serverless, etc.

FinOps
Démarche, Optimisations, Outillage (Showback, etc.)

Architecture
Cloud / OnPremise / Hybride, Traditionnelle / BigData, etc.

Conseil & Audits
BDD, performance, passage à l'échelle, DevOps, etc.

Data Science
Extraction, Stockage, Intelligence artificielle, etc.

Formations
Cloud, Data Science, BigData, Databases, etc.

POUR EN SAVOIR PLUS : CONTACT@DATA-AND-CLOUD.COM · 06.86.98.15.80

PORTFOLIO DATA SCIENCE / DATA ENGINEERING / CLOUD

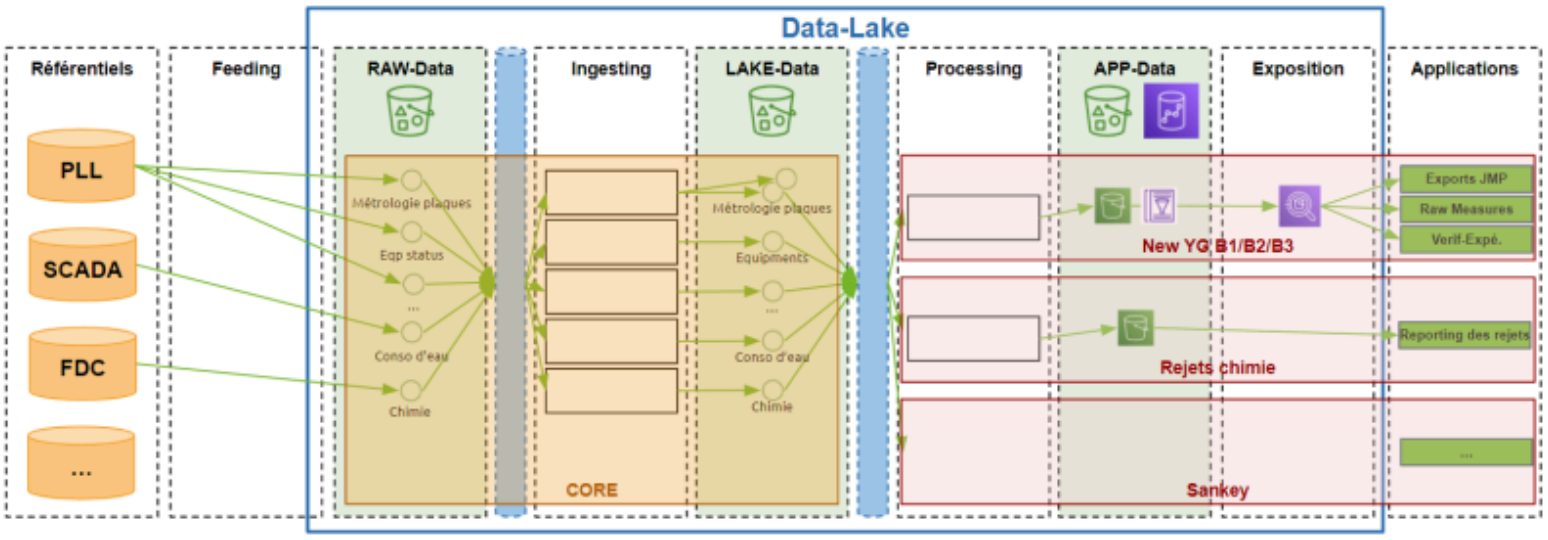
Exemples de projets *Data Science*, *Data Engineering* et *Cloud* menés ces 5 dernières années.

DATALAKE DANS LE CLOUD AWS

2022

AWS S3, STEP FUNCTION, LAMBDA, CLOUDFORMATION, ATHENA, GLUE, SNS, CLOUDWATCH, ORACLE

Structuration du Datalake sur le Cloud AWS, fédérant et centralisant diverses datasources SOITEC (MES, SCADA, SCP, ERP, etc.), et premiers use-cases d'analyse des rendements (Athena, SAS JMP), et des détracteurs dans la consommation d'eau (Redshift, PowerBI).

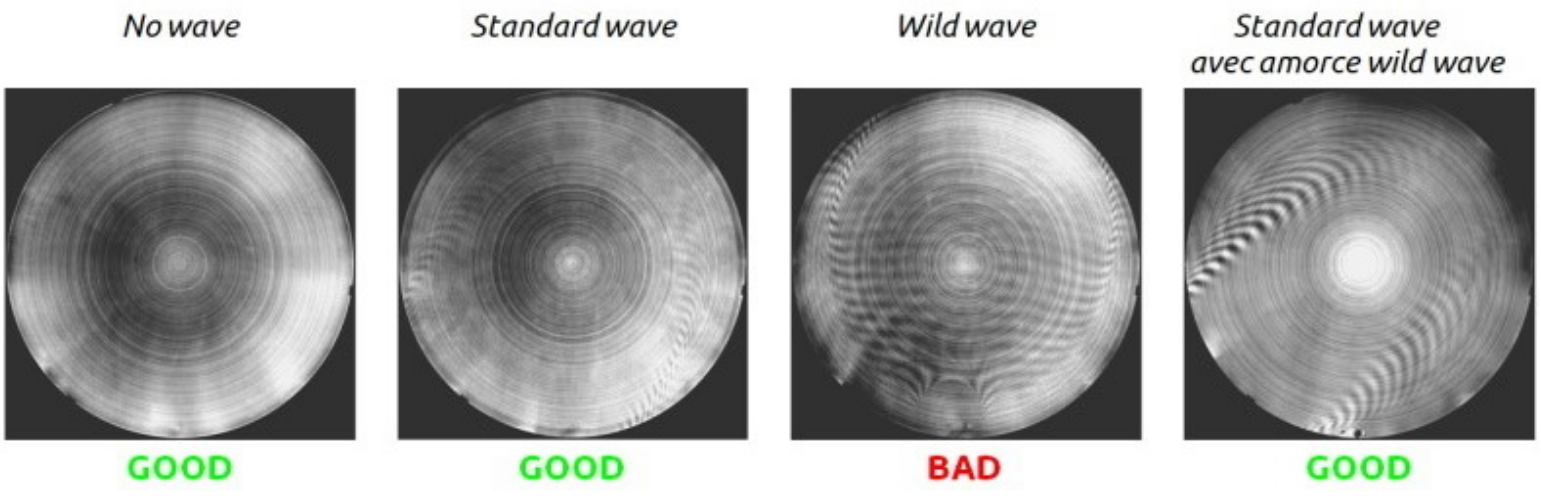


DÉTECTION DE DÉFAUTS

2022

DEEP LEARNING (TENSORFLOW, TF.DATA)

Classification d'images pour une détection automatique de défauts rares et/ou peu marqués.

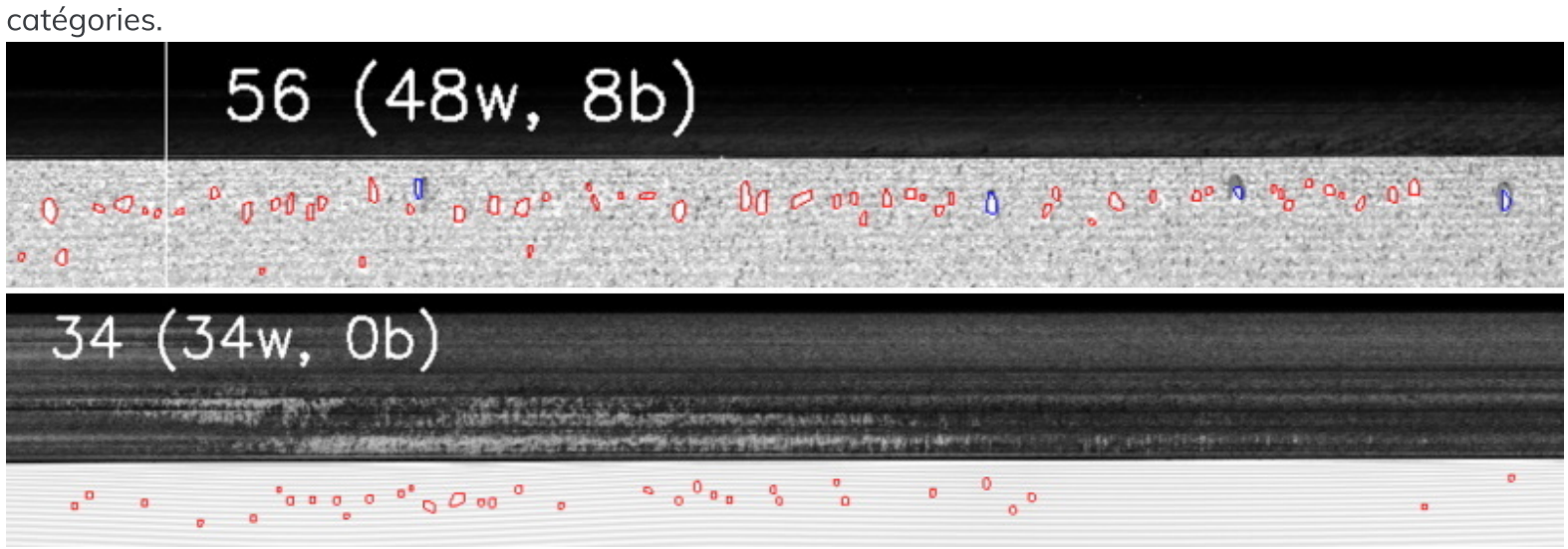


DÉTECTION DE DÉFAUTS

2022

TRAITEMENT D'IMAGES (CONVOLUTION, SOBEL, OPEN/CLOSE)

Analyse d'images de contrastes variés, afin d'évaluer automatiquement le nombre de picots de diverses catégories.

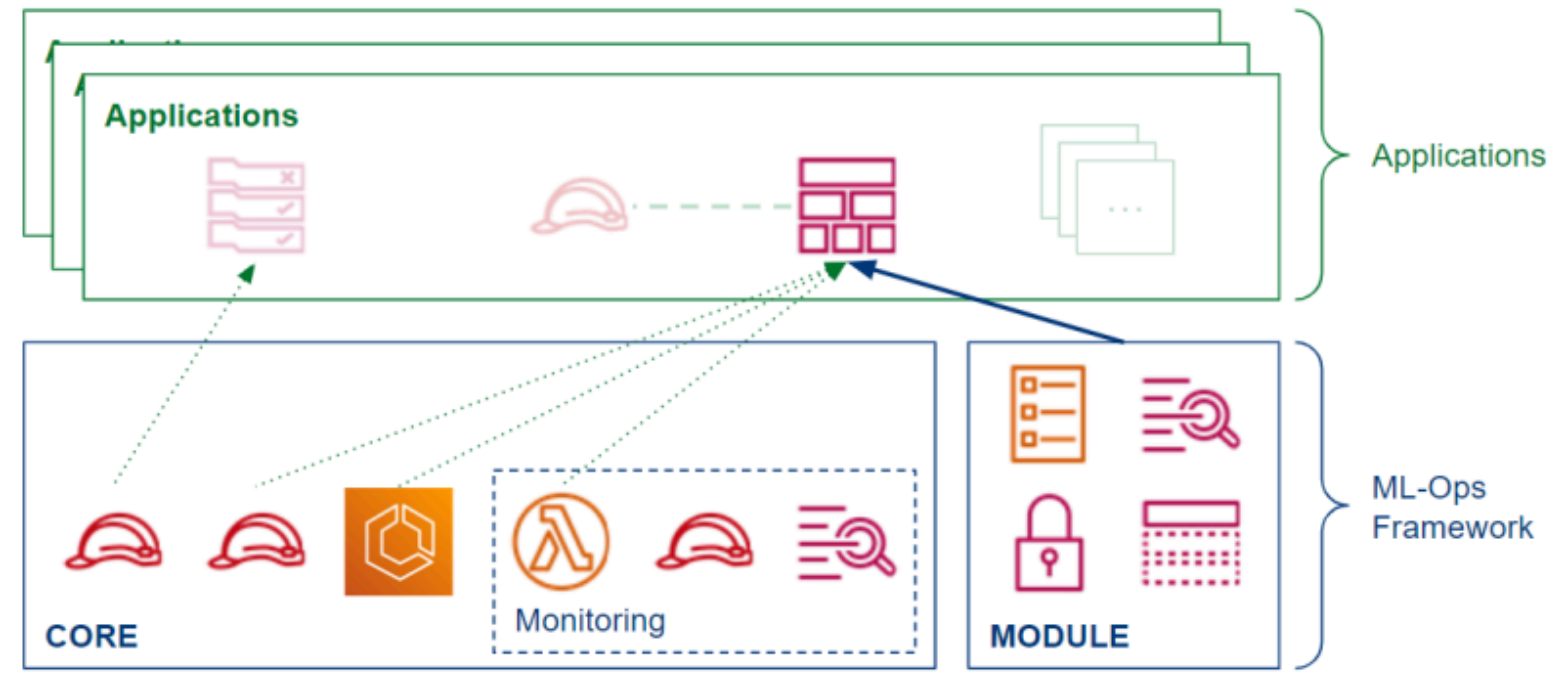


DÉVELOPPEMENT D'UN FW DE ML-OPS POUR LE CLOUD AWS

2021

MLFLOW, DOCKER, AWS (ECS, FARGATE, CLOUDFORMATION)

Socle technique partagé à destination des applications de Data Science et des besoins serverless, visant à accélérer les déploiements, homogénéiser les architectures techniques, imposer certains standards d'industrialisation (supervision, sécurité, etc.), tout en réduisant les coûts de BUILD et de RUN. Ce framework est compatible avec ML-Flow, et permet de mutualiser les ressources de Tracking.

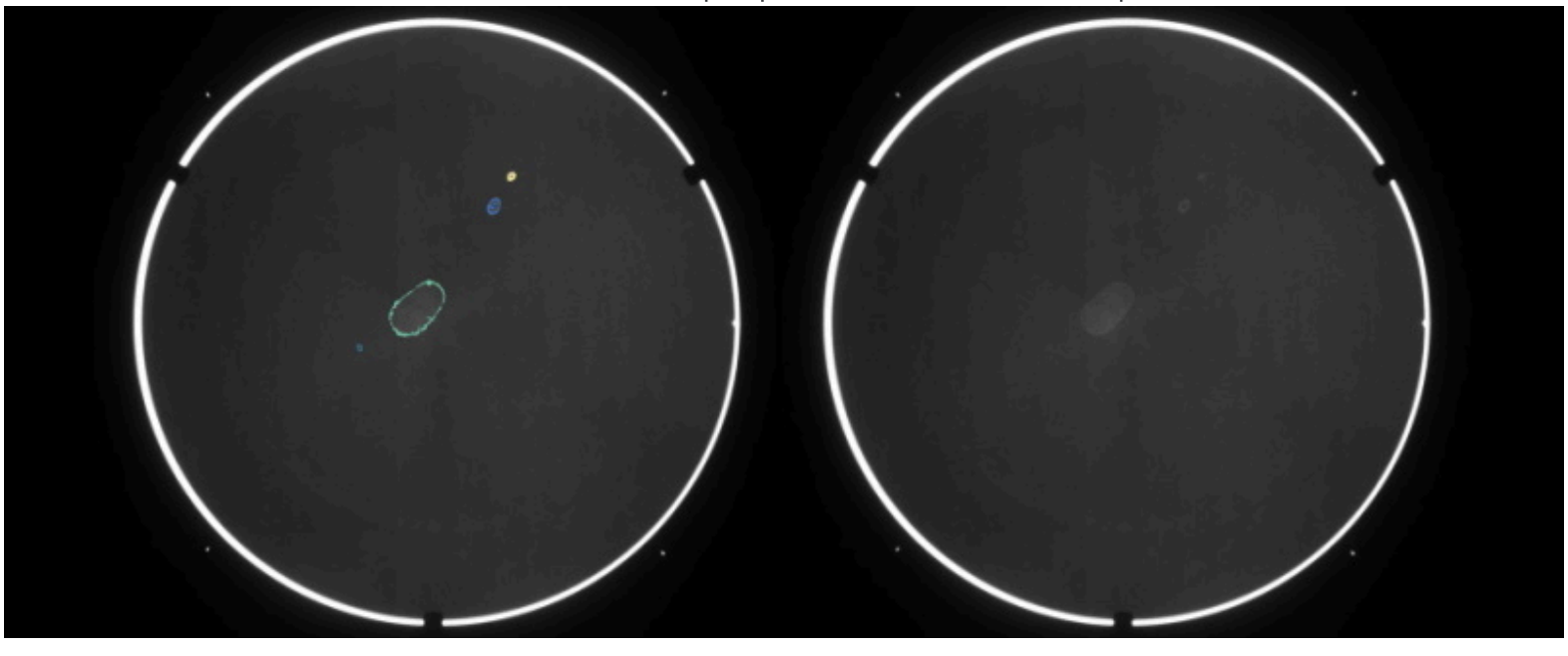


DÉTECTION DE BULLES

2021

TRAITEMENT D'IMAGES (CANNY), APPROCHE STATISTIQUE & CLUSTERING (DBSCAN)

Automatisation de la détection de bulles sur des plaques afin de limiter les inspections manuelles.

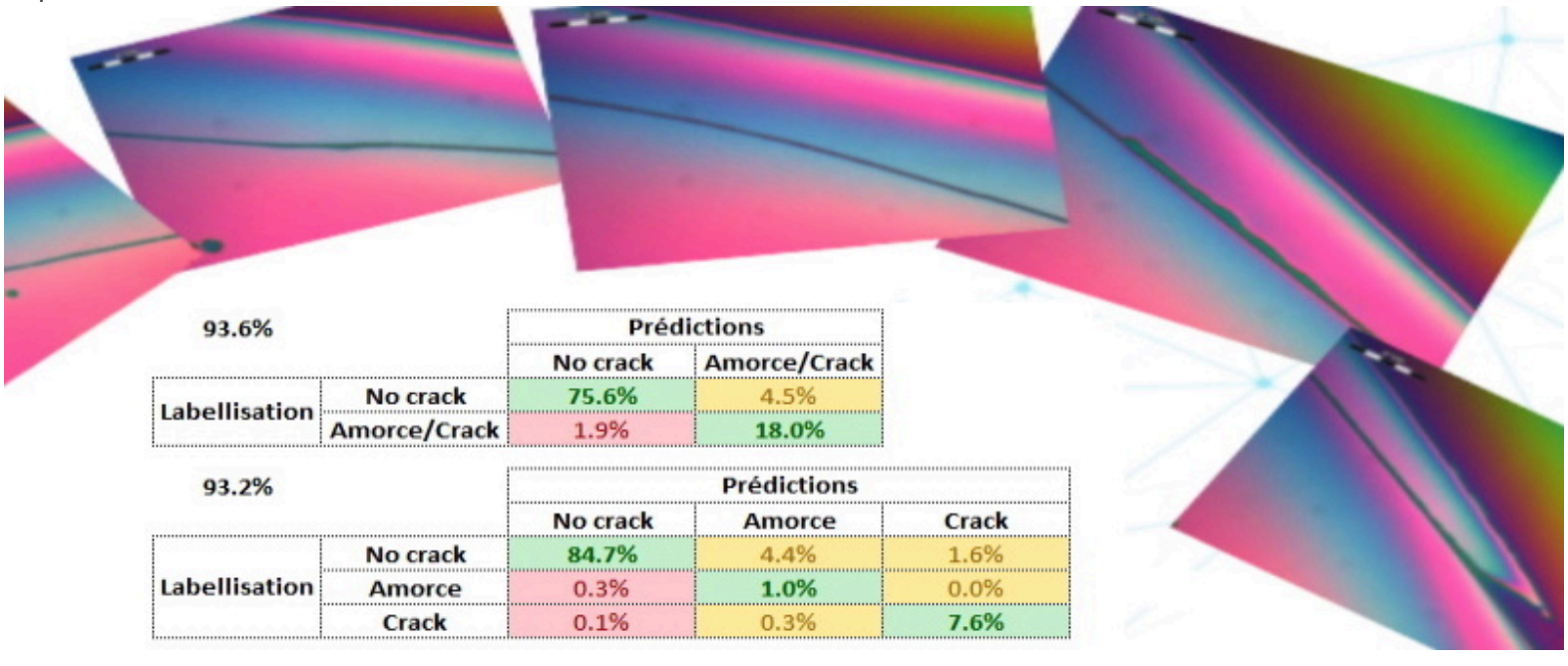


CLASSIFICATION DE PLAQUES

2021

MACHINE LEARNING (RANDOMFOREST), DEEP LEARNING (CNN)

Transcription de cartographies de défauts en "features" servant d'entrée à des algorithmes de classification supervisés.



93.6%

		Prédictions	
		No crack	Amorce/Crack
Labellisation	No crack	75.6%	4.5%
	Amorce/Crack	1.9%	18.0%

93.2%

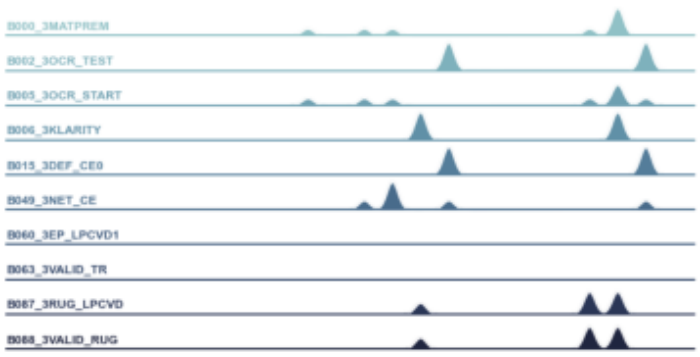
		Prédictions		
		No crack	Amorce	Crack
Labellisation	No crack	84.7%	4.4%	1.6%
	Amorce	0.3%	1.0%	0.0%
	Crack	0.1%	0.3%	7.6%

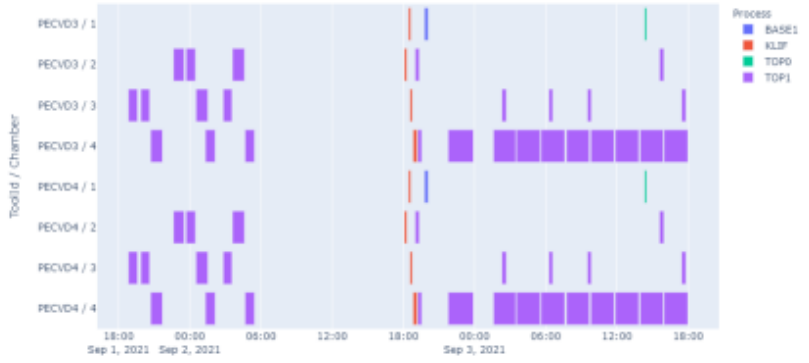
ANALYSE DE CONTENTIONS

2021

SEABORN, PLOTLY

Accompagnement des métiers dans leurs analyses, via des travaux d'extraction et de visualisation de la data.



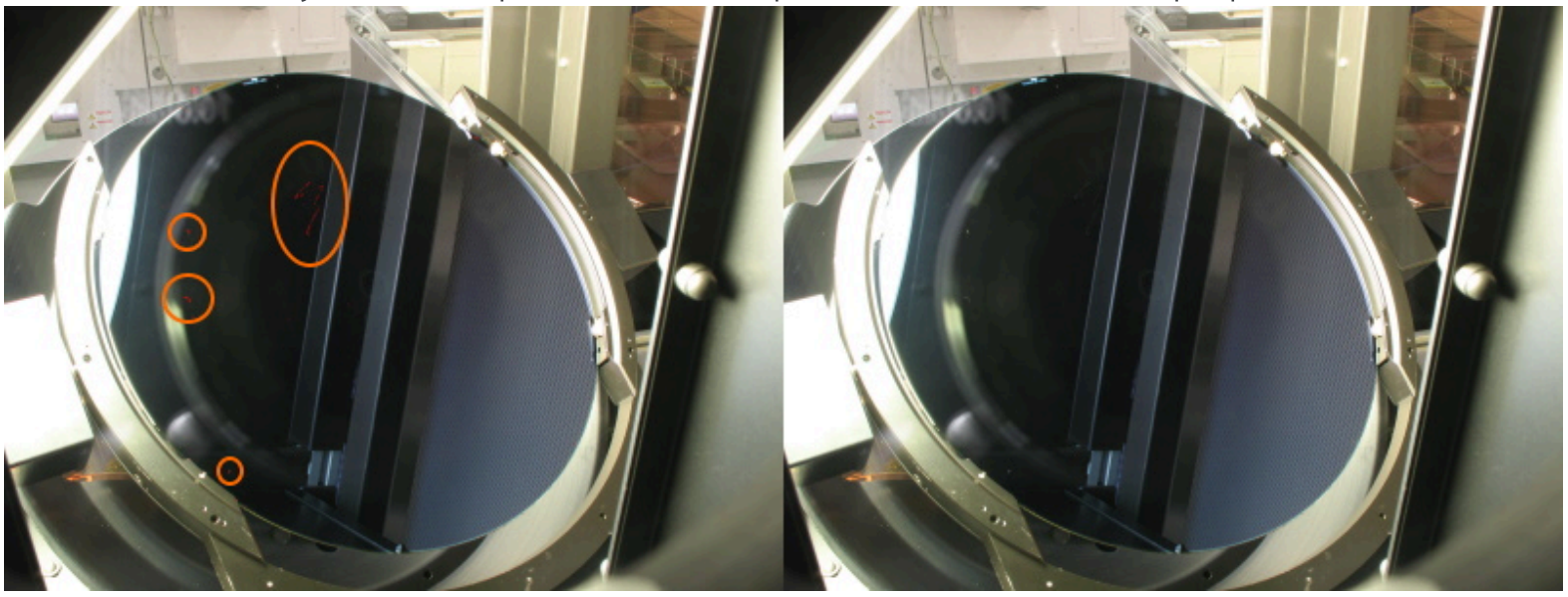


DÉTECTION DE RAYURES

2021

TRAITEMENT D'IMAGES & CLUSTERING (DBSCAN)

Détection de fines rayures sur des prises de vues obliques en lumière naturelle de plaques.

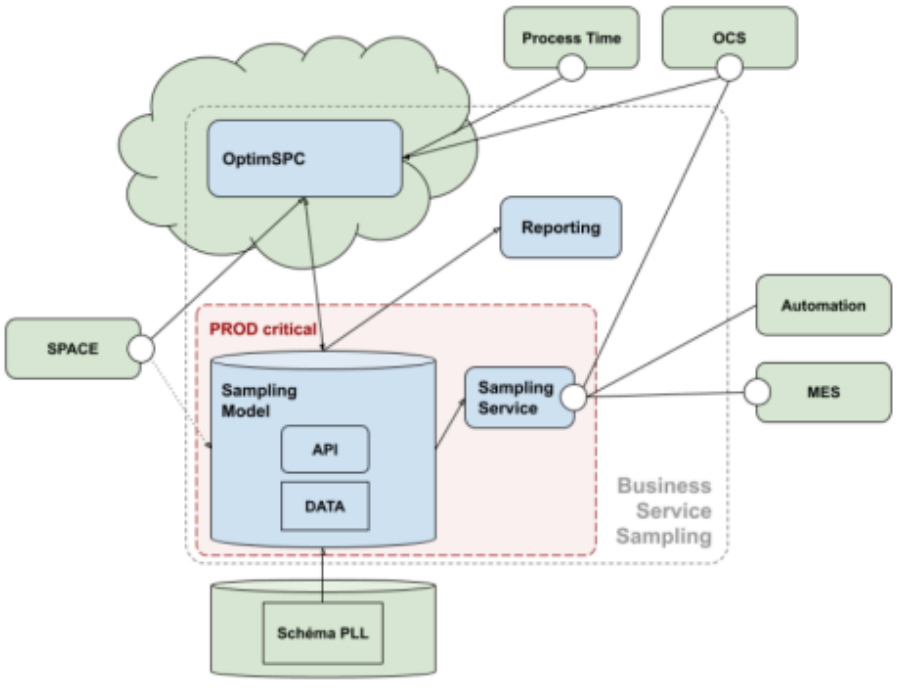


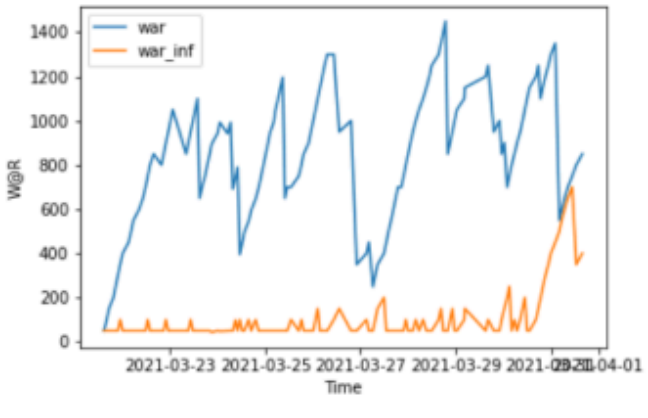
SMART SAMPLING

2021

ORACLE, AWS STEP FUNCTION, LAMBDA, CLOUDFORMATION, SNS, CLOUDWATCH

Architecture et industrialisation d'un prototype de Sampling intelligent des plaques dans une solution mixte on-premise / Cloud AWS.



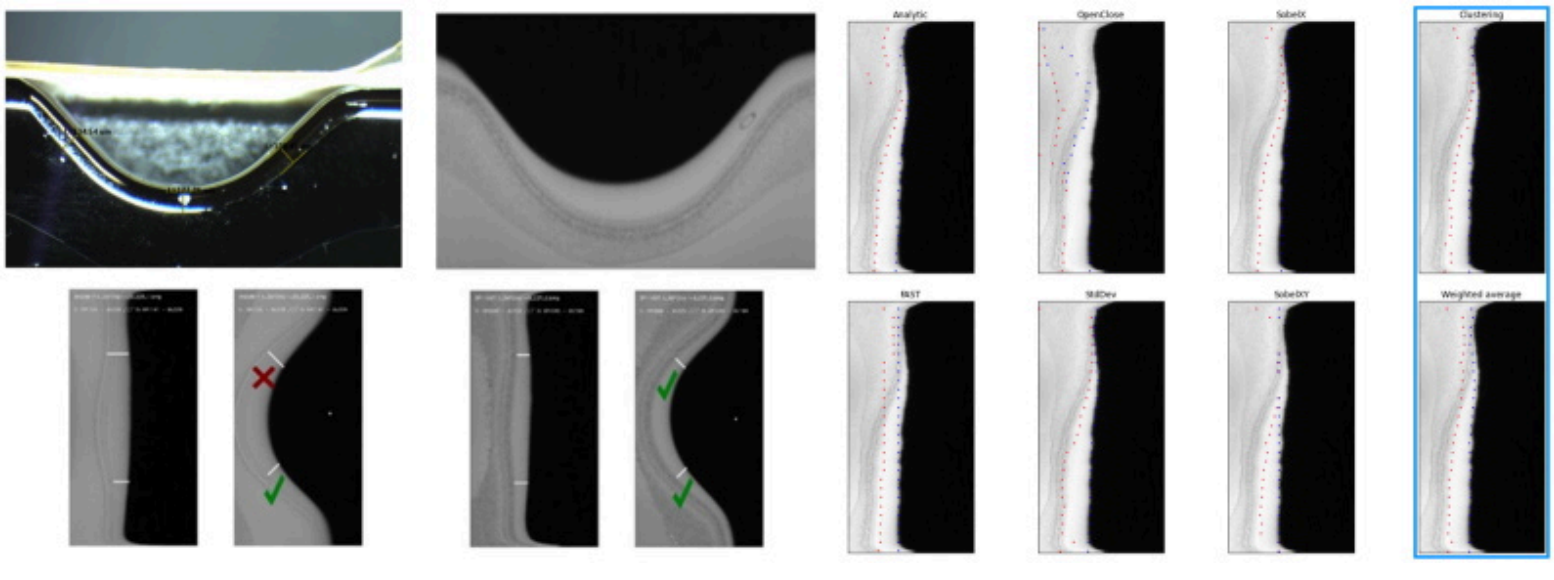


VALIDATION DU POLISSAGE

2020

DEEP LEARNING, MACHINE LEARNING, CLUSTERING, TRAITEMENT D'IMAGE, DATA AUGMENTATION

Étude de l'automatisation de la mesure du polissage des plaques. Diverses méthodes de traitement d'images (Sobel, Open/Close, Kernel, etc.) ou de Deep Learning (CNN) pour extraire l'information de la limite de polissage depuis les images, puis utilisation de cette information pour alimenter des algorithmes de Machine Learning, supervisés (classification du polissage) ou non (sélection de mesures).



ANALYSE DE SIMILARITÉ SUR UNE BASE PARTITIONNÉE

2020

DATABRICKS, SPARK-ML, TF-IDF, DISTANCE COSINUS

Utilisation de Spark-ML pour traiter des données à l'échelle (ex. : TF-IDF, calcul de similarité pour des systèmes de recommandation, etc.).

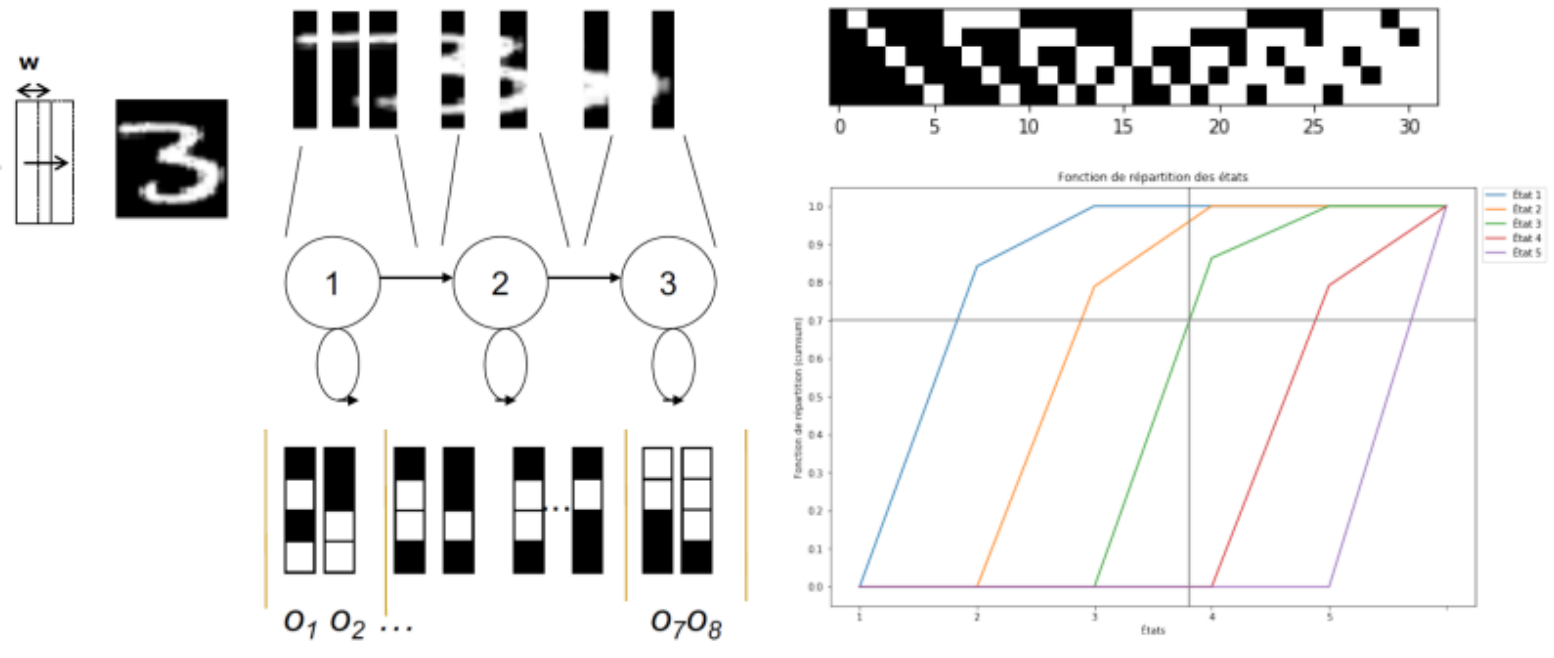
```
+-----+-----+-------------------+
|user1|user2|         sim_cosinus|
+-----+-----+-------------------+
|  128|  211| 0.09226097313475812|
|  154|  349|0.028194193721730043|
|    1|  354|  0.2838671546501407|
|  347|  428| 0.18073922282301275|
|  182|  531| 0.09549253299510274|
+-----+-----+-------------------+
only showing top 5 rows

== Physical Plan ==
*(1) InMemoryTableScan [user1#2009L, user2#2010L, sim_cosinus#2011]
   +- InMemoryRelation [user1#2009L, user2#2010L, sim_cosinus#2011], StorageLevel(disk, memory, 1 replicas)
         +- *(1) Project [i#2003L AS user1#2009L, j#2004L AS user2#2010L, value#2005 AS sim_cosinus#2011]
            +- Scan ExistingRDD[i#2003L,j#2004L,value#2005]
```

RECONNAISSANCE D'IMAGES VIA DES CHAÎNES DE MARKOV

2020

RÉSEAUX BAYÉSIENS - CHAÎNES DE MARKOV CACHÉES

Établissement d'une matrice de probabilités de transitions, puis inférence basée sur le calcul des vraisemblances de séquences d'observation.

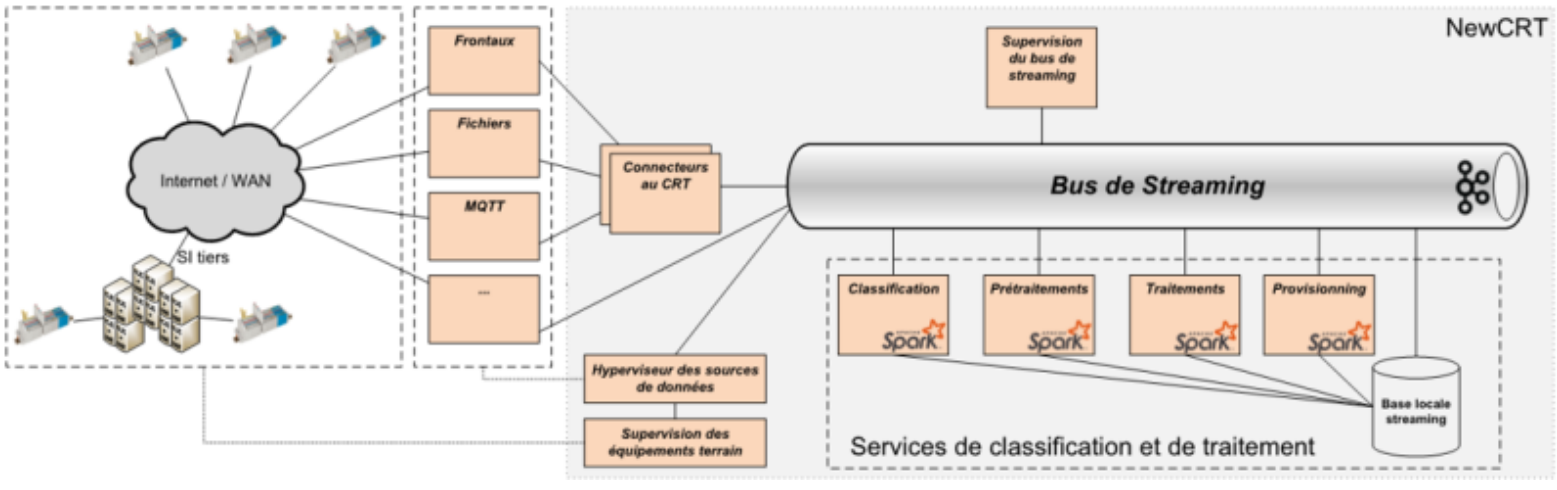


STREAMING BIG DATA

2020

KAFKA, ZOOKEEPER, SPARK STREAMING, ORACLE

Refonte architecturale, conception et développement, d'une solution historique vers une solution Big Data pour assurer le passage à l'échelle (gestion de plusieurs milliers d'événements par seconde) et la résilience.

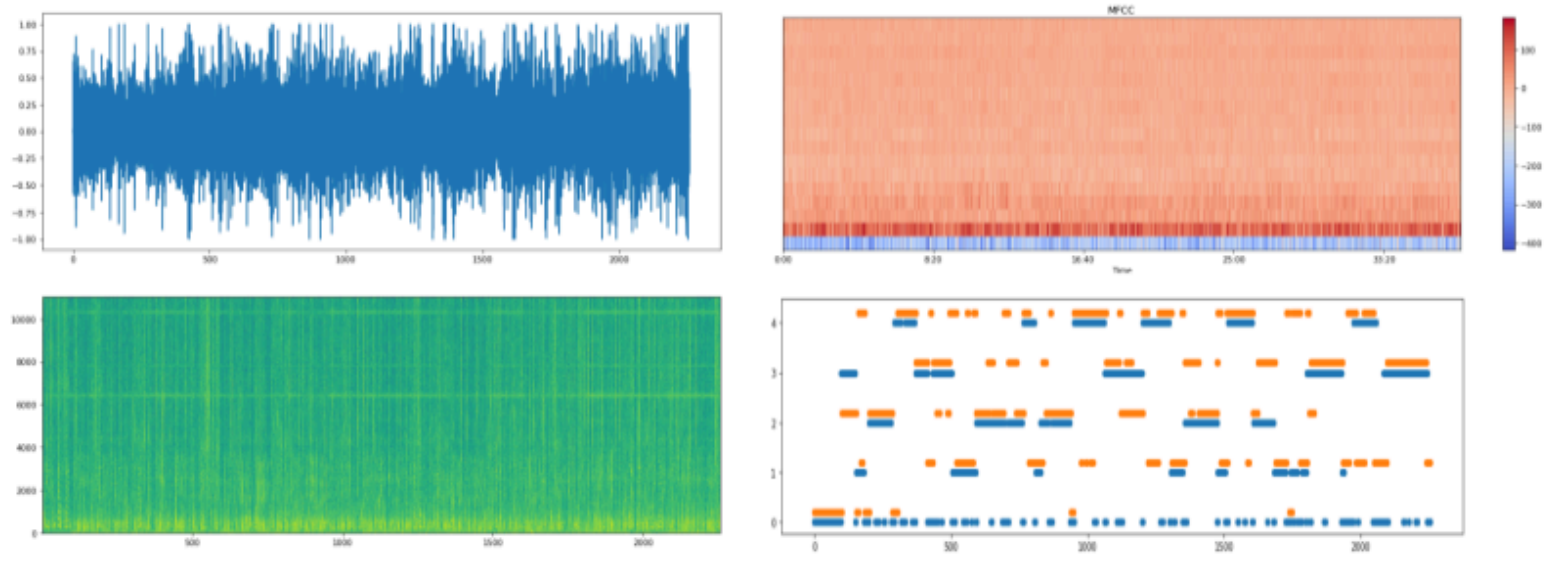


IDENTIFICATION DE LOCUTEURS

2020

SPECTROGRAMME, MFCC (LIBROSA), CLUSTERING

À partir d'une bande son, reconnaissance des différents locuteurs au fil du temps. Extraction d'informations depuis un signal audio, alimentant un algorithme d'apprentissage non-supervisé.



ANALYSE D'ARTICLES WEB

2019

NLP (SPACY)

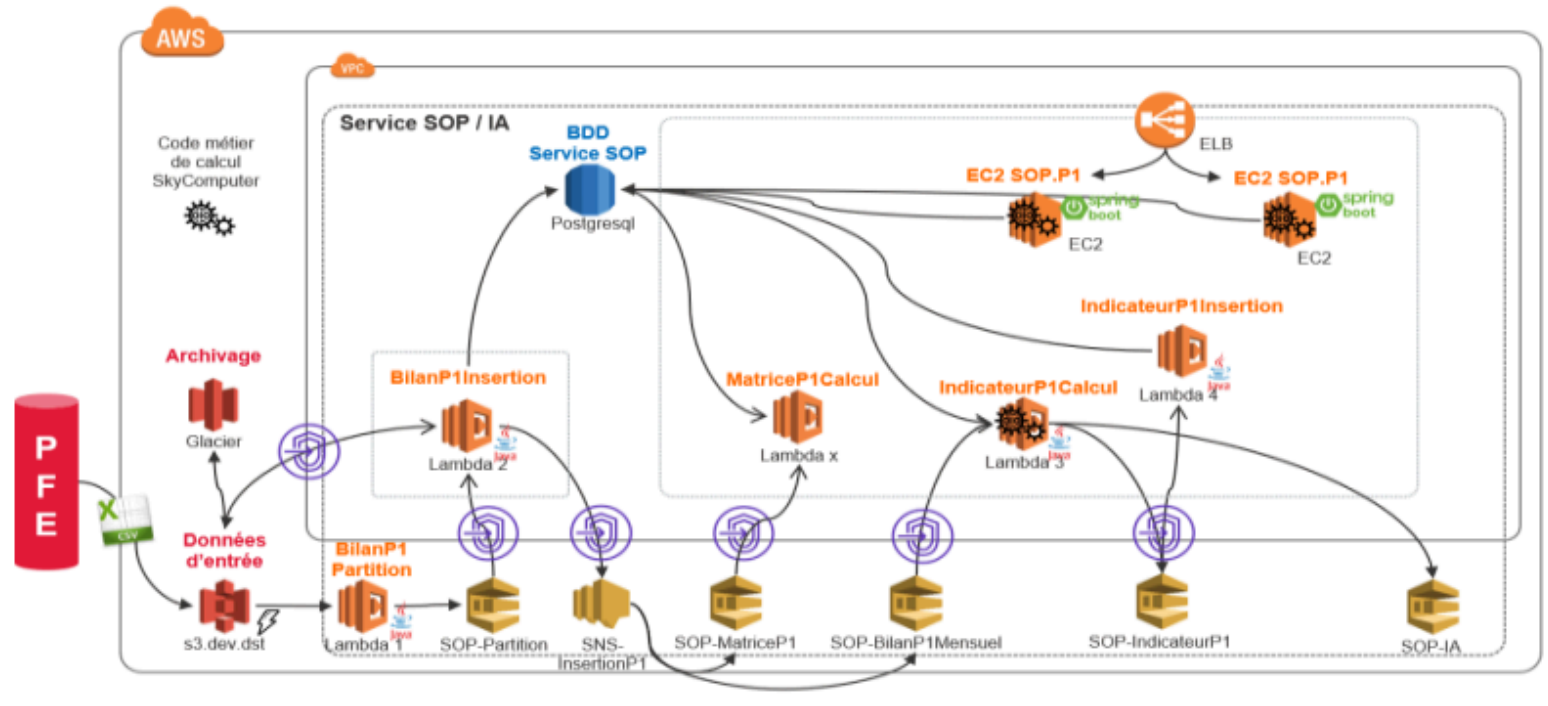
Extraction automtisée de qualificatifs/catégories des articles de la base Wikipedia.

REPORTING WEB

2019

KAFKA, SPRING BOOT, CASSANDRA, CLOUD AWS SERVERLESS

Migration et fiabilisation d'une application vers une solution Cloud AWS Serverless pour réduire ses coûts.



MIGRATION & PARTITIONNEMENT ORACLE

2019

ORACLE RAC EXADATA

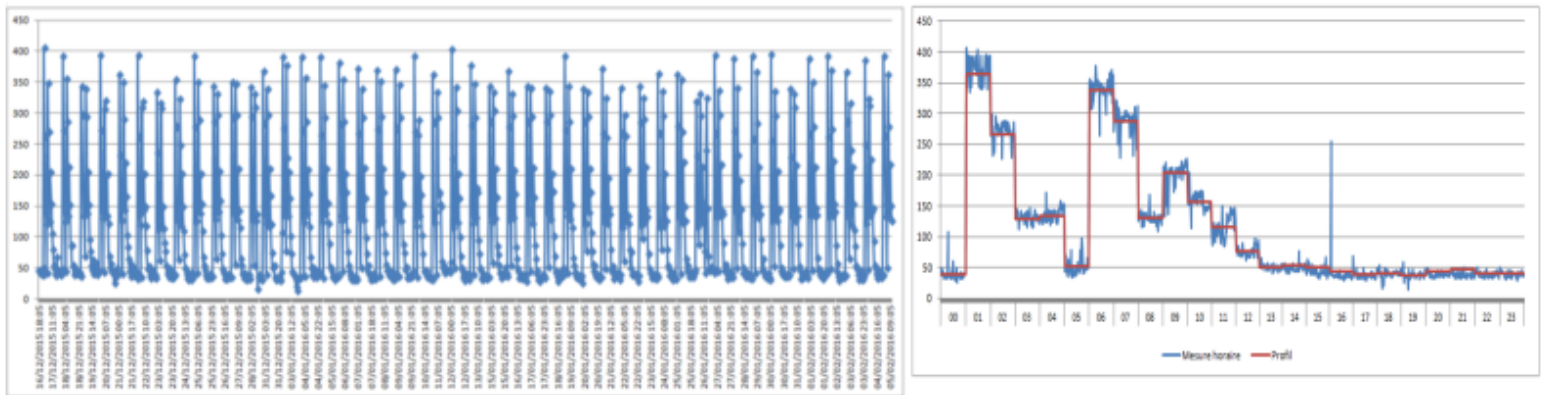
Partitionnement et montée de version du SGBD Oracle (10 To de données et tables de 40 milliards d'enregistrements) vers un cluster Oracle RAC Exadata.

SUPERVISION APPLICATIVE

2018

TIMESERIES, STATISTIQUES, JAVA

Supervision applicative d'un SCADA basée sur une analyse statistique de l'activité de l'application.

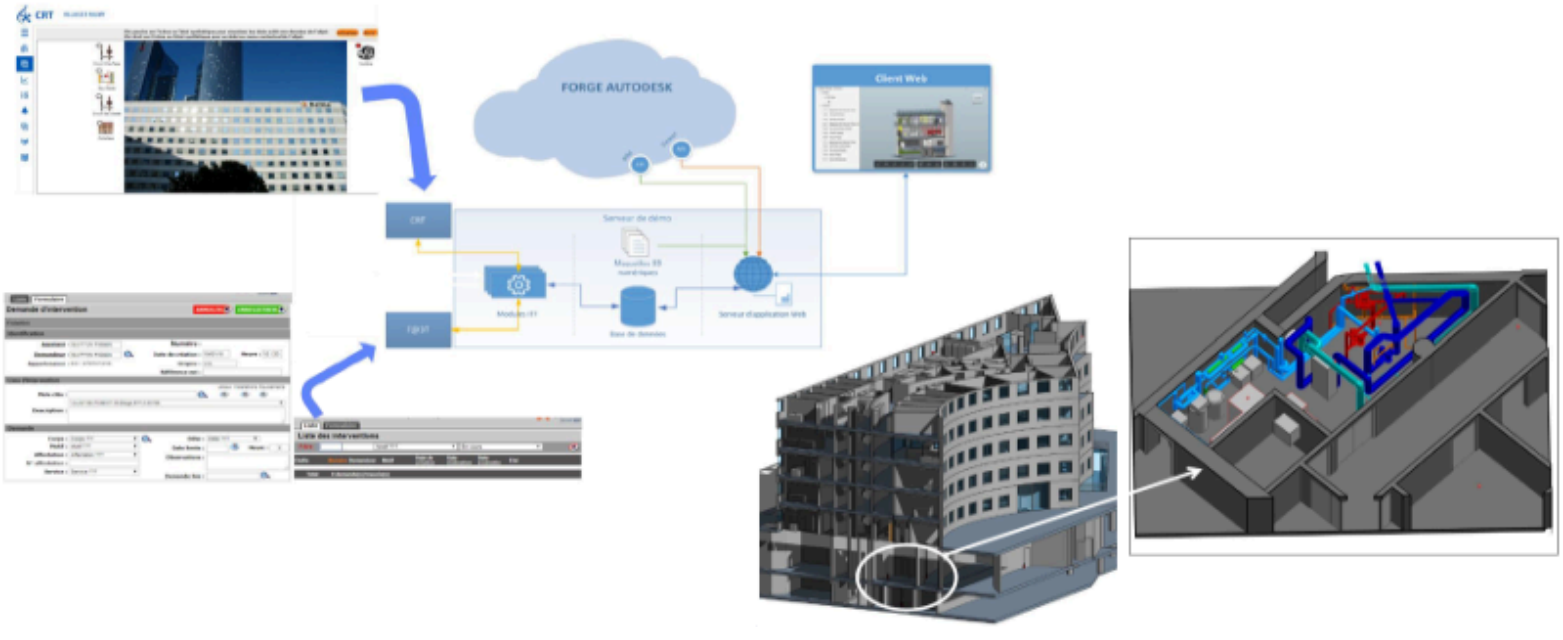


BIM - BUILDING INFORMATION MODELING

2018

AUTODESK API VIEWER, SPRING BOOT, AWS RDS POSTGRESQL

Démonstrateur pour un salon de l'interfaçage d'une solution de BIM avec des outils de GMAO.



AUDIT DE PERFORMANCE ET CONTENTIONS DU SGBD ORACLE

2018

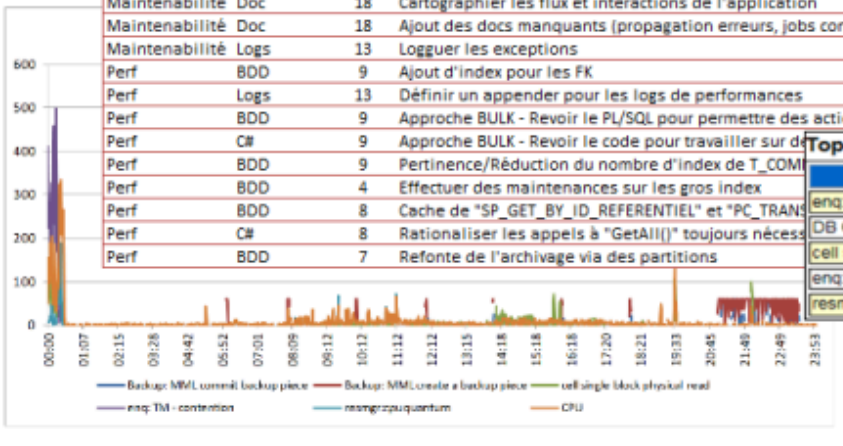
ORACLE RAC, .NET C#, QUARTZ, ELK

Analyse de divers composants applicatifs (Database Oracle, code C#, ordonnanceur Quartz), et mise en avant de contentions dans l'utilisation d'Oracle.

Axe	Domaine	Slide	Préconisation	Gain attendu	Complexité	Coût
Fiabilité	C#	14	Logguer les Horodates/Threads/Id pour recherche de parallélisme des traitement	↑	✓	✓
Fiabilité	Logs	15	Activer les logs du scheduler du ND.FWK (debugAppender sur sortie std)	↑	!	✓
Fiabilité	C#	12	Définir un timeout sur les connexions	↑	!	!
Fiabilité	Quartz	10	Revoir le déploiement sur 3 serveurs	↑	✗	!
Fiabilité	Logs	22	Vérification des logs de ALL_NDI	→	!	✓
Fiabilité	C#	17	Eviter les appels synchrones bloquants dans les traitements	→	!	!
Fiabilité	BDD	17	Vérifier la propagation de ID_CLE_TECHNIQUE_TMS	↓	✓	✓
Maintenabilité	Logs	16	Activer les logs Kibana sur l'environnement de production	↑	!	✓
Maintenabilité	Logs	16	Remettre en place les logs sur la PROD & Gestion mémoire	↑	!	!
Maintenabilité	C#	12	Activer les compteurs Oracle	→	!	✓
Maintenabilité	Doc	18	Cartographier les flux et interactions de l'application	→	!	✓
Maintenabilité	Doc	18	Ajout des docs manquants (propagation erreurs, jobs console, ND.FWK, etc.)	→	!	!
Maintenabilité	Logs	13	Logguer les exceptions	→	!	!
Perf	BDD	9	Ajout d'index pour les FK	↑	✓	✓
Perf	Logs	13	Définir un appender pour les logs de performances	↑	!	✓
Perf	BDD	9	Approche BULK - Revoir le PL/SQL pour permettre des actions en masse	↑	!	✗
Perf	C#	9	Approche BULK - Revoir le code pour travailler sur de			
Perf	BDD	9	Pertinence/Réduction du nombre d'index de T_COM			
Perf	BDD	4	Effectuer des maintenances sur les gros index			
Perf	BDD	8	Cache de "SP_GET_BY_ID_REFERENTIEL" et "PC_TRANS			
Perf	C#	8	Rationaliser les appels à "GetAll()" toujours néces			
Perf	BDD	7	Refonte de l'archivage via des partitions			

Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
enq: TX - row lock contention	28	47,373	1691893	68.03	Application
DB CPU		10,354		14.87	
cell single block physical read	613,285	4,164	7	5.98	User I/O
enq: TM - contention	399	3,458	8666	4.97	Application
resmgr:cpu quantum	334,021	1,109	3	1.59	Scheduler



COMPÉTENCES & CERTIFICATIONS

Thématique	Domaine	Principales technologies
Data Science	Apprentissage & Inférence	Machine/Deep Learning (SK-Learn, TensorFlow), Timeseries (ARIMA), Bayésien (pyAgrum)
	Extraction d'informations	Traitement d'images (OpenCV, SK-Image), NLP (Spacy, NLTK), Scraping (Beautiful Soup)
	Data Viz	Frameworks Python (Matplotlib, Seaborn, Altair, Plotly) et Web (D3.js), Tableau, PowerBI
Data Engineering	DevOps & MLOps	CI/CD (GIT, Gitlab, Jenkins, Sonar, Nexus), IaC (CloudFormation), Docker, MLFlow, DVC
	Ingestion & Processing	Kafla, dbt, Databricks Spark, Stack Hadoop (Zookeeper, Airflow, Hive, ...), Beam, Talend
	Stockage SQL	OLTP (Oracle, Postgresql), OLAP (Snowflake, Lakehouse, Redshift, Synapse, BigQuery)
	Stockage NOSQL	Colonne (Cassandra, HBase), Document (MongoDB), Graphe (Neo4J), Clef/valeur (Redis)
	Développement	Nombreux langages (Python, Java, .Net C#, C++, SQL, PL/SQL), Tests (xUnit, TDD), Patterns
	Monitoring	CloudWatch Logs / LogsInsight, ELK & OpenSearch, TIG (Telegraf, InfluxDB, Grafana)
Cloud	Amazon Web Services	Lambda, Step Functions, Glue, Kinesis, EMR, S3, DynamoDB, Athena, Fargate, Sagemaker
	Microsoft Azure	Data Factory, Event Hubs, Synapse, Stream Analytics, Azure Databricks, ML Studio
	Google Cloud Platform	Cloud SQL, Cloud Spanner, Firestore, Bigtable, BigQuery, Dataflow, Dataproc, Composer
	Management équipe/projet	Management RH, Gestion de projets (budgets, KPI, etc.), Méthodes agiles (Scrum, Kanban)