

Löst DIPC ODI, Golden Gate und Data Quality ab?

Holger Dressing

Oracle Deutschland BV & Co. KG

Schlüsselworte

Data Integration Plattform Cloud, DIPC, Cloud, Datenintegration, Oracle Data Integrator, ODI, Golden Gate, Data Quality, EDQ

Einleitung

Data Integration Plattform Cloud oder kurz DIPC ist die Cloud-Lösung für die Datenintegration. Eigentlich müßte man die Frage nach den Datenintegrationsprodukten noch ausweiten: Zunächst hat man in OWB und Streams investiert, anschließend in ODI, Golden Gate und Data Quality. Nach nur wenigen Jahren bzw. unter Berücksichtigung des Supports nach nur wenigen Monaten bringt Oracle schon wieder etwas neues. Schließlich handelt es sich um Entwicklungssysteme, bei denen sich die Investitionskosten durch die Entwicklung und die Einbettung in die Produktion noch erheblich steigern lassen.

Aber in der IT gibt es ein neues Paradigma und das heißt Cloud. Auch Oracle hat darauf reagiert und die Einzelprodukte ODI, Golden Gate und Data Quality in einer Cloudlösung zusammengefaßt. Darüber wird schon ziemlich lange diskutiert: Als Kunde möchte ich nicht über die Einzelfeatures mit Oracle diskutieren, sondern anwendungsbezogen die fachlich benötigten Features einsetzen können ohne sich mit Lizenzfragen zu belasten bzw. einzuschränken. Genau auf diese Punkte zielt DIPC. Mit DIPC stehen ODI, Golden Gate und Data Quality 1:1 auch in der Cloud zur Verfügung. Entwicklung und Produktion können wechselweise on Premise als auch in der Cloud durchgeführt werden. Es gibt noch weitere Möglichkeiten: Auch hybride Ansätze werden z. B. durch remote Agents unterstützt. Neben den Entwicklungswerkzeugen, die häufig eher für den Software-Entwickler gedacht waren, stehen zusätzlich Werkzeuge zur Verfügung, die auch von Fachanwendern genutzt werden können.

Data Integration Plattform Cloud im Überblick

Die DIPC fasst viele unterschiedliche Anforderungen aus dem Bereich Datenintegration zusammen: transformieren, bereinigen, integrieren, analysieren und steuern von Daten. Es lassen sich batch- und real-time basierte Lösungen in der Cloud als auch On-Premise erstellen, die die Konsistenz der Daten ohne Fehlertoleranzen belastbar verarbeiten. Dazu gehört

- das Verarbeiten von Massendaten im Batch- oder Real-Time in der Cloud,
- das Synchronisieren und Kopieren von großen Datenmengen in eine Cloud mit einfachen Zugriff-, Prüf- und Transformationsmethoden,
- das Streamen von Daten in Real-Time in neue Datenquellen, Anstoßen von Data Analytics Verfahren mit gestreamten Daten auf einer Vielzahl von zu synchronisierenden Datenquellen,
- das Replizieren von Daten nach Kafka und das Laden von „Stream Analytics“ basierenden Daten,
- das Transformieren von Massendaten mit Szenarien zum Import und zum Ausführen von Mappings,
- das Importieren von Daten aus Flat Files oder On-Premise Daten in Oracle Object Storage Data Lake sowie
- das Integrieren von Big Data Technologien.

Damit verbunden sind Anwendungsfälle wie das vereinfachte Befüllen von Data Warehouses, das Automatisieren der Erstellung von Data Marts, Zero down tim (ZDT) Datenmigration, die Big Data Integration, die Datensynchronisation, das Prüfen von Dateninhalten, Data Profiling und das Validieren von Daten.

Die Architektur von DIPC

Die Architektur von DIPC basiert auf der Infrastruktur und der Funktionalität der Oracle Public Cloud, integriert ist auch die Identity Management Cloud für Benutzer und die Zugriffsrechte über alle Oracle Cloud Services. In der Vergangenheit mussten ETL Entwickler und Datenbankadministratoren die komplexen Operationen für Real-time Analysen aufsetzen und die notwendigen Schemata in der Datenbank anlegen. Dazu mussten sie die komplexen Strukturen und Workflows kennen und ggf. auch eine Governance aufsetzen. In der folgenden Abb. werden die Zusammenhänge zwischen den einzelnen ETL-Features und der Data Governance aufgezeigt.

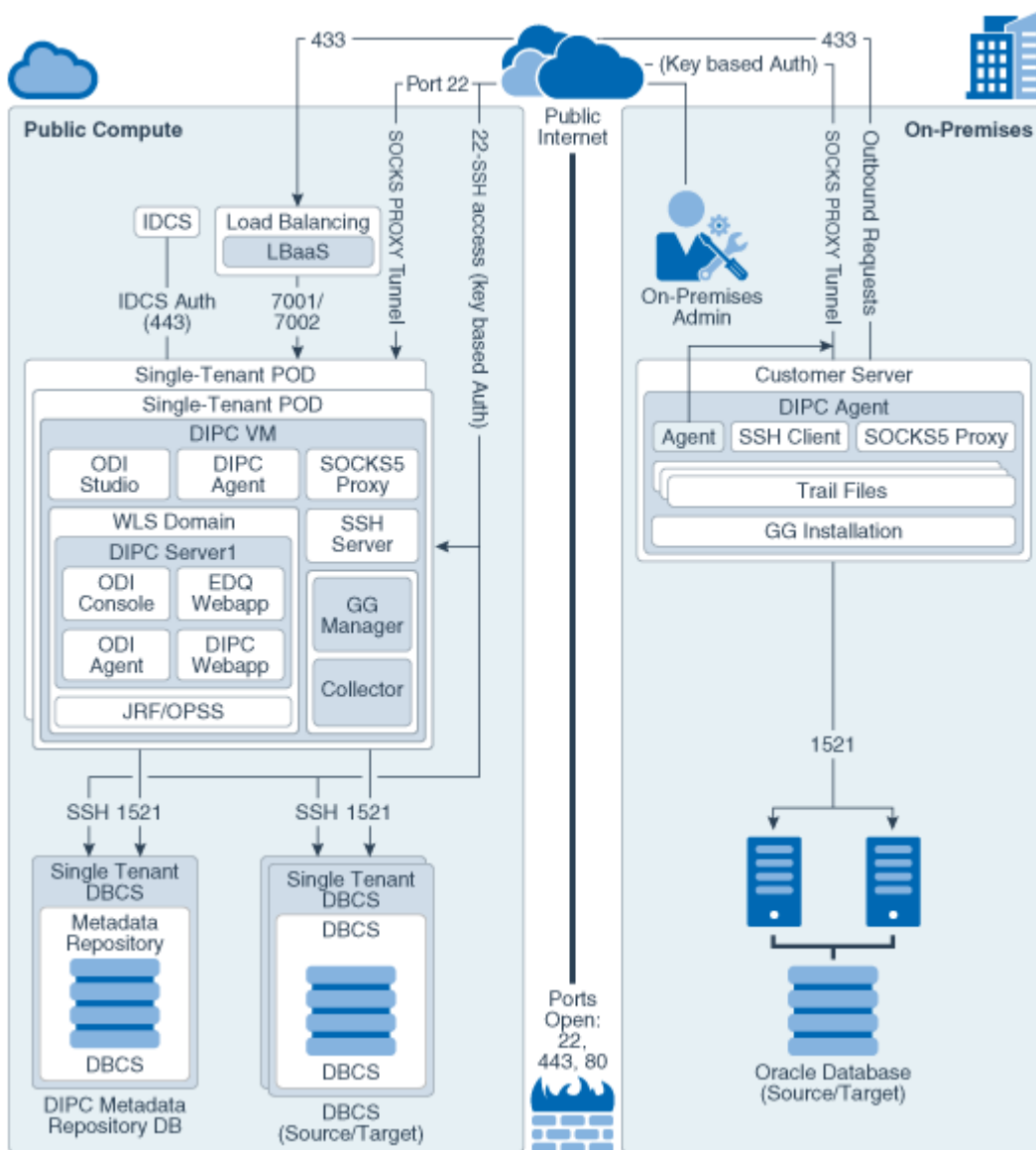


Abb. 1: DIPC Architekturdiagramm

Die Integration verschiedener Komponenten mit Massendatenverwaltung für ETL und zeitnahe Replikation kann äußerst komplex sein. DIPC nimmt diese Komplexität und hilft ETL Entwicklern und Datenbankadministratoren, quasi auf Knopfdruck eine komplexe Umgebung mit allen bislang eingeführten Integrationsprodukten aufzubauen.

User-managed oder autonomous Variante

DIPC wird in einer autonomous und einer user-managed Variante angeboten. Die Data Integration Platform Cloud Console ist bei beiden Varianten gleich. Bei der autonomous Variante wird die Administration automatisiert, so dass sich der Anwender auf die Integrationsausgaben beschränken kann. Dazu gehört das Patchen, Upgrade, Rollback, Start/Stop/Restart sowie das Backup. Wenn spezielle Datensynchronisierungsaufgaben mit Änderungen in den Standard-Parametern bei ODI oder OGG erforderlich sind, sollte die user-managed Variante gewählt werden. Beide Varianten basieren auf der Oracle Cloud Infrastruktur (OCI). Bei der user-managed Variante wird ein Cloud Agent direkt mit aufgesetzt. Falls Integrationsprozesse auf Systeme On-Premise zugreifen sollen, können in beiden Varianten Remote Agents aufgesetzt werden. Diese müssen jedoch manuell heruntergeladen und installiert werden. In der user-managed Variante kann der Benutzer die Zugriffsrechte selbst verwalten und hat per SSH auf ODI Studio, die ODI Console und die Golden Gate Applicationen in der VM Zugriff. Ebenso kann per IP-Adresse auf die DIPC Instanz sowie die damit verbundenen Datenbanken und deren Speicher zugegriffen werden. Es besteht ein Zugriff auf die Fusion Middleware Control Console und die WebLogic Server Console. In der autonomous Variante ist nur ein Zugriff über ODI Studio möglich, das auf einer user-managed Instanz läuft oder On-Premise. Die Prozesse können per Datei importiert/exportiert werden. Es besteht die Möglichkeit, Dateien an die Administration zu übergeben. VPN als Service ist für die autonomous Instanzen nicht verfügbar, wohl aber für die user-managed. Autonomous DIPC Instanzen benötigen erweiterte Database Cloud Service Instanzen für die Speicherung des Repositories. Bei user-managed Instanzen wird eine Database Cloud Service Instanz benötigt, auf der selbstständig das Repository angelegt werden muß. Als Zähler für die Abrechnung dient bei der autonomous Variante die pro Stunde verarbeitete Datenmenge. Aktuell wird nach 1 GB/5 GB/10 GB/20 GB Daten pro Stunde abgerechnet. Bei der user-managed Variante wird nach „shapes“ abgerechnet. Shapes sind vordefinierte Ressourcen-Profile mit einer Anzahl von OCPUs, Speicher und Plattenkonfigurationen. Optionen in der user-managed Variante sind OC1M – 1.0 OCPU – 15.0 GB Memory oder OC3M – 4.0 OCPU – 60.0 GB Memory. Identity Cloud Service ist für beide Varianten verfügbar, um damit User und Rollen zu verwalten. Die Kombination mit BIG Data Features ist in der autonomous Variante an den Big Data (OGG) Agent verbunden. In der user-managed Variante können OGG Big Data sowie MySQL binaries für heterogene Szenarien aufgesetzt werden. Ein Up-/Downscaling ist nur in der user-managed Variante möglich. Ebenso ist Stream Analytics nur in der user-managed Variante möglich. Load Balancing Optionen sind in beiden Varianten verfügbar, für Anwender administrierbar ist das auch nur in der user-managed Variante.

Aktuell gibt es alle 3 Monate ein neues Release, daher sind die unterstützten Systeme nur eine Momentaufnahme. Bitte prüfen Sie die entsprechenden Informationen auf der Oracle Website.

Editionen von DIPC

Es gibt 3 Editionen von DIPC und diese geben einen Hinweis auf ODI, OGG und EDQ.

Standard	Extrahieren, Transformieren und Laden von Massendaten. Dies sind Funktionen, die mit ODI verknüpft sind.
Enterprise	Zusätzlich: Erweiterte Integrationsfunktionen wie das Replizieren und Streamen von Daten, Streams Analytics und Big Data Technologien. Diese Funktionen sind verknüpft mit OGG oder Big Data Optionen von OGG und ODI.
Governance	Zusätzlich: Data Quality, Data Profiling and Data Governance Funktionen. Diese Funktionen sind verknüpft mit Enterprise Data Quality.

Beim Betrachten der Architektur und der Editionen wird klar, dass die jeweiligen Produkte aus der on Premise Welt in DIPC integriert sind. Damit ergibt sich beispielsweise, dass die Entwicklungen weiterhin On Premise erfolgen können und dann ein Deployment in die Cloud erfolgen kann. Es gibt aber auch eine neue browserbasierte Oberfläche in DIPC, mit der alle Schritte von der Entwicklung bis zur Produktion durchgeführt werden können.

Die nachfolgend dargestellten Umgebungen zeigen die Möglichkeiten auf, eine genaue Beschreibung ist in den unter dem Abschnitt „Links“ dargestellten Oracle By Example Demos dargestellt.

ODI und DIPC

Mit ODI Studio erstellte Szenarien können in DIPC ausgeführt und zentral gemonitort werden. Szenarien sind eingefrorener Code für alle Arten von Prozessen wie Mappings, Packages oder Procedures. Dazu gehört, dass die passenden Parameter auf Quellen und Ziele (Topology in ODI) sowie die Repositories (Development Repositories) aufgesetzt sind, wie man das aus der on Premise Variante kennt. Der ODI kann in der Cloud installiert werden und beispielsweise via VNC ist ein Zugriff auf die Cloud möglich.

Ebenfalls wird ein ODI Agent benötigt. In DIPC gibt es eine weitere Variante des Agents, der Remote Agent genannt wird. Es wird die DIPC Infrastruktur in der Cloud benutzt, um dort Szenarios auszuführen und der Remote Agent führt Szenarios on Premise aus und schreibt die Protokollinformationen in das Repository von DIPC. Ein Remote Agent muß von den Anwendern selbst heruntergeladen, installiert und gestartet werden.

OGG und DIPC

OGG läuft in DIPC genauso wie man das in der on Premise Variante kennt. Es besteht die Möglichkeit, mit Golden Gate von on Premise in die Cloud oder von Cloud nach Cloud zu replizieren. Ebenso wie bei ODI kann hier der Remote Agent eingesetzt werden, gemischte Umgebung on Premise und in der Cloud aufzubauen. Zusätzlich gibt es die Möglichkeit, nach Kafka on Premise zu replizieren.

EDQ und DIPC

EDQ ist ein Werkzeug, das aus einem GUI, einem Repository und einem Business Layer besteht. Aus Entwickler- und Anwendersicht ist der Director von EDQ das wichtigste Werkzeug. Die EDQ Werkzeuge können ebenso wie on Premise aufgerufen und benutzt werden.

Zusammenfassung

DIPC ist zwar ein neues Produkt mit einer eigenen Oberfläche. Aber es besteht weiterhin die Möglichkeit, mit den vorhandenen on Premise Produkten zu entwickeln und dann die erstellten Integrationslösungen auf DIPC zu betreiben. Ziel war die Integration der on Premise und der Cloud

Lösungen zu beschreiben, aber DIPC bietet eine integrierte GUI ohne die bisherige Unterscheidung nach Produkten. Bislang gibt es noch Einschränkungen in der Benutzeroberfläche von DIPC hinsichtlich der Funktionalität. Der Abstand in den Möglichkeiten wird sich jedoch in den kommenden Release verringern. Bislang gibt es noch die Möglichkeit, aus den graphischen Tools von ODI, OGG und EDQ auszuweichen. Aus der Sicht der Lizenzen gibt es die Möglichkeit, mit „Bring your own Licence“ die on Premise Lizenzen mit DIPC zu kombinieren. Man muß sich zwar hinsichtlich der Lizenzierung unterscheiden, aber die technische Realsierung und der Betrieb können ausgehend von ODI, OGG und EDQ auch in der Cloud mit DIPC weitergeführt werden.

Links

DIPC Produktüberblick: https://cloud.oracle.com/en_US/data-integration-platform, Zugriff am 08.11.2018

DIPC Dokumentation vom Oktober 1998: <https://docs.oracle.com/en/cloud/paas/data-integration-platform-cloud/index.html>, Zugriff am 08.11.2018

Oracle by Example/Do it yourself tasks for ODI/OGG/EDQ: <https://docs.oracle.com/en/cloud/paas/data-integration-platform-cloud/index.html>, Zugriff am 08.11.2018

Kontaktadresse:

Dr. Holger Dresing
Oracle Deutschland B.V. & Co. KG
Kühnehöfe 5
D-22761 Hamburg

Telefon: +49 (0) 89091-056
Fax: +49 (0) 89091-056
E-Mail: holger.dresing@oracle.com
Internet: www.oracle.de