

Oracle DBA & Developer Days 2011

日本オラクル、今年最大の技術トレーニングイベント

2011年11月9日(水)～11月11日(金) シェラトン都ホテル東京



ORACLE®

サポートエンジニアが語る！ RAC 環境のトラブルシューティング

日本オラクル株式会社 カスタマーサポートサービス統括
吉村 隆

以下の事項は、弊社の一般的な製品の方向性に関する概要を説明するものです。また、情報提供を唯一の目的とするものであり、いかなる契約にも組み込むことはできません。以下の事項は、マテリアルやコード、機能を提供することをコミットメント(確約)するものではないため、購買決定を行う際の判断材料になさらないで下さい。オラクル製品に関して記載されている機能の開発、リリースおよび時期については、弊社の裁量により決定されます。

OracleとJavaは、Oracle Corporation 及びその子会社、関連会社の米国及びその他の国における登録商標です。文中の社名、商品名等は各社の商標または登録商標である場合があります。

本セミナーの目的

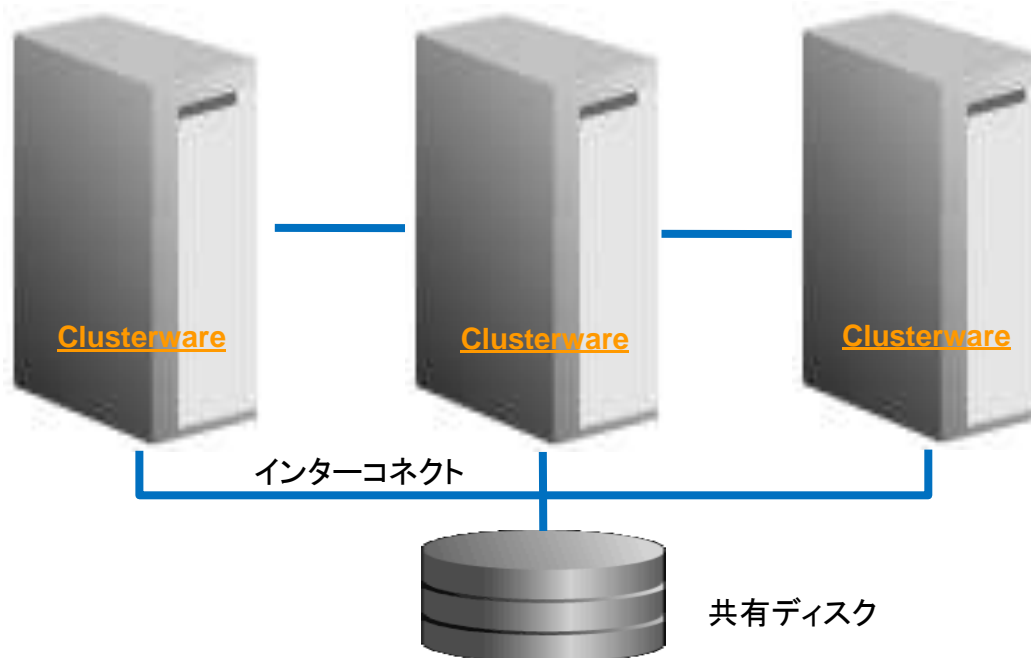
- Oracle Real Application Clusters 構成の理解
 - クラスタ・システムの理解
 - Grid Infrastructure
 - Cluster Database
- RAC 環境におけるトラブルシューティング
 - ノード・メンバーシップに問題が発生したケース
 - リソース・メンバーシップに問題が発生したケース
 - RAC Database のパフォーマンス



Oracle Real Application Clusters 構成の理解

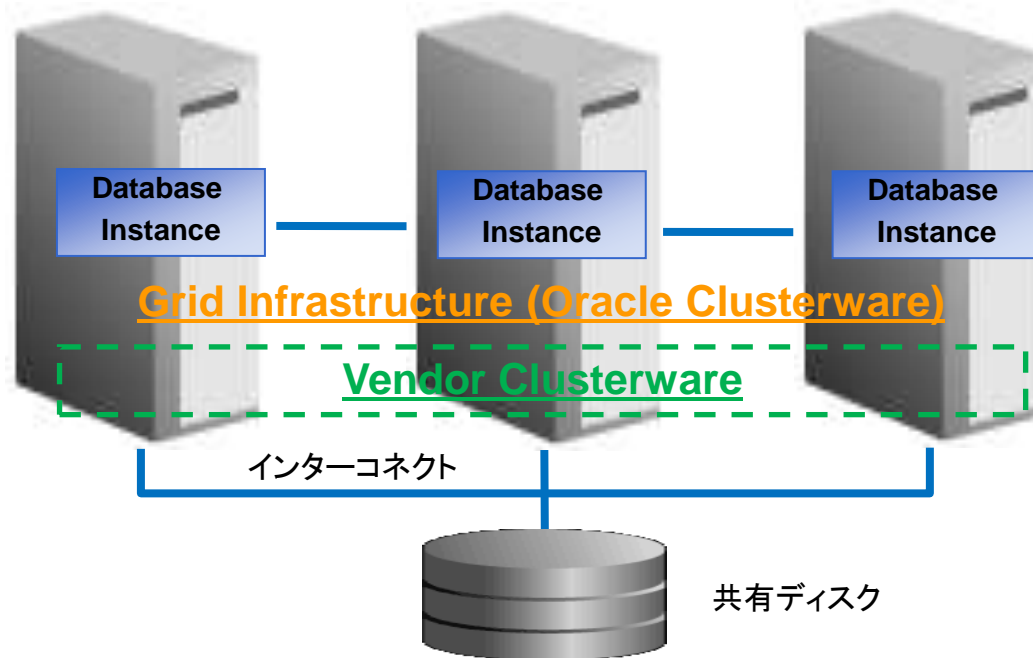
クラスタ・システムの理解

- 相互接続され、1つのサーバーのように動作するノード群
 - インターコネクトを介したメンバーシップの管理が必要
 - 全てのサーバーから、共有ディスク上のデータにアクセス可能



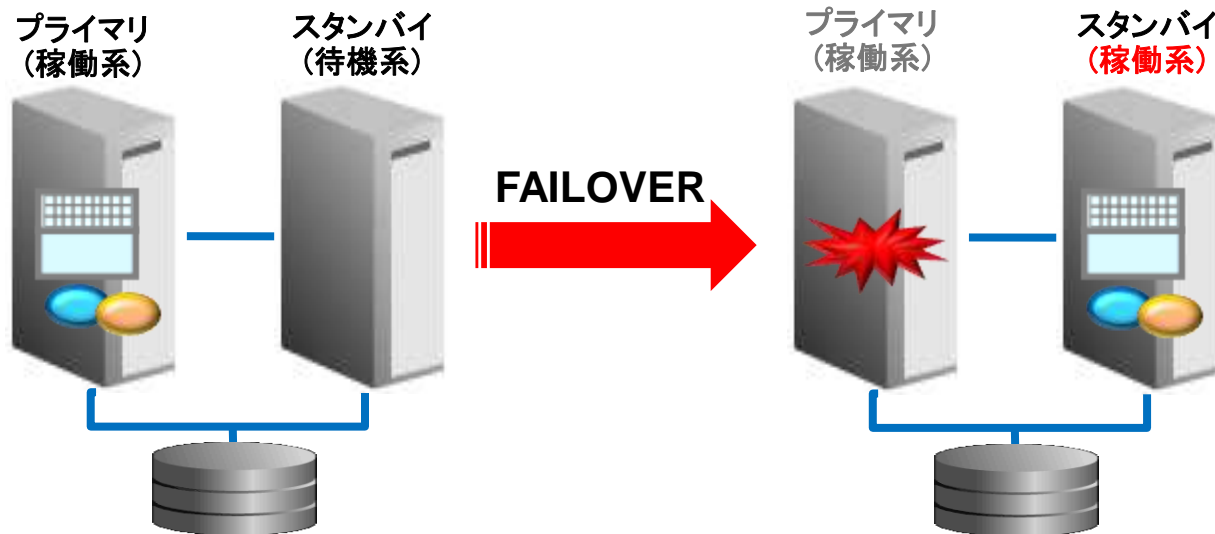
Oracle Real Application Clusters とは

- Oracle 製のクラスタウェアを提供
 - システムがクラスタ化されている必要がある
 - ノードやストレージの管理機能を提供 (Grid Infrastructure)
- クラスタ・データベースの機能を提供
 - Database Instance を分散配置
 - アプリケーションからはクラスタを意識する必要なし



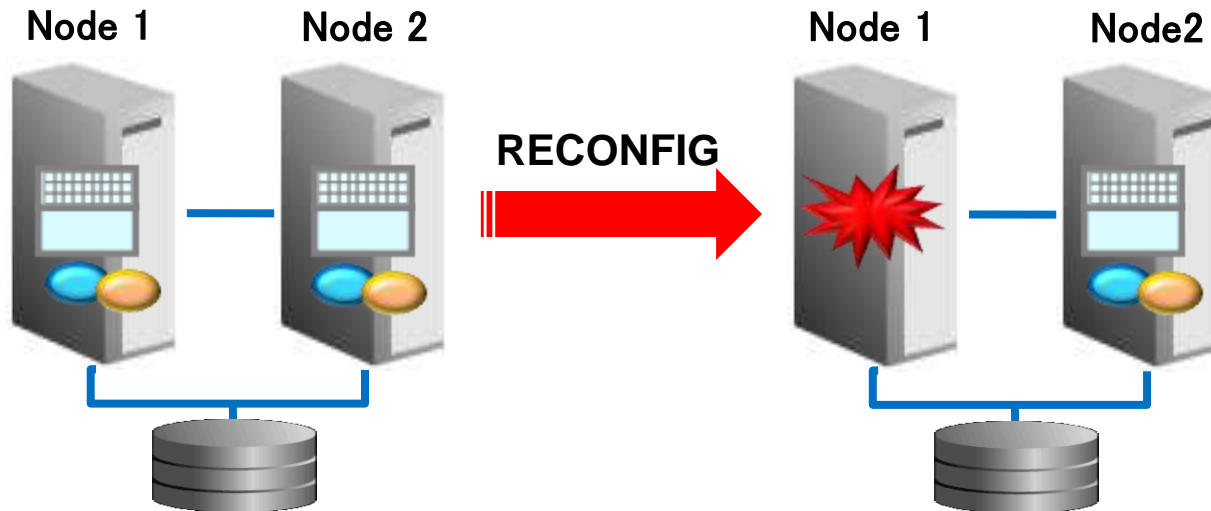
High Availability 構成と RAC

- HA (High Availability) 構成は、スタンバイ・サーバを準備し、障害発生時にサーバを切り替える構成
 - 障害発生時にフェイルオーバーを行う



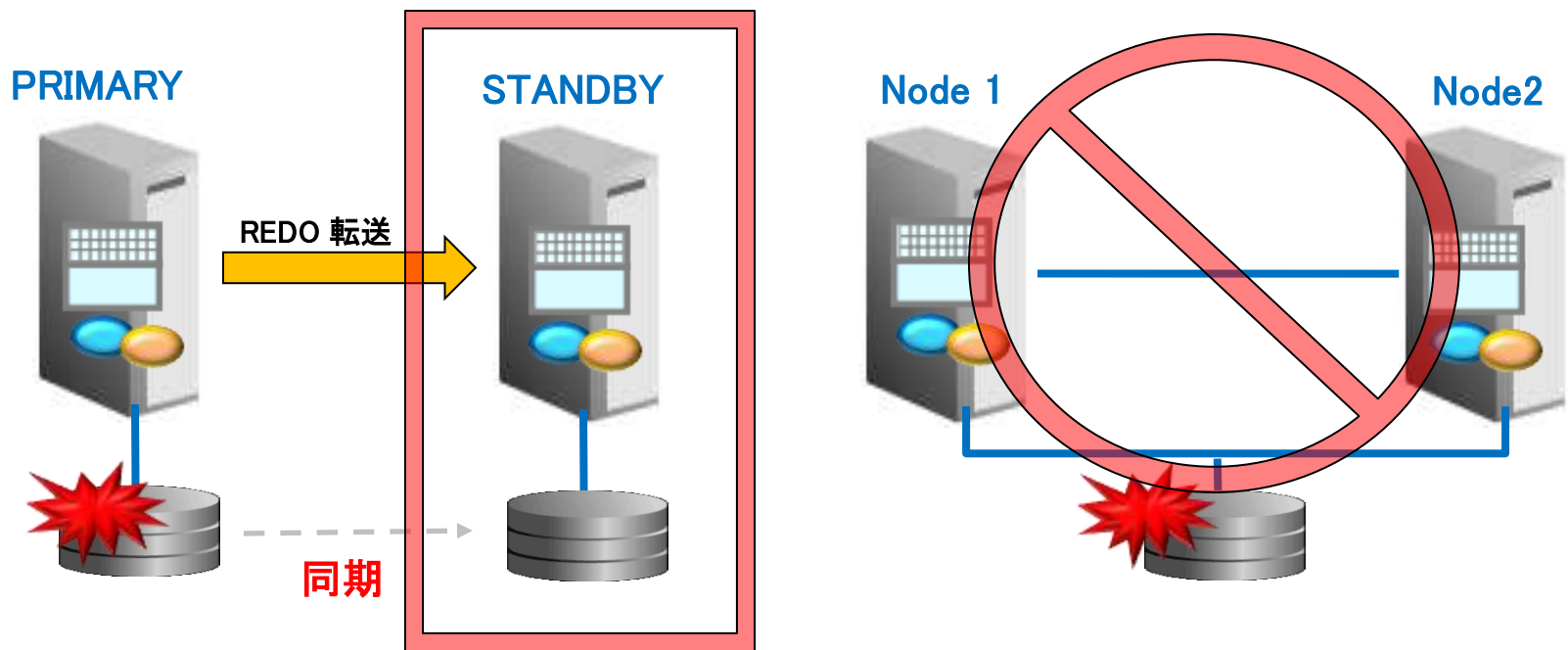
High Availability 構成と RAC

- RAC 構成では、全てのノードが並列的に構成されているため、全てのサーバでサービスの提供が可能
 - サーバ・リソースの有効活用
 - ノード障害が発生した場合でも、短時間で再構成が可能



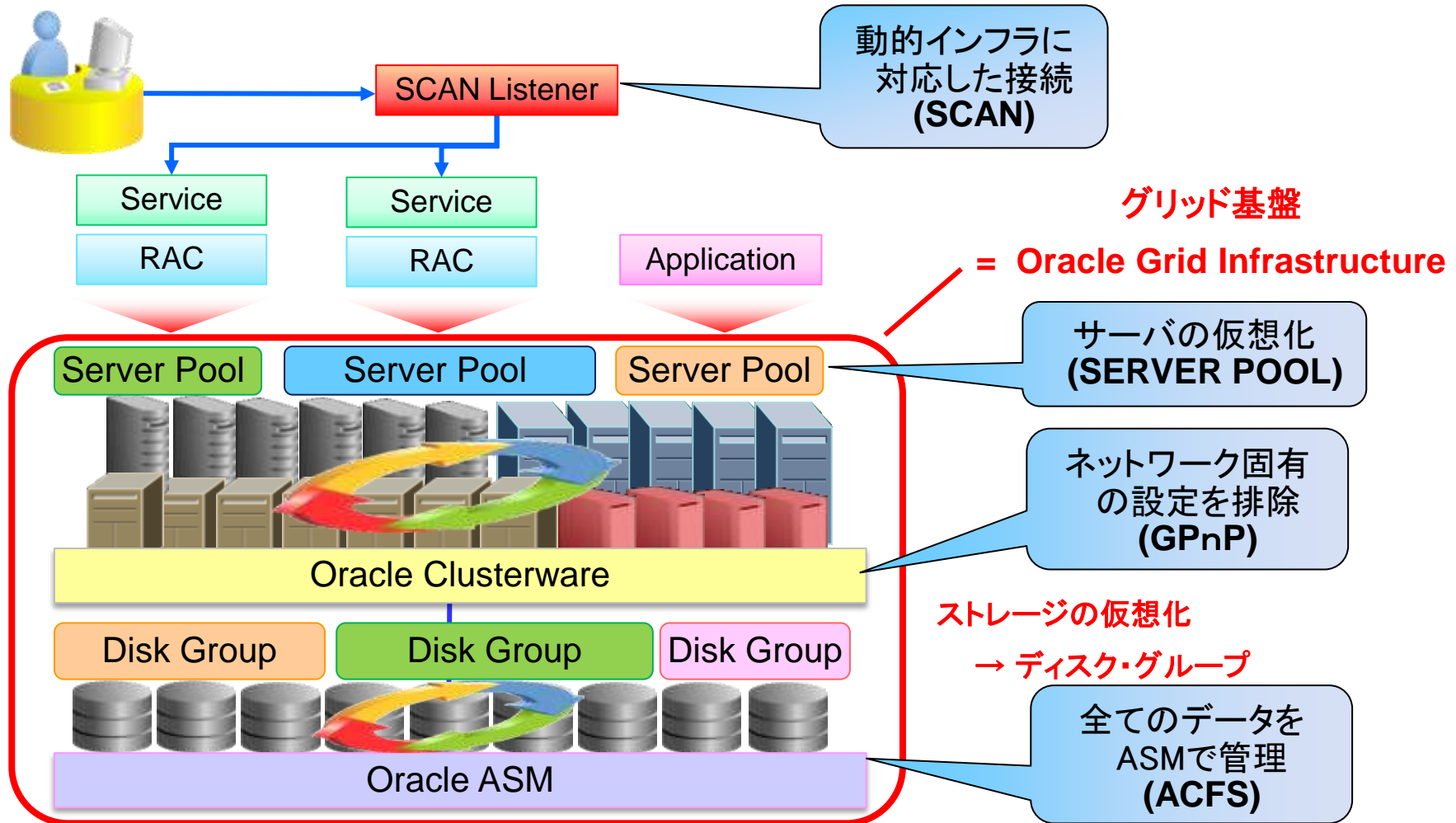
Oracle Data Guard と RAC

特徴	Data Guard	RAC
構成	2つ以上のDatabaseで構成	1つのDB環境を複数のインスタンスで共有
目的	データ保護、障害時ダウンタイムの減少	処理の分散化、ノード・インスタンスの可用性向上
アクセス	基本的にプライマリ・サイトを使用	全てのインスタンスにアクセス可能



Grid Infrastructure の全体像

- リソースの物理的な存在を意識せずにサービスとして利用可能

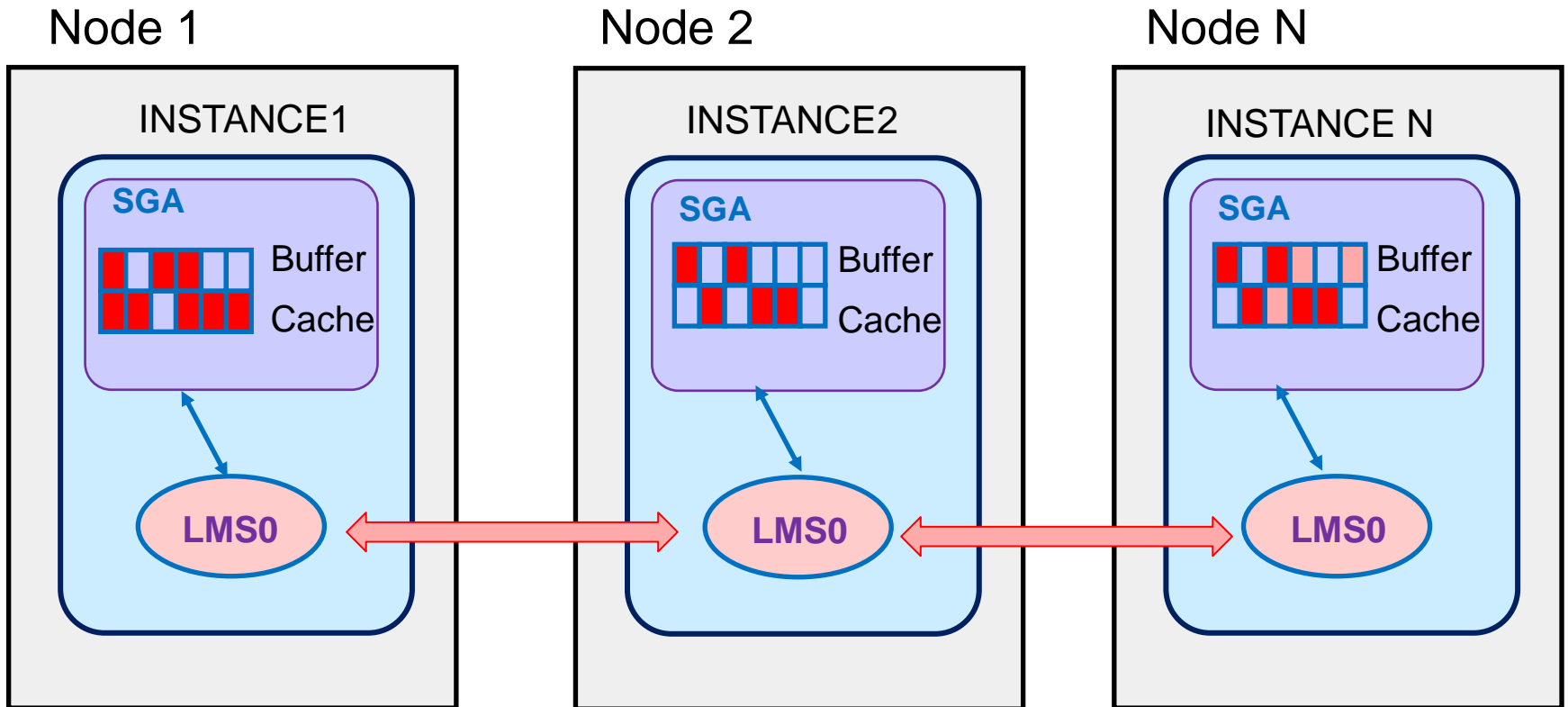


Grid Infrastructureの主要機能

- Single Client Access Name (SCAN)
 - 動的インフラに対応した接続形式
- Server Pool
 - サーバ・リソースの構成を最適化するインフラ
- Grid Plug and Play (GPnP)
 - ネットワーク固有の設定を排除したドメインの構築
- Oracle ASM Cluster File System (ACFS)
 - どのサーバからもアクセス可能なファイルシステム

Cluster Database とは

- 複数インスタンスからの同時アクセスを制御する機能を提供
 - 複数のインスタンスが同じブロックに対するアクセスを要求した場合でも、矛盾や更新の欠落が起きないようにする機能

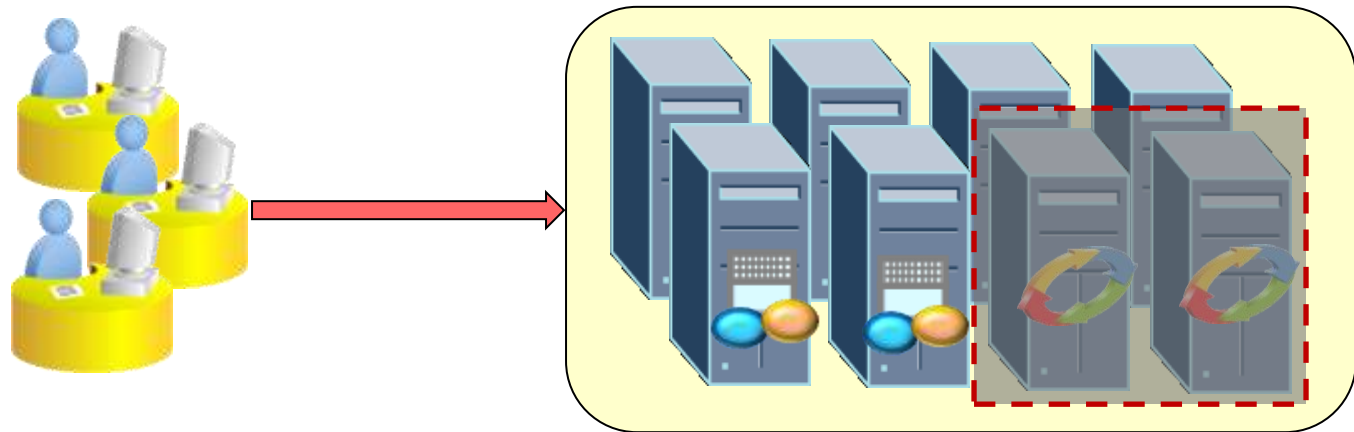


RAC 環境における トラブルシューティング

- ノード・メンバーシップに
問題が発生したケース

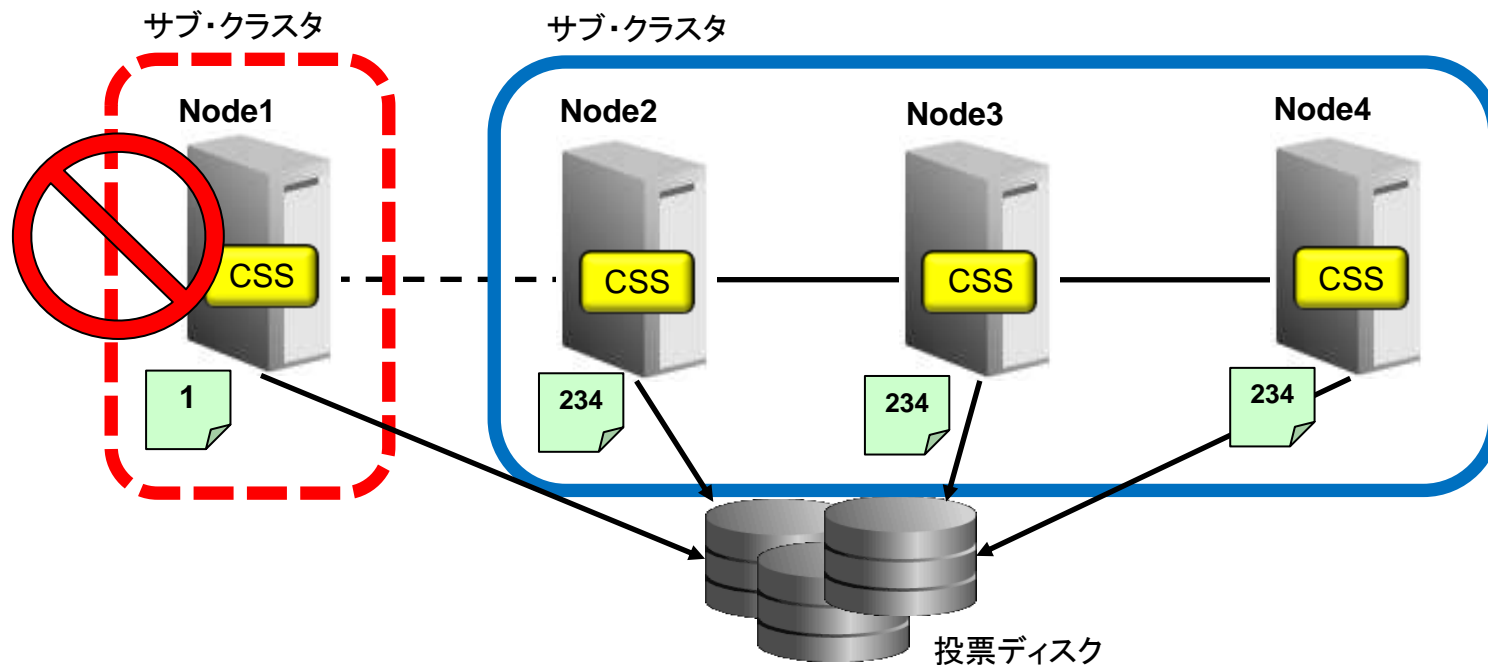
ノード・メンバーシップに問題が発生したケース

- ノード・メンバーシップとは、クラスタを構成する各ノードの生存情報
 - クラスタ環境では、**すべての**ノードのメンバーシップを管理する必要がある
- ノード・メンバーシップに問題が発生すると、ノードはクラスタから離脱
 - 従来のリリースでは、**ノード再起動**が発生
 - 再起動が発生したノード上で稼働している、**すべてのアプリケーションが停止**
- 11.2.0.2 のリリースより、**Reboot-less Node Fencing** の導入
 - インターコネクト障害、投票ディスク障害発生時、全てのプロセスの停止が可能であれば、Oracle Clusterware の停止 (ノードは起動したまま)
 - Oracle Clusterware の管理下でないアプリケーションは継続して稼働
 - 上記以外の障害、もしくは上記障害でプロセスの停止が行えない場合、ノード再起動が発生



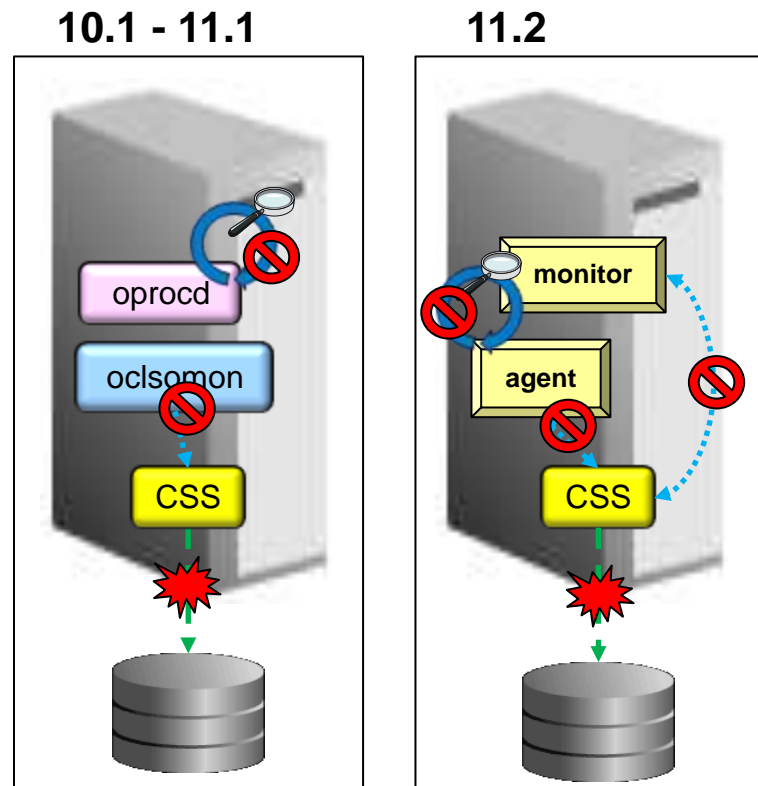
ノード・メンバーシップに問題が発生したケース

- インターコネクト障害 (Split Brain) が発生すると、通信できるノードのみクラスタに残存させる
 - 投票ディスクを使用した投票
 - 最大サブ・クラスタのルール



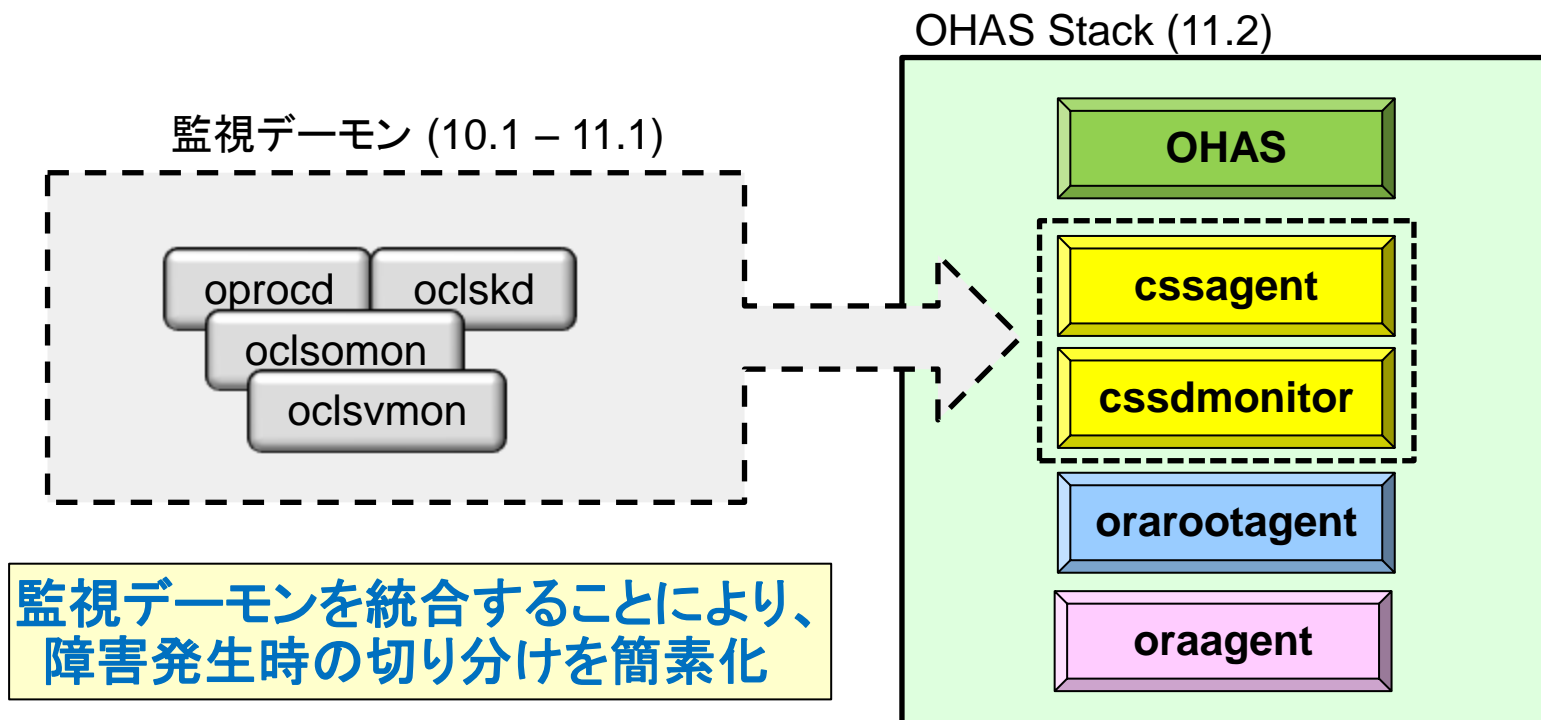
ノード・メンバーシップに問題が発生したケース

- ノード・メンバーシップが確認できなくなるケースは、ノード間通信の遮断 (Split Brain) のみではない
 - OS / Kernel レベルでのハング
 - oprocd** による監視 (10.1 – 11.1)
 - cssdmonitor / cssdagent** による監視 (11.2)
 - メンバーシップ管理機能の (CSS デーモン) のハング
 - oclsomon** による監視 (10.1 – 11.1)
 - cssdmonitor / cssdagent** による監視 (11.2)
 - (Split Brain 発生時に) ディスクパスの遮断により、投票が行えないケース
 - CSS デーモン**による監視

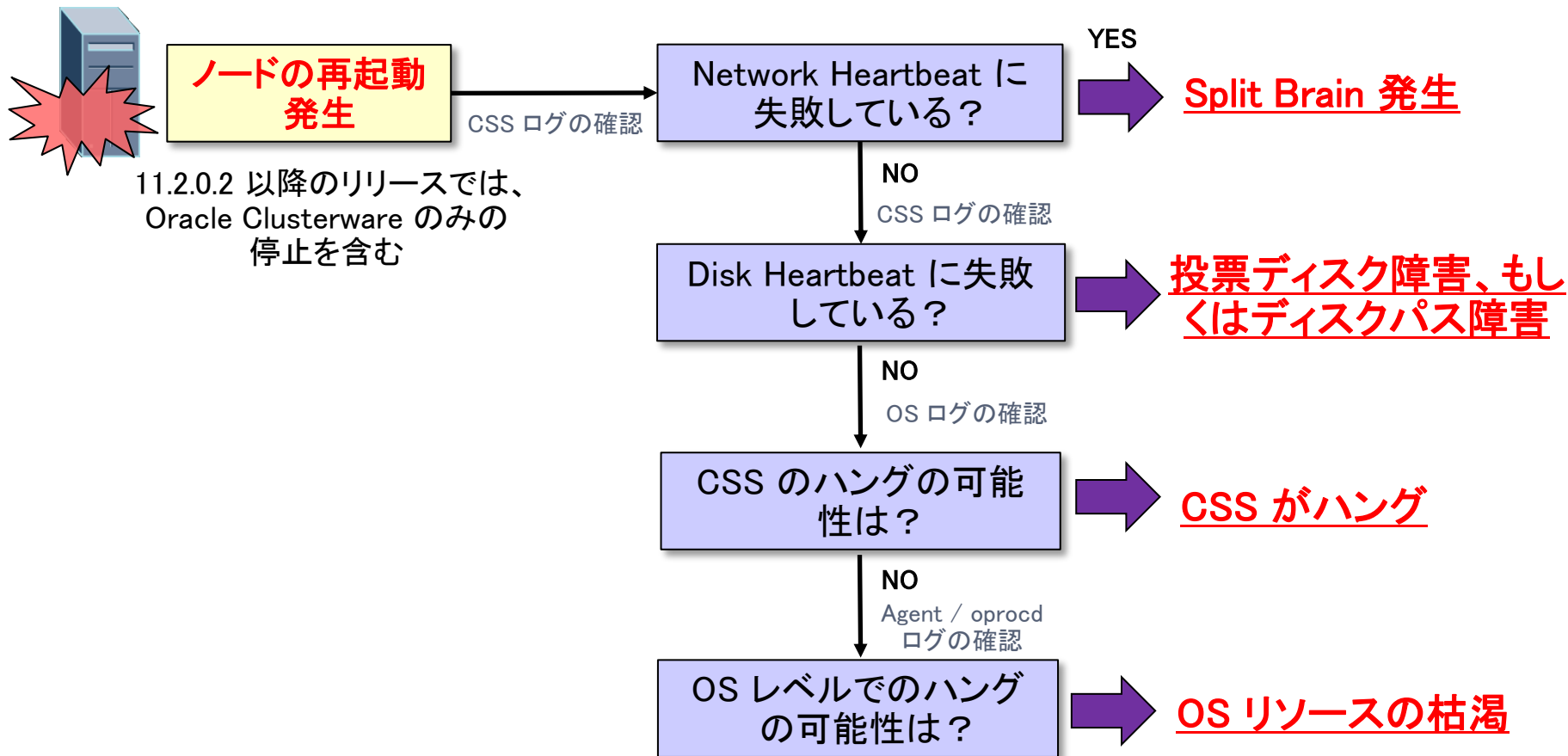


ノード・メンバーシップに問題が発生したケース

- 11.2 のリリースより、監視デーモンの統合が行われている
 - OHAS (Oracle High Availability Services) スタックの導入



ノード・メンバーシップに問題が発生したケース



ノード・メンバーシップに問題が発生したケース

- ノード再起動が発生した場合には、原因となった機能を管理しているデーモン、エージェントのログを確認する必要がある
 - Split Brain の発生 (インターコネクトの遮断)
 - **CSS デーモン**による監視
 - \$ORA_CRS_HOME/log/<hostname>/cssd/ocssd.log
 - \$ORA_CRS_HOME/css/log/ocssd.log (10.1 のリリースのみ)

KEY WORD

- heartbeat fatal
- Eviction started
- Aborting local node to avoid splitbrain

Eviction は、排除する側、される側で開始されている可能性がある。
結果として排除された場合、Aborting local node の出力。

ノード・メンバーシップに問題が発生したケース

- ノード再起動が発生した場合には、原因となった機能を管理しているデーモン、エージェントのログを確認する必要がある
 - ディスクパスの遮断により、投票が行えないケース
 - **CSS デーモン**による監視
 - \$ORA_CRS_HOME/log/<hostname>/cssd/ocssd.log
 - \$ORA_CRS_HOME/css/log/ocssd.log (10.1 のリリースのみ)

KEY WORD

- voting device hang
- stale disk
- Aborting, <Y> of <X> voting disks unavailable

各ノードの CSS は、クラスタに構成されている過半数の投票ディスクにアクセスできなければならない。

ノード・メンバーシップに問題が発生したケース

- ノード再起動が発生した場合には、原因となった機能を管理しているデーモン、エージェントのログを確認する必要がある
 - メンバーシップ管理機能の (CSS デーモン) のハング
 - **oclsomon** による監視 (10.1 – 11.1)
 - \$ORA_CRS_HOME/log/<hostname>/cssd/oclsomon/oclsomon.*
 - **cssdmonitor / cssdagent** による監視 (11.2)
 - \$ORA_CRS_HOME/log/<hostname>/agent/ohasd/oracssdagent_root/*
 - \$ORA_CRS_HOME/log/<hostname>/agent/ohasd/oracssdmonitor_root/*

KEY WORD

- clssomon: Timeout waiting for CSS response
- Oracle CRS failure. Rebooting for cluster integrity
- Oracle clsomon failed with fatal
- Impending reboot at 90% of limit

1番目のキーワードは、oclsomon.* に出力される。2,3番目は OS のログ。4番目は Agent のログ。

ノード・メンバーシップに問題が発生したケース

- ノード再起動が発生した場合には、原因となった機能を管理しているデーモン、エージェントのログを確認する必要がある
 - OS / Kernel レベルでのハング
 - **oproc** による監視 (10.1 – 11.1)
 - /etc/oracle/oprocd/\$HOST.oprocd.*
 - /var/opt/oracle/oprocd/\$HOST.oprocd.*
 - messages / syslog
 - **cssdmonitor / cssdagent** による監視 (11.2)
 - \$ORA_CRS_HOME/log/<hostname>/agent/ohasd/oracssdagent_root/*
 - \$ORA_CRS_HOME/log/<hostname>/agent/ohasd/oracssdmonitor_root/*

KEY WORD

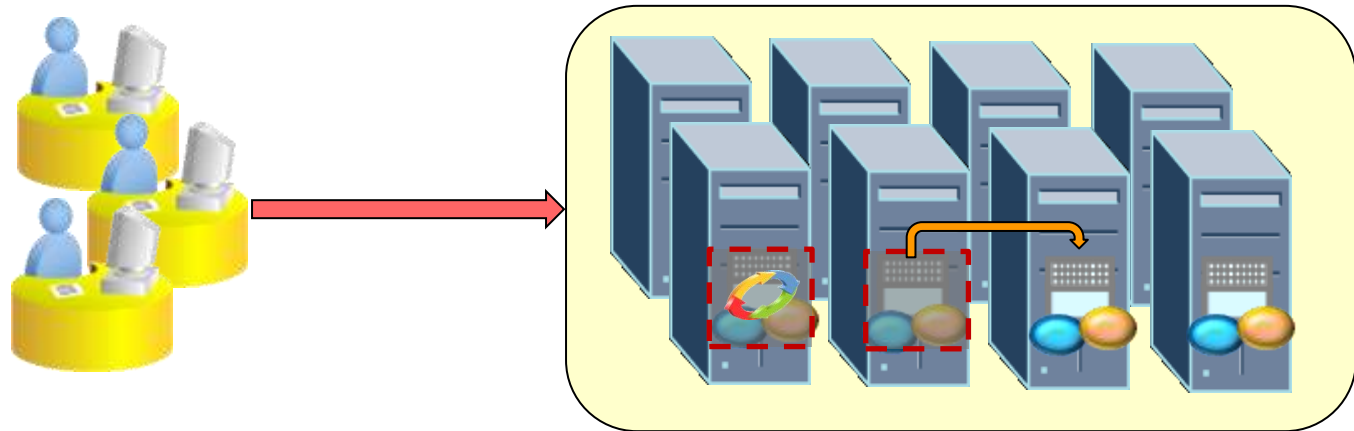
- timeout(XXXXX msec) exceeds ... Rebooting NOW.
- Rebooting after limit XXXXX exceeded

1番目のキーワードは、oprocd.* に出力される。
2番目は Agent のログ。

- リソース・メンバーシップに
問題が発生したケース

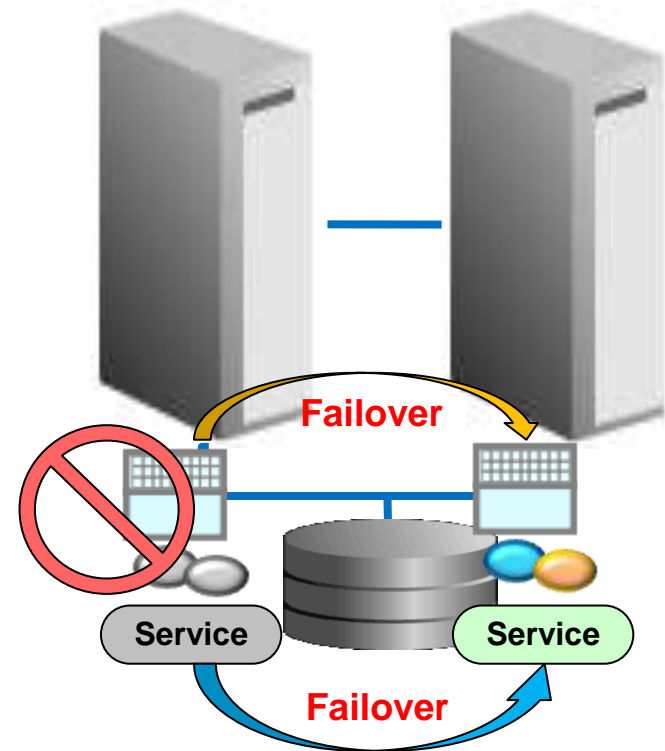
リソース・メンバーシップに問題が発生したケース

- リソース・メンバーシップとは、Oracle Clusterware で管理されているアプリケーションの生存情報
 - CRS リソース : Oracle で提供しているアプリケーションを管理するリソース
 - カスタム・アプリケーション・リソース : ユーザーが任意で登録したアプリケーションを管理するリソース
- リソース・メンバーシップに問題が発生すると、リソースの再起動や、フェイルオーバーが行われる
 - 同一ノードでの再起動を試行
 - 再配置可能ノードへのフェイルオーバーを試行



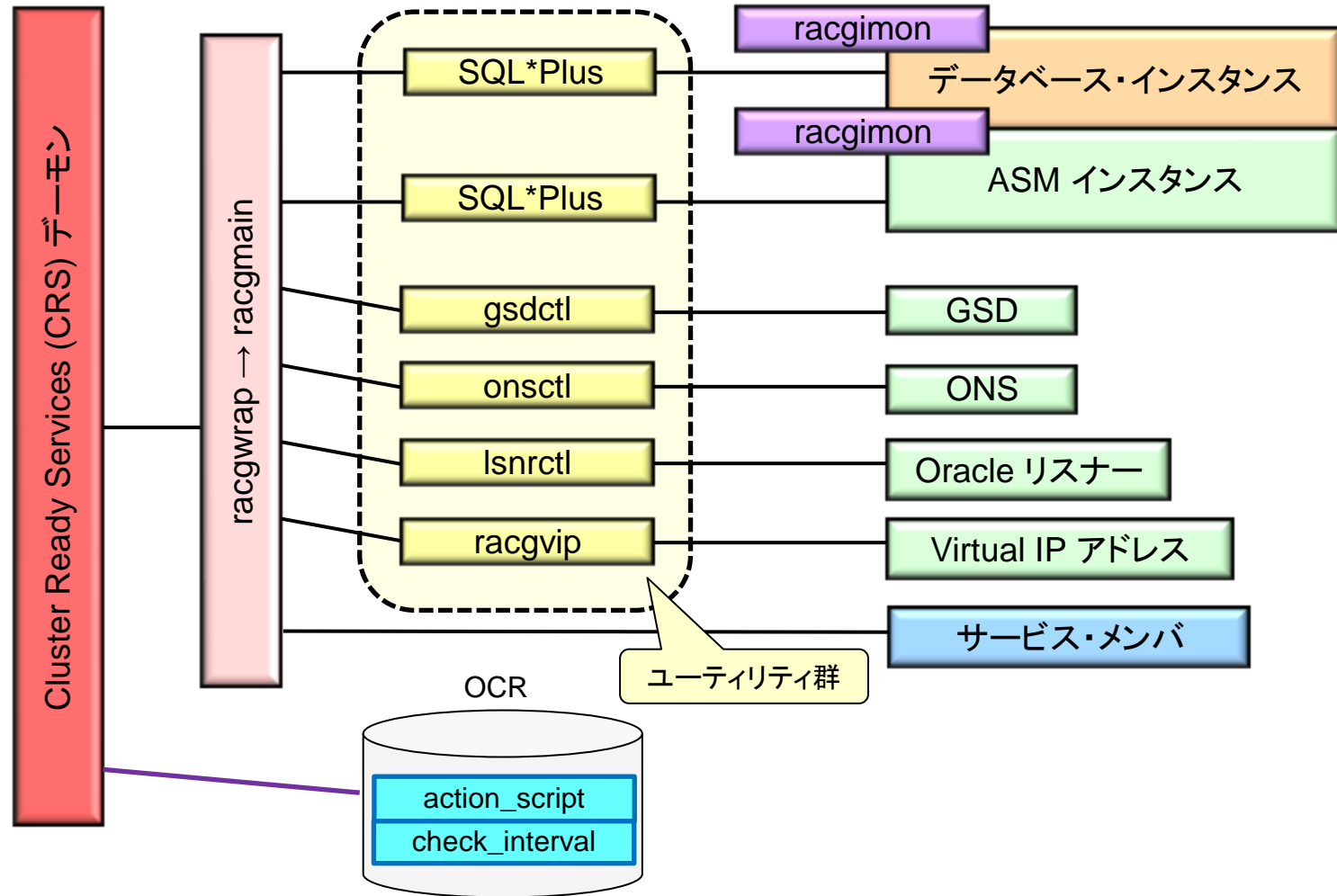
リソース・メンバーシップに問題が発生したケース

- フェイル・オーバー可能なリソース
 - 10.1 - 11.1
 - Virtual IP
 - サービス
 - 11.2
 - Virtual IP
 - Scan VIP, SCAN Listener
 - 管理者管理データベース
 - サービス
 - ポリシー管理データベース
 - サービス (SINGLETON)
 - RAC Database Instance



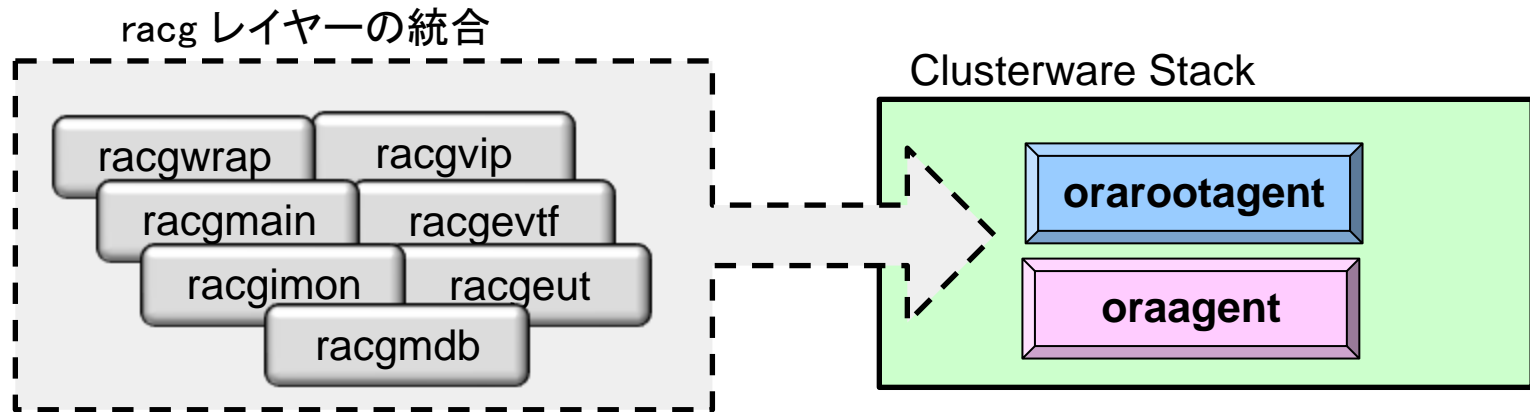
リソース・メンバーシップに問題が発生したケース

- リソースの可用性管理 (10.1 – 11.1)



リソース・メンバーシップに問題が発生したケース

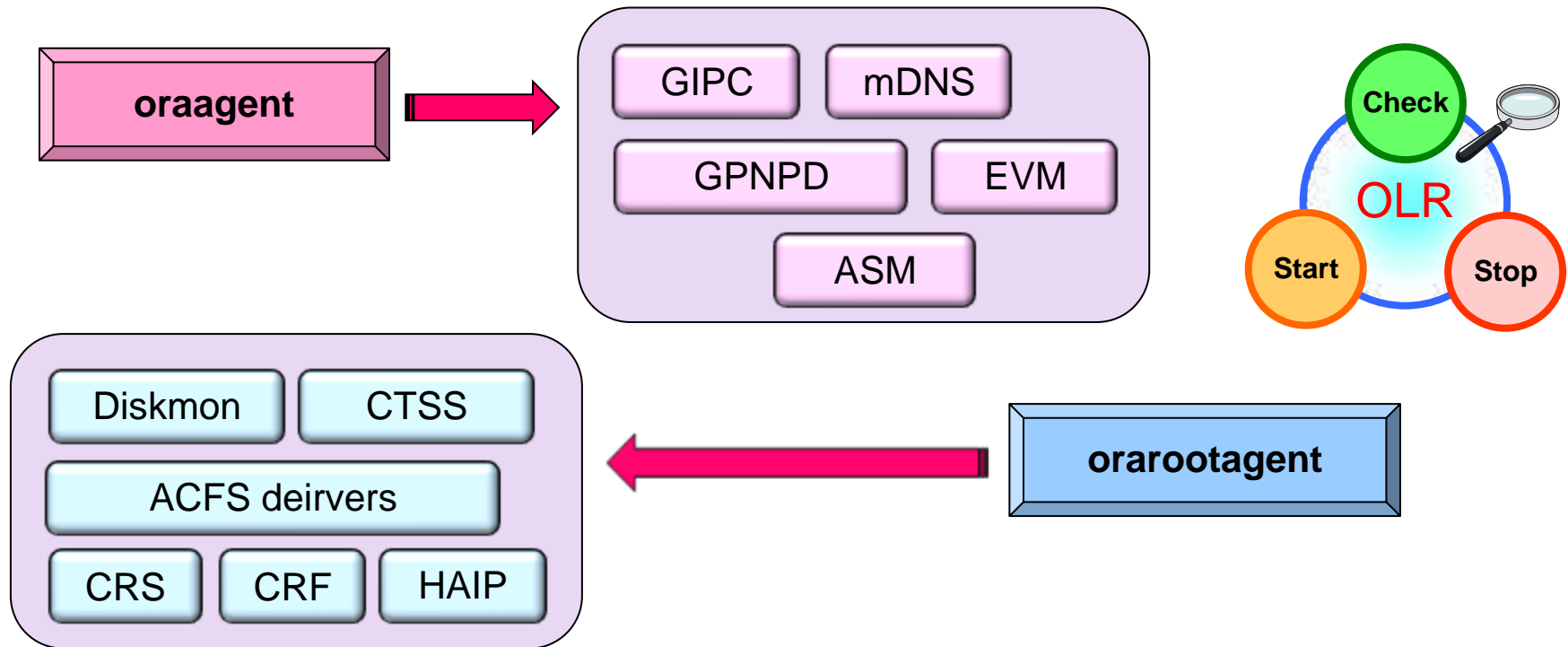
- 11.2 のリリースよりデーモンの統合が行われている



リソースごとに異なっていたチェック方法を統合することにより、障害発生時の切り分けが簡素化

リソース・メンバーシップに問題が発生したケース

- 11.2 のリリースでは、一部のクラスタウェア・デーモンについても、デーモン・リソースとして管理
- CRS リソースと同様の管理フレームワークを適用



リソース・メンバーシップに問題が発生したケース

- リソースの再起動やフェイルオーバーが発生した場合には、リソースを管理しているデーモン、エージェントのログを確認する必要がある
 - 全てのリソースの管理の大元は、CRS デーモン
 - \$ORA_CRS_HOME/log/<hostname>/crsd/crsd.log
 - \$ORA_CRS_HOME/crs/log/crsd.log (10.1 のリリースのみ)

KEY WORD

- OFFLINE unexpectedly
- Attempting to stop `ora.NODE2.vip` on member `NODE2`
- Attempting to start `ora.NODE2.vip` on member `NODE1`

障害が発生したリソースを、別のノードで起動している。

これがフェイルオーバー

リソース・メンバーシップに問題が発生したケース

- リソースの再起動やフェイルオーバーが発生した場合には、リソースを管理しているデーモン、エージェントのログを確認する必要がある
 - エージェント
 - \$ORA_CRS_HOME/log/<hostname>/agent/crsd/oraagent_<DB_User>
 - \$ORA_CRS_HOME/log/<hostname>/agent/crsd/oraagent_<GI_User>
 - \$ORA_CRS_HOME/log/<hostname>/agent/crsd/orarootagent_root
 - \$ORA_CRS_HOME/log/<hostname>/agent/crsd/scriptagent_<GI_User>

KEY WORD

- state changed from: ONLINE to: FAILED
- state changed from: ONLINE to: OFFLINE
- unplanned state change

管理者が停止コマンドを実行した時間帯でなければ、障害が発生して停止した可能性がある。

リソース・メンバーシップに問題が発生したケース

- 11.1 までのリリースでは、リソースの再起動やフェイルオーバーが発生した場合には、racg レイヤのデーモンのログの確認も必要
 - racg (11.1 まで)
 - \$ORACLE_HOME/log/<hostname>/racg/*.log
 - \$ORA_CRS_HOME/log/<hostname>/racg/*.log

KEY WORD

- Invalid parameters, or failed to bring up VIP
- IP address is not specified
- Invalid option
- IP xxx.xxx.xxx.xxx is already up in the network

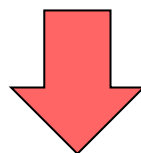
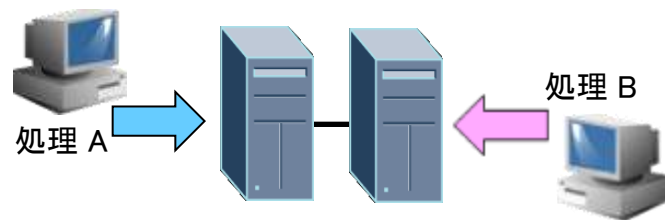
ここに挙げられているキーワードは、VIP に関するメッセージ。VIP の挙動に関しては、racg のログを確認するケースが多い。



- RAC Database の パフォーマンス

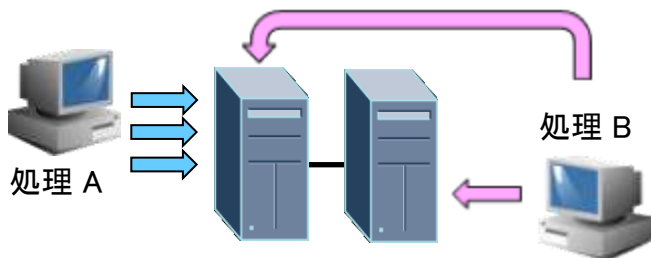
RAC Database のパフォーマンス

- RAC Database のパフォーマンスを上げるためには、アプリケーション・パーティショニングが必要
 - オブジェクトごとに処理ノードを限定
 - ノードごとに、参照処理と更新処理を分ける



GI 環境では対応できない
場合もある

- 11g からは RAC Database で、自動的に最適化処理が行われる
 - オブジェクトごとに、処理の多いノードを自動識別
 - 参照処理が多い環境では、参照ノードの分散配置に対応した最適化



RAC Database のパフォーマンス

```
SQL> select a.data_object_id, a.gc_mastering_policy, a.current_master,
2      b.owner, b.object_name, b.object_type
3      from v$gcspfmaster_info a, dba_objects b
4      where a.data_object_id = b.object_id(+)
5      order by data_object_id;
```

DATA_OBJECT_ID	GC_MASTERING_POLICY	CURRENT_MASTER	OWNER	OBJECT_NAME	OBJECT_TYPE
95903	Affinity	0	SCOTT	DEPT	TABLE
95905	Affinity	0	SCOTT	EMP	TABLE
95907	Read mostly	1	SCOTT	BONUS	TABLE
95908	Read mostly	1	SCOTT	SALGRADE	TABLE

Instance1 (Current_Master + 1)
において、処理が高速になるよう
に最適化済み

全てのノードにおいて、参照
処理のみが高速になるよう
に最適化済み

※ ここに表示されないオブジェクトについては、特別な最適化は行われていない

まとめ

- Oracle Real Application Clusters 構成の理解
 - クラスタ・システムの理解
 - Grid Infrastructure
 - Cluster Database
- RAC 環境におけるトラブルシューティング
 - ノード・メンバーシップに問題が発生したケース
 - リソース・メンバーシップに問題が発生したケース
 - RAC Database のパフォーマンス

補足：技術文章のご案内

サポート契約を締結されているお客様は、以下の技術文書 (KROWN) も確認いただけます。

文章例

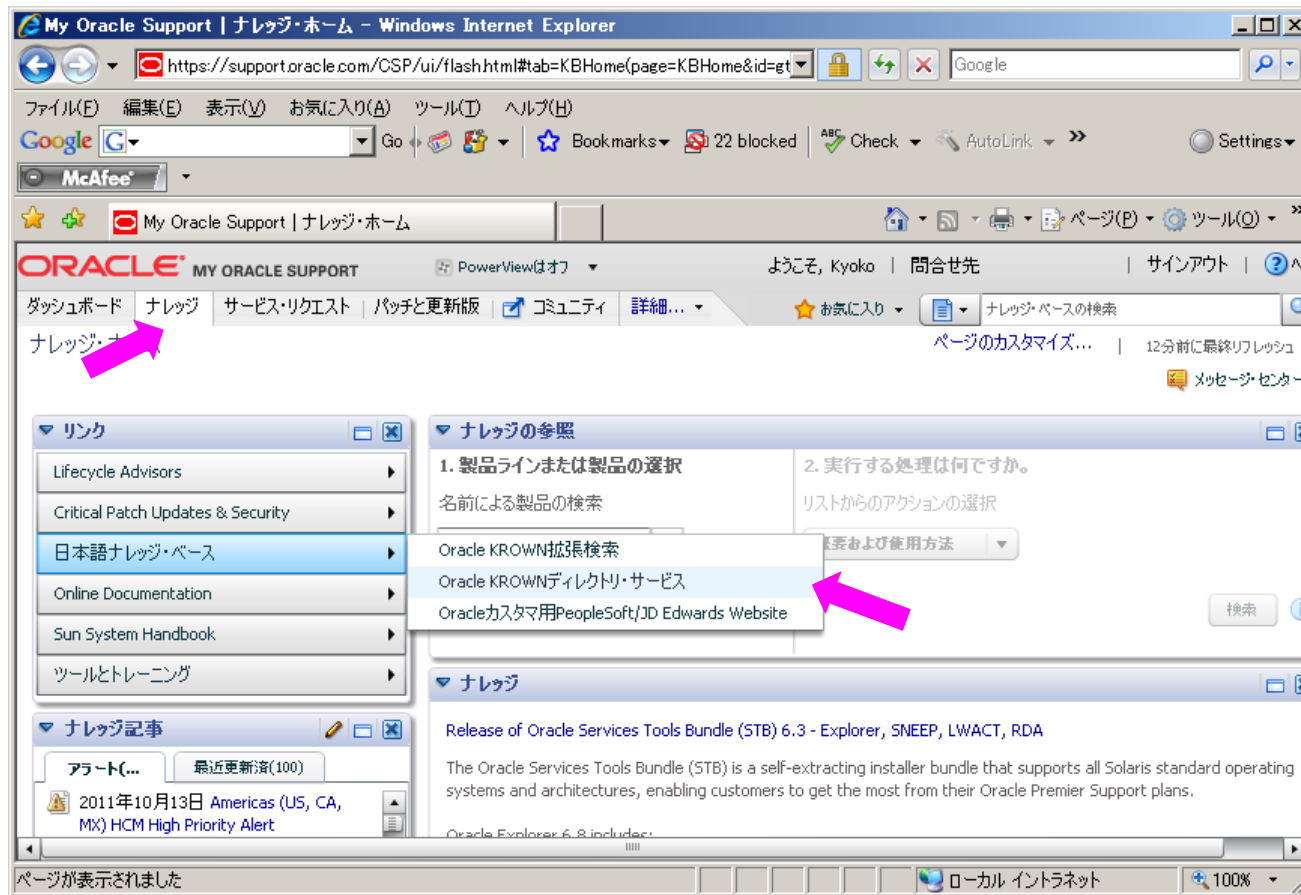
- 145837: 11gR2 RAC : 11gR2 Clusterware and Grid Home - What You Need to Know
- 152202: 11.2 Grid Infrastructure FAQ
- 140037: [11gR2 新機能] RAC : Grid Infrastructure Single Client Access Name (SCAN) について
- 137922: [11gR2 新機能] RAC : Policy-Managed と Administrator-Managed の変換方法について
- 153443: 11gR2 RAC : hagsuser グループに追加する必要がある OS ユーザーについて (AIX)
- 145766: 11gR2 RAC : AIX Power6 で LPAR を使用する場合の制限事項 (AIX)
- 152490: 11gR2 RAC : root.sh、rootupgrade.sh 実行時の注意点
- 140159: 11gR2 RAC : Grid Infrastructure の起動に対するトラブルシューティング
- 140036: 11gR2 RAC: Clusterware Node Evictions(Reboot)の発生に対するトラブルシューティング

補足：技術文章のご案内

My Oracle Support (<https://support.oracle.com/CSP/ui/flash.html>)

⇒ ナレッジ

⇒ 日本語ナレッジ・ベース ⇒ Oracle KROWN ディレクトリ・サービス



補足：技術文章のご案内

Krown Directory Service

Krown Directory Service 2 : Real Application Clusters - Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)

oracle.co.jp https://krown.oracle.co.jp/techinfo/NodeListController?nid=173

McAfee

Krown Directory Service 2 : Real ...

ブラウズ [ツリー表示]

- KROWNディレクトリ・サービス
 - >> データベース
 - >> 高可用性(RAC・OFS・データガード等)
 - >> Real Application Clusters
- 10gR2 RAC : 10.2 情報 (236)
- 10gR1 RAC (10.1.0.4 情報) (32)
- インストールと初期設定 (63)
- トラブルシューティング (9)
- 使い方/TIPS(HowTo) (30)
- 技術情報 (124)
- 障害情報 (153)

リンク

- Oracle Technology Network [LINK](#)
- Oracle AppsNet [LINK](#)
- MetaLink [LINK](#)

Real Application Clusters

KROWNディレクトリサービス >> データベース >> 高可用性(RAC・OFS・データガード等) >> [Real Application Clusters](#)

11gR2 RAC : 11.2.0 情報

11.2 新機能

- #145837 11gR2 RAC : 11gR2 Clusterware and Grid Home - What You Need to Know
- #152202 11.2 Grid Infrastructure FAQ
- #140037 [11gR2 新機能] RAC : Grid Infrastructure Single Client Access Name (SCAN) について
- #137922 [11gR2 新機能] RAC : Policy-Managed と Administrator-Managed の変換方法について

SCAN(Single Client Access Name) に関する情報は以下も参照して下さい。

- SCAN [KDS](#)

インストール関連情報

インストール全般

- #149432 11gR2 : [マニュアルバグ] 11.2.0.1 から 11.2.0.2 へのアップグレードに必要なパッチについて **ALERT**
- #152490 11gR2 RAC : root.sh, rootupgrade.sh 実行時の注意点
- #152506 Grid Infrastructure をインストールするために必要なメモリ要件について
- #152943 11.2.0.2 RAC : マルチキャストの要件により Grid Infrastructure のインストール・アップグレードが

https://krown.oracle.co.jp/techinfo/NodeListController?nid=173

OTNセミナーオンデマンド

コンテンツに対する
ご意見・ご感想を是非お寄せください。

OTNオンデマンド 感想



http://blogs.oracle.com/oracle4engineer/entry/otn_ondemand_questionnaire

上記に簡単なアンケート入力フォームをご用意しております。

セミナー講師/資料作成者にフィードバックし、
コンテンツのより一層の改善に役立てさせていただきます。

是非ご協力をよろしくお願いいたします。

OTNセミナーオンデマンド

日本オラクルのエンジニアが作成したセミナー資料・動画ダウンロードサイト

掲載コンテンツカテゴリ(一部抜粋)

Database 基礎

Database 現場テクニック

Database スペシャリストが語る

Java

WebLogic Server/アプリケーション・グリッド

EPM/BI 技術情報

サーバー

ストレージ



超入門! Oracle データベースって何

再生時間: 60分

100以上のコンテンツをログイン不要でダウンロードし放題

データベースからハードウェアまで充実のラインナップ

毎月、旬なトピックの新作コンテンツが続々登場

例えばこんな使い方

- 製品概要を効率的につかむ
- 基礎を体系的に学ぶ/学ばせる
- 時間や場所を選ばず(オンデマンド)に受講
- スマートフォンで通勤中にも受講可能



毎月チェック!



コンテンツ一覧 はこちら

<http://www.oracle.com/technetwork/jp/ondemand/index.html>

新作&おすすめコンテンツ情報 はこちら

<http://oracletech.jp/seminar/recommended/000073.html>

OTNオンデマンド



オラクルエンジニア通信

オラクル製品に関わるエンジニアの方のための技術情報サイト

オラクルエンジニア通信 - 技術資料、マニュアル、セミナー

Oracleエンジニアのための技術情報サイト by Oracle Japan

新着情報を知りたい

技術資料を探したい

セミナーを受けたい

About

Oracleエンジニアの方がスキルアップしていただくために、厳選した情報をお届けしています

技術資料	<p>インストールガイド・設定チュートリアルetc. 欲しい資料への最短ルート</p>	アクセスランキング	<p>他のエンジニアは何を見ているのか？人気資料のランキングは毎月更新</p>
特集テーマ Pick UP	<p>性能管理やチューニングなど月間テーマを掘り下げて詳細にご説明</p>	技術コラム	<p>SQLスクリプト、索引メンテナンスetc. 当たり前運用/機能が見違える!?</p>

<http://blogs.oracle.com/oracle4engineer/>

オラクルエンジニア通信



The screenshot shows the top navigation bar of the oracletech.jp website. It features the 'oracletech.jp' logo in red and black, with the tagline '好奇心が、エンジニア人生を豊かにする。' below it. To the right is the 'ORACLE' logo, a search bar, and social media icons for Twitter, Facebook, LinkedIn, YouTube, and RSS. Below these is a red navigation bar with five buttons: '製品/技術情報', 'スキルアップ', 'セミナー', 'キャンペーン', and 'ちょっと一息'.

製品/技術
情報



Oracle Databaseっていくら？オプション機能も見積れる簡単ツールが大活躍

セミナー



基礎から最新技術までお勧めセミナーで自分にあった学習方法が見つかる

スキルアップ



ORACLE MASTER !
試験頻出分野の模擬問題と解説を好評連載中

Viva!
Developer



全国で活躍しているエンジニアにスポットライト。きらりと輝くスキルと視点を盗もう

<http://oracletech.jp/>

oracletech



あなたにいちばん近いオラクル



Oracle Direct

まずはお問合せください

Oracle Direct



システムの検討・構築から運用まで、ITプロジェクト全般の相談窓口としてご支援いたします。
システム構成やライセンス/購入方法などお気軽にお問い合わせ下さい。

Web問い合わせフォーム

専用お問い合わせフォームにてご相談内容を承ります。
http://www.oracle.co.jp/inq_pl/INQUIRY/quest?rid=28

※フォームの入力にはログインが必要となります。
※こちらから詳細確認のお電話を差し上げる場合がありますので
ご登録の連絡先が最新のものになっているかご確認下さい。

フリーダイヤル

0120-155-096

※月曜～金曜
9:00～12:00、13:00～18:00
(祝日および年末年始除く)

ORACLE

Hardware and Software Engineered to Work Together

ORACLE®