DATENMIGRATION ZUR STUFENWEISEN MODERNISIERUNG

VON HISTORISCH GEWACHSENEN SYSTEMLANDSCHAFTEN

Oracle Forms gehört zu den Pioniertechnologien des Rapid Developments. In vielen Unternehmen ist es ein wesentlicher Bestandteil der Datenverarbeitung und untrennbar mit der Datenhaltung verbunden.

Gerade diese tiefe Einbettung wurde in vielen Fällen zu einer großen strategischen Hürde – die Java-Applet-Technologie, auf der Forms lange basierte und oft noch im Einsatz ist, ist veraltet und Experten, die Oracle Forms entwickeln können, sind zunehmend schwerer zu finden. Dazu kommt, dass die Geschäftslogik oft teils im Forms-Code und teils in verschiedenen Stored Procedures liegt, die über Trigger oder Jobs aufgerufen werden. Das erhöht die Komplexität einer Ablösung erheblich. Daher bietet sich eine Migration in Etappen an – und dabei unterstützen wir in mehreren Bereichen.

HERAUSFORDERUNG

Metro Digital beauftragte uns mit der Datenmigration für ein solches Projekt. Alt- und Zielsystem mussten in einer Koexistenzphase parallel laufen, um einen riskanten Big-Bang-Wechsel zu vermeiden.

Dafür synchronisierten wir zwei komplexe Oracle-Landschaften bidirektional und unterstützten die Migration von Teradata auf BigQuery in der Google Cloud. Die Anforderungen entstanden parallel zu unserer Lösung und änderten sich laufend.

Die Übertragung musste korrekt und für die Wirtschaftsprüfung belegbar sein. Da beide Systeme die Prozesse unterschiedlich abbildeten, musste die Datenqualität den Vorgaben und der Validierung beider Systeme genügen. Zu korrigierende Datensätze wurden automatisch erkannt, damit die Datenanreicherung parallel zur Migration laufen konnte.

LÖSUNGSANSATZ

Wir entwickelten eine ETL-Plattform, um Migrationspipelines schnell umzusetzen, zu orchestrieren und Fehler für die Analyse zu isolieren.

Das Fundament bildete Apache Spark mit Java in einer serverlosen Google-Cloud-Umgebung. Die ursprünglichen Abstraktionen blieben fast unverändert, die Pipelines wuchsen mit den Anforderungen: neue Validierungsregeln, systemübergreifende Lookups und geänderte Mappings. Über 400 Versionen der Mappingregeln liefen per CI/CD mit minimalem Aufwand aus.

Ein Quarantänesystem versah fehlerhafte Datensätze mit sprechenden Meldungen zur gezielten Fehlerbehebung. Apache Airflow orchestrierte die unabhängigen Pipelines, Zusätze wie E-Mail-Trigger waren reine Konfiguration.

UMSETZUNG

1 CLOUD-INFRASTRUKTUR

Per Infrastructure-as-Code bauten wir ein Spoke-Projekt in der Google Cloud Platform auf der Infrastruktur des Kunden auf, inklusive Netzwerkanbindung an die OCI-Systeme (Oracle Cloud Infrastructure) und Verschlüsselung bei Übertragung und Speicherung. So waren Datensicherheit und Compliance gewährleistet.

2 SKALIERUNG NACH BEDARF

Mit Serverless Apache Spark in der Google Cloud Platform skalierten einzelne Workloads automatisch. Große Datenmengen aus einem Data Warehouse ließen sich so etwa in einem größeren Cluster deutlich schneller migrieren. Da wir keinen eigenen Server betreiben mussten, sanken die Betriebskosten in diesem Projekt erheblich.

METRO digital × POSEDIO

Do it RIGHT.

3 MIGRATIONS-KONZEPT

Passend zum ETL-Prozess definierten wir 2 Umgebungen: die Landing Zone als Ergebnis der Extraktion und die Curated Zone als Ergebnis der Transformation inklusive Validierung. Je nach Datenvolumen realisierten wir sie in AlloyDB oder als Parquet-Dateien in Google Cloud Storage, damit die Kosten überschaubar blieben.

4 FLEXIBLE ARCHITEKTUR

Eine durchdachte Java-Architektur nach dem Prinzip Composition over Inheritance und ein Testkonzept mit Testcontainers für echte Spark-Integrationstests trugen die ETL-Plattform. Ob CSV, Teradata, Oracle oder statische Daten – das war nur noch eine Frage der Konfiguration.

5 BAU DER PIPELINES

Über GitHub CI/CD konnten Domänenexperten Mappingregeln als SQL-Dateien einreichen, die direkt in die Plattform eingespielt wurden. Das automatisierte Releasemanagement stellte sicher, dass jede Pipeline-Ausführung fest an eine Version der Mappingregeln gebunden war. Meist genügten ein paar Parameter, um eine neue Pipeline lauffähig bereitzustellen.

6 FEHLERANALYSE

Die sprechenden Meldungen des Quarantänesystems machten Widersprüche und Lücken in den Migrationskonzepten sichtbar. Gemeinsam mit den Domänenexperten des Kunden lösten wir diese auf.

7 OBSERVABILITY & LOGGING

Airflow schickte im Fehlerfall Benachrichtigungen an Microsoft Teams, sodass wir Probleme in den Pipelines früh erkannten und rasch reagieren konnten. Durchsuchbare Log-Dateien im Parquet-Format zeigten die Migrationshistorie jedes Datensatzes und führten so schneller zur Ursache eines Problems.

ERGEBNISSE

ZENTRALE KOMPONENTE

Unsere ETL-Plattform ist das Herzstück eines großen Migrationsprojekts. Neben der Oracle-Migration versioniert sie die Mappingregeln und dient als Nachweis für die Wirtschaftsprüfung.

FLEXIBILITÄT

Flexibilität war früh eines der wichtigsten Designziele. So konnten wir die Domänenexperten technisch begleiten und neue Anforderungen gemeinsam entwickeln, oft samt Tests in weniger als einer Stunde in Produktion.

SICHERE KOEXISTENZ

Während der Koexistenzphase orchestriert die Plattform die regelmäßigen Datenabgleiche zwischen Alt- und Zielsystem, bis alle Metro-Filialen in Österreich an das Zielsystem angeschlossen sind.

EFFIZIENZGEWINN

Unterm Strich steht ein deutlicher Effizienzgewinn: Die Plattform und ihre wiederverwendbaren ETL-Bausteine bilden die Basis für ca. 300 ETL-Pipelines und einen wiederholbaren Migrationsprozess über alle österreichischen Filialen hinweg.

Dieses Projekt hat erneut gezeigt: Eine One-Size-Fits-All Lösung gibt es nicht. Wer Kundenanforderungen versteht und schon früh in das System- und Anwendungsdesign einbettet, spart später Kosten und Aufwand. Komplexe Projekte verlangen Agilität – und mit unserer ETL-Plattform waren wir dafür bestens gerüstet.

WEITERE INFOS

www.posedio.com
sales@posedio.com
+43 660 915 44 19

METRO digital × POSEDIO

Do it RIGHT.