

DATA REPLY est une filiale du groupe Reply offrant une large gamme de services pour le Big Data et l'Intelligence Artificielle. Nous opérons dans différents secteurs et fonctions commerciales, en travaillant directement avec des professionnels de haut niveau et des directeurs généraux pour leur permettre d'obtenir des résultats significatifs grâce à une utilisation efficace des données. Nous aidons les entreprises à concevoir et à mettre en œuvre des produits de données centrés sur l'être humain afin de transformer des idées embryonnaires en solutions d'entreprise. Notre objectif est de trouver des compromis entre les exigences commerciales et les contraintes techniques, de développer des architectures Big Data holistiques et d'assurer la cybersécurité ainsi que la protection des données.

STORM REPLY est spécialisé dans la conception et la mise en œuvre de solutions et de services basés sur le cloud. AWS Premier Consulting Partner depuis 2014, Storm Reply a aidé un grand nombre de clients en Europe à exécuter leurs systèmes et applications sur la plateforme AWS. Grâce à une expertise bien établie dans les architectures cloud IaaS, PaaS et SaaS, Storm Reply fournit un ensemble de services de bout en bout pour l'adoption du cloud, le développement de nouvelles applications et la gestion de services complets dans l'environnement cloud. Fort de son expertise et de ses compétences poussées dans la gestion de projets cloud complexes, Storm Reply aide ses clients à mettre en œuvre leur stratégie cloud. Plus spécifiquement, Storm Reply propose des services de stratégie cloud et de migration vers le cloud, de développement d'applications cloud natives et de gestion de services cloud.

MONTE TITOLI: ANALYTIQUE AVANCÉE ET METHODES DE MACHINE LEARNING (ML) SUR AWS

Storm Reply et Data Reply ont accompagné Monte Titoli dans la définition et le développement d'une stratégie d'adoption d'architecture cloud, de gouvernance des ressources et de modèles de machine learning (ML).

MONTE TITOLI, appartenant au London Stock Exchange Group, fournit des services efficaces et sécurisés de règlement, de garde, d'administration des actifs, de gestion de garanties et d'émission auprès de clients nationaux et internationaux. La société gère une large gamme d'instruments financiers. En 2019, elle comptait 3 320 milliards d'euros d'actifs en dépôt et a traité plus de 44 millions de transactions avec un taux de règlement de fin d'année de 96 %.

LE BESOIN – UNE PLATEFORME INNOVANTE POUR L'APPRENTISSAGE AUTOMATIQUE ET L'ANALYTIQUE AVANCÉE

Monte Titoli, dépositaire central italien (CSD), appartenant au London Stock Exchange Group et fournisseur de services post-marché en Europe, recherchait un moyen rapide et sécurisé de valoriser ses données, autrefois gérées sur site selon un modèle traditionnel d'entrepôt de données.

En 2018, Monte Titoli a décidé d'utiliser des solutions d'analytique avancée et de machine learning sur AWS pour renforcer l'efficacité de ses processus internes et améliorer l'expérience client.

Monte Titoli a fait confiance à Amazon Web Services pour sa plateforme cloud et à Reply pour la définition et le développement d'une stratégie d'adoption d'architecture cloud, de gouvernance des ressources et de modèles de machine learning. Reply est AWS Premier Consulting Partner depuis 2014.

L'objectif final du client était d'obtenir une plateforme cloud flexible et évolutive pour l'introduction de pipelines d'analyses

prescriptives et prédictives avancées, en vue de se développer plus rapidement en concevant et en exécutant des projets de data science sur Amazon Web Services et d'accélérer l'adoption du machine learning au sein de l'organisation.

Compte tenu des politiques de sécurité interne très rigoureuses et à la lumière de l'importance des données gérées, nous avons tout particulièrement veillé à concevoir une solution des plus sûre. Le défi impliquait aussi une conception de réseau hautement sécurisée et la réplication du service initial, puisqu'il s'agissait du premier projet cloud natif du client.

LA SOLUTION - UNE PLATEFORME DATA LAKE ET MACHINE LEARNING SANS SERVEUR ÉVOLUTIVE ET SÉCURISÉE SUR AWS

UNE ARCHITECTURE DATA LAKE POUR DES DÉCISIONS OPÉRATIONNELLES FONDÉES SUR L'INTELLIGENCE ARTIFICIELLE.

L'architecture lancée début 2019 s'appuie principalement sur l'approche sans serveur AWS à des fins d'orchestration, d'automatisation et de surveillance, conférant à l'infrastructure l'agilité nécessaire pour évoluer à la demande en fonction des besoins opérationnels.

L'architecture lancée début 2019 s'appuie principalement sur l'approche sans serveur AWS à des fins d'orchestration, d'automatisation et de surveillance, conférant à l'infrastructure l'agilité nécessaire pour évoluer à la demande en fonction des besoins opérationnels. Cette approche a abouti à une infrastructure facile à gérer avec un coût d'exploitation extrêmement faible comparé aux autres solutions data lake « standards », qui

s'accompagnent généralement de frais de licence élevés. Une infrastructure sandbox de machine learning basée sur Amazon SageMaker et Glue a été ajoutée au data lake pour fournir un environnement de travail à l'équipe de data science. Un pipeline Machine Learning automatise le modèle et le processus de prévision produit à l'aide des services ETL (Extraction, Transformation et Chargement) et ML AWS Services, apportant chaque

jour de nouvelles données exploitables à l'organisation via l'outil de Business Intelligence.

Plus de 70 lots de flux d'ingestion bruts composés de plusieurs Go de données de transactions financières sont chargés chaque jour dans le data lake, déclenchant

Titoli a pu concevoir et livrer une solution conforme et performante, adaptée à ses besoins opérationnels, tout en optimisant la valeur ajoutée pour ses clients internes et externes. Le projet de data lake a été développé et déployé à l'aide de la méthodologie Agile et de DevOps et mis en service en moins de

Le **processus complet**, de la collecte des données aux tableaux de bord finaux et aux prédictions, prend **moins d'une heure**



ainsi un processus ETL automatique (orchestré et synchronisé via Lambda et des tables DynamoDb) qui fournit des données utilisables à l'entrepôt de données. Les demandes d'agrégation sont traitées sur les Datamarts qui produisent des données opérationnelles, présentées via le logiciel Tableau software sous forme de tableaux de bord consommés par l'entreprise.

Des processus ETL supplémentaires traitent automatiquement des données fraîchement ingérées et des Datamarts préexistants, produisant des prévisions automatiques également présentées via le logiciel Tableau software pour anticiper les tendances de données. Le processus complet, de l'ingestion aux tableaux de bord finaux et aux prédictions, prend moins d'une heure.

En s'appuyant sur l'expertise de Storm Reply et d'Amazon Web Services, Monte

6 mois. Actuellement, de nouveaux flux sont ingérés et un ensemble croissant de tableaux de bord et de prévisions est produit et livré chaque jour à l'aide d'un jeu de données contenant plus de deux ans et demi d'informations sur les échanges boursiers. Une équipe de data scientists travaille actuellement de manière indépendante sur des données partagées, chacun utilisant son instance de notebook Sagemaker et ses tâches Glue personnalisées via des déclencheurs personnalisés pour contrôler la consommation des ressources. L'architecture conçue par Storm Reply peut évoluer à la demande, n'implique aucun frais de licence et demande très peu d'efforts de gestion, car elle a principalement recours à des services gérés et sans serveurs. Le design d'architecture proposé est devenu une architecture de référence au sein du groupe LSEG pour les data lakes et les datalabs.

UNE PLATEFORME MACHINE LEARNING POUR CONCEVOIR ET EXECUTER DES PROJETS DE BOUT-EN-BOUT

Data Reply a accompagné Monte Titoli dans la mise en œuvre d'une véritable plateforme de machine learning sur AWS, avec Amazon SageMaker comme composant central, permettant aux data scientists de concevoir et d'exécuter des projets d'apprentissage automatique de bout-en-bout à partir de données structurées et non structurées..

La nouvelle architecture sans serveur s'appuie sur Amazon Glue pour les activités ETL liées à la préparation des données et aux résultats de modèles après traitement ; sur Amazon SageMaker pour construire, former et déployer des modèles de machine learning via des conteneurs Docker ; et sur Amazon Lambda pour automatiser la formation des modèles, le réglage des hyperparamètres.

AMAZON GLUE

Activités ETL: préparation des données et post-traitement des résultats des modèles

AMAZON SAGEMAKER

Construire, former et déployer des modèles via des conteneurs Docker

AMAZON LAMBDA

Automatiser la formation des modèles, le réglage des hyperparamètres

Elle offre ainsi une conformité parfaite avec les strictes exigences de sécurité interne. De plus, un système de surveillance personnalisé contrôle en permanence les performances des modèles de machine learning en production et envoie des alertes en cas de problème de qualité des données ou de performance des modèles. Toutes les données prédictives sont affichées sur des tableaux de bord personnalisés alimentés par des résultats de modèles stockés dans un compartiment S3.

Data Reply a proposé de concevoir un workflow de sciences des données adapté en adoptant des services cloud natifs: De la préparation des données à la mise en œuvre des modèles, la formation, le service et la surveillance.

PRÉPARATION DES DONNÉES ET MISE EN ŒUVRE DES MODÈLES

FORMATION

SERVICE

SURVEILLANCE

L'équipe de data scientists de Monte Titoli, aidé par Data Reply, a mis en œuvre des modèles de classification et de régression visant à améliorer les processus post-marché et à prédire l'efficacité des règlements.

L'intégration fluide entre les instances de notebook SageMaker, les tâches ETL Spark Glue et les terminaux des modèles déployés a permis de simplifier les tâches telles que le nettoyage, la fusion et l'exploration des données, la conception de fonctionnalités, la modélisation, la formation et les tests, l'optimisation des hyperparamètres, la comparaison des modèles, l'évaluation et le déploiement en production. Les modèles exploitent

environnement de production, générant des lots quotidiens de données de prévisions et des rapports personnalisés via des tableaux de bord sur Tableau Software.

Cette nouvelle infrastructure a permis la mise en œuvre, la formation, le réglage des hyperparamètres, le déploiement et la surveillance des applications de machine learning, réduisant ainsi de 50 % le temps nécessaire pour concevoir et exécuter un flux de travail de machine learning. En outre, une surveillance des modèles personnalisés contrôle la performance des modèles de machine learning en production et permet une nouvelle formation uniquement en cas de dérive des données. Cela permet jusqu'à 75 % de réduction des coûts de mise à jour des modèles.

Cette plateforme flexible permet le développement et le traitement d'un nombre illimité de modèles de machine learning prédictifs et prescriptifs. Elle peut servir un grand nombre d'utilisateurs



-50%

IL EST TEMPS DE CONCEVOIR ET D'EXÉCUTER UN WORKFLOW ML COMPLET



-75%

MODÈLES DE MISE À JOUR DES COÛTS

les algorithmes conteneurisés intégrés et personnalisés TensorFlow 2.0 de deep learning d'AWS. Les pipelines de formation et d'inférence ont été déployés dans un

internes et externes grâce à des paramètres de prévisions en temps réel et en lots, et évolue à la demande en fonction des besoins opérationnels.