

DATA REPLY è la società del gruppo Reply che offre servizi di eccellenza per Big Data e Artificial Intelligence. Operando sulla gran parte delle Industry e delle funzioni aziendali, supportiamo professionisti di livello esecutivo e Chief Officers per trarre valore dai dati. Costruiamo Data Platform, definiamo e implementiamo modelli di ML e AI in modo efficiente, replicabile e scalabile, attraverso persone con alte competenze in Big Data Engineering, Data Science e Intelligent Process Automation. Sempre attivi sulle innovazioni, stiamo applicando algoritmi Quantistici a supporto dell'ottimizzazione di processi con alte necessità computazionali.

STORM REPLY è l'azienda del gruppo Reply specializzata nella progettazione e implementazione di soluzioni e servizi cloud innovativi. Attraverso competenze consolidate e molti anni di esperienza nella creazione e gestione di servizi cloud Infrastructure as a Service (IaaS), Software as a Service (SaaS) e Platform as a Service (PaaS), Storm Reply offre il proprio supporto a importanti aziende in Europa e in tutto il mondo nell'implementazione di sistemi e applicazioni basati sul cloud.

MONTE TITOLI: ANALISI AVANZATA E MACHINE LEARNING SU AWS

Storm Reply e Data Reply hanno supportato Monte Titoli nel definire e sviluppare una strategia finalizzata all'adozione di architetture cloud, la gestione delle risorse e i modelli di Machine Learning.

MONTE TITOLI, parte del London Stock Exchange Group (LSEG), fornisce servizi sicuri ed efficienti di liquidazione, custodia, asset servicing, gestione collaterale e agli emittenti per clienti nazionali e internazionali. L'azienda gestisce un'ampia gamma di strumenti finanziari, con 3,32 trilioni di euro di asset in custodia nel 2019, elaborando oltre 44 milioni di transazioni con un tasso di liquidazione a fine anno del 96%.

OBIETTIVO: SVILUPPARE UNA PIATTAFORMA INNOVATIVA DI ANALISI AVANZATA E MACHINE LEARNING

Monte Titoli, depositario centrale per il mercato finanziario italiano (CSD), parte del London Stock Exchange Group e provider leader in Europa nei servizi di post-trading, necessitava di un sistema rapido e sicuro per generare valore dai dati, precedentemente gestiti internamente con un approccio classico di data warehouse (DWH).

Per aumentare l'efficienza dei processi interni e migliorare l'esperienza complessiva dei clienti, nel 2018 ha scelto di utilizzare le soluzioni di analisi avanzata e Machine Learning di AWS. Monte Titoli ha scelto di affidarsi ad Amazon Web Services per la propria piattaforma cloud e a Reply, dal 2014 AWS Premier Consulting Partner, per definire e sviluppare la strategia finalizzata all'adozione di architetture cloud, la gestione delle risorse e i modelli di Machine Learning.

Gli obiettivi finali di Monte Titoli: disporre di una **piattaforma cloud flessibile e scalabile** in grado di supportare l'introduzione di pipeline di analisi

predittive e prescrittive avanzate, **scalare più velocemente** elaborando ed eseguendo progetti di data science su Amazon Web Services e **accelerare l'adozione del Machine Learning in tutta la propria organizzazione**.

Considerata l'importanza dei dati trattati e in virtù delle policy di sicurezza interne adottate da Monte Titoli, per Reply è stato fondamentale garantire un elevato livello di protezione nel corso della progettazione della soluzione. In particolare, trattandosi del primo progetto cloud nativo del cliente, è stato necessario sviluppare una rete con il massimo grado di sicurezza e un whitelisting di servizio iniziale.

LA SOLUZIONE: UNA PIATTAFORMA DATALAKE E ML SERVERLESS, SCALABILE E SICURA

UN'ARCHITETTURA DATALAKE PERMETTE DI PRENDERE DECISIONI COMMERCIALI BASANDOSI SULL'INTELLIGENZA ARTIFICIALE

L'architettura Datalake, in produzione dall'inizio del 2019, si basa essenzialmente sull'approccio serverless di AWS per l'orchestrazione, l'automatizzazione e il monitoraggio, e consente all'infrastruttura l'agilità necessaria per scalare on-demand in base ai requisiti aziendali.

Si tratta di una soluzione che utilizza servizi di AWS completamente gestiti on-demand e senza licenza. Attraverso l'approccio serverless è stato possibile sviluppare un'infrastruttura di semplice gestione e con un costo totale di possesso notevolmente inferiore rispetto ad altre soluzioni Datalake "standard", i cui costi di licenza sono tipicamente elevati. Per offrire un ambiente di lavoro al team di Data Science, all'architettura Datalake è stata aggiunta un'infrastruttura sandbox ML basata su SageMaker e Glue di Amazon. Una pipeline ML automatizza il training del modello e il processo di previsione eseguiti con i servizi Extract, Transform and Load (ETL) e ML di AWS,

rendendo disponibili ogni giorno nuove informazioni dettagliate per il business grazie al tool di BI. Ogni giorno, oltre 70 flussi di inserimento di dati non elaborati in blocco, costituiti da svariati GB di dati di transazioni finanziarie, vengono caricati sull'architettura Datalake, innescando un processo ETL automatico (orchestrato e sincronizzato tramite tabelle Lambda e DynamoDb) per fornire al DWH dati utilizzabili.

Le query di aggregazione vengono eseguite sui data smart business-oriented di produzione dei dati, che sono presentati attraverso Tableau Server sotto forma di dashboard a disposizione dell'azienda.

Ulteriori processi e trasformazioni ETL vengono eseguiti automaticamente sia sui dati appena inseriti, sia sui data smart preesistenti, generando previsioni ML, presentate anch'esse attraverso Tableau, in modo da prevedere i trend dei dati. Grazie ad Amazon Web Services e all'esperienza di Storm Reply, Monte Titoli ha potuto sviluppare e mettere in opera

a un dataset alimentato da informazioni di market trading raccolte nel corso di più di due anni e mezzo. Un team di data scientist sta lavorando separatamente su dati condivisi, in cui ciascuno utilizza la propria istanza notebook SageMaker personalizzata e processi Glue invocati tramite trigger personalizzati per eseguire il controllo sul consumo delle

L'intero processo, dall'inserimento alle dashboard e alle previsioni finali, è completato in **meno di un'ora**.



una soluzione efficiente, che risponde alle esigenze aziendali di creazione di valore per i clienti interni ed esterni. Il progetto Datalake è stato sviluppato e distribuito adottando la metodologia **Agile** e l'approccio **DevOps**, rendendone possibile il lancio in meno di 6 mesi. Giornalmente vengono inseriti nuovi flussi, mentre un insieme sempre più grande di dashboard e previsioni è prodotto e fornito grazie

risorse. **L'architettura sviluppata da Storm Reply**, utilizzando perlopiù servizi gestiti e serverless, **può essere scalata su richiesta, non ha costi di licenza ed è estremamente facile da gestire**. L'architettura proposta è diventata un punto di riferimento all'interno del gruppo LSEG per quanto riguarda Datalake e Datalab.

ML PLATFORM TO DESIGN AND EXECUTE END-TO-END ML PROJECTS

Data Reply ha supportato Monte Titoli nell'implementazione di una piattaforma completa di ML su AWS, con Amazon SageMaker quale componente principale, consentendo ai data scientist di sviluppare ed eseguire progetti di ML end-to-end sia su dati strutturati, sia su dati non strutturati.

L'architettura serverless recentemente sviluppata impiega le soluzioni **Amazon Glue** per le attività ETL responsabili della preparazione dei dati e della post-elaborazione dei risultati

dei modelli, **Amazon SageMaker** per realizzare, addestrare e distribuire modelli di Machine Learning tramite contenitori Docker, e **Amazon Lambda** per automatizzare il training, la regolazione degli iperparametri e l'interferenza dei modelli, fornendo una piena conformità con i requisiti di sicurezza interni.

AMAZON GLUE	AMAZON SAGEMAKER	AMAZON LAMBDA
Attività ETL: preparazione dei dati e post-elaborazione dei risultati dei modelli	Realizzare, addestrare e distribuire modelli di ML tramite contenitori Docker	Automatizzare il training, la regolazione degli iperparametri e l'interferenza dei modelli

Inoltre, un processo di monitoraggio dei modelli personalizzato controlla continuamente le prestazioni dei modelli di Machine Learning in produzione e invia segnalazioni nell'eventualità di problemi in termini di qualità dei dati o prestazioni dei modelli. Tutte le informazioni dettagliate predittive vengono mostrate su dashboard aziendali personalizzati, alimentati dai risultati dei modelli memorizzati in un bucket S3.

Data Reply ha proposto lo sviluppo di un workflow di data science idoneo adottando servizi cloud nativi:

PREPARAZIONE ED IMPLEMENTAZIONE DEI DATI

TRAINING

SERVIZIO

MONITORAGGIO

Il team di data science di Monte Titoli, supportato dai data scientist di Data Reply, ha implementato i modelli di classificazione

e regressione atti a migliorare i processi di post-trading e prevedere l'efficienza dei regolamenti. Operazioni come la pulizia, l'unione e l'esplorazione dei dati, lo sviluppo di funzionalità, il modeling, il training e il testing, l'ottimizzazione degli iperparametri, il confronto, la valutazione e la messa a punto fino alla produzione dei modelli sono diventate attività semplici, grazie alla perfetta integrazione tra le istanze notebook di SageMaker, i processi ETL di Spark Glue e gli endpoint dei modelli distribuiti. I modelli vengono realizzati sfruttando sia gli algoritmi integrati di AWS sia gli algoritmi personalizzati di deep learning contenuti in TensorFlow 2.0.

Nell'ambiente di produzione, sia le pipeline di training sia le pipeline di interferenza sono attive e funzionanti, generando giornalmente informazioni dettagliate delle previsioni in blocco e report aziendali personalizzati tramite i dashboard di Tableau. Tale infrastruttura innovativa ha reso possibile l'implementazione, il training, la regolazione degli iperparametri, la distribuzione e il monitoraggio di applicazioni ML, comportando un

risparmio del 50% del tempo necessario per sviluppare ed eseguire un workflow ML completo. Inoltre, un monitoraggio personalizzato dei modelli controlla le prestazioni dei modelli ML durante la produzione, consentendone il riaddestramento soltanto quando si presenta una deriva dei dati e riducendo i costi di aggiornamento dei modelli fino al 75%.



-50%

TEMPO NECESSARIO PER SVILUPPARE ED ESEGUIRE UN WORKFLOW ML COMPLETO



-75%

COSTI DI AGGIORNAMENTO DEI MODELLI

Questa piattaforma flessibile consente lo sviluppo e l'esecuzione di un numero illimitato di modelli di Machine Learning predittivi e prescrittivi. È possibile servire molti utenti interni ed esterni grazie agli endpoint delle previsioni sia in blocco sia in tempo reale, con una scalabilità in funzione dei requisiti aziendali on-demand.