

Do's und Don'ts mit der Amazon Oracle RDS - Erfahrungen mit einer hochverfügbaren SaaS Anwendung

Frank Szilinski
esentri software GmbH
Ettlingen

Schlüsselworte

Amazon Cloud, AWS, RDS, SaaS, Datenbank

Einleitung

Für das SaaS Produkt Social PM betreibt esentri seit 1,5 Jahren eine produktive Oracle Datenbank 11g mit den Amazon Relational Database Services, die Teil der Amazon Web Services sind. Die Software ist auf viele Tausend User ausgelegt und hat höchste Anforderungen an Performance und Skalierbarkeit. Neben der Datenbank wird auch ein Middleware-Cluster auf der Amazon Infrastruktur betrieben.

Der Vortrag geht praxisnah auf die Herausforderungen ein, die bezüglich Hochverfügbarkeit, Sicherheit und Backup mit der Amazon Infrastruktur zu meistern sind und zeigt auf, wie sich Amazon RDS optimal nutzen lässt. Dabei wird explizit auf alltägliche Probleme in der Praxis eingegangen und exemplarisch aufgezeigt, wie sich der Betrieb der Datenbank optimieren lässt und was mit den von Amazon zur Verfügung gestellten Services möglich ist.

Die richtige Plattform finden

Um für die eigene Anwendung die richtige Hosting Plattform zu finden sind deren Anforderungen genau zu definieren. Welche Konfiguration werden minimal zum Betrieb benötigt, welche wären optimal und welche Kompromisse lassen sich eingehen, um bestimmten Anforderungen gerecht zu werden.

Im konkreten Fall gilt es für eine Konfiguration einer 3-Tier-Lösung, konkret einer modernen Software as a Service Anwendung, die optimale Plattform zu finden. Im Einzelnen lassen sich die Anforderungen an die verschiedenen Schichten wie folgt stichpunktartig beschreiben:

1. Frontend mit High-Availability-Proxy: hohe Verfügbarkeit im Netzwerk, Ausfallsicherheit, Failover Konfiguration eher geringere Last, statisches Backup nach Konfigurationsänderung ausreichend
2. Middleware mit Jetty: hohe Verfügbarkeit, Auslegung auf Hohe Last, einfache Skalierbarkeit, nach Konfigurationsänderung ausreichend
3. Datenbank: hohe Verfügbarkeit, Ausfallsicherheit, Auslegung auf hohe Last, einfache Skalierbarkeit, permanentes Backup oder Replication Betrieb

Außerdem soll die Hardwarebasis möglichst flexibel gestaltbar sein, um schnelles Wachstum und Zeiträume mit hoher Last schnell bedienen zu können.

Aus allen möglichen Hostingoptionen, die für die eigene Anwendung richtige Infrastruktur zu finden ist daher nicht trivial. und im folgenden wird daher der Weg zur Entscheidung für die Amazon Cloud aufgezeigt.

Erste Alternative: eigenes Hosting

Generell hierzu notwendig sind die eigene Hardwarebeschaffung und eine langfristige Planung. Je niedriger die beschaffte Hardware ausgelastet ist, umso ineffektiver ist deren Einsatz.

Die Netzwerkanbindung muss ausreichend dimensioniert sein. Bei einer SaaS Lösung ist eine hohe Kapazität notwendig, die teuer in der Beschaffung ist. Eine Redundanz bei Ausfall ist nur sehr aufwendig planbar und für kleine und mittlere Unternehmen nur bedingt sinnvoll. Ein Vorteil ist jedoch, dass die Datensicherheit hier vollumfänglich selbst kontrolliert werden. Durch Zugriff auf die eigene Hardware kann eine Backupstrategie frei gewählt werden.

Zweite Alternative: Root Server Hosting

Die Anmietung von klassischen Rootservern oder Virtual Private Server ist meist an eine längerfristige Bindung an die Hardware beziehungsweise der vordefinierten Ressourcen gekoppelt. Eine Skalierung nach oben ist typischerweise nicht kurzfristig möglich, nach unten ist häufig mit kostenpflichtigem Vertragswechsel gekoppelt.

Dahingegen ist die Netzwerkanbindung zumindest bei größeren Providern gut und redundant ausgelegt.

Die Verschlüsselung von kritischen Daten ist hier wichtig, da Unbefugte über möglichen Hardwarezugriff vollen Zugang auf die eigenen Daten erlangen können. In Summe sind die Sicherheitsvorkehrungen in vielen Rechenzentren allerdings als besser einzustufen, als diese im eigenen Unternehmen oft gewährleistet werden könnte. Backups können per Netzwerk übertragen werden und sind je nach Datenmenge über die eigene Infrastruktur realisierbar.

Cloud Server Hosting hat viele Vorteile

Es gibt keinen einheitlichen Standards und auch keine gemeinsame Definition was Cloud Computing ist. Daher ist ein Anbietervergleich immer individuell.

Generell ist die flexible Skalierung hier vorteilhaft. Angemietete Ressourcen können einfach und schnell erweitert werden, umso kritischer müssen hier jedoch die Themen Ausfallsicherheit und Datensicherheit betrachtet werden.

Eine Entscheidung pro Amazon wird durch viele verschiedene Faktoren vereinfacht und es können natürlich die generellen Vorteile von Cloud Computing ausgespielt werden. Ein wichtiger Punkt ist, dass Verfügbarkeitszonen überall auf der Welt angeboten werden. Dieser Punkt hat jedoch auch juristische Relevanz, wenn es beispielsweise um personenbezogene Datenverarbeitung geht. Amazon bietet hierfür auch einen Standort innerhalb der EU an.

Die Angebotene API zur Cloud Verwaltung ist relativ umfangreich und intuitiv. Zu einem schnellen Einstieg werden sowohl vorkonfigurierte Instanzen für verschiedene Betriebssysteme als auch für spezielle Datenbankkonfigurationen angeboten, darunter auch Oracle Datenbanken. Die seit einiger Zeit angebotenen RDS Datenbank Instanzen ergänzen die Auswahlmöglichkeit von Amazon Services zusätzlich. Bei diesem Angebot wird neben dem Betrieb auch die Wartung und das Backup der Daten übernommen.

Für ein Cloudbasiertes Datenbank Hosting muss zunächst die grundlegende Entscheidung getroffen werden, ob man ein eigenes Datenbank Server Image in der Cloud betreiben möchte oder auf die bereits angesprochenen Services aufsetzt. Neben der Hardware-Skalierung behält man damit aber auch alle Nachteile einer selbst gemanagten Umgebung, wie man diese beispielsweise mit einem Rootserver hat. Allerdings behält man die Kontrolle über die Parametrisierung des eigenen Systems und kann auch im Feintuning der Datenbank eingreifen. Je nachdem welche Ressourcen hierfür bei

Amazon gemietet werden, sind hier vor allem lizenzrechtliche Fragen relevant (die DOAG hatte ausführlich über die Lizenzierungsproblematik berichtet). Typische Fragen sind z.B.: darf ich mit meiner vorhandenen Lizenz die doppelte Anzahl virtueller CPUs überhaupt nutzen?

Mit einer RDS (Relation Database Service) Instanz hat man zumindest in der “Standard Edition One” Version die Möglichkeit die Lizenzkosten im Stundensatz der Instanznutzung mit inkludiert zu erhalten. Dabei gibt es folgende unterschiedliche Optionen bei Amazon:

- Standard Edition One - bring-your-own license / license included
- Standard Edition - bring-your-own license
- Enterprise Edition - bring-your-own license

Nachteil aller Lösungen mit Amazon RDS sind der fehlende SYS Zugang und die Festlegung auf vordefinierte Standardkonfigurationen. Eine Individuelle Anpassung ist hier nicht möglich, was natürlich in gewissen Anwendungsfällen einen enormen Nachteil mit sich bringt.

Die Nutzung der RDS Datenbank bietet sich jedoch besonders an, wenn der Umfang der benötigten Ressourcen stark schwankt, man aufgrund gegebener Volatilität sich nicht langfristig binden möchte und eine Konfiguration grundsätzlich ausreicht.

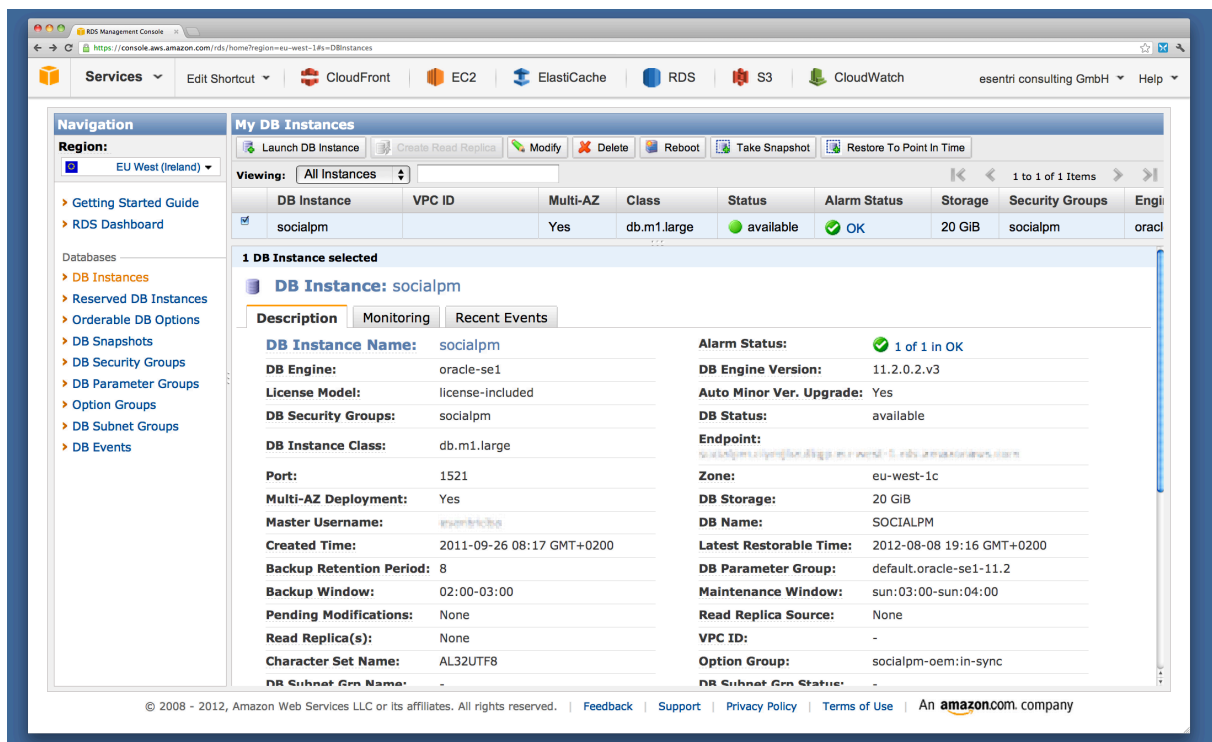


Abbildung 1: Managementkonsole für die Amazon RDS Instanzen

Für die Bereitstellung unserer dynamischen, sich weiterentwickelnden und wachsenden SaaS-Plattform überwiegen in Summe die Vorteile, vor allem getrieben durch die hohe Skalierbarkeit.

Everything fails all the time

Zunächst einmal ist die Verschlüsselung kritischer Daten als sehr wichtig anzusehen. Hierfür gibt es verschiedene Möglichkeiten. Bei einer Instanz mit Datenbank oder eigenem System kann z.B. der Tablespace verschlüsselt werden. In einer RDS Instanz bleibt hingegen nur die Möglichkeit die Daten selbst zu verschlüsseln - die Kommunikation zwischen Middleware und RDS ist generell leider nicht verschlüsselbar, bei korrekter Konfiguration sollten aber alle Datenpakete sicher den richtigen Empfänger im internen virtuellen Netzwerk erreichen.

Die beste Verschlüsselung ist jedoch nutzlos bei Verlust der Daten, daher hat das Thema Datenverfügbarkeit oberste Priorität. Optimal wäre hier der Einsatz einer Read Replica oder eines Multi-AZ-Deployments, um eine echte Redundanz mit Failoverfunktion nutzen zu können. Read-Replicas stehen zur Zeit nur für MySQL Datenbanken zur Verfügung wobei das Multi-AZ-Deployment auch mit Oracle Datenbanken funktioniert. Dass eine gute Backup-Strategie mehr als notwendig ist, zeigte ein Ereignis im Juli 2011 sehr anschaulich, als im Amazon Rechenzentrum die Verfügbarkeitszone EU-West-1a von Amazon einem größeren Ausfall betroffen war. Man sollte sich demnach nicht blind auf den Anbieter verlassen, sondern frühzeitig selbst zusätzliche Konzepte für Backup & Recovery entwickeln.

Zusammenfassung

Amazon bietet mit Oracle RDS eine sehr attraktive und skalierbare Datenbanklösung an. Insbesondere die Möglichkeit zu wachsen, schnell Ressourcen hinzuzubuchen oder bei fehlender Auslastung Kosten zu sparen, sind große Vorteile, die für ein Cloud basiertes Hosting sprechen. Damit am Ende des Tages jedoch auch die Verfügbarkeit und Datensicherheit gewährleistet wird, empfiehlt es sich regelmäßige Backups zu fahren und sich für alle Eventualitäten vorzubereiten. Diese werden im Vortrag konkret an der Live-Instanz vorgestellt. Jeder Amazon Administrator und Entwickler kann jedoch gespannt sein welche Features die nächsten sind. So kann bspw. seit Anfang des Jahres per Knopfdruck den Oracle Enterprise Manager aktivieren und seit August 2012 ist APEX im Paketumfang enthalten.

Kontaktadresse:

Frank Szilinski
esentri software GmbH
Pforzheimer Straße 132
76275 Ettlingen

Telefon: +49 (0) 7243-354 90 0
E-Mail frank.szilinski@esentri.com
Internet: www.esentri.com