

# 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®

### 実践!! 高可用性システム構築 ～RAC詳細編～

日本オラクル株式会社 製品事業統括 テクノロジー製品事業統括本部  
エンジニア 小林 史郎

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。

# Agenda

- Real Application Clusters概要
- 高拡張性①
  - キャッシュの一貫性
- 高拡張性②
  - パラレル処理
- 高可用性
- まとめ

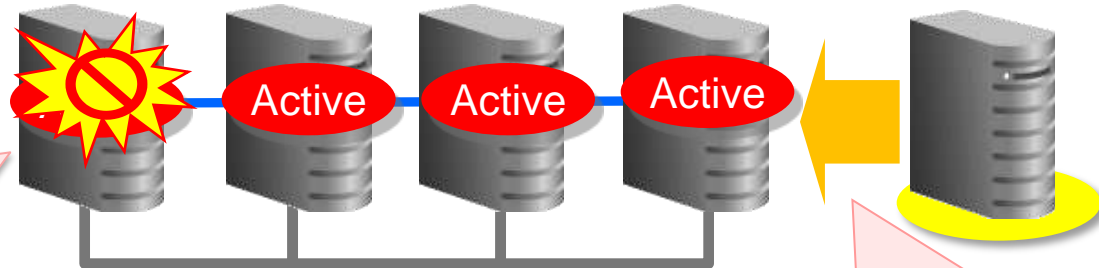
# Real Application Clusters概要



# Real Application Clusters (RAC) とは

- Oracle Real Application Clustersとは  
システム障害によるダウンタイムを最小に抑え、データ保護と連続的なサービス環境を実現するOracle Databaseのクラスタ技術
- 特徴
  - ✓ 高拡張性: サーバの追加によって容易にスケールアウトが可能
  - ✓ 高可用性: Active-Activeのクラスタ構成により、障害時サービスが停止しても残りのサーバーで処理を継続できる

障害によるサーバ停止があっても、システムは残りのサーバーが維持できる



1つのデータベースに複数のサーバ(ノード)から同時にアクセスできる

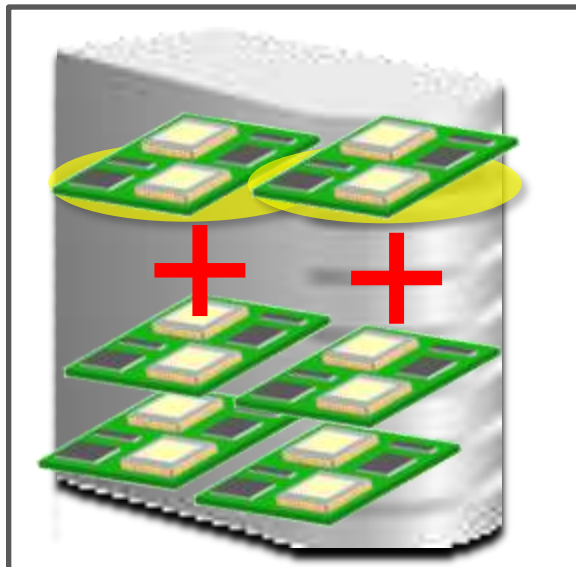
共有ディスク

処理量の増加に合わせ、容易に拡張可能

# RACの特徴① 高拡張性の実現

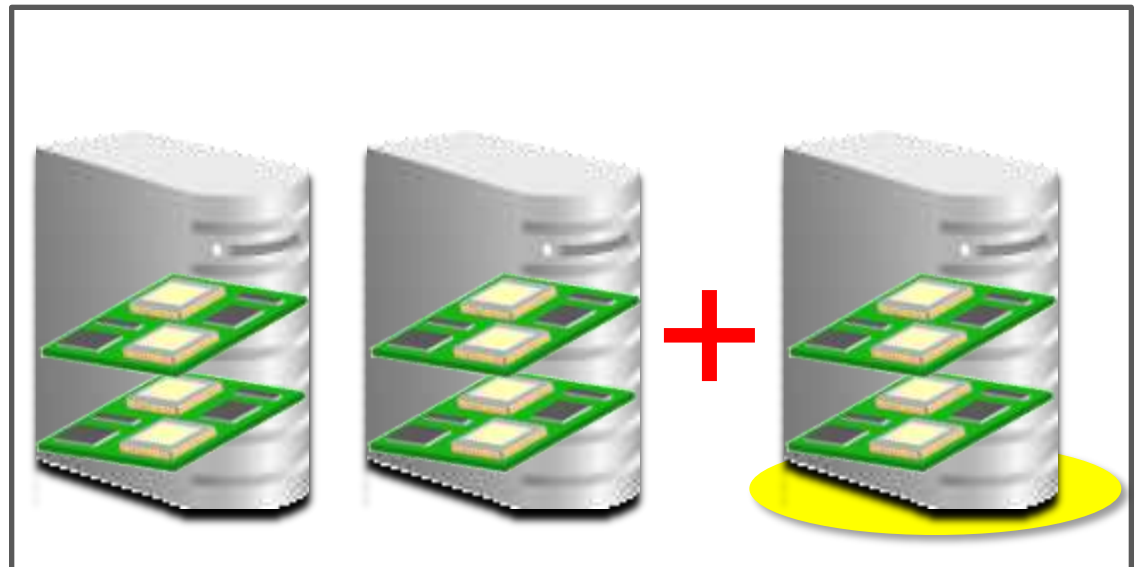
- サーバーの処理性能を高めるためには...
  - CPUコアを増設して同時実行性を向上
  - CPUコアを増設するアプローチは2つ: スケールアップ / スケールアウト

スケールアップ



サーバーにCPUを増設

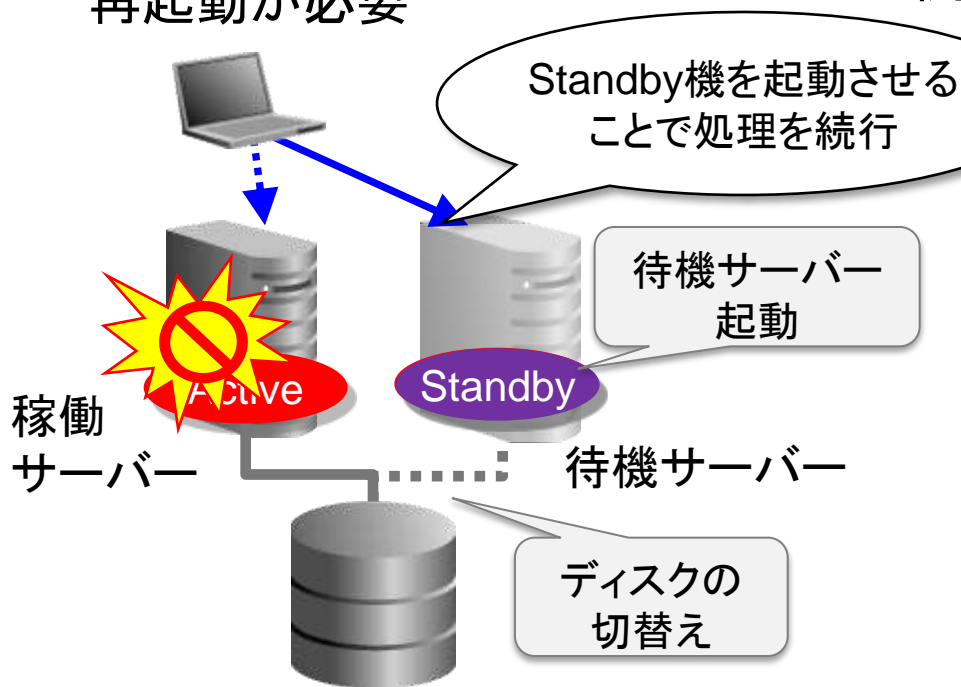
スケールアウト



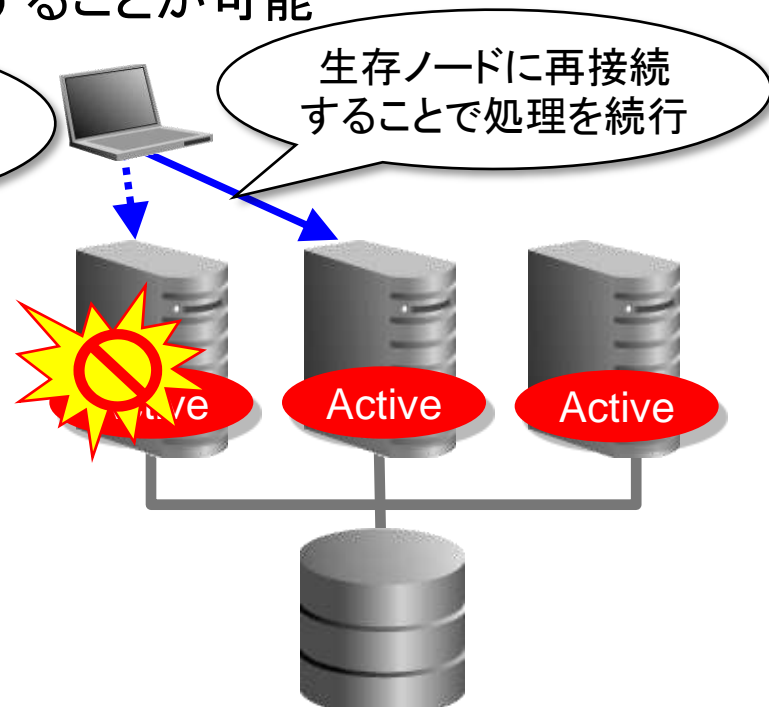
ノードを追加してCPUを増設

## RACの特徴② 高可用性の実現

- HA(Active-Standby)構成では、障害時にディスクの切り替えやデータベース・インスタンスの再起動が必要
- RAC(Active-Active)構成では、全てのサーバーが稼働しているためすぐに別のサーバーですぐに処理を再開することが可能



Active-Standby構成

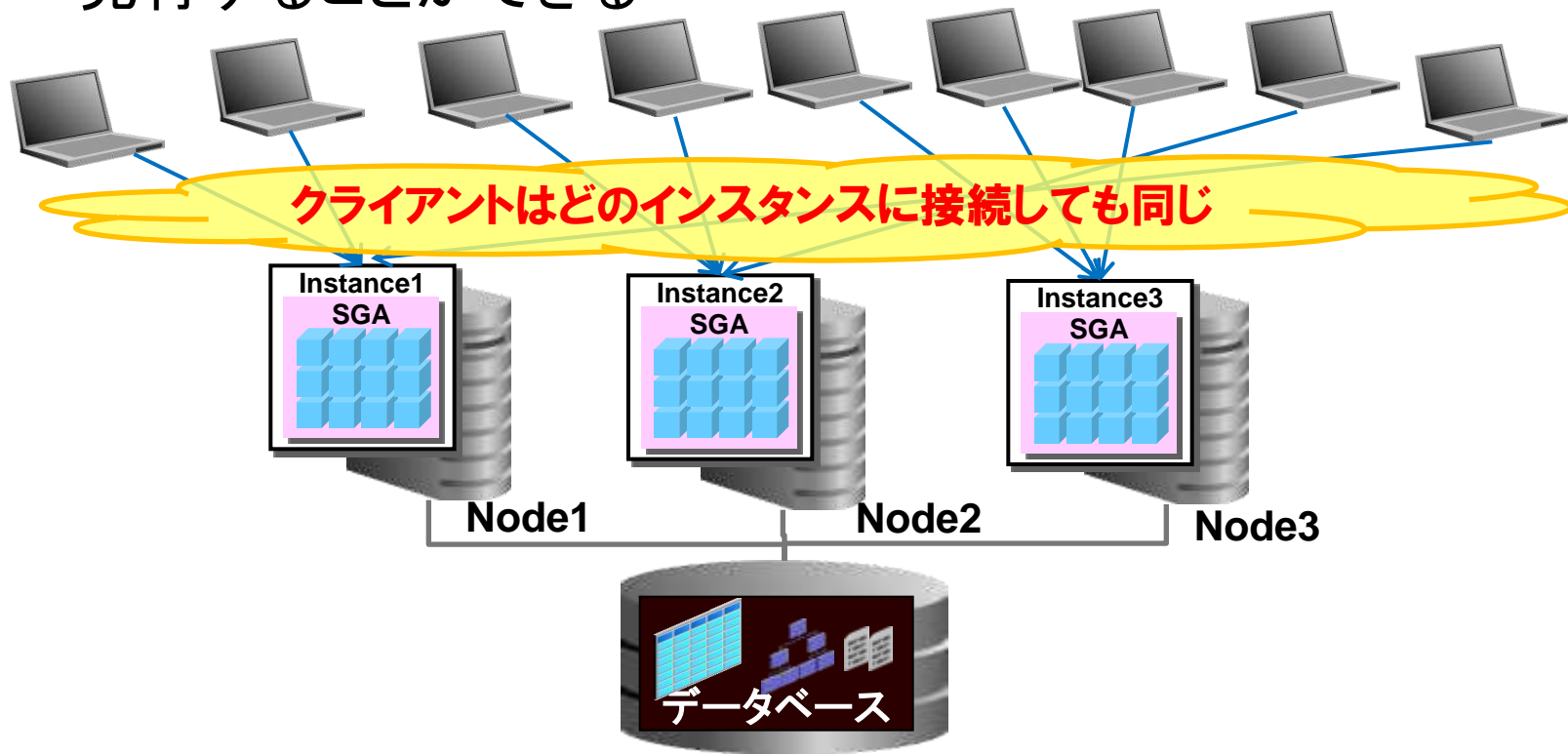


RAC構成

ORACLE

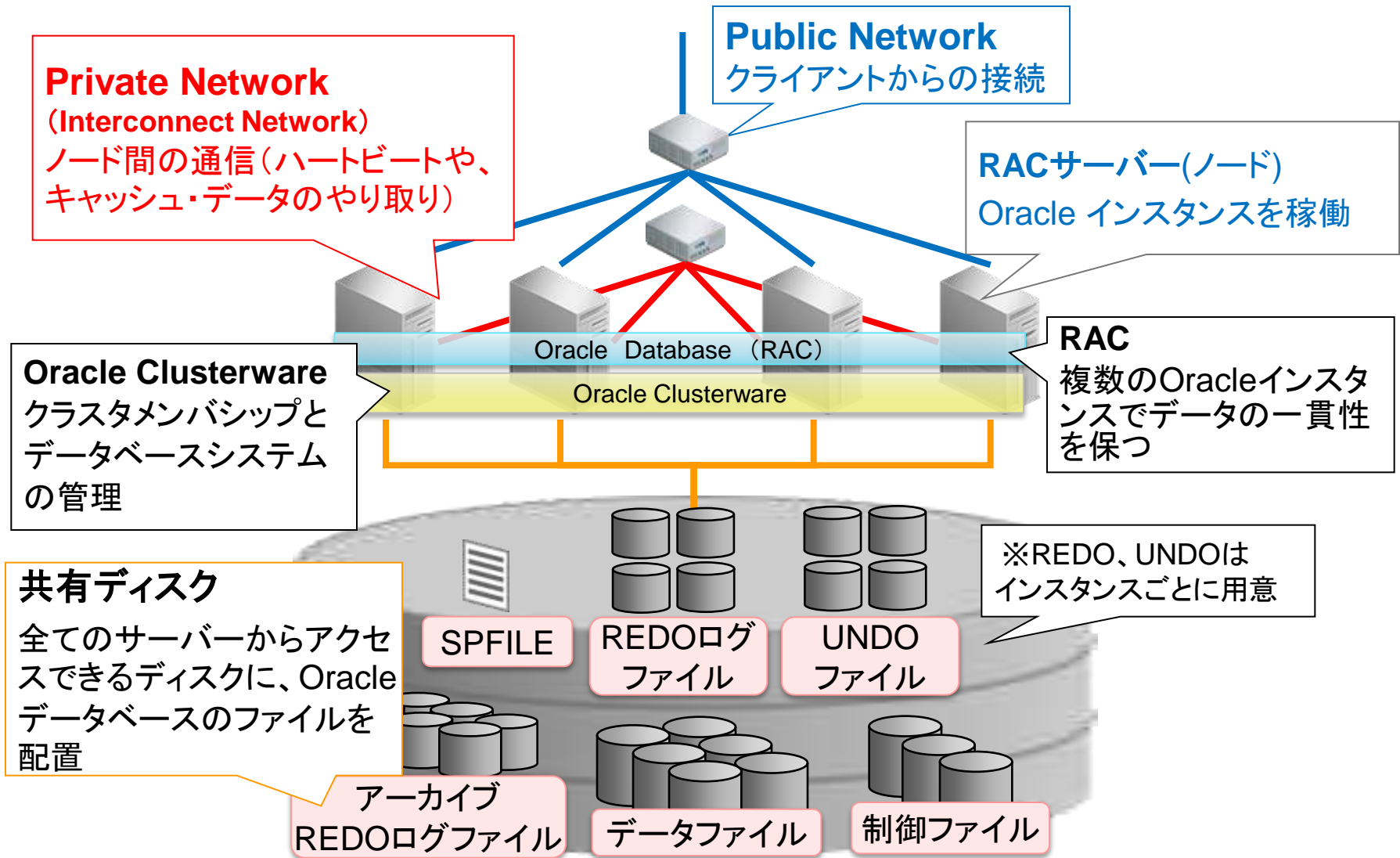
# RAC データベースとインスタンスの関係

- 複数のインスタンスが同じデータベースにアクセスする
- クライアントはどのインスタンスに接続をしても同じSQLを発行することができる



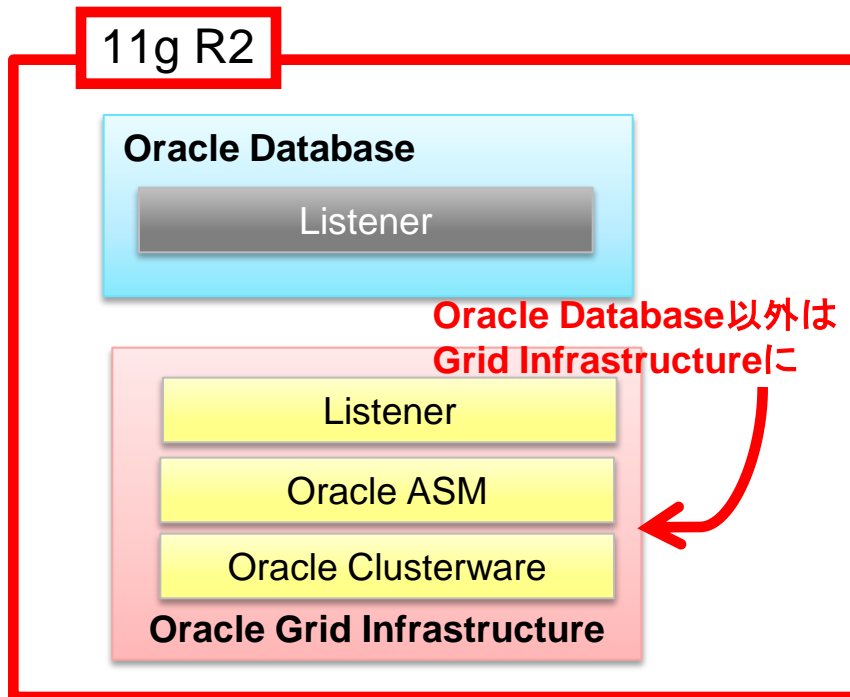
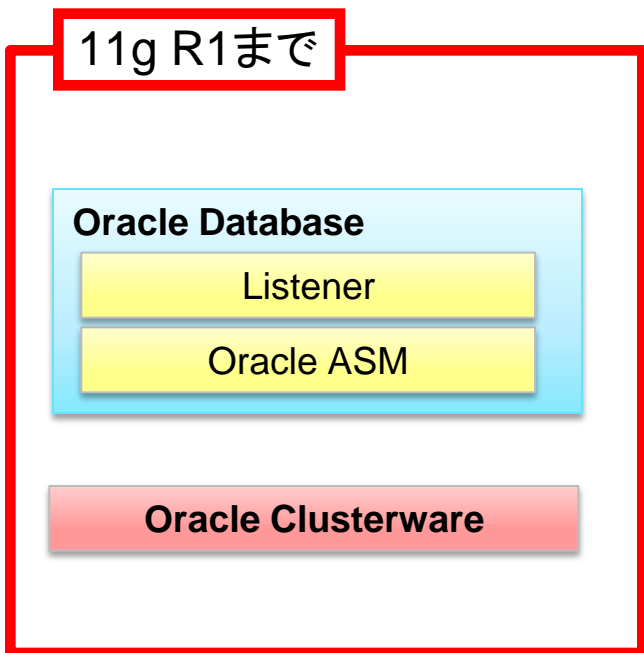


# RACの構成



# RACのソフトウェアとコンポーネントの関係

- RACを構成するためには Oracle Clusterwareが必要
- 11g R2から Oracle Clusterwareと Oracle ASMが統合され Oracle Grid Infrastructureの構成要素となった
  - データベース統合基盤となるコンポーネントを集約



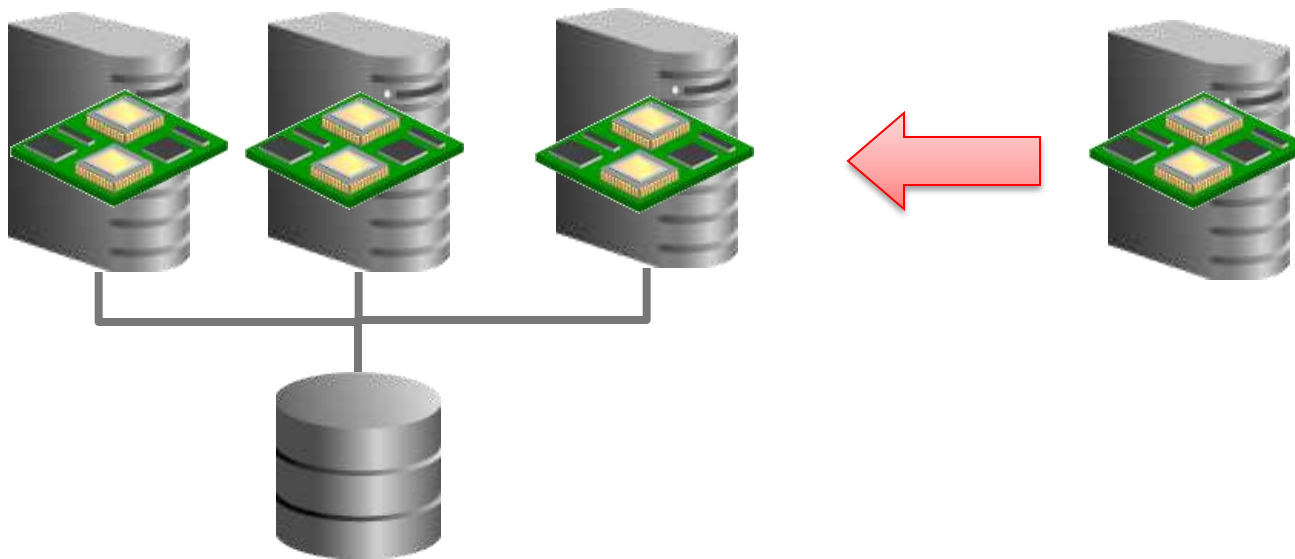
Oracle ASM ... Oracle Automatic Storage Management

# RACの高拡張性 キャッシュの一貫性



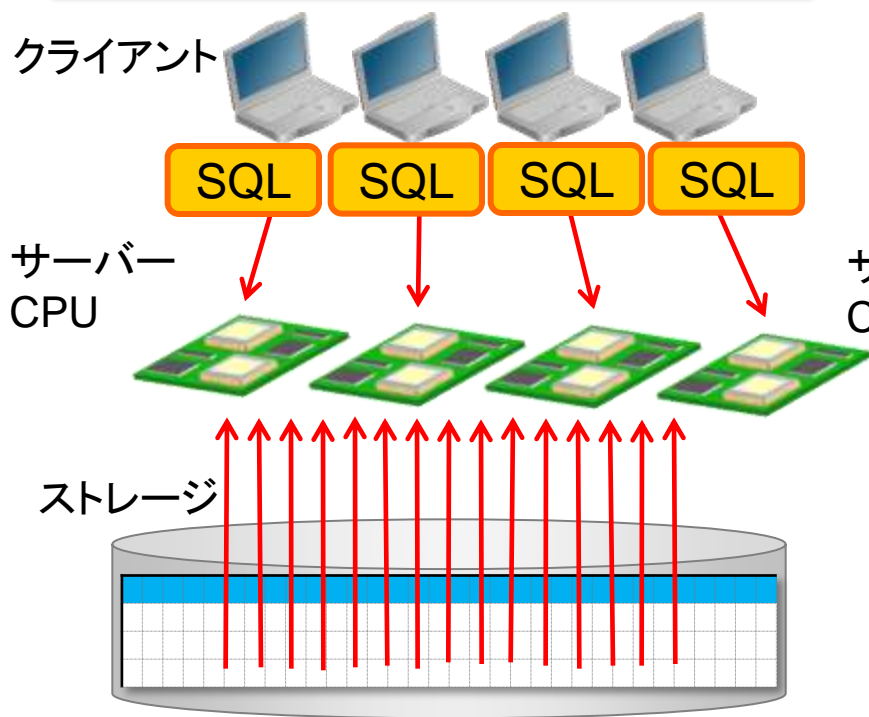
# 「ノード」を追加するということとは？

- RACにノードを追加するのはCPUを追加するということ
  - 同時に実行可能な処理を増やすことができる
  - 1つの処理を並列化することができる

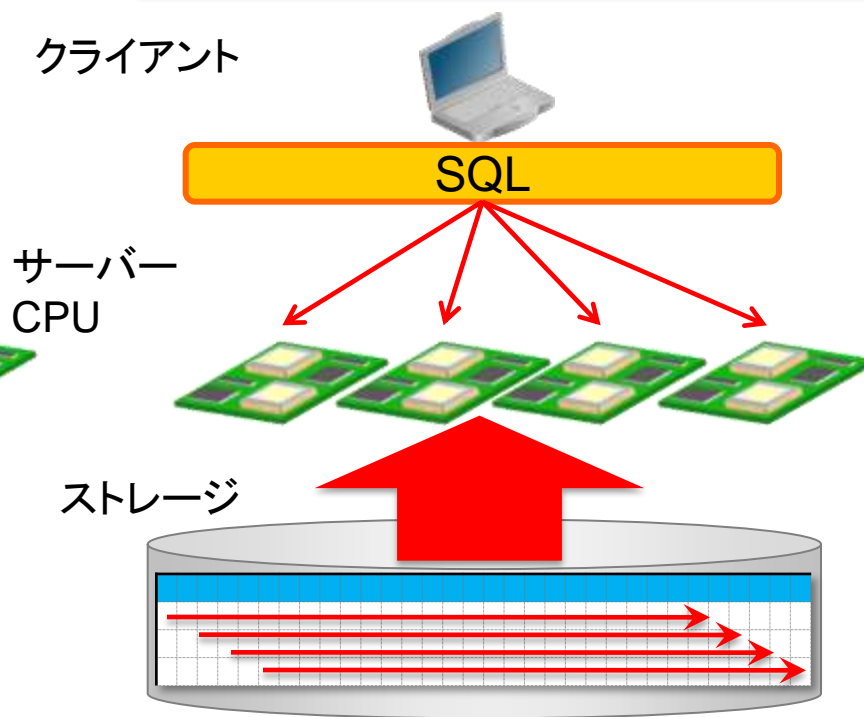


# 複数あるCPUの利用

複数の処理を同時実行  
(スループットを向上)



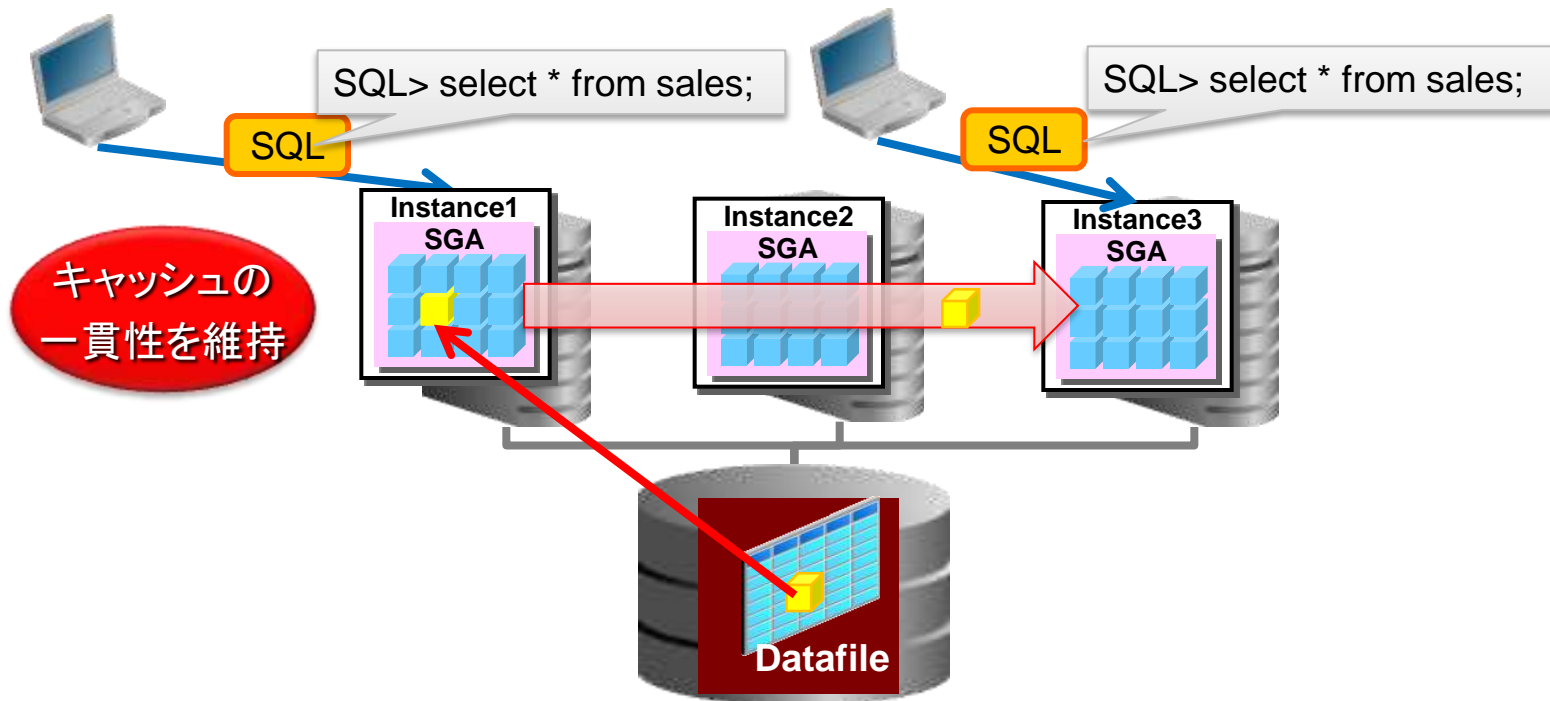
1つの処理を並列実行  
(レスポンスタイムを向上)



# キャッシュの一貫性を保持する仕組み

## • Cache Fusion

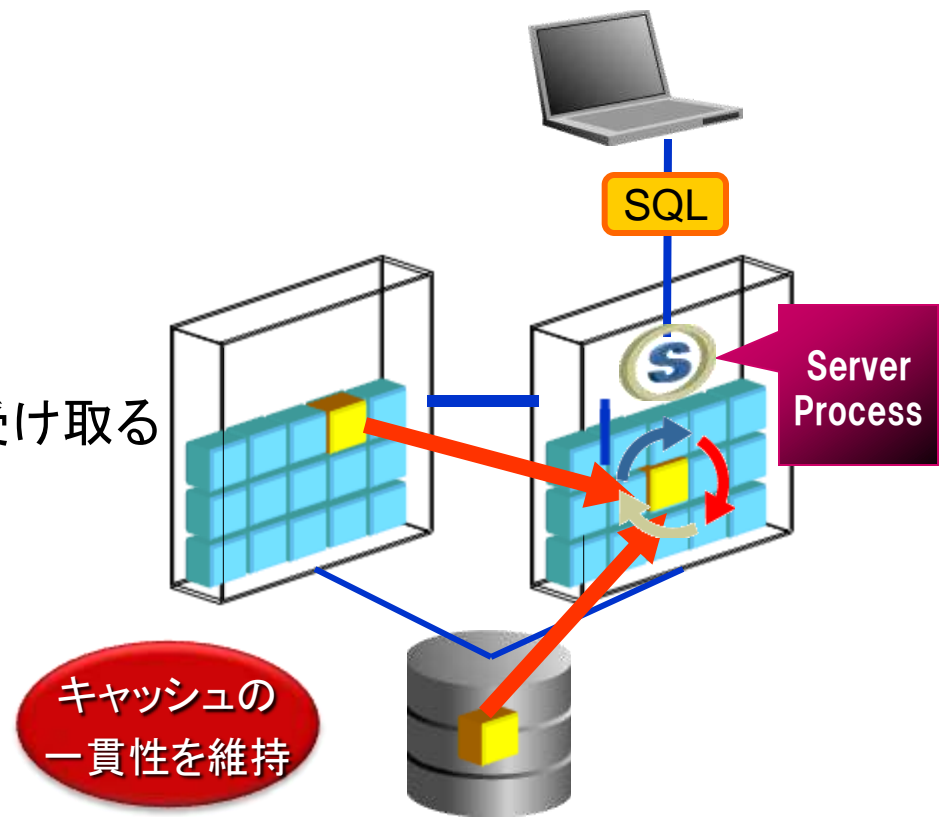
- 各ノード間でキャッシュの一貫性を維持する仕組み
- 要求されたOracle データ・ブロックが他のノードで更新されていても、最新のデータ・ブロックを取得することが可能



# RACのデータアクセス

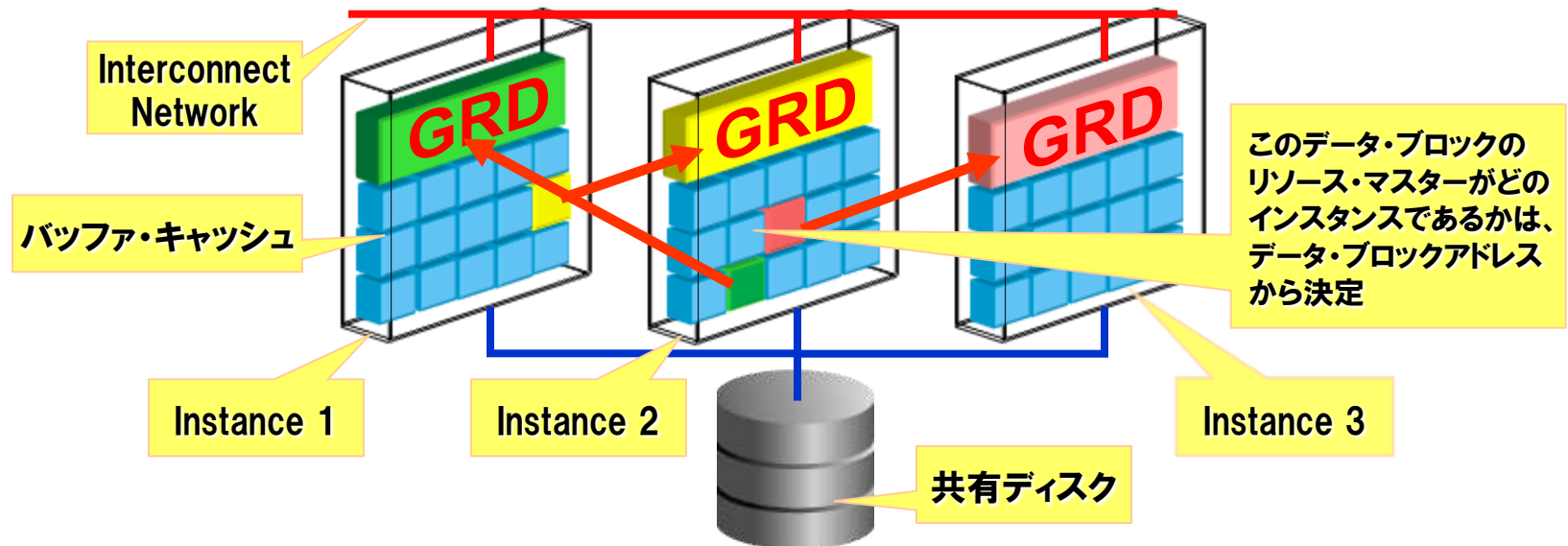
## Cache Fusion

- キャッシュ・ヒットした場合
  - キャッシュのデータを使用
- キャッシュ・ミスした場合
  - シングル・インスタンス
    - ディスクから読み込む
  - RAC
    - 他ノードのキャッシュから受け取る
    - ディスクから読み込む



# Global Resource Directory (GRD)

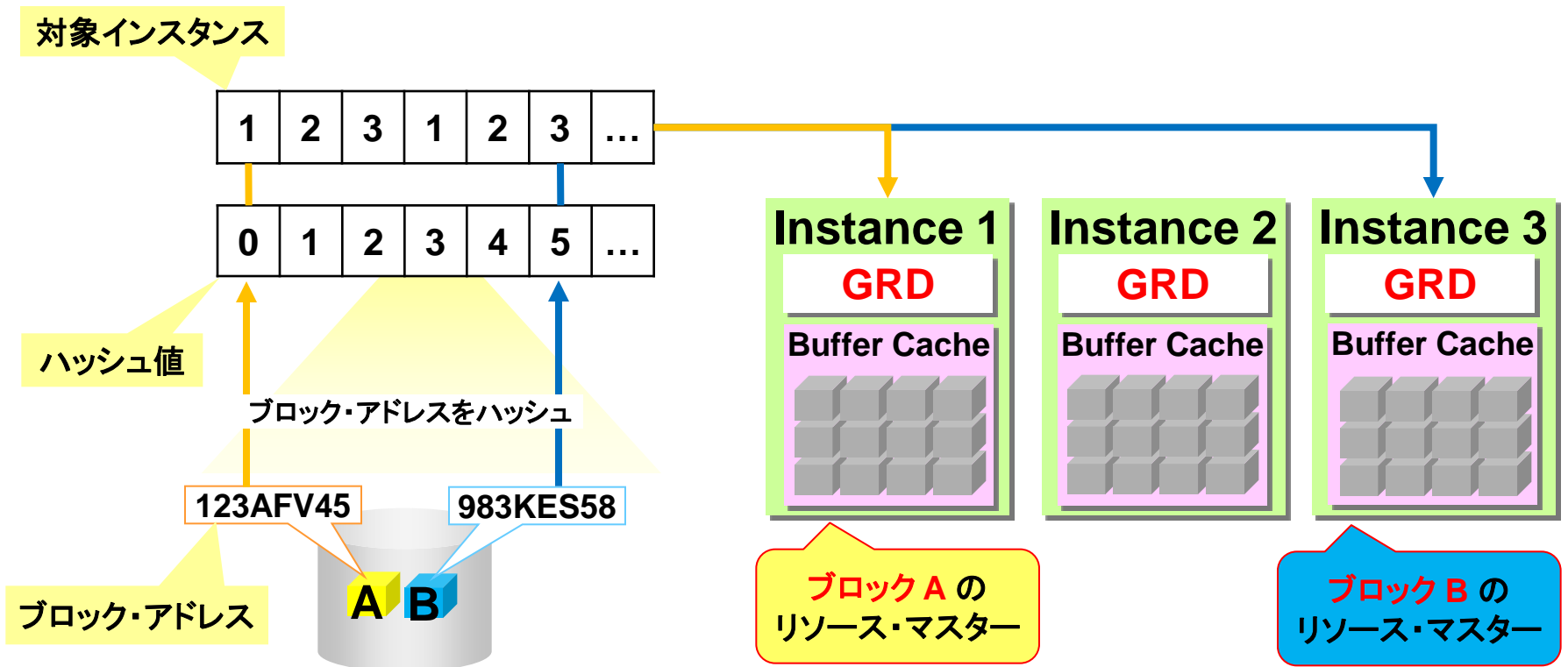
- Oracle データ・ブロックをどのインスタンスがどのロック状態でキャッシュしているかを管理する領域
- RACの全Oracle インスタンスのSGA上に分散配置
  - あるデータ・ブロックのロックの状態に責任をもつGRDは、クラスタのノードうち1つあり、そのデータ・ブロックのリソース・マスターと呼ぶ



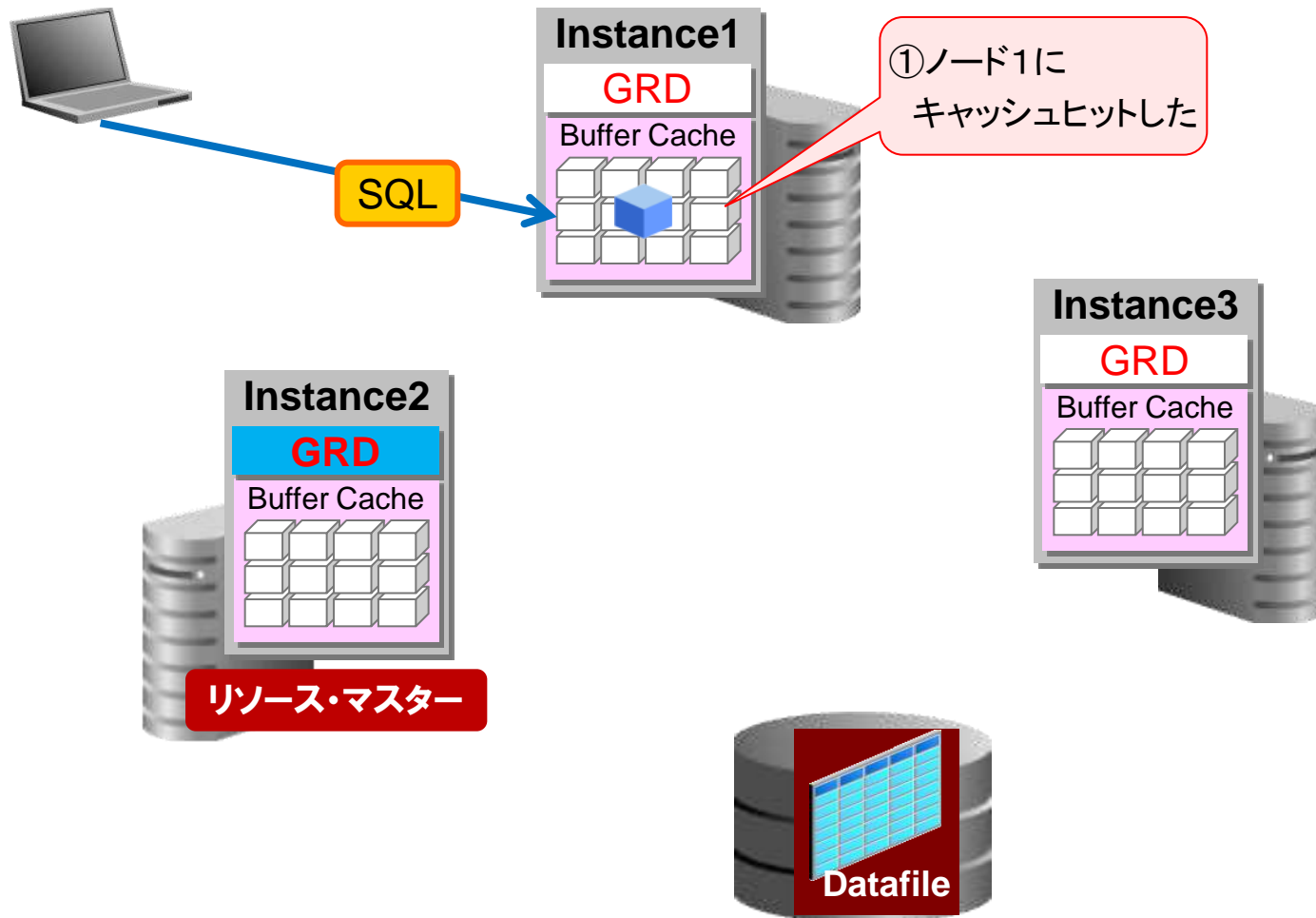


# リソース・マスターの割り当て

- あるデータ・ブロックのリソース・マスターとなるインスタンスは、そのブロック・アドレスをハッシュした値に基づき決定する

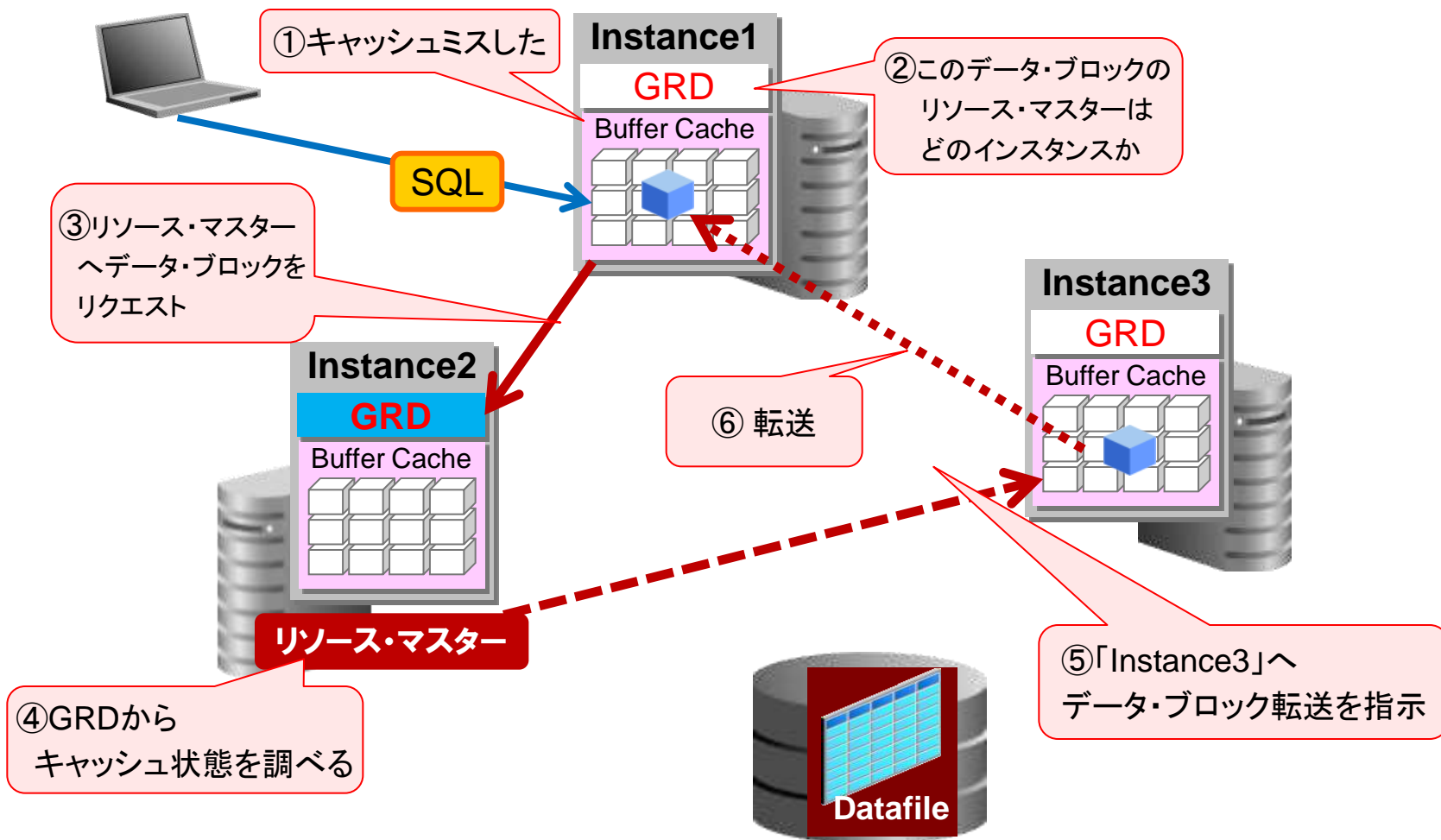


# キャッシュ・ヒットした場合



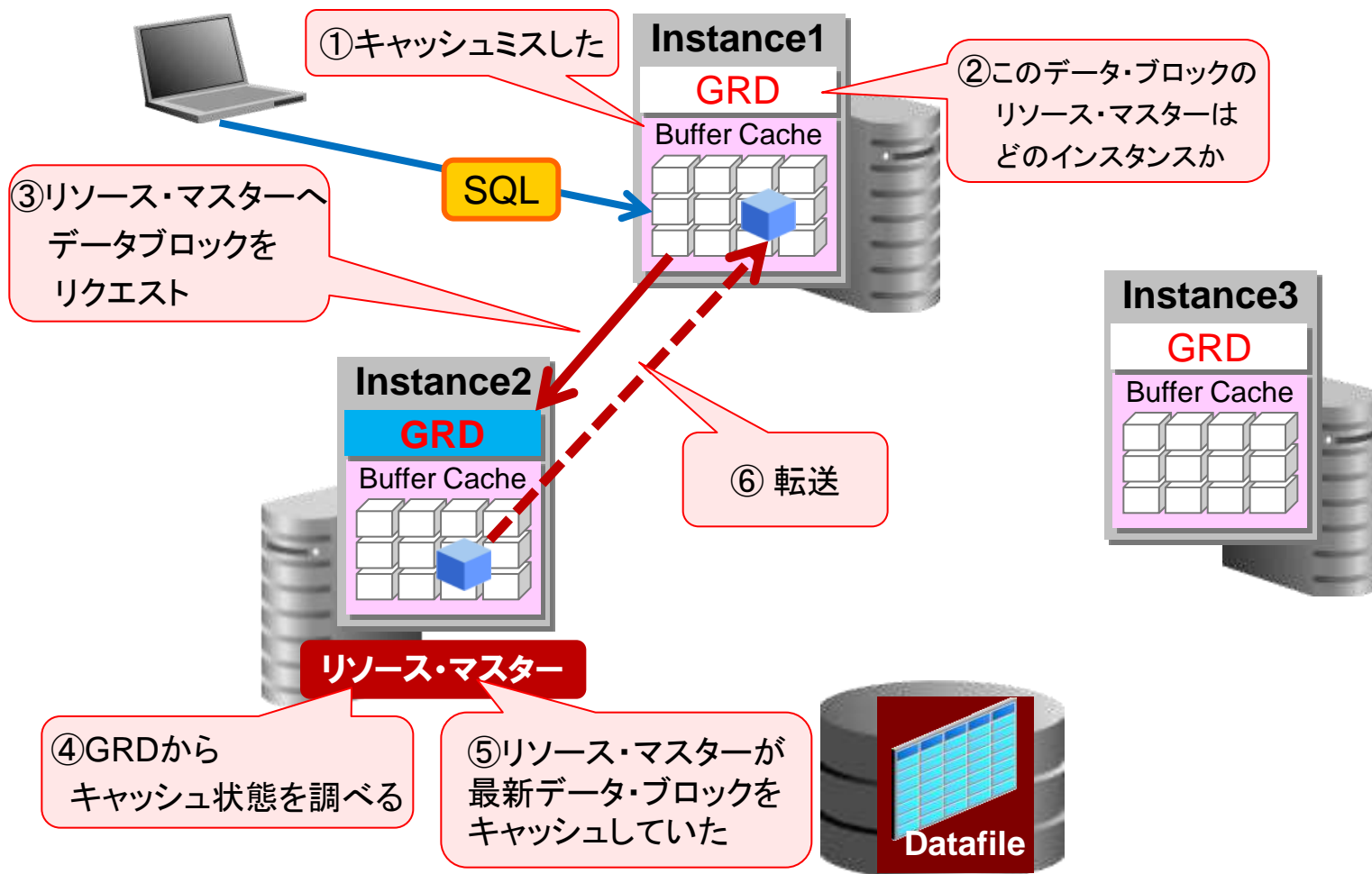
# キャッシュ・ミスした場合①

## リモート・キャッシュから取得



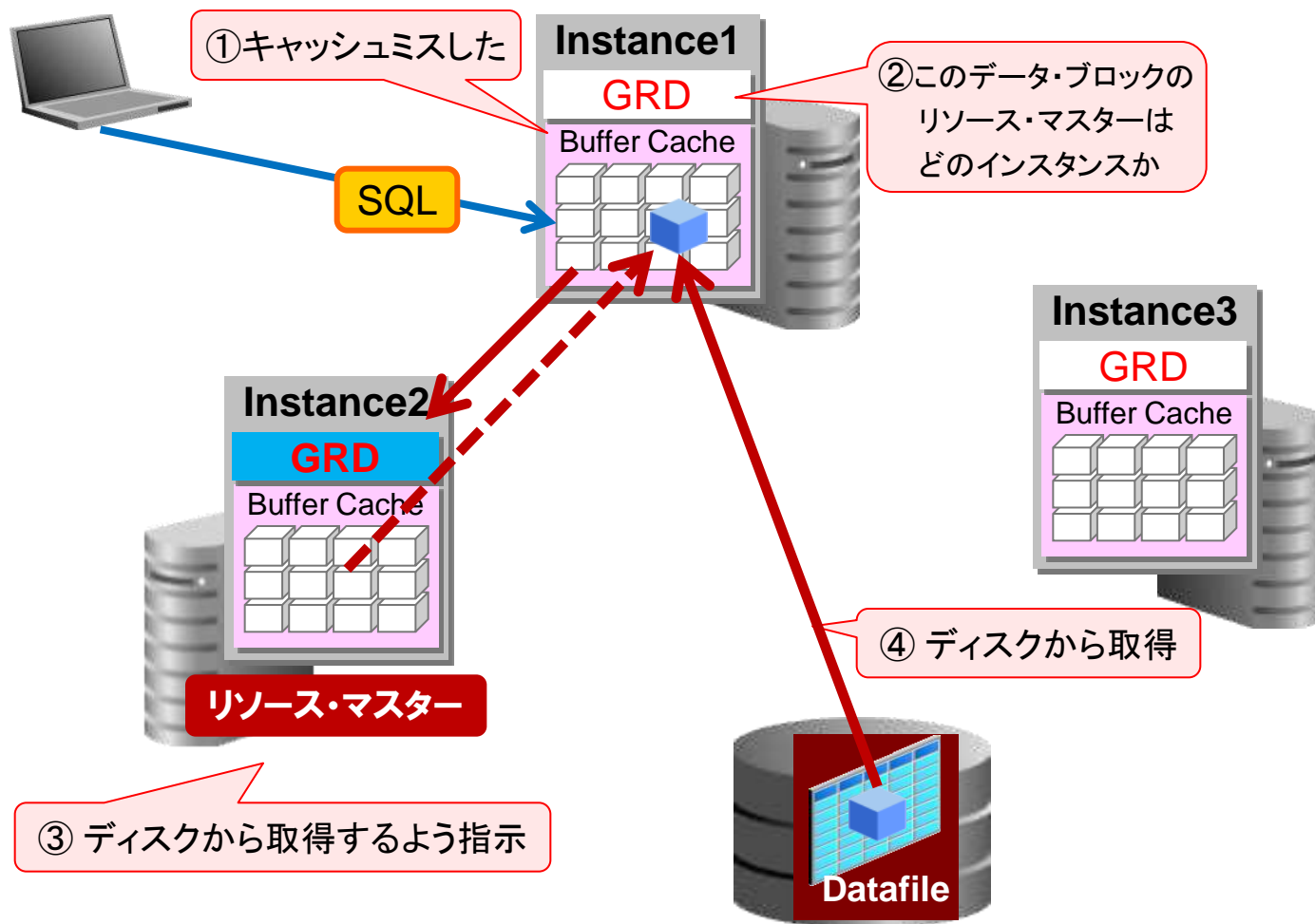
# キャッシュ・ミスした場合②

## リモート・キャッシュから取得



# キャッシュ・ミスした場合③

## ディスクから取得

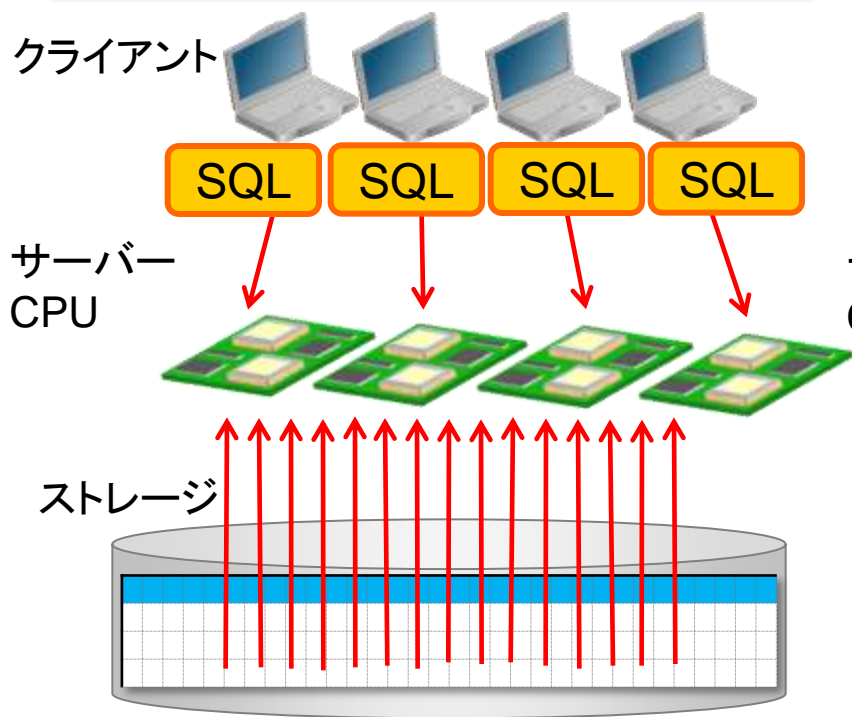


## RACの高拡張性② パラレル処理

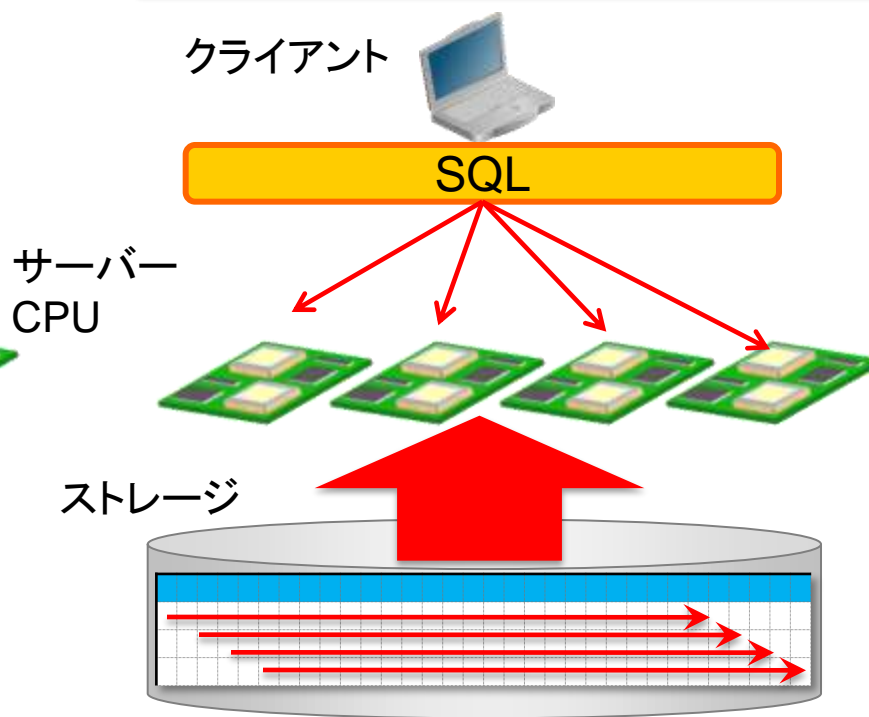


# 複数あるCPUの利用

複数の処理を同時実行  
(スループットを向上)

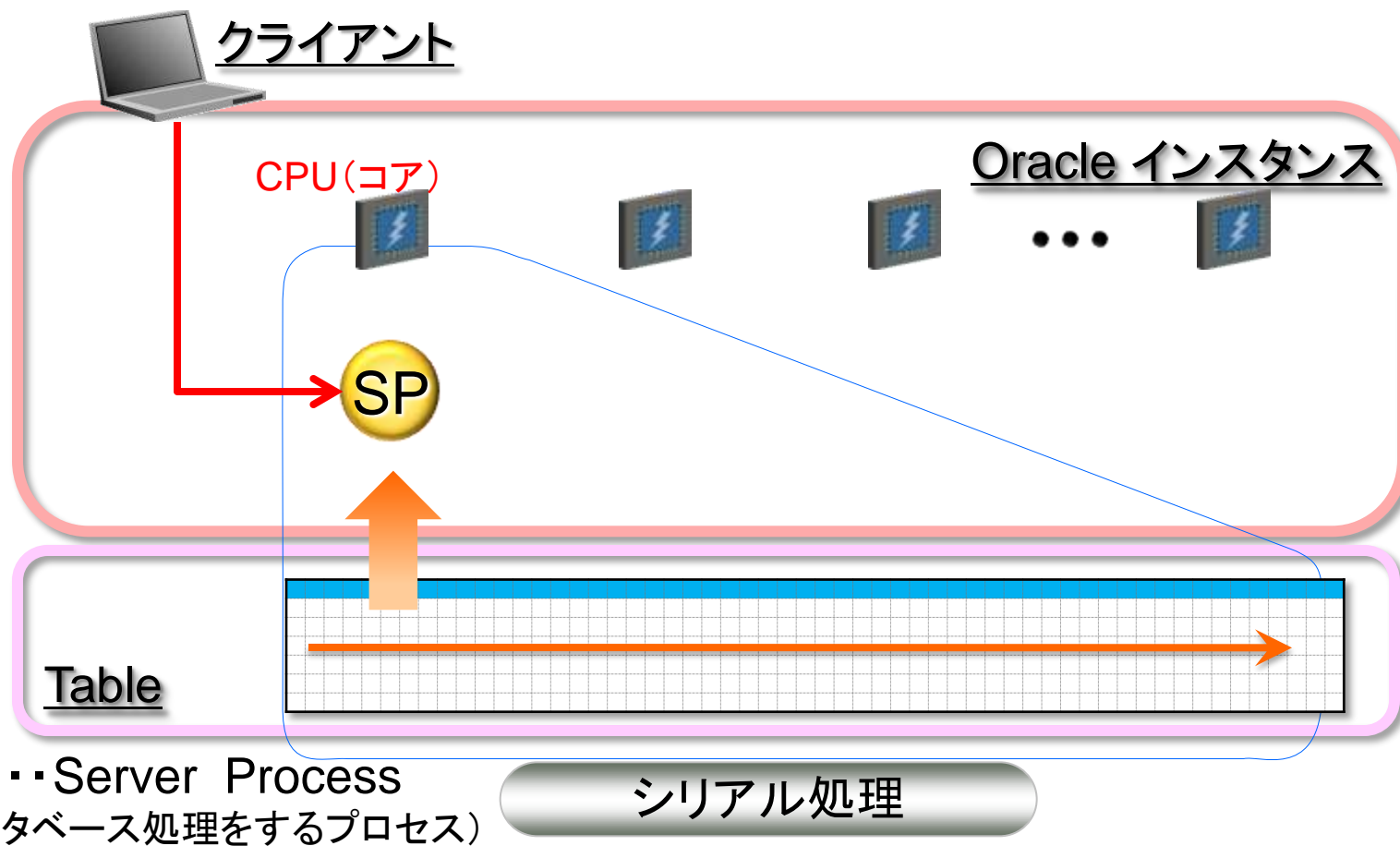


1つの処理を並列実行  
(レスポンスタイムを向上)



# 1つのCPUを使用したデータベース処理

大量データを集計するようなSQLをシリアル実行した場合

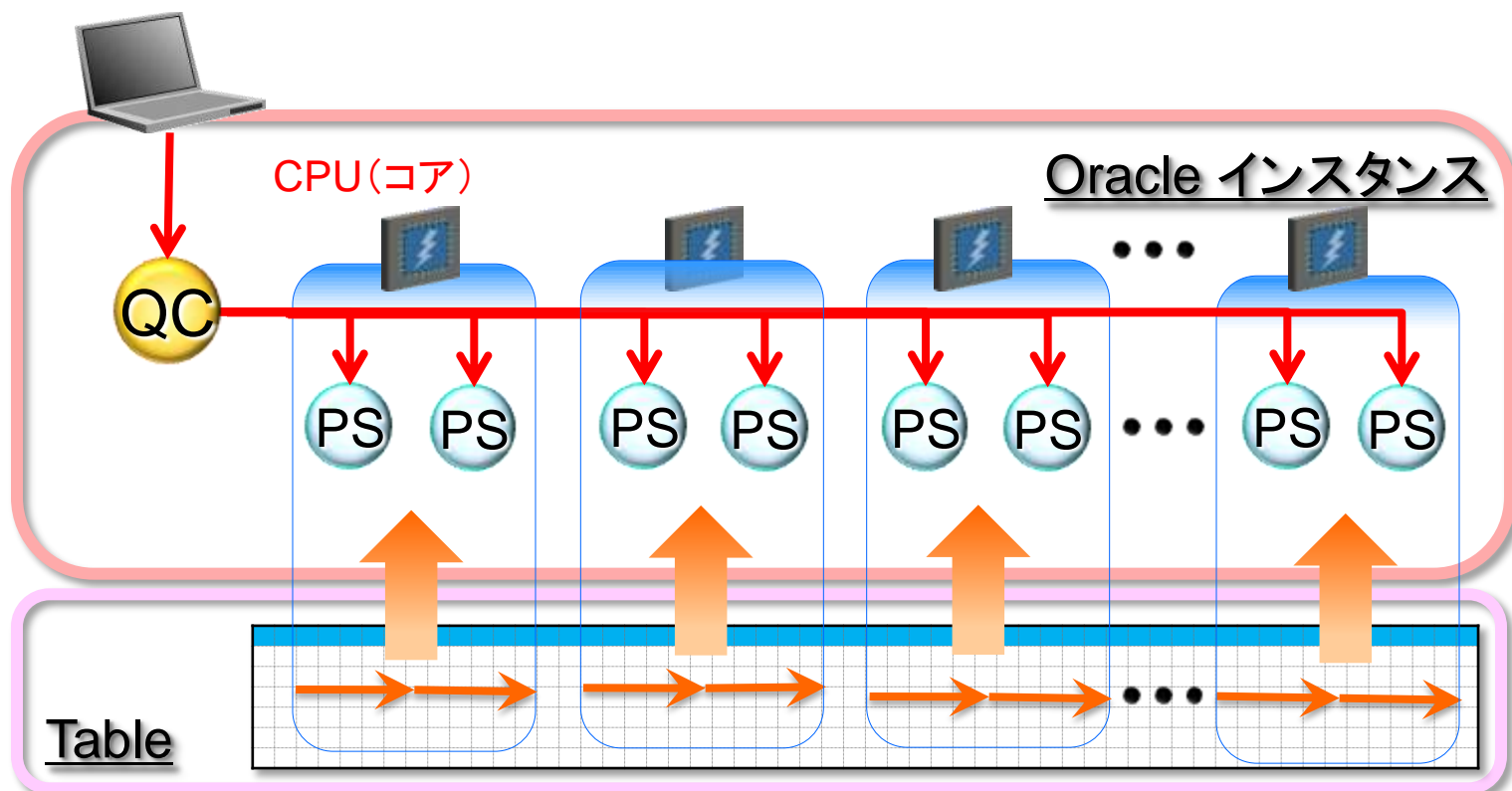




# パラレル処理による複数CPUの利用

## 1つのSQLを自動並列化

パラレル処理により搭載しているCPUにデータを分散して短時間で結果を返す。



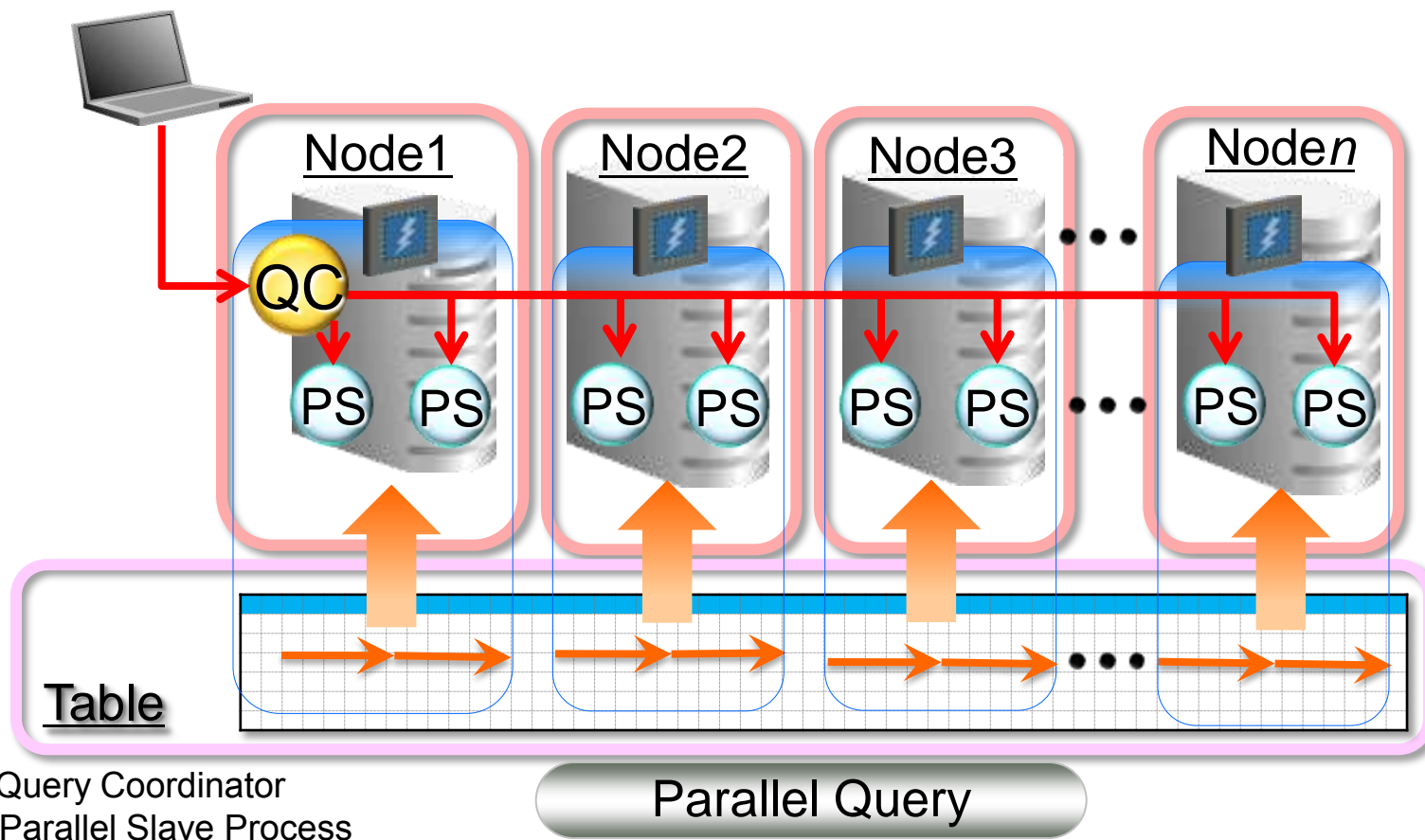
QC...Query Coordinator  
PS ...Parallel Slave Process

Parallel Query

# パラレル処理による複数CPUの利用

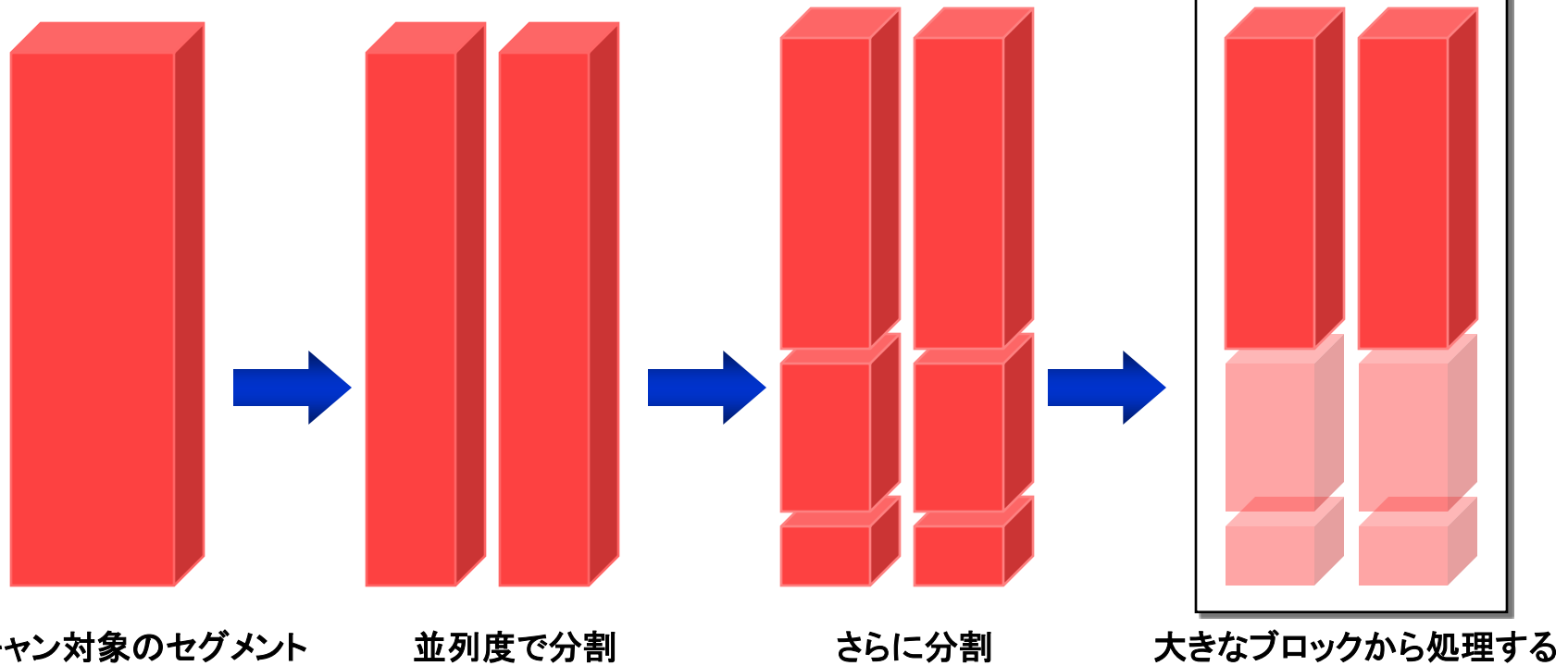
## 1つのSQLを自動並列化

パラレル処理により全ノードに搭載しているCPUにデータを分散して短時間で結果を返す。



# Parallel Queryロードバランス

- スキャン範囲の担当を動的に決定する
  - 各スレーブ・プロセスは、異なるブロックを担当
  - スレーブ・プロセスの実行時間を均等にする



# パラレル・クエリーの設定

## Oracle Database 11g Release 2からの新たなパラレル度設定の方法

11.2 New

- 自動パラレル

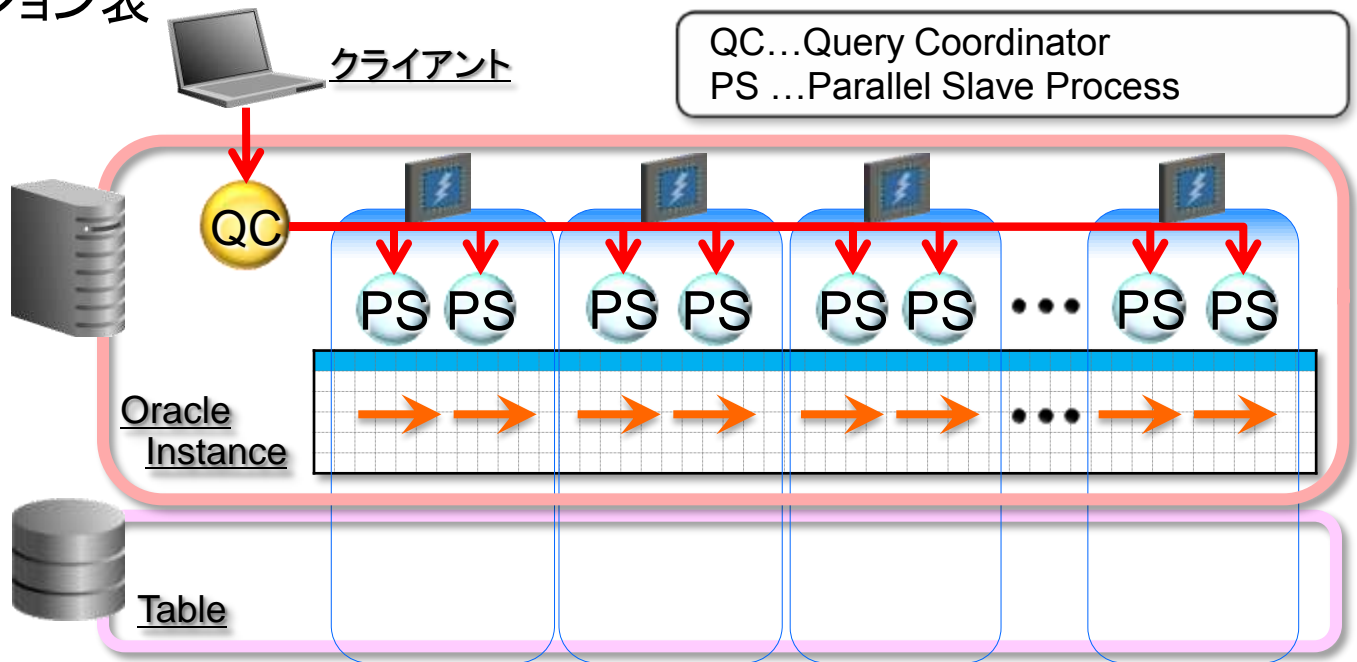
- 各SQLの最適なパラレル度を自動的に設定

- アプリケーション側での設定は不要
- 初期化パラメータ PARALLEL\_DEGREE\_POLICYの設定
  - RAC環境では、全てのインスタンスに同じ設定をする
- I/O測定情報の収集
  - DBMS\_RESOURCE\_MANAGER.CALIBRATE\_IOプロシージャの実行
- パラレル実行となるSQLの基準
  - PARALLEL\_MIN\_TIME\_THRESHOLD初期化パラメータの設定
    - このパラメータで設定されている値(単位:秒)以上の時間がシリアル実行で必要と判断された場合、自動パラレル実行の対象となる  
デフォルト値:AUTO (10秒)

# In-Memory Parallel Execution による さらなる高速化

11.2 New

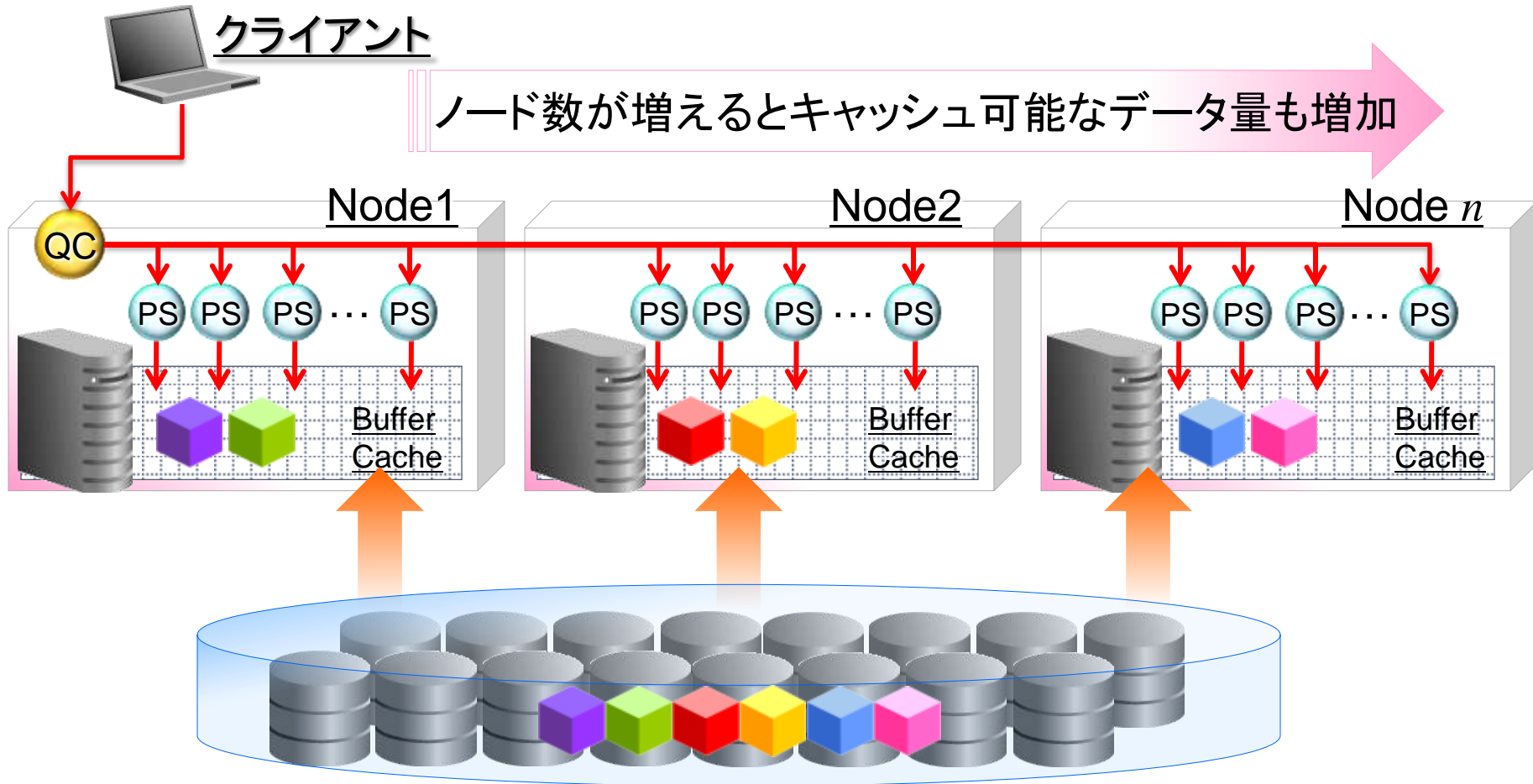
- パラレル実行でもOracleインスタンスにキャッシュされたデータ・ブロックを使用
- 近年のサーバは搭載可能な物理メモリの上限も増加している
- アクセスするデータ量を削減
  - パーティション表
  - 圧縮



ORACLE

# In-Memory Parallel Execution

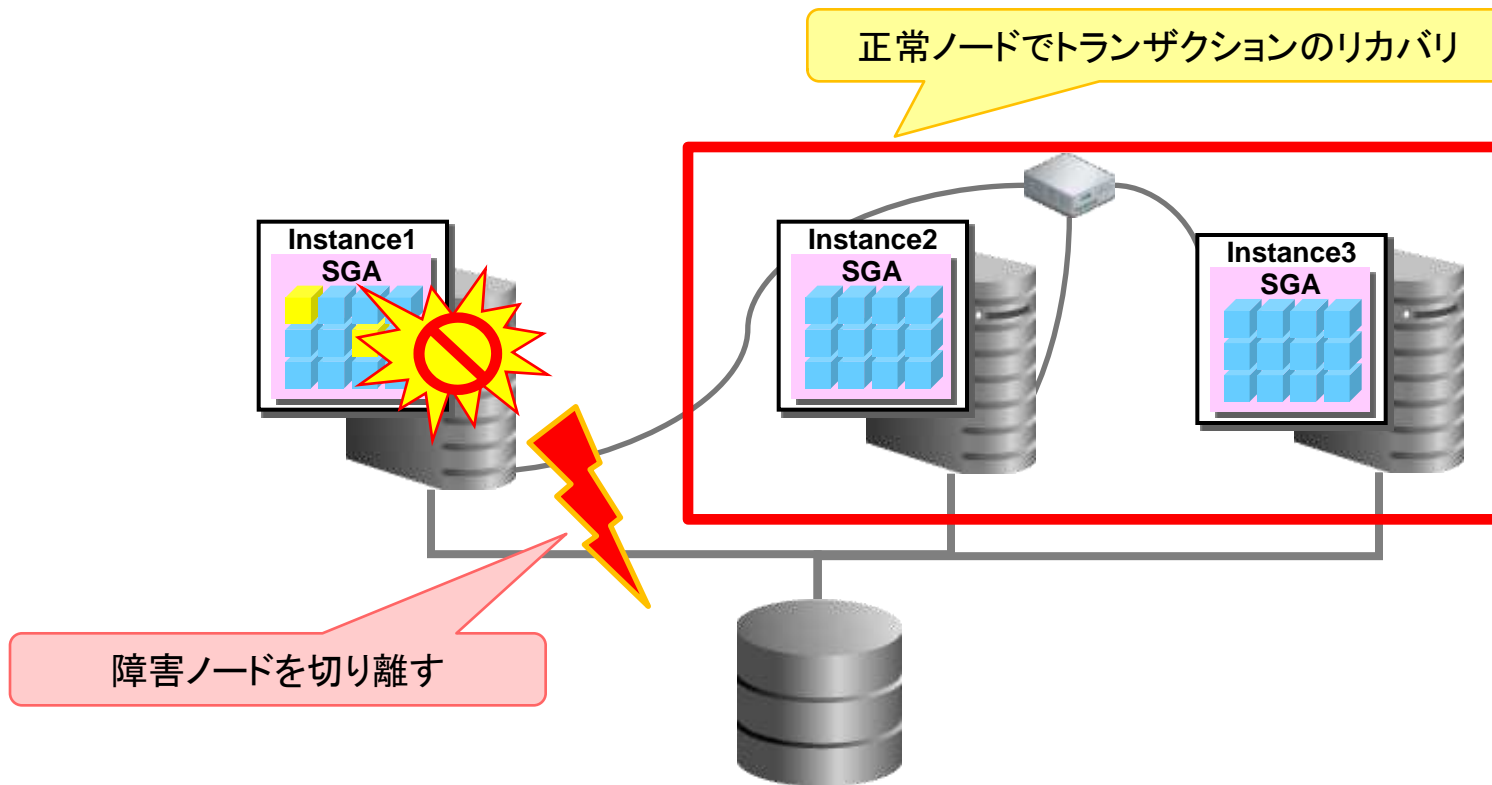
## RACの場合





# RACの可用性

- Oracle Clusterwareによる障害ノードの切り離し
- 正常ノードでトランザクションのリカバリ（インスタンス・リカバリ）





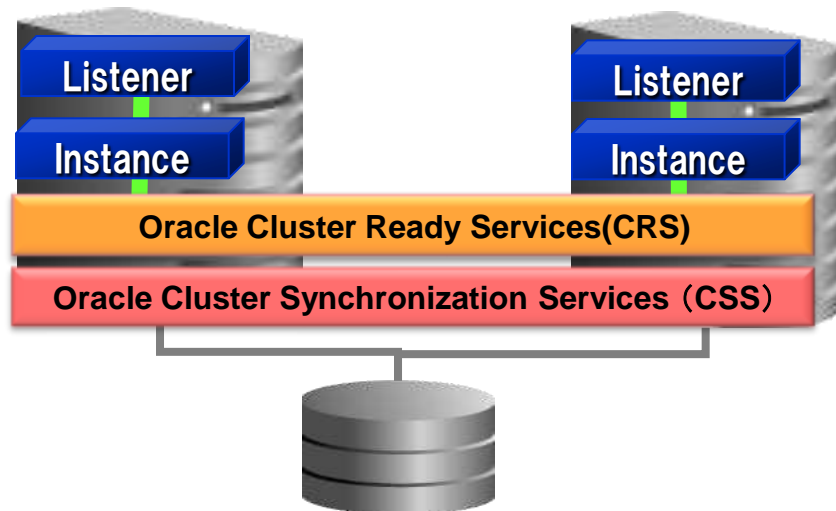
# Oracle Clusterwareの役割

## Oracle Cluster Synchronization Services (CSS)

- クラスタにおけるノード構成の管理
  - ノード間で相互に生存確認
  - 障害ノード切り離し

## Oracle Cluster Ready Services(CRS)

- 監視対象のリソース (CRS リソース) の管理
  - リソースの監視
  - 障害時のリソース再起動
  - 障害時のフェイルオーバー



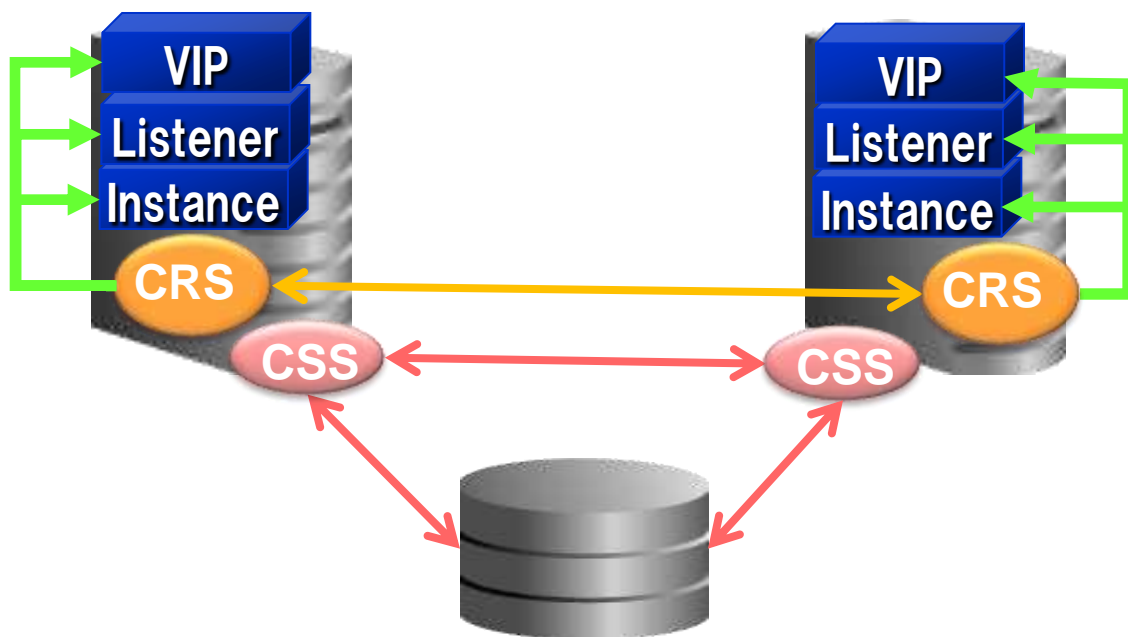
# Oracle Clusterwareによる監視の仕組み

## Cluster Synchronization Services (CSS)

- クラスタ・メンバーシップを管理
- 他ノードの生存を監視
  - 共有ディスクのハートビート
  - インターコネクットのハートビート

## Cluster Ready Services (CRS)

- インスタンス、VIP、リスナー等のリソースを監視
- リソースの停止、起動、フェイルオーバーなどの操作を行う

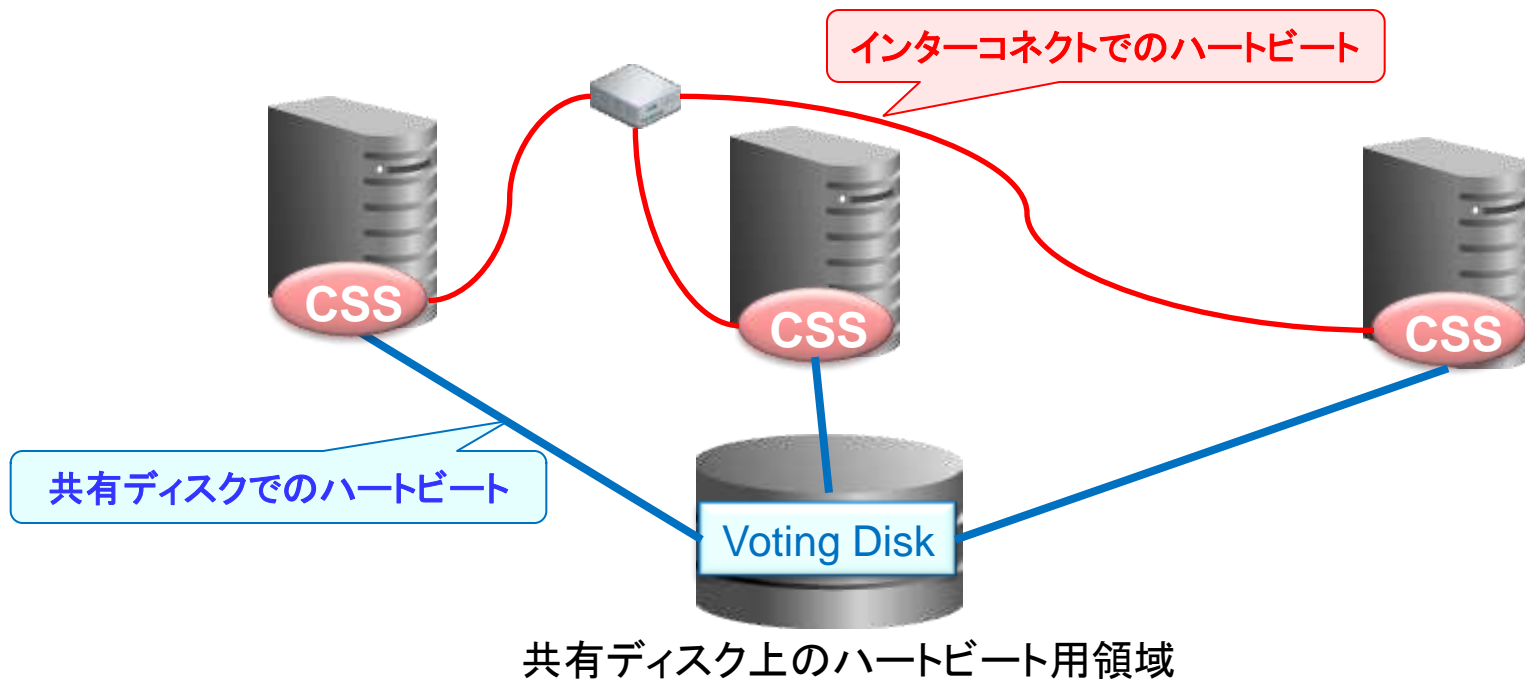


## CRSが管理する主なリソース

- データベース・インスタンス
- リスナー
- サービス
- ASMインスタンス
- 仮想IPアドレス(VIP)
- SCAN IP

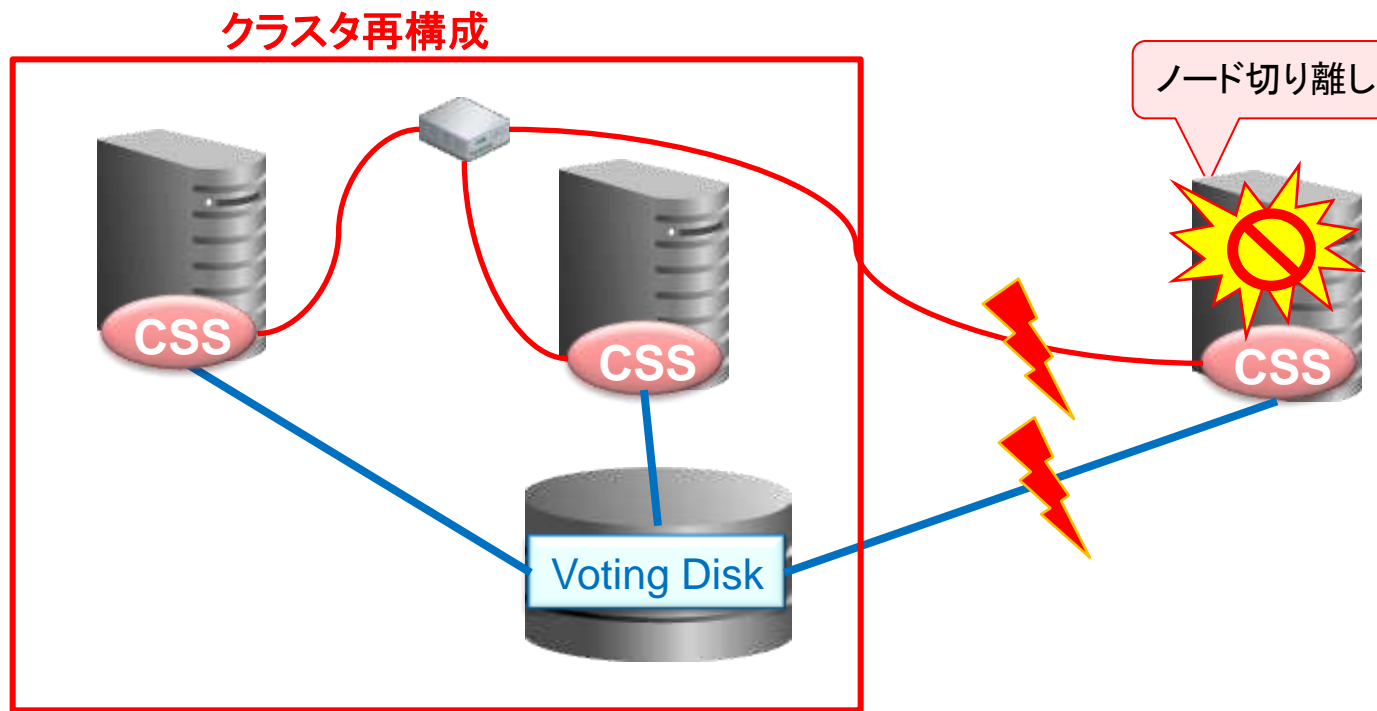
# Cluster Synchronization Services (CSS)

- クラスタ・メンバーシップの管理
  - ハートビートを行い生存確認
  - ハートビートが途絶えると、障害ノードを切り離す



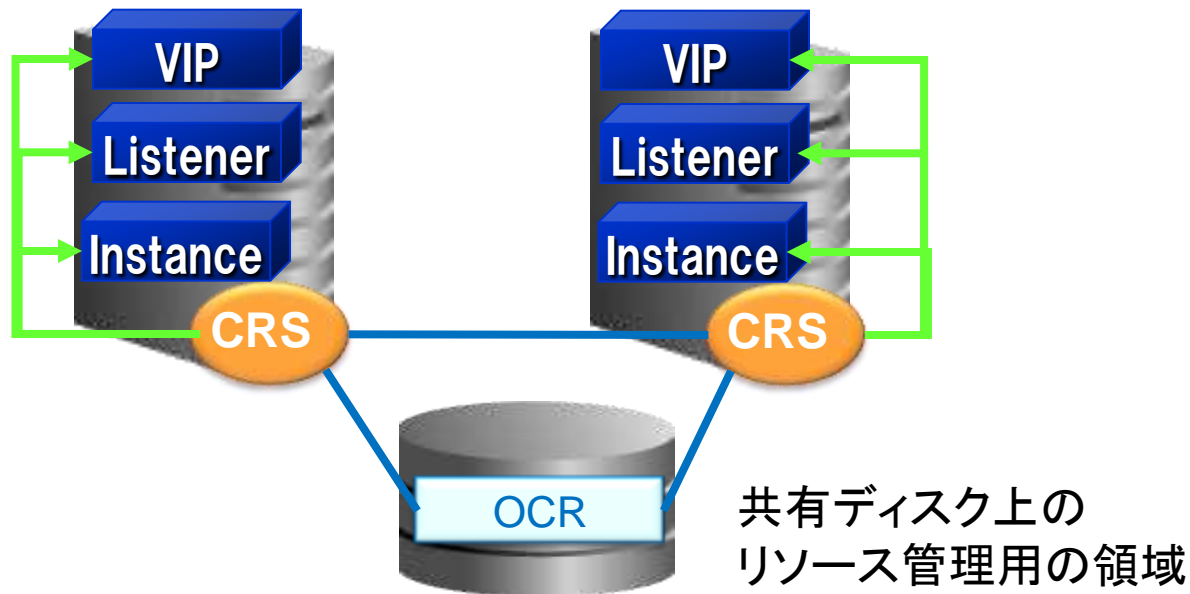
# CSSによるノード障害の検知

- ハートビートが一定時間途切れたら、CSSが障害として検知
- 最もメンバシップが多いグループでクラスタを再構成し、障害グループを切り離す



# Cluster Ready Services (CRS)

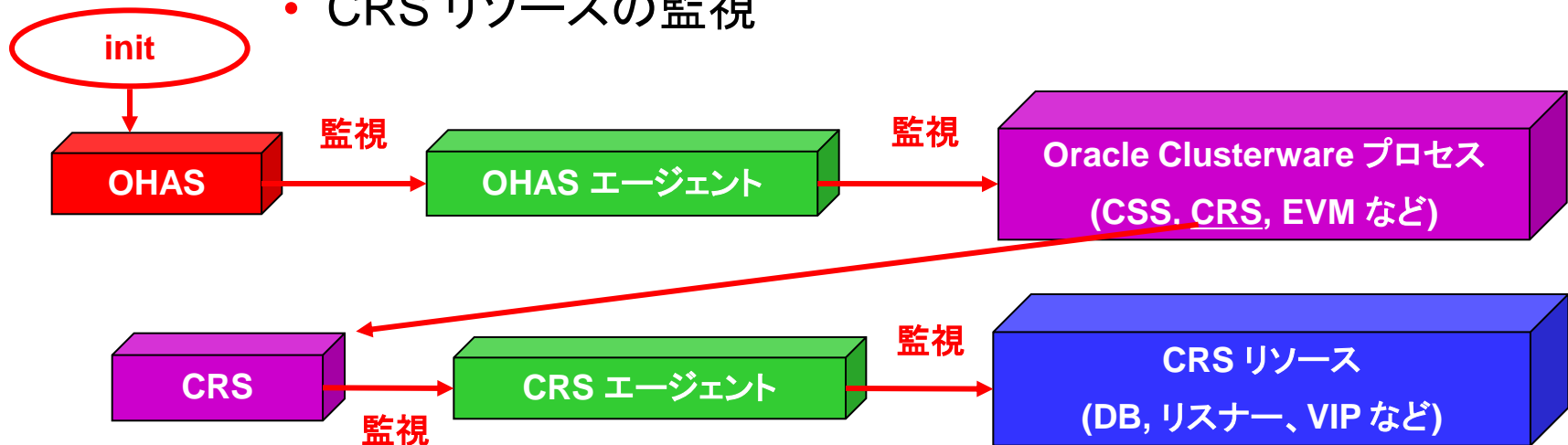
- CRSリソースの起動、停止、監視
  - 障害を検知すると、リソースの再起動又はフェイルオーバーを試行する
  - CRSリソースの構成及びステータスはOracle Cluster Registry(OCR)に記録されている



# Oracle Clusterware によるリソース管理

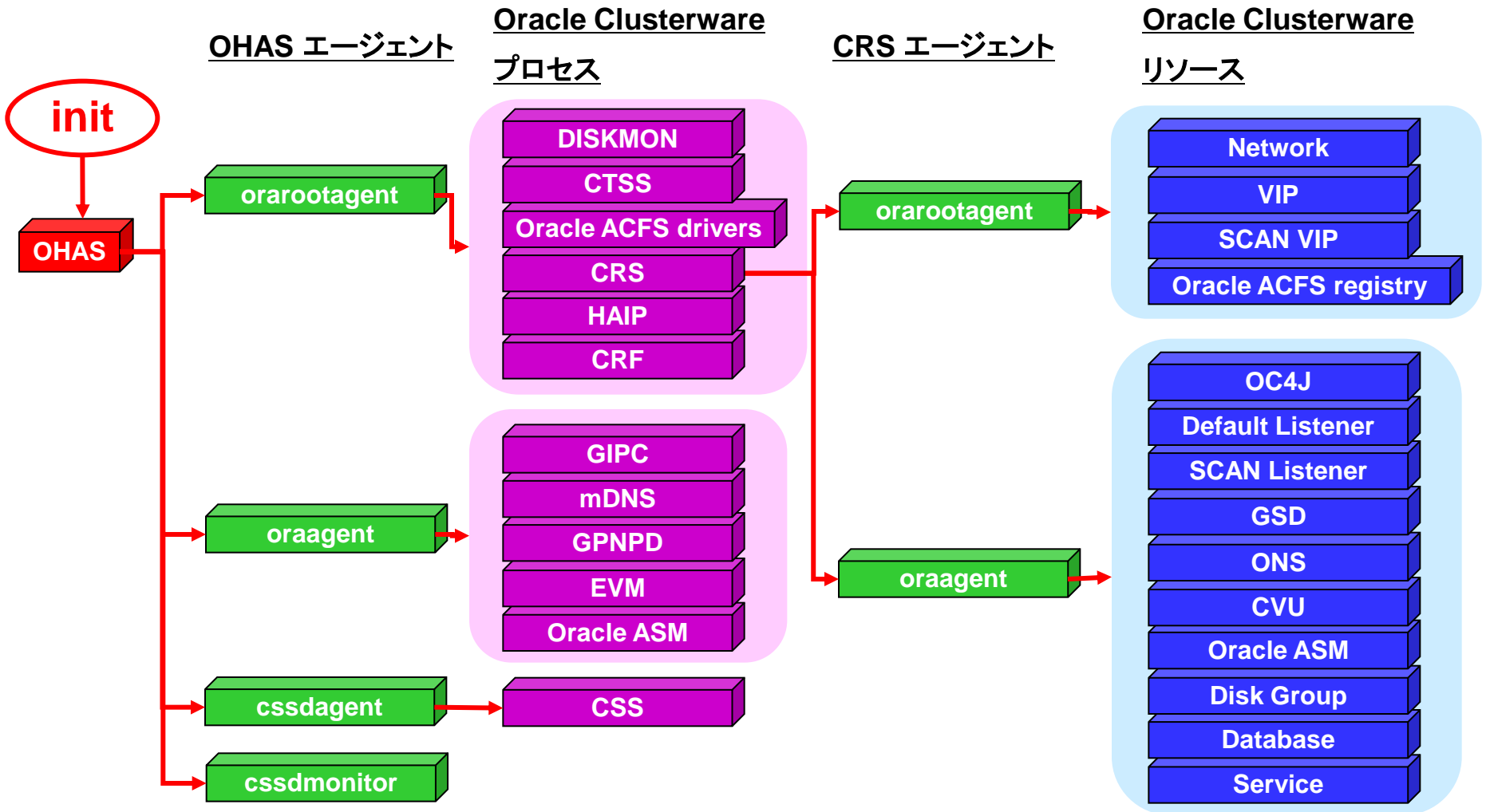
## エージェントによるリソースの監視

- 11g Release 2 では、エージェントが Oracle Clusterware のプロセスおよびリソースを監視
  - エージェントは、高可用性サービスを提供する常駐プロセス
    - OHAS エージェント
      - Oracle Clusterware プロセスの監視
    - CRS エージェント
      - CRS リソースの監視



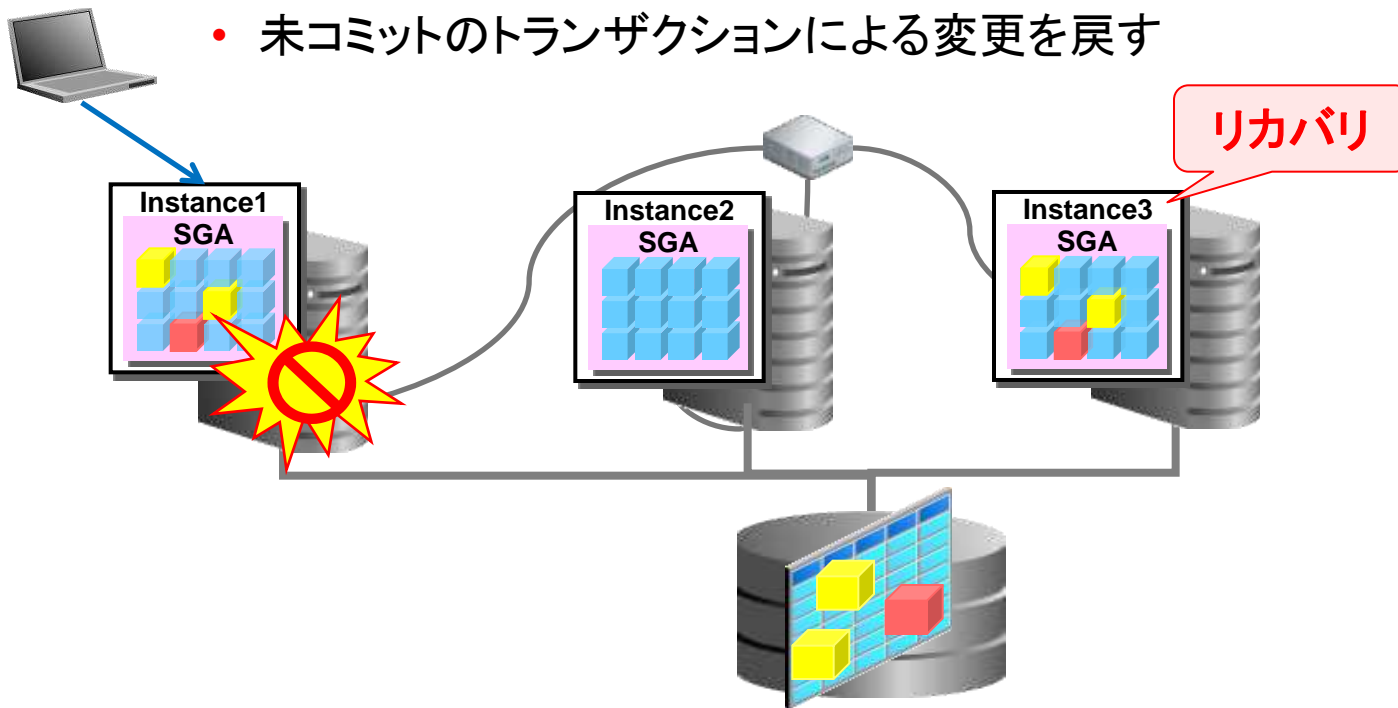
# エージェントおよびプロセスとリソースの一覧

## 11g R2 (11.2.0.3)



# インスタンス・リカバリ

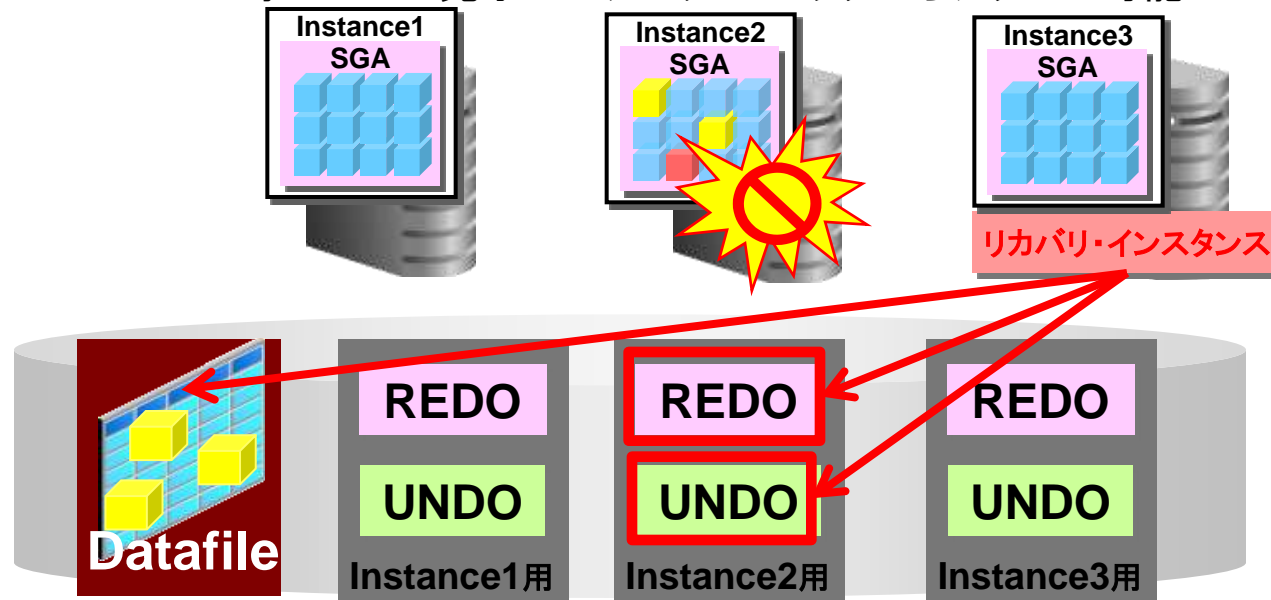
- 障害インスタンスで実行されていたトランザクションのリカバリ
  - ロールフォワード
    - コミット済みのトランザクションによる変更をデータファイルに反映する
  - ロールバック
    - 未コミットのトランザクションによる変更を戻す





# RACのインスタンス・リカバリ

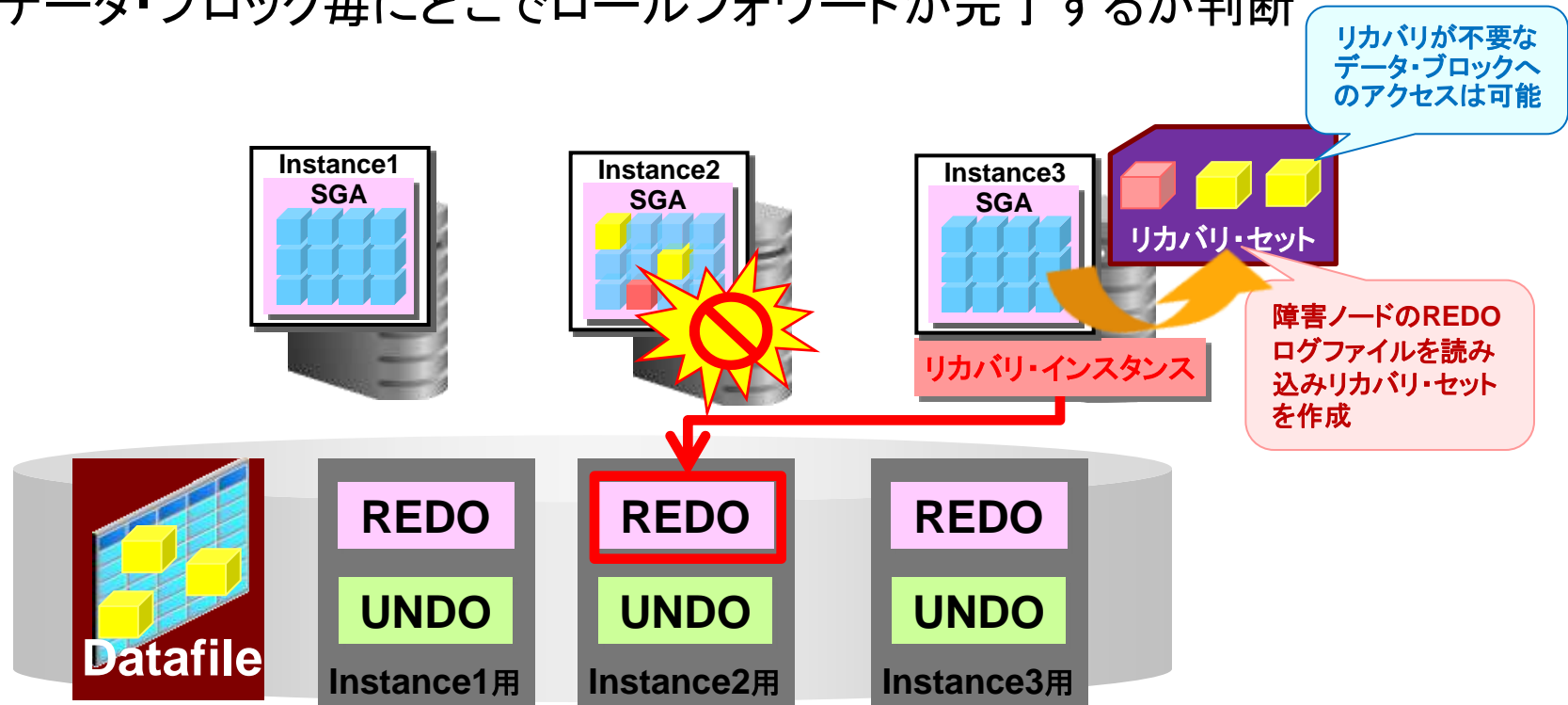
1. GRDを再構成し、リカバリを実行するインスタンスを決定
2. リカバリが必要なデータ・ブロックを識別 (Fast Pass Log Read)
  - データ・ブロック毎にどこでロールフォワードが完了するか判断
  - リカバリ不要なデータ・ブロックはアクセス可能
3. リカバリが必要なデータ・ブロックをリカバリ (Second Pass Log Read)
  - ロールフォワードが完了したデータ・ブロックからアクセス可能



# Fast Pass Log Read

## リカバリが必要なデータ・ブロックを識別

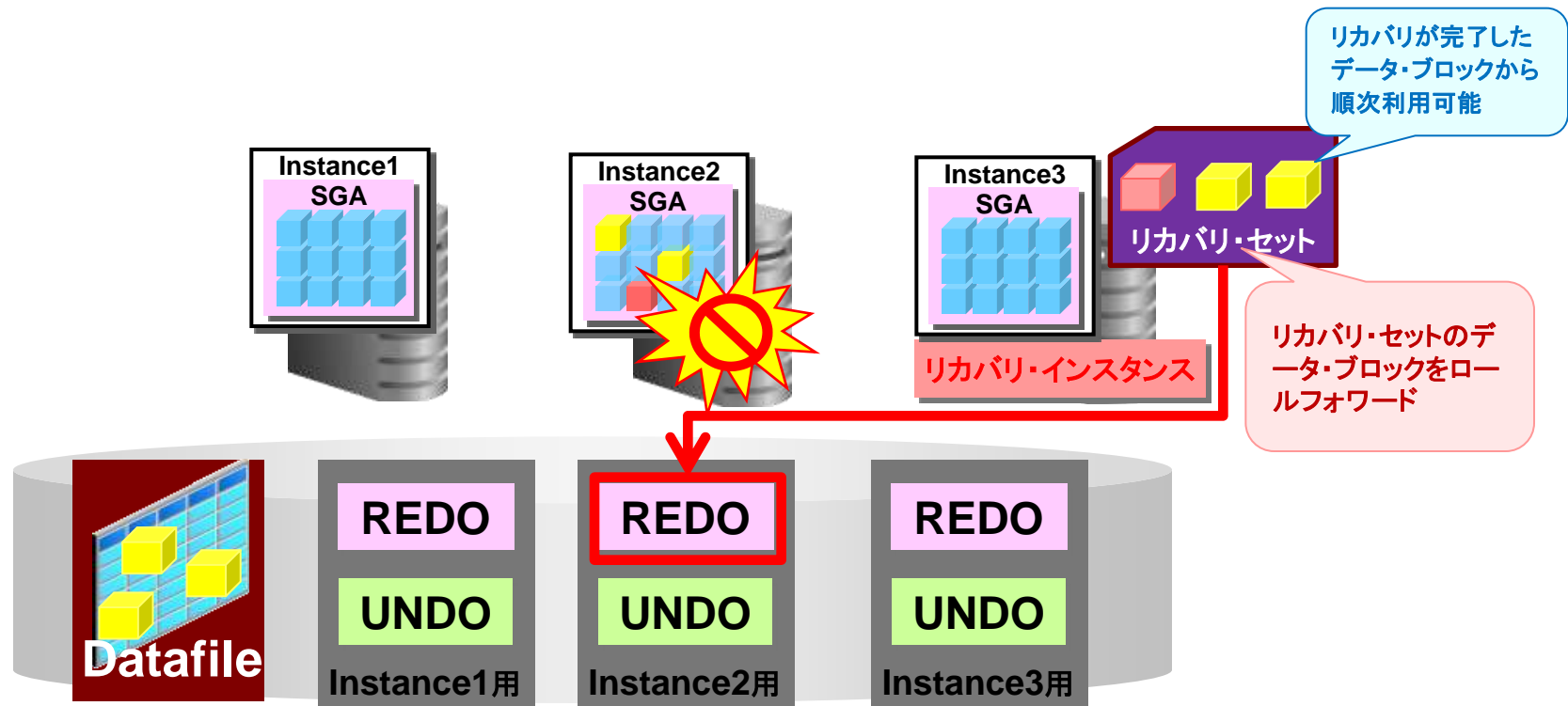
- 障害が起きたインスタンスのREDOログを読み込み、リカバリが必要なブロックのリスト(リカバリ・セット)を作成
- リカバリ不要なデータ・ブロックはアクセス可能になる
- データ・ブロック毎にどこでロールフォワードが完了するか判断



# Second Pass Log Read

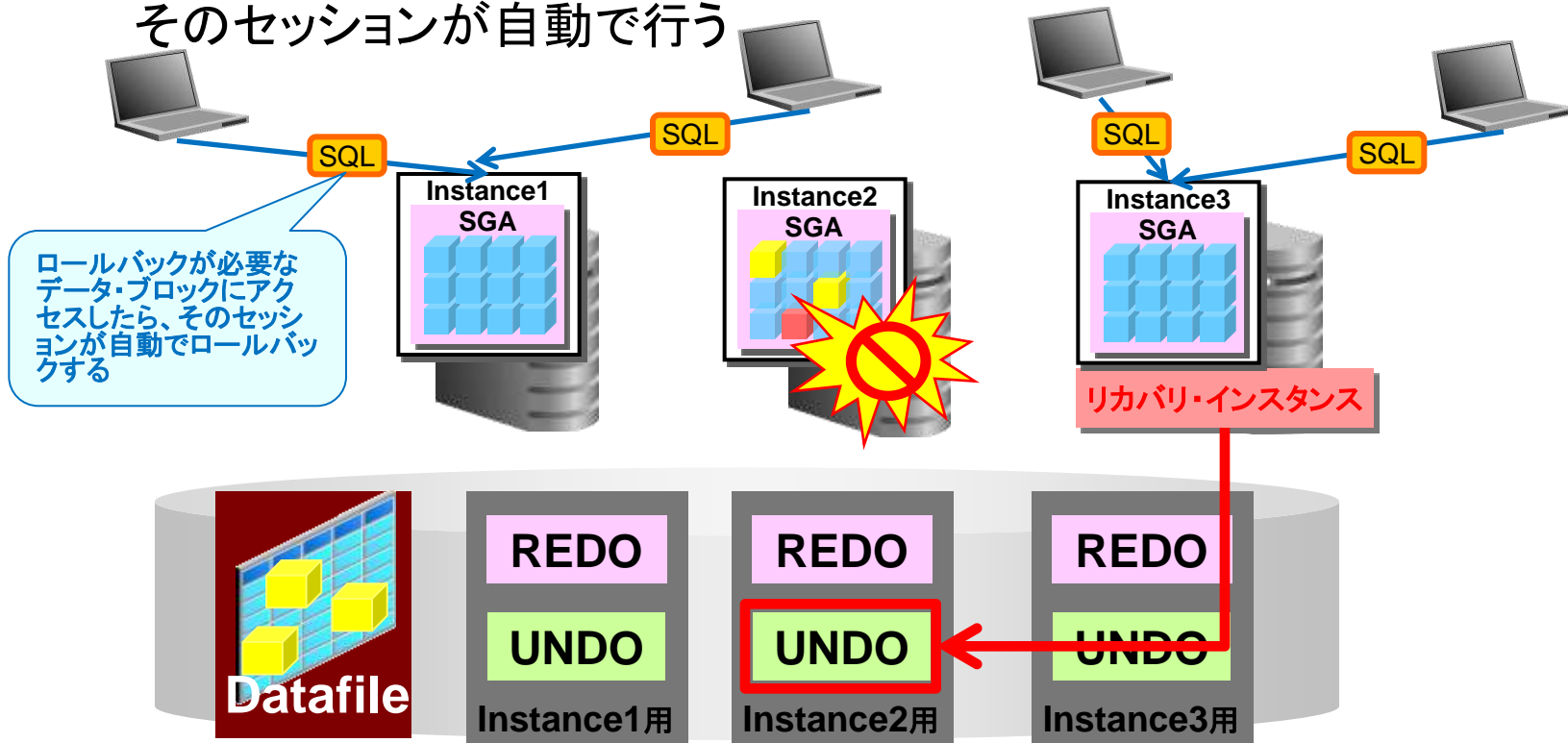
## リカバリが必要なデータ・ブロックのみをロールフォワード

- リカバリ・セットから特定されたデータ・ブロックをロールフォワード
- ロールフォワードが完了したデータ・ブロックから順次、使用可能となる
  - SQLの処理が進行する



# ロールバック

- ロールフォワードが完了したデータ・ブロックから順次、使用可能となる
  - SQLの処理が進行する
- SQLがアクセスするデータ・ブロックにロールバックが必要なら、そのセッションが自動で行う

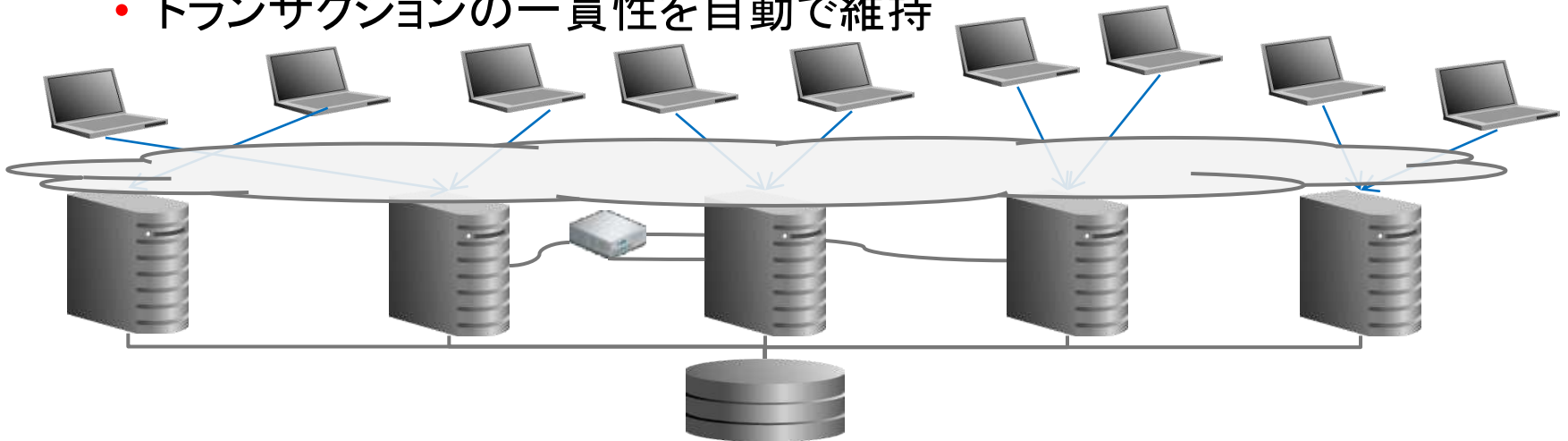




# まとめ

# Oracle Real Application Clusters

- 高拡張性
  - 全てのOracleインスタンスは同じデータベースへアクセスする
  - 同時に実行可能な処理を増やすことができる
  - 1つの処理を並列化することができる
- 高可用性
  - 障害ノードを切り離して、正常インスタンスで処理が可能
  - トランザクションの一貫性を自動で維持



# 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](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



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

アクセスランキング



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

特集テーマ Pick UP



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

技術コラム



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!?

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

オラクルエンジニア通信





The screenshot shows the top section of the oracletech.jp website. On the left is the 'oracletech.jp' logo with the tagline '好奇心が、エンジニア人生を豊かにする。'. On the right is the 'ORACLE' logo, a search bar, and social media icons for Twitter, Facebook, Ustream, 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](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®**