

**DATA REPLY** ist innerhalb der Reply-Gruppe das Unternehmen, das Dienstleistungen zur Wertschöpfung aus Daten anbietet. Dabei verwendet es Big Data-Technologien, künstliche Intelligenz und Quantum Computing. Data Reply arbeitet mit leitenden Angestellten wie CIOs, CMOs, CDOs, Qualitäts- und Produktionsleitern in verschiedenen Branchen zusammen, um die Geschäftsfunktionen Vertrieb und Marketing, Betrieb, Sicherheit, IoT und Industrie 4.0 zu unterstützen. Data Reply bietet die Kompetenzen und Erfahrungen seiner Dateningenieure und Datenwissenschaftler, um Datalakes auf AWS zu erstellen und zu verwalten, Echtzeitanalysen zu generieren und zu verarbeiten, vertikale KI/ML-Projekte und vom Quantum Computing inspirierte Anwendungen zu implementieren.

**STORM REPLY** ist das Unternehmen der Reply Gruppe, das sich auf die Entwicklung und Implementierung innovativer cloudbasierter Lösungen und Dienste spezialisiert hat. Mit seinem fundierten Fachwissen und seiner langjährigen Erfahrung in der Erstellung und Verwaltung von Cloud-Lösungen in den Bereichen Infrastructure as a Service (IaaS), Software as a Service (SaaS) und Platform as a Service (PaaS) unterstützt Storm Reply bedeutende Unternehmen in Europa und auf der ganzen Welt bei der Implementierung von Cloud-basierten Systemen und Anwendungen.

# MONTE TITOLI: FORTSCHRITTLICHE ANALYTIK UND MASCHINELLES LERNEN MIT AWS

Storm Reply und Data Reply halfen Monte Titoli dabei eine Cloud-Architektur zu entwerfen, die das Ressourcen-Management und ML-Modelle unterstützt.



**MONTE TITOLI**, teil der London Stock Exchange Group (LSEG), bietet nationalen und internationalen Kunden effiziente und sichere Services in den Bereichen Abwicklung, Verwahrung, Vermögenswertdienste, Sicherheitenverwaltung und Emittentendienste. Das Unternehmen verwaltet eine breite Palette von Finanzinstrumenten mit einem verwahrten Vermögen von 3,32 Billionen Euro. Im Jahr 2019 wurden über 44 Millionen Transaktionen

# DIE HERAUSFORDERUNG – BEDARF FÜR EINE INNOVATIVE PLATTFORM FÜR FORTSCHRITTLICHE ANALYTIK UND MASCHINELLES LERNEN

Monte Titoli, der italienische Zentralverwahrer, ist Teil der London Stock Exchange Group und führender Anbieter von Post-Trade-Diensten in Europa. Die Gesellschaft suchte nach einer schnellen und sicheren Möglichkeit zur Wertschöpfung aus ihren Daten, die zuvor vor Ort mit einem klassischen Data Warehouse-Ansatz verwaltet wurden.

Im Jahr 2018 entschied sich Monte Titoli, AWS-Lösungen für fortschrittliche Analytik und maschinelles Lernen zu verwenden, um die Effizienz interner Prozesse zu steigern und das Gesamterlebnis der Kunden zu verbessern. Monte Titoli entschied sich bei seiner Cloud-Plattform für Amazon Web Services. Reply, das seit 2014 AWS Premier Consulting Partner ist, wurde damit beauftragt, die Einführungsstrategie für Cloud-Architekturen, die Ressourcensteuerung und Modelle für maschinelles Lernen zu definieren und zu entwickeln.

Das Endziel des Kunden bestand darin, eine **flexible und skalierbare Cloud-Plattform zu schaffen**. Sie sollte die

Einführung fortschrittlicher Pipelines für **prädiktive und präskriptive Analysen unterstützen**, durch Entwurf und Ausführung von **Data Science-Projekten auf Amazon Web Services** eine schnellere Skalierung ermöglichen und die Einführung von maschinellem Lernen im gesamten Unternehmen beschleunigen. Aufgrund der strengen internen Sicherheitsrichtlinien und der Bedeutung der verarbeiteten Daten wurde mit großer Sorgfalt an der Entwicklung einer sicheren Lösung gearbeitet. Da dies das erste cloud-native Projekt des Kunden war, mussten zudem ein hochsicheres Netzwerkdesign und eine erste Whitelist für den Service entwickelt werden.

# DIE LÖSUNG – EINE SERVERLOSE, SKALIERBARE UND SICHERE DATALAKE- & ML-PLATTFORM AUF AWS

## EINE DATALAKE-ARCHITEKTUR FÖRDERT GESCHÄFTSENTSCHEIDUNGEN AUF DER GRUNDLAGE KÜNSTLICHER INTELLIGENZ

Die Datalake-Architektur, die seit Anfang 2019 in Produktion ist, stützt sich stark auf den serverlosen AWS-Ansatz für Orchestrierungs-, Automatisierungs- und Überwachungszwecke. Dadurch kann die Infrastruktur je nach Bedarf flexibel auf die Geschäftsanforderungen skaliert werden.

Die Lösung nutzt vollständig verwaltete und lizenzfreie bedarfsorientierte AWS-Services. Dieser Ansatz führte zu einer einfach zu verwaltenden Infrastruktur mit extrem niedrigen Gesamtbetriebskosten im Vergleich zu anderen „Standard“-Datalake-Lösungen, die normalerweise hohe Lizenzkosten verursachen. Dem Datalake wurde eine ML-Sandbox-Infrastruktur hinzugefügt, die auf Amazon SageMaker und Glue basiert und eine Arbeitsumgebung für das Data Science-Team bietet. Eine ML-Pipeline automatisiert den Modellschulungs- und Vorhersageprozess, der mithilfe von Extraktions-, Transformations- und Ladevorgängen (ETL) sowie ML AWS

Services durchgeführt wird, und stellt dem Unternehmen täglich neue Erkenntnisse über das BI-Tool zur Verfügung. Täglich werden über 70 Batches Raw-Ingestion-Streams, die aus mehreren GB Finanztransaktionsdaten bestehen, in den Datalake geladen, wodurch ein automatisierter ETL-Prozess (orchestriert und synchronisiert über Lambda- und DynamoDb-Tabellen) ausgelöst wird, der dem Data Warehouse letztendlich Verbrauchsdaten liefert.

Aggregationsabfragen werden für die Daten ausgeführt, um geschäftsorientierte Datamarts zu erstellen, die über Tableau Server in Form von Dashboards

angezeigt werden, die das Unternehmen verwendet. Zusätzliche ETL-Jobs und -Transformationen werden automatisch sowohl für neu aufgenommene Daten als auch für bereits vorhandene Datamarts ausgeführt. Dabei werden ML-Prognosen erstellt, die zur Vorhersage von Datentrends auch über Tableau angezeigt werden.

werden immer mehr Dashboards und Vorhersagen erzeugt und bereitgestellt. Dabei wird ein Datensatz verwendet, in dem Markthandelsinformationen aus mehr als zweieinhalb Jahren enthalten sind. Ein Team von Datenwissenschaftlern arbeitet derzeit getrennt an gemeinsam genutzten Daten. Dabei verwendet jeder seine personalisierten Sagemaker-

**Der gesamte Prozess**, von der Datenaufnahme bis zu den endgültigen Dashboards und Vorhersagen, **dauert weniger als eine Stunde.**



Mithilfe des Fachwissens von Amazon Web Services und Storm Reply konnte Monte Titoli eine leistungsfähige und konforme Lösung für seine Geschäftsanforderungen entwerfen und umsetzen, was zur Wertsteigerung für interne und externe Kunden beitrug. Das Datalake-Projekt wurde mit **agiler Methodik** und dem **DevOps-Ansatz** entwickelt und implementiert und in weniger als sechs Monaten in Betrieb genommen. Derzeit werden neue Streams aufgenommen und täglich

Notebook-Instanzen und Glue-Jobs, die über benutzerdefinierte Trigger aufgerufen werden, um den Verbrauch von Kontrollressourcen zu erzwingen. Die von **Storm Reply entworfene Architektur kann nach Bedarf skaliert werden, hat keine festen Lizenzkosten und einen sehr geringen Verwaltungsaufwand, da sie hauptsächlich verwaltete und serverlose Dienste verwendet.** Das vorgeschlagene Architekturdesign wurde zu einer Referenzarchitektur innerhalb der LSEG-Gruppe für Datalakes und Datalabs.

## ML-PLATTFORM ZUM ENTWERFEN UND AUSFÜHREN VON DURCHGÄNGIGEN MACHINE LEARNING-PROJEKTEN

Data Reply unterstützte Monte Titoli bei der Implementierung einer vollwertigen Plattform für maschinelles Lernen in AWS mit



Amazon SageMaker als Kernkomponente. Datenwissenschaftler können damit durchgängige Projekte für maschinelles Lernen für strukturierte und unstrukturierte Daten entwerfen und ausführen.

Die neu entwickelte serverlose Architektur nutzt **Amazon Glue** für ETL-Aktivitäten, die sich mit der Datenaufbereitung und Nachbearbeitung von Modellergebnissen befassen: auf **Amazon SageMaker**, um Modelle für maschinelles Lernen über Docker-Container zu erstellen, zu schulen und bereitzustellen, auf **Amazon Lambda**, um die Modellschulung, die Optimierung von Hyperparametern und die Inferenz zu automatisieren. All dies bei einer vollständigen Einhaltung strenger interner Sicherheitsanforderungen.

AMAZON GLUE	AMAZON SAGEMAKER	AMAZON LAMBDA
Für ETL-Aktivitäten, die sich mit der Datenaufbereitung und Nachbearbeitung von Modellergebnissen	Modelle für maschinelles Lernen über Docker-Container zu erstellen, zu schulen und bereitzustellen	Um die Modellschulung, die Optimierung von Hyperparametern und die Inferenz

Darüber hinaus steuert ein benutzerdefinierter Modellüberwachungsjob kontinuierlich die Leistung von Modellen für maschinelles Lernen in der Produktion. Er sendet Warnungen, wenn Probleme mit der Datenqualität oder der Modelleistung auftreten. Alle prädiktiven Erkenntnisse werden in benutzerdefinierten Geschäfts-Dashboards angezeigt, die von in einem S3-Bucket gespeicherten Modellergebnissen gespeist werden.

Data Reply schlug vor, einen geeigneten Data Science-Workflow mit cloud-nativen Diensten zu entwerfen:

**DATENAUFBEREITUNG ÜBER DIE IMPLEMENTIERUNG**

**SCHULUNG**

**BEREITSTELLUNG**

**ÜBERWACHUNG**

Das Data Science-Team von Monte Titoli implementierte zusammen mit Datenwissenschaftlern von Data Reply Klassifizierungs- und Regressionsmodelle, um die Post-Trading-Prozesse zu verbessern und Vorhersagen zur Abwicklungseffizienz zu treffen. Aufgrund der nahtlosen Integration zwischen SageMaker-Notebook-Instanzen, Spark Glue ETL-Jobs und bereitgestellten Modellendpunkten können folgende Prozesse unkompliziert abgewickelt werden:

Datenbereinigung, -zusammenführung und -erkundung, Merkmalskonzeption, -modellierung, -schulung und -prüfung, Hyperparameteroptimierung, Modellvergleich, -bewertung und -bereitstellung für die Produktion. Die Modelle werden unter Verwendung von AWS-integrierten und benutzerdefinierten containerisierten Deep Learning TensorFlow 2.0-Algorithmen erstellt. Sowohl Schulungs- als auch Inferenz-Pipelines sind in der Produktionsumgebung betriebsbereit und generieren tägliche Batch-Vorhersagen und benutzerdefinierte

Geschäftsberichte über Tableau-Dashboards. Diese neue Infrastruktur ermöglichte die Implementierung, Schulung, Hyperparameteroptimierung, Bereitstellung und Überwachung von ML-Anwendungen, was zu einer Zeitersparnis von 50% beim Entwerfen und Ausführen eines vollständigen ML-Workflows führte. Darüber hinaus steuert eine benutzerdefinierte Modellüberwachung die Leistung des ML-Modells in der Produktion, so dass es nur bei einer Datendrift neu geschult werden muss und die Kosten für die Modellaktualisierung um bis zu 75% reduziert werden.



**Diese flexible Plattform ermöglicht die Entwicklung und Ausführung einer unbegrenzten Anzahl prädiktiver und präskriptiver Modelle für maschinelles Lernen.** Die Plattform kann von einer großen Anzahl interner und externer Benutzer sowohl über Batch- als auch über Echtzeit-Vorhersageendpunkte genutzt werden und ist bei Bedarf entsprechend den Geschäftsanforderungen skalierbar.