

BIG DATA UND ASAM ODS

Test- und Simulationsdatenanalyse mit **Big Data**

Die Menge der Daten, die bei Tests und Simulationen gesammelt wird, ist in den letzten Jahren massiv gewachsen und es ist kein Ende dieser Entwicklung in Sicht. Bisherige Techniken zur Analyse stoßen an ihre Grenzen. Die Kombination bestehender Standards mit Big Data bietet einen Ausweg.

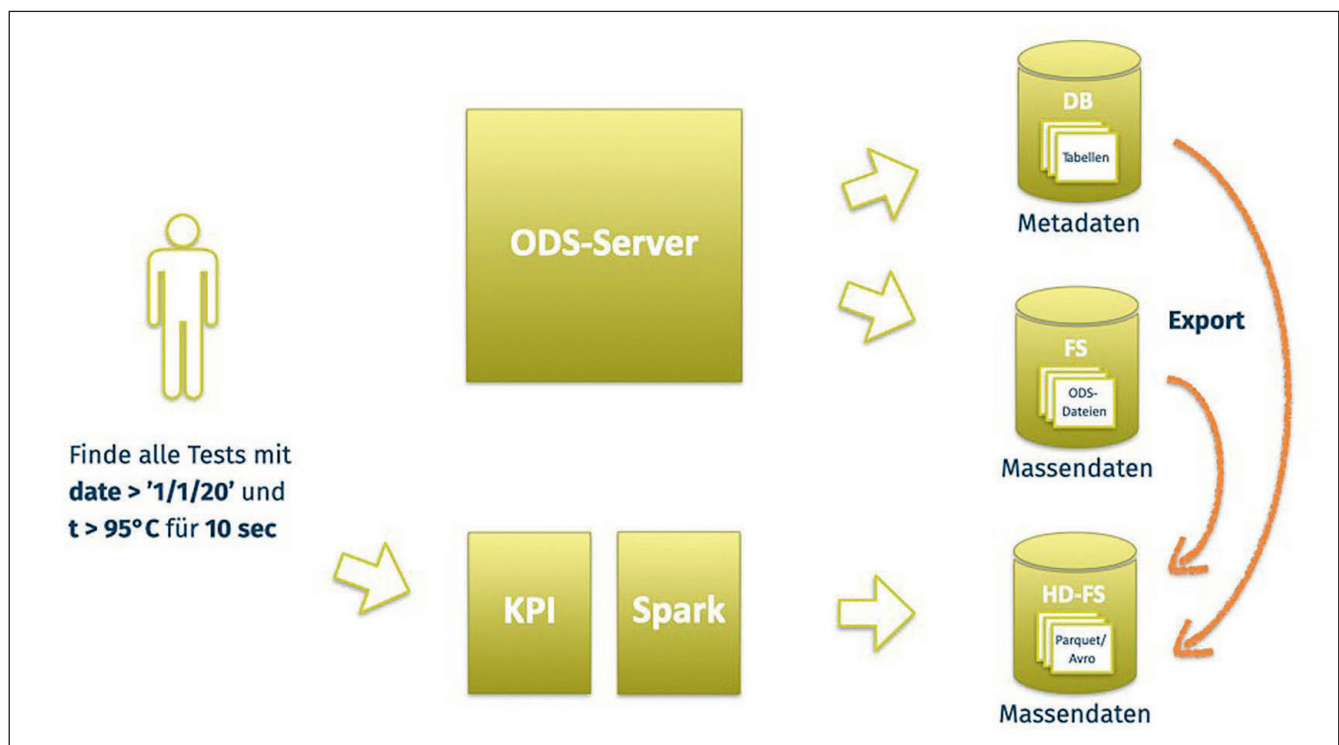


Bild 1: Integration durch Export und Konvertierung der Daten. © Karakun

Während die bei Tests und Simulationen anfallende Datenmenge immer größer wird, können einzelne Computer nicht im gleichen Maße erweitert werden.

Damit erreicht die bisher übliche Praxis, Daten zur Analyse auf einen Arbeitsplatzrechner zu kopieren, ihr Limit. Hinzu kommt, dass Transfer und Kopieren der Daten selbst zu einem Zeit- und Kostenfaktor wird.

Die Automobilbranche ist mit dem Problem, immer größere Datenmengen verarbeiten zu müssen, nicht allein. Zur

Lösung sind in anderen Bereichen Techniken entwickelt worden, die mit dem Begriff „Big Data“ zusammengefasst werden. Unglücklicherweise entstammt ASAM ODS, ein von der Automobilindustrie entwickelter Standard zur Verwaltung von Testdaten, aus einer Zeit vor Big Data und wurde bislang unabhängig weiterentwickelt. Erst seit kurzem gibt es erste Bemühungen, ODS und Big Data zusammenzubringen. Die verschiedenen Möglichkeiten werden in diesem Artikel beleuchtet.

Neben dem bereits erwähnten Vorteil, dass bestehende Verfahren mit Hilfe von Big Data auch auf sehr große Datenbestände angewandt werden können, ermöglicht die Kombination von Big Data mit ODS auch neue Einsatzmöglichkeiten.

Ein wesentlicher Anwendungsfall ist die kontinuierliche Berechnung und Anzeige von KPIs (Key Performance Indicators) in Applikationen und Dashboards. So kann sichergestellt werden, dass wesentliche Indikatoren immer im Auge behalten werden und bei Abweichun-

gen schnell reagiert werden kann.

Ein weiterer Vorteil der Nutzung von Big Data zur Ermittlung von KPIs ist die firmenweit standardisierte Berechnung. So können Fehler und Abweichungen aufgrund unterschiedlicher Implementierungen vermieden werden.

Automatisierte Analysen und Berechnungen

Per se wird die Automatisierung von Berechnungen durch Big Data weder vereinfacht noch erschwert. Allerdings wird die regelmäßige Ausführung von Analysen auf großen Datenbeständen erst durch Big Data wirklich praktikabel.

ODS bietet bereits eine Suche über Meta-Informationen, wie z. B. dem Testdatum oder der getesteten Einheit. Möchte man aber die gemessenen Daten selbst in die Suche einbeziehen (beispielsweise möchte man alle Tests suchen, bei denen die Temperatur für einen Zeitraum in einem kritischen Bereich war) helfen einem die Standardfunktionen von ODS nicht weiter. Eine solche Anfrage kann jedoch mit Hilfe von Big Data relativ zügig beantwortet werden. Deutlich komplexere Suchen sind natürlich auch möglich.

Auch die Suche nach Fehlern kann mit Big Data verbessert werden. Üblicherweise werden neue Daten einmalig beim Import überprüft. Dies hat den Nachteil, dass neue Erkenntnisse nur mühevoll auf alte Datensätze angewandt werden können. Erfolgt die Fehlersuche mit Hilfe eines auf Big Data ba-

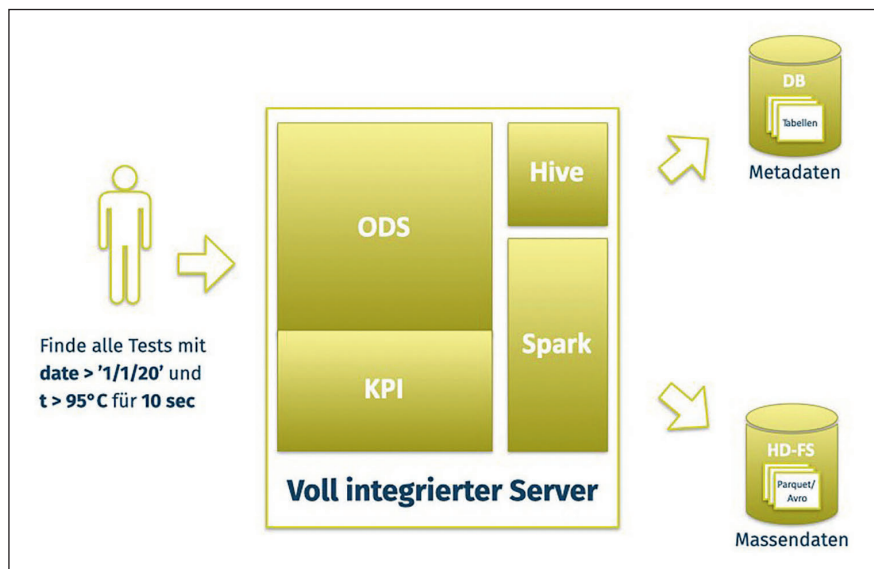


Bild 3: Ein voll integrierter Server als langfristige Lösung. © Karakun

sierenden Ansatzes, können sämtliche Daten zu jedem Zeitpunkt erneut geprüft werden. Ein weiterer Vorteil dieses Ansatzes ist wiederum, dass die Algorithmen zur Fehlerkontrolle zentral und somit standardisiert abgelegt und ausgeführt werden, was wiederum Inkonsistenzen und Fehler vermeidet.

Export der Daten

Der einfachste Ansatz zur Integration von ODS und Big Data besteht im Exportieren der Daten. Für diese Möglichkeit gibt es bereits einen Associate Standard der ASAM ODS Working Group. Daten werden zur Analyse vom ODS-Server in einem Format exportiert, das für den Einsatz in einem Hadoop-

Cluster optimiert ist. Bild 1 skizziert diesen Ansatz. Vor- und Nachteile dieser Lösung liegen auf der Hand. Sowohl Implementierung als auch Einführung erfordern wenig Aufwand. Durch das optimierte Dateiformat können Analysen effizient berechnet werden.

Dafür ist die Lösung umständlich, denn der Export stellt einen zusätzlichen, meist manuellen Schritt dar. Zusätzlich müssen die Daten doppelt gespeichert werden. Soll das Ergebnis der Analyse im ODS-Server gesichert werden, muss es in einem weiteren Schritt wieder importiert werden. Als Übergangslösung ist dieser Ansatz eine Möglichkeit, auch wenn langfristig der umständliche Workflow und die doppelt gehaltenen Daten einen gravierenden Nachteil darstellen und eine tiefere Integration angestrebt werden sollte.

Apache Spark Connector

Apache Spark kann Daten nicht nur aus einem Hadoop Dateisystem lesen, sondern auch aus anderen Datenquellen. Bild 2 zeigt, wie die Massendaten über einen speziellen Connector direkt aus den ODS-Dateien ausgelesen werden können.

Ein großer Vorteil dieser Lösung ist ebenfalls, dass Umsetzung und Einführung wenig Aufwand erfordern. Auch im täglichen Betrieb ist diese Lösung sehr einfach, da der Zugriff auf die Daten vom Benutzer unbemerkt geregelt wird. Hinzu kommt, dass die Daten bei diesem Ansatz nicht dupliziert werden

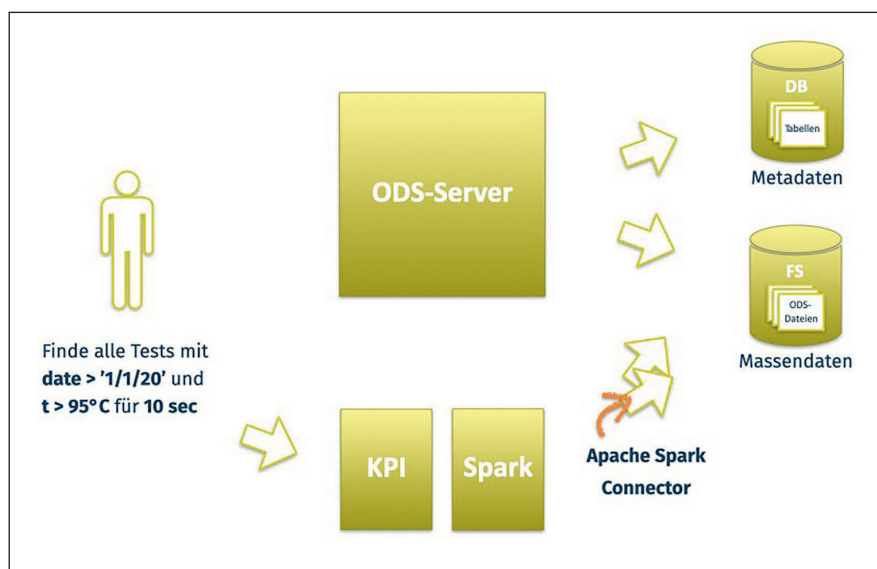


Bild 2: Integration mit Hilfe eines Apache Spark Connectors. © Karakun

müssen. Nachteil dieser Lösung ist, dass die Daten bei jedem Zugriff von Apache Spark erst konvertiert und in der Big Data Infrastruktur verteilt werden müssen. Dadurch können die Stärken von Big Data nicht vollständig ausgespielt werden und Analysen laufen weniger effizient.

Auch dieser Ansatz empfiehlt sich als Übergangslösung. Bei weiterwachsender Datenmenge wird sich der Performanzverlust aber früher oder später bemerkbar machen.

Voll integrierter Server

Die tiefste Integration erreicht man mit einem speziell entwickelten ODS-Server, der direkt mit Daten eines Big Data Clusters umgehen kann. Die Daten werden im Hadoop Dateisystem abgelegt und sowohl der ODS-Server als auch Apache Spark greifen direkt auf die Daten zu, wie aus Bild 3 ersichtlich. Dieser Ansatz bietet die effizienteste Lösung. Daten werden nur einmal ge-

speichert, liegen im optimierten Format für Analysen vor und es sind keine zusätzlichen Arbeitsschritte erforderlich. Hinzu kommt, dass bei dieser Lösung ODS- und Big Data Funktionalität vermischt werden kann, da beides voll integriert ist. Unglücklicherweise bietet keines der aktuell am Markt befindlichen Tools derzeit eine so tiefe Integration. Hinzu kommt, dass bestehende Daten zunächst migriert werden müssen, bevor sie als Teil dieser Lösung genutzt werden können.

Nur diese Lösung ist effizient genug, um auch mit zukünftigen Datenmengen umgehen zu können. Der Nachteil ist, dass die Migration bestehender Daten einen erheblichen Aufwand bedeuten kann.

Zusammenfassung

Keiner der drei vorgestellten Lösungsansätze für sich ist perfekt. Möchte man zeitnah und ohne großen Migrationsaufwand Big Data Analysen auf be-

stehenden Test- und Simulationsdaten ausführen, sollte der Datenexport oder ein Connector für Apache Spark in Betracht gezogen werden. Langfristig führt aber kein Weg an einem vollständig integrierten Server vorbei, denn nur so kann sichergestellt werden, dass auch die in Zukunft anfallenden Datenmengen effizient verwaltet und analysiert werden können.

Bei der Karakun AG verfolgt man deshalb mit Lyzium (<https://lyzium.com>) einen hybriden Ansatz bestehend aus einem voll integrierten Server inklusive einem Apache Spark Connector. Damit stehen für bestehende Daten sofort die vollen Möglichkeiten moderner Big Data Technologien zur Verfügung, aber auch für die Zukunft mit noch größeren Datenmengen und komplexeren Analysen ist man gerüstet. ■ (oe)

www.karakun.com

Michael Heinrichs ist Mitgründer der Karakun AG, Basel.

RIGOL
Possibilities and More

Die nächste Generation kompakter Speicheroszilloskope

UltraVision II
TECHNOLOGY



ab € 8.999,-
plus MwSt.



DS8000-R

1 HE 19" Einschub-Digital-Speicheroszilloskop

- 350 MHz bis 2 GHz Bandbreite
- 4 analoge Kanäle (Standard),
flexibel erweiterbar auf bis zu 512 Kanäle
- Hochgeschwindigkeits-Schnittstelle 10 GE SFP+
- Bis zu 10 GS/sek. Echtzeit-Abtastrate
- Bis zu 500 Mpts. Speichertiefe
- FFT @ 1 Mio. Abtastpunkte und Spitzenwertsuche
- Inklusive Datenerfassungssoftware ULTRADAQ
- Optionen: Decoding RS232, I2C, SPI, I2S, CAN, LIN, MIL, Flexray etc.
- Echtzeit-Augendiagramm und Jitter-Analysesoftware (Option)

3 Jahre Garantie – verlängerbar!

Registrieren
Sie sich
für unseren
Newsletter →



www.rigol.eu



SCOPE



SPECTRUM



AWG



DVM



COUNTER



PROTOCOL



RECORD



PASS/FAIL



POWER
ANALYSIS

9-IN-1