

**// ADASTRA**

**Case Study**

# **Adastras Big Data Lake**

Adastras innovativer Big Data Lake bringt  
Ordnung in das Datenchaos,  
Cloud-basiert oder on-premise.



Eine ungeheure Menge digitaler Informationen durchdringt jeden Geschäftsbereich großer Unternehmen. Nie war die Verfügbarkeit von Daten größer. Darin schlummert gewaltiges wirtschaftliches Potenzial. Adastra konzipierte eines von Deutschlands größten Big Data Lakes und definierte so die Best Practices für rund 200 Entwickler einer der größten deutschen Banken. Heute ermöglicht das innovative Projekt nicht nur gewinnbringende Analysen im Regulatory Reporting und Customer Analytics; die Implementierung auf der Cloud bringt sogar noch weitere Chancen in diversen Anwendungsgebieten mit sich.



## **Big Data - eine Mammutaufgabe?!**

Seit geraumer Zeit ist der Begriff in aller Munde: Big Data. Die zahlreichen Daten bieten ein scheinbar unendlich großes Potential, denn in ihnen steckt viel mehr, als man auf den ersten Blick sieht. Durch die großen Mengen an Kundendaten können Firmen unter anderem ihre Prozesse optimieren, Produkte an die Bedürfnisse ihrer Zielgruppe anpassen und dadurch ihre Gewinne steigern. Doch fällt der Begriff Big Data, wird nicht selten im selben Atemzug festgestellt, vor welche Herausforderung diese Mammutaufgabe die Unternehmen stellt. Wohin mit all den Daten und zu welchem Zweck eigentlich?



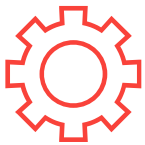
## **Vorreiter im Big Data Engineering**

Mit allein in Deutschland fast 13 Millionen Kunden verfügt auch eine der größten deutschen Banken über zahlreiche Daten, die sie für die Verbesserung ihrer Angebote nutzen kann. Jedoch ist es dafür zunächst notwendig, die Daten aus verschiedenen Quellen zu sammeln, zusammenzuführen und in einem Big Data Lake aufzubereiten.

Als Vorreiter in diesem Bereich hat Adastra gemeinsam mit der Bank einen innovativen Big Data Lake entwickelt. Da der Markt über keine fertigen Ressourcen für einen solchen Big Data Lake verfügte, wurde es eigens in Adastras Data Engineering und AI Labs konzipiert und dann gemeinsam mit dem Kunden entwickelt. Der technologische Durchbruch gelang im Jahr 2014 durch die Migration einer SAS Engine auf Apache SPARK, wodurch die Berechnung der Prozesse deutlich verkürzt werden konnte.

Neben Apache SPARK basiert das Framework auch auf Apache Hive, Kafka, SPARK Streaming zur Verarbeitung sowie zum Streamen der Daten aus allen Geschäftsbereichen, sodass diese für weitere Analysen und Verarbeitungen zur Verfügung gestellt werden.

Der Data Lake basiert auf wiederverwendbaren zentralen Komponenten für das Laden der realen Daten in unterschiedlichen Schichten des Data Lakes einerseits wie auch für die Steuerung der Log-Daten und die Steuerung der Ladeprozesse andererseits. Es bildet die Grundlage für die strukturierte und einheitliche Arbeit von etwa 200 Entwicklern, indem es bestimmte Prozesse als Best Practices definiert.



## **Adastra entwickelte einen der größten Data Lakes Deutschlands**

Der Big Data Lake ist nun seit vier Jahren im Einsatz und wird stets weiter optimiert. Er wird in diversen Bereichen eingesetzt, wie z. B. im Regulatory Reporting, Customer Analytics sowie Zahlungsverkehr Analytics. Nach der Konzeptionierung und Mitentwicklung des Projekts betreut Adastra auch den weiteren Verlauf. Durch die enge Zusammenarbeit während der gesamten Projektphase konnte durch das gemeinsame Wissen und die starke Expertise der Spezialisten von Adastra und der IT-Abteilung der Bank eine solide Grundlage für diverse Folgeprojekte geschaffen werden. Der Big Data Lake ermöglicht die Beantwortung zeitaktueller Fragestellungen, die durch die enormen Datenmengen und ihre umfassende Analyse und Verarbeitung beantwortet werden können.



## **Die Implementierung des Big Data Lakes auf die Cloud bringt neue Chancen**

Drei Jahre nach der Entwicklung des innovativen Big Data Lakes hat sich der Markt weiterentwickelt und neben zahlreichen Herausforderungen auch hervorragende Chancen eröffnet. Die Speicherung von Daten auf der Cloud bietet eine agile Alternative zu herkömmlichen Speichermethoden. Cloud Technologien vereinfachen Prozesse und überzeugen durch ihre unbegrenzten Speicherkapazitäten.

Adastra implementiert die modernen Technologien, die auch im Big Data Lake eingesetzt werden, bei einem der größten Automobilhersteller Deutschlands auf der Cloud. Durch die Anwendung von AWS S3, AWS EMR, AWS Redshift, AWS Athena und AWS Glue sowie Apache Spark können die Datenmengen sowohl in den Bereichen Logistik und Produktion als auch für Connected Cars geordnet, aufbereitet und gewinnbringend eingesetzt werden.

