# Hochverfügbarkeit für die Datenbank – was ist zu beachten

**Jochen Kutscheruk**
**merlin.zwo InfoDesign GmbH & Co. KG**
**Karlsruhe**

**Schlüsselworte**
Oracle Hochverfügbarkeit, Grid Infrastructure, Real Application Cluster, DataGuard, Standby, Image Copy

## Einleitung

Oracle DataGuard und Real Application Cluster - obwohl immer prominent hervorgehoben - können nur ein Baustein in einer wirklichen Datenbank-Hochverfügbarkeitslösung sein. Dieser Vortrag behandelt die Probleme und Stolperfallen bei der Konzeption, dem Aufbau und dem Betrieb einer Hochverfügbarkeitslösung. Folgende Punkte werden detailliert betrachtet:

- Welche Arten von Hochverfügbarkeit gibt es?
- Wie kann eine Datenbank für die Anwendung und die Anwender hochverfügbar gemacht werden?
- Welche Randbedingungen, die nicht unbedingt augenfällig sind, sind zu beachten?
- Ist Hochverfügbarkeit auch ohne DataGuard und ohne Real Application Cluster zu erreichen?
- Welche SLAs können ohne Risiko angeboten oder erfüllt werden?
- Worüber muss man sich womöglich gar keine Gedanken machen?

Es werden Strategien erläutert mit denen dafür gesorgt werden kann, dass der GAU nicht zum Desaster wird.

## Arten der Hochverfügbarkeit

Generell ist die Definition von Hochverfügbarkeit abhängig von den Anforderungen. In einem Fall kann eine Verfügbarkeit von 99,9% als Hochverfügbar gelten, während im anderen Fall 99,999% notwendig sind. Zusätzlich muss noch unterschieden werden zwischen „Verfügbarkeit der Daten“ (kein Datenverlust) und „Verfügbarkeit der Anwendung“ (keine Ausfallzeit).

Zwei Beispiele hierzu:
Die Tourismus-Webseite eines Landes muss 24x7x365 erreichbar sein, da diese Webseite weltweit rund um die Uhr aufgerufen wird. Allerdings wäre es kein großes Problem, wenn die dort abrufbaren Werbetexte eventuell vom Vortag stammen und die allerletzten Aktualisierungen nicht verfügbar sind. Ein (geringer) Datenverlust wäre akzeptabel.

Bei einer Hochregallagersteuerung mit chaotischer Lagerhaltung hingegen bedeutet bereits ein geringer Datenverlust von wenigen Minuten oder Sekunden, dass nicht mehr zuverlässig bestimmt werden kann, in welchem Regal welche Ware gelagert wurde und welche Ware gerade im Hochregallager wo unterwegs ist.
Dagegen wäre ein Ausfall der Anlage im Bereich von 15 Minuten bis zu einer Stunde zwar „unschön“, aber nicht wirklich tragisch. Falls nachts nicht gearbeitet wird wäre sogar ein mehrstündiger Ausfall unproblematisch.

**Definition der Hochverfügbarkeit**
„Ein System wird als verfügbar bezeichnet, wenn es in der Lage ist, die Aufgaben zu erfüllen, für die es vorgesehen ist. Als Verfügbarkeit wird die Wahrscheinlichkeit bezeichnet, dass ein System innerhalb eines spezifizierten Zeitraums funktionstüchtig (verfügbar) ist. Die Verfügbarkeit wird als Verhältnis aus fehlerbedingter Stillstandszeit (= Ausfallzeit) und Gesamtzeit eines Systems bemessen.“ (Wikipedia)

Die Verfügbarkeit eines Gesamtssystems wird in Verfügbarkeitsklassen eingeteilt.
**Verfügbarkeitsklasse 2**:
99% = 87,7 Stunden = 3 Tage 15:39:36 Stunden Ausfall pro Jahr
**Verfügbarkeitsklasse 3**:
99,9% = 8:45:58 Stunden Ausfall pro Jahr
**Verfügbarkeitsklasse 4**:
99,99% = 52:36 Minuten Ausfall pro Jahr
**Verfügbarkeitsklasse 5**:
99,999% = 5:16 Minuten Ausfall pro Jahr
**Verfügbarkeitsklasse 6**:
99,9999% = 31,6 Sekunden Ausfall pro Jahr

Diese Klassifizierung ist sicherlich die gebräuchlichste Einteilung. Alternativ kann auch die „Availability Environment Classification“ (AEC) der Harvard Research Group verwendet werden:

**AEC-0**: (Conventional) Funktion kann unterbrochen werden, Datenintegrität ist nicht essentiell
**AEC-1**: (Highly Reliable) Funtkion kann unterbrochen werden, Datenintegrität gewährleistet
**AEC-2**: (High Availability) Funktion darf nur innerhalb festgelegter Zeiten oder zur Hauptbetriebszeit minimal unterbrochen werden
**AEC3**: (Fault Resilient) Funktion muss innerhalb festgelegter Zeiten oder während der Hauptbetriebszeit ununterbrochen aufrecht erhalten werden
**AEC-4**: (Fault Tolerant) Funktion muss ununterbrochen aufrechterhalten werden, 24/7 Betrieb muss gewährleistet sein
**AES-5**: (Disaster Tolerant) Funktion muss unter allen Umständen verfügbar sein

Zweifellos wird sich jeder (insbesondere das Management) für die unternehmenskritische Anwendung die Verfügbarkeitsklasse 6 bzw. AES-5 wünschen. Ungünstigerweise steigt der notwendige finanzielle Aufwand für eine höhere Verfügbarkeitsklasse nicht linear sondern exponentiell. Dieser finanzielle Aufwand betrifft dabei nicht nur die einmalige Investition zum Aufbau einer Hochverfügbarkeitslösung, sondern auch den laufenden Betrieb (aufwendigere Aus- und Weiterbildung des für den Betrieb notwendigen Personals, mehr Ausgaben für Wartung, …).

Aus diesem Grund reduziert das Management sehr schnell wieder die Anforderungen. Stattdessen muss ein brauchbarer Kompromiss gefunden werden, welcher finanziell darstellbar ist und dennoch die notwendigen Anforderungen hinreichend erfüllt.

**Wie wird die Verfügbarkeit berechnet?**

Um gegenüber dem Management oder den Kunden Aussagen über die Verfügbarkeit eines Systems treffen oder bestimmte SLAs eingehen zu können, muss zunächst die mögliche Verfügbarkeit korrekt berechnet werden. Oftmals ist zu beobachten, dass die Verfügbarkeit eines Systems aus dem Durchschnitt der Verfügbarkeit der Einzelkomponenten berechnet wird. Dies ist jedoch grundlegend falsch.

Für die folgenden Beispiele wurde der Einfachheit halber für jede beteiligte Komponente eine Verfügbarkeit von 99,9% angenommen.

**Serienschaltung abhängiger Komponenten**

Bei der Serienschaltung sind voneinander abhängige Komponenten hintereinander geschaltet. Der Ausfall einer Komponente führt zum Ausfall des Gesamtsystems. Hier ein durchaus gängiger Aufbau:

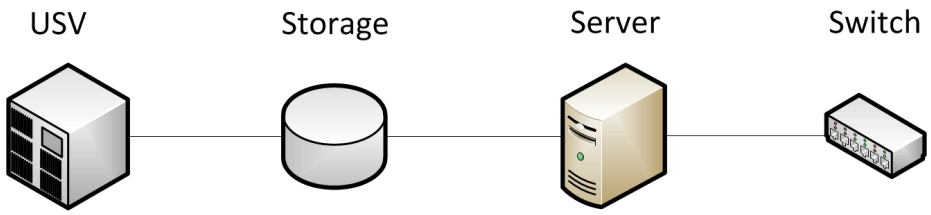



*Abbildung 1: Serienschaltung abhängiger Komponenten*

Die Gesamtverfügbarkeit errechnet sich hierbei aus dem Produkt der Verfügbarkeit der Einzelkomponenten:

99,9% USV x 99,9% Storage x 99,9% Server x 99,9% Switch = 99,6% Gesamt!

Dies bedeutet, dass das Gesamtsystem nicht nur 8,76 Stunden im Jahr ausfallen kann, sondern beachtliche 35,04 Stunden! Dies ist eine vollkommen andere Größenordnung.

**Parallelschaltung von Komponenten**

Umgangssprachlich ist damit „Redundanz“ gemeint. Das Gesamtsystem ist verfügbar, solange noch eine der parallel geschalteten Komponenten verfügbar ist. Ein typisches Beispiel hierfür ist ein Real Application Cluster von Oracle. Bereits in einer Konfiguration mit 2 Serverknoten erhält man eine deutlich höhere Verfügbarkeit. Die Gesamtverfügbarkeit errechnet sich hierbei aus dem Produkt der Ausfallzeiten (nicht Verfügbarkeit!) der Einzelkomponenten:

0,1% Server1 x 0,1% Server2 = 0,01% RAC.

Oder anders ausgedrückt: die Verfügbarkeit steigert sich drastisch von 99,9% (8,76 Stunden Ausfall pro Jahr) auf 99,99% (52,6 Minuten Ausfall pro Jahr)! Bei einer 4-Knoten Konstellation erreicht man bereits eine Verfügbarkeit von 99,9999%, d.h. lediglich 31,6 Sekunden Ausfall pro Jahr.

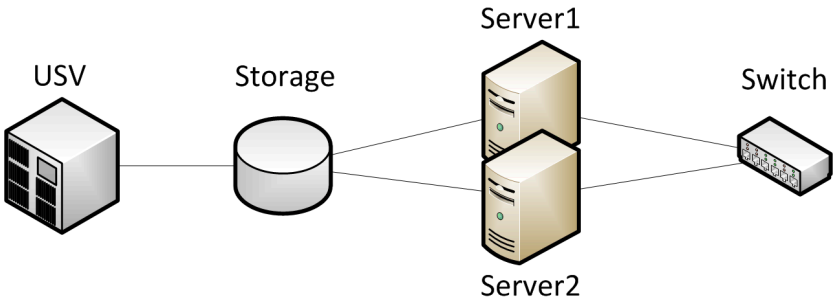



*Abbildung 2: Parallelschaltung von Komponenten*

Unglücklicherweise wurde in diesem Beispiel lediglich die Verfügbarkeit der Server berechnet. Die sonstigen seriellen Komponenten wurden nicht berücksichtigt. Bezieht man diese in die Berechnung mit ein, so erhält man folgende Verfügbarkeit:

99,9% USV x 99,9% Storage x 99,99% RAC x 99,9% Switch = 99,69% Gesamt!

Dies bedeutet: Obwohl eine einzelne Komponente des Gesamtsystems jetzt deutlich verfügbarer ist, ändert dies an der Gesamtverfügbarkeit nur relativ wenig.

**Serienschaltung abhängiger redundanter Komponenten**

Für die Praxis bedeutet dies, dass alle beteiligten Komponenten mindestens einfach redundant ausgelegt werden müssen, um eine deutliche Verbesserung der Verfügbarkeit zu erreichen:

99,99% red. USV x 99,99% red. Storage x 99,99% RAC x 99,99% Switch = 99,96% Gesamt!

In dieser Konstellation wird die Ausfallzeit auf  3,5 Stunden reduziert.

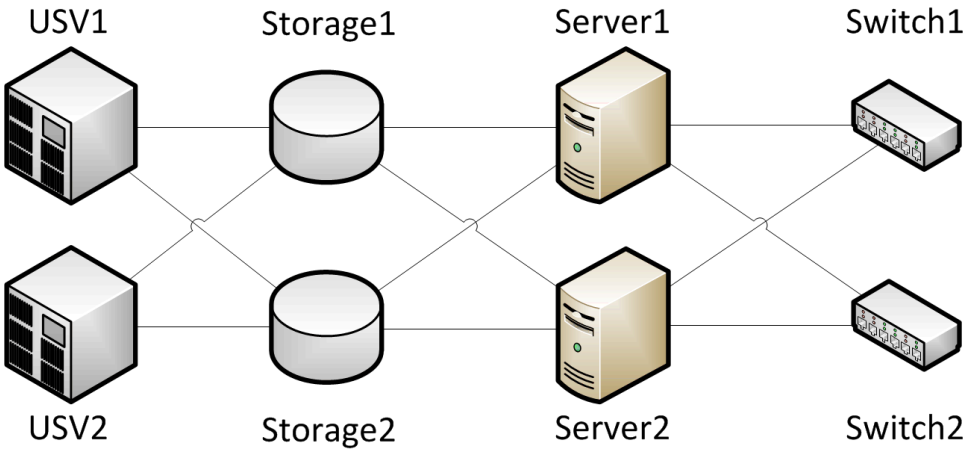



*Abbildung 3:  Serienschaltung abhängiger redundanter Komponenten*

Ziel dieser Schaltung ist es, den Single Point of Failure (SPOF) zu eliminieren. Aus langjähriger Praxis kann ich jedoch versichern, dass sich dieser SPOF niemals wirklich eliminieren, sondern immer nur verschieben lässt. Und schlimmer noch: je weiter man ihn verschiebt, umso teurer wird es.

Ein kleines Beispiel zum Nachdenken: Sie müssen natürlich auch die Verfügbarkeit des Serverraums in die Berechnung mit einbeziehen (z.B. Brand). Sie errichten also einen zweiten Serverraum, in welchem die redundanten Komponenten aufgebaut werden. Haben Sie dabei auch an eine redundante Internetleitung gedacht oder gibt es diesen Anschluss nur im ersten Serverraum? Wie lange benötigen Sie, um das alternative Routing - sowohl von Intern als auch von Extern – wieder korrekt am Laufen zu haben? Steht überhaupt das für die Umschaltung notwendige Fachpersonal zur Verfügung?

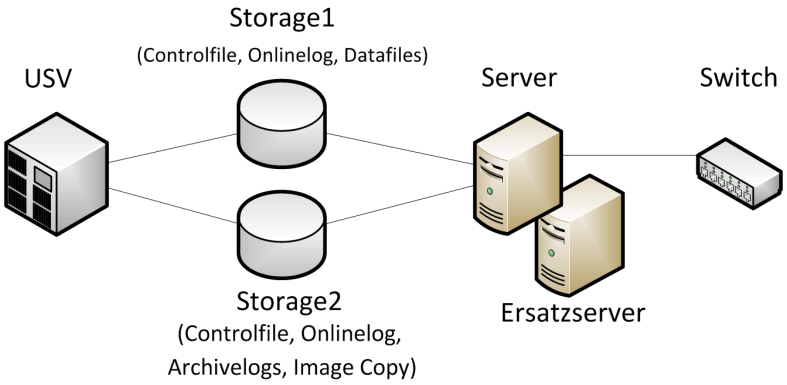
**Zusätzliche Ausfallzeiten durch gestiegene Komplexität**

Durch die Steigerung der Ausfallsicherheit steigt auch automatisch die Komplexität des Gesamtsystems. Im Extremfall kann das bedeuten: wo vorher ein Single Server mit lokaler Storage jahrelang problemlos seinen Dienst verrichtet hat (etwas Glück gehört auch dazu), treten nach der Umstellung plötzlich gehäuft Fehler auf. Diese Fehler resultieren aus der gestiegenen Komplexität des Systems (z.B. RAC-Ausfall aufgrund eines Firmware-Updates in der Storage).

**Funktionieren einfache Lösungen auch?**

Diese Frage ist nicht ganz einfach zu beantworten, denn es hängt wie immer von den konkreten Anforderungen an die Ausfallsicherheit ab. Eine kombinierte RAC/DataGuard-Lösung auf entsprechend ausfallsicherer Hardware im passenden Kontext wird sicherlich unschlagbar sein – sowohl was die Ausfallsicherheit und den Komfort, aber auch die Kosten betrifft.

Eventuell genügt jedoch auch eine simplere Lösung, die näherungsweise ähnliche Verfügbarkeiten gewährleistet. Beispielsweise könnten zwei unabhängige, günstige Storages an einen Server angeschlossen werden. Diese Storages werden nicht gespiegelt, sondern auf der ersten Storage liegen die Datenbankfiles mit einem Satz Onlinelogs und einer Kopie des Controlfiles.
Auf der zweiten Storage liegen ebenfalls eine Kopie des Controlfiles, der zweite Satz Onlinelogs und die Archivelogs. Dazu kommt noch eine Imagekopie der Datenbank, welche nachts während der Sicherung inkrementell über RMAN fortgeschrieben wird.




*Abbilung 4: Steigerung der Verfügbarkeit durch Image Copy*

Sollte nun z.B. die erste Storage ausfallen, so kann relativ einfach auf die zweite Storage geswitcht werden (RMAN switch database to copy). Kurz noch die letzten Archivelogs recovered, das fehlende Controlfile und den fehlenden Satz Redologs in der Konfiguration angepasst – schon kann es weiter gehen. Für einen erfahrenen Oracle DBA sollte das nur eine kleine Fingerübung sein. Aber: er muss verfügbar sein und dieses Verfahren fehlerfrei durchführen. Und: durch die zusätzliche Storage wird die Verfügbarkeit im ersten Schritt nicht erhöht, sondern gesenkt! Es handelt sich nicht um eine redundante Einheit, sondern eine weitere abhängige Komponente!

In dieser Richtung kann man sich noch dutzende andere Verfahren einfallen lassen, welche – immer bezogen auf die konkrete Anforderung – die geforderte Ausfallsicherheit gewährleisten können. Es muss also tatsächlich nicht immer RAC und DataGuard sein.

Allen diesen Lösungen ist jedoch gemein, dass manuelle Aktionen erforderlich sind und dafür das entsprechende Fachpersonal benötigt wird (Sie erinnern sich: der SPOF lässt sich nicht eliminieren, immer nur verschieben – in diesem Fall auf den DBA).

**Weitere bedenkenswerte Aspekte**
Ab welchem Zeitpunkt ist ein System eigentlich nicht mehr verfügbar? Es muss nicht immer ein Hardwareausfall sein. Ein System, welches zwar prinzipiell läuft, jedoch für die Beantwortung einer einfachen Bestandsabfrage statt 30ms plötzlich 2,5 Stunden benötigt, kann nicht als verfügbar bezeichnet werden.

Und Sie sollten auf keinen Fall den menschlichen Faktor vergessen. Die meisten Ausfälle werden inzwischen nicht mehr von defekter Hardware, sondern durch die darum herum befindlichen Menschen verursacht.
Konkret bedeutet dies, dass durch die am Gesamtsystem beteiligten Personen die Datenbank beschädigt wird – nicht physisch, sondern inhaltlich. Es wird versehentlich eine falsche Tabelle gedroppt und gepurged („…ich dachte ich bin auf dem Testsystem…"), es wird ein Update ohne where-Bedingung durchgeführt („…das ist mir ja noch nie passiert…"), …
Der Datenbank und der Hardware geht es also noch prima, lediglich der Inhalt ist nicht mehr brauchbar. Auch dies ist ein Ausfall des Gesamtsystems! Und genau für diesen Fall hatte leider niemand einen Notfallplan in der Schublade.

Sie sehen also: Das Thema Hochverfügbarkeit ist keineswegs trivial. Es gibt sehr viele Aspekte, die abgewogen und im jeweiligen Kontext bewertet werden müssen. Und: es empfiehlt sich immer, hierfür externe Berater hinzuzuziehen und deren Erfahrungsschatz zu nutzen. Es geschehen in der Realität Dinge, welche man mit gesundem Menschenverstand niemals als Problem in Erwägung ziehen würde.

**Kontaktadresse:**
Jochen Kutscheruk
merlin.zwo InfoDesign GmbH & Co. KG
Taglöhnergärten 43
D-76228 Karlsruhe

| | |
|---|---|
| Telefon: | +49 (0) 7052 50898-0 |
| Fax: | +49 (0) 7052 50898-50 |
| E-Mail | jochen.kutscheruk@merlin-zwo.de |
| Internet: | www.merlin-zwo.de |