



**ENGINEERED
FOR INNOVATION**



ORACLE®

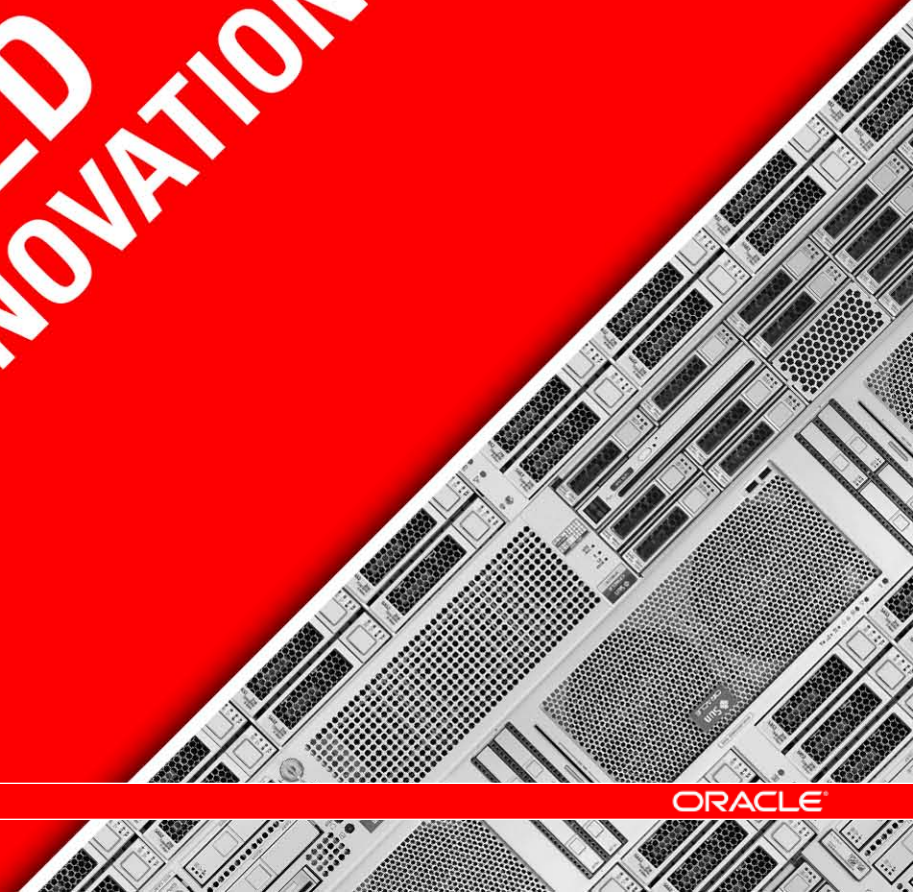
**非構造化データの世界と構造化データの世界を繋ぐ！
－ ビッグデータのためのオラクル製品と技術 －**

製品戦略統括本部 戦略製品ソリューション本部

下道 高志



**ENGINEERED
FOR INNOVATION**



ORACLE DEVELOP

Russia

17–18 April 2012

India

3–4 May 2012



San Francisco

September 30–October 4, 2012

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。

Big Dataの捉え方



Hadoop を有名にした代表的な事例

- New York Times 社の”TimesMachine”

- 目的

- 1851-1922年発行の新聞イメージのオンラインサービス化
 - 4TBのTIFFイメージデータを1,100万ページのPDFに変換

- Amazon EC2 を100インスタンス利用

- Hadoop “アプリケーション”によって24時間で処理

- 特徴

- (2007年当時としては)巨大なサイズの非定型データが対象
 - 「PDFへ変換」という全く同じ処理を72年分繰り返したこと
 - 他のデータとのマッチングや条件検索等は殆ど無し

米政府が2億ドルの”BIG DATA”イニシアチブを開始



Office of Science and Technology Policy
Executive Office of the President
New Executive Office Building
Washington, DC 20502

FOR IMMEDIATE RELEASE

March 29, 2012

Contact: Rick Weiss 202 456-6037 rweiss@ostp.eop.gov
Lisa-Joy Zgorski 703 292-8311 lisajoy@nsf.gov

OBAMA ADMINISTRATION UNVEILS “BIG DATA” INITIATIVE: ANNOUNCES \$200 MILLION IN NEW R&D INVESTMENTS

Aiming to make the most of the fast-growing volume of digital data, the Obama Administration today announced a “Big Data Research and Development Initiative.” By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some of the Nation’s most pressing challenges.

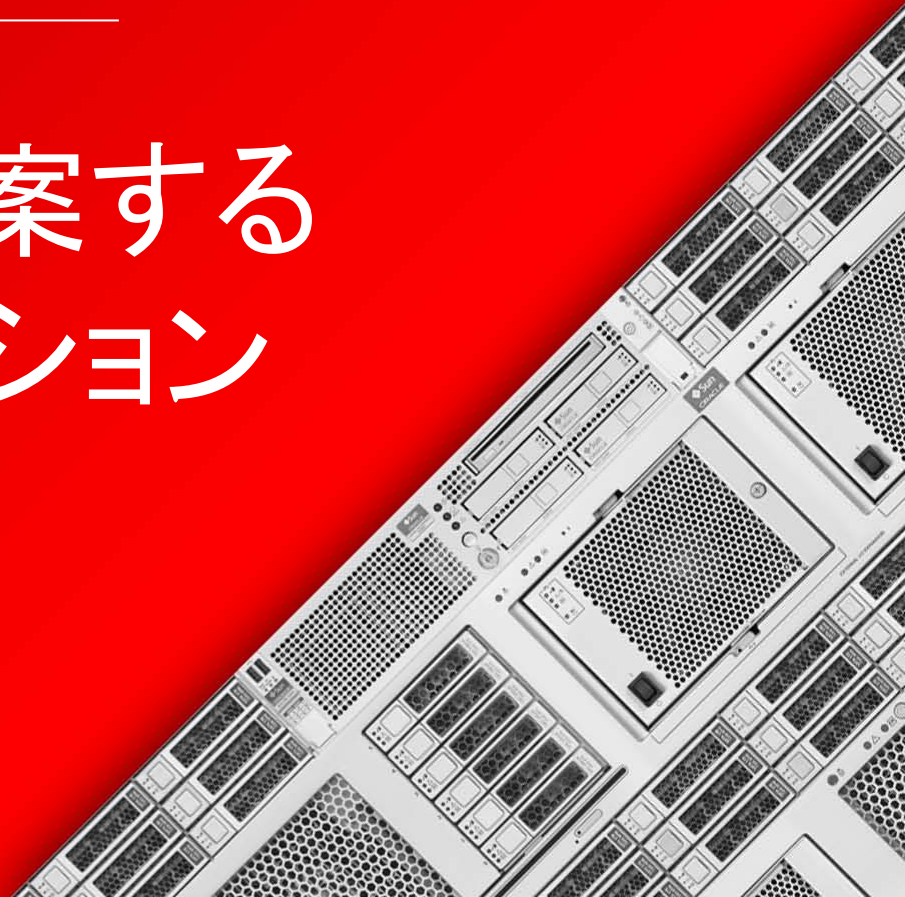
センサー機器、情報端末の急増が引き起こす情報爆発



Big Dataの特徴

バッチ指向	リアルタイム
“使う”ためにデータを加工	サービスを提供
大容量のストレージ	特定のデータへ高速アクセス
Write once, read all	Read, write, delete update

オラクルが御提案する Big Dataソリューション



オラクルが考える “Big Data の3大要件”

“エンタープライズ用途”に求められる要素

- ✓ エンタープライズ・クラスのセキュリティ
- ✓ 高可用性
- ✓ ソフトウェアとハードウェアを融合して提供する超高性能

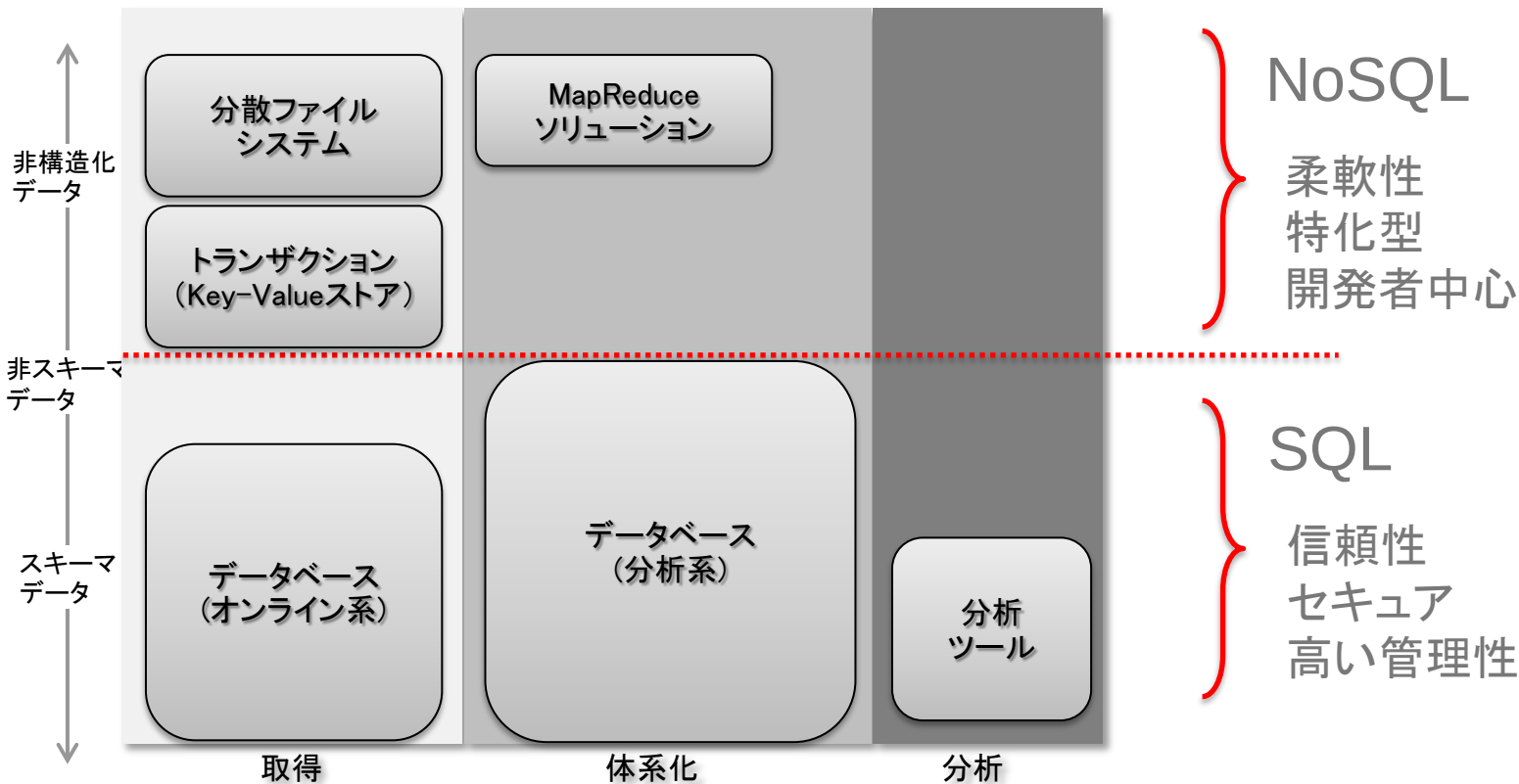
莫大な非構造化データの取り扱い、構造化データと非構造化データを体系化

- ✓ 非構造化データ格納・管理: “分散ファイルシステム”
- ✓ 非構造化データと構造化データを繋ぐ: “Connectors”
- ✓ ビッグデータに対する高度な分析機能、直観的なインターフェース

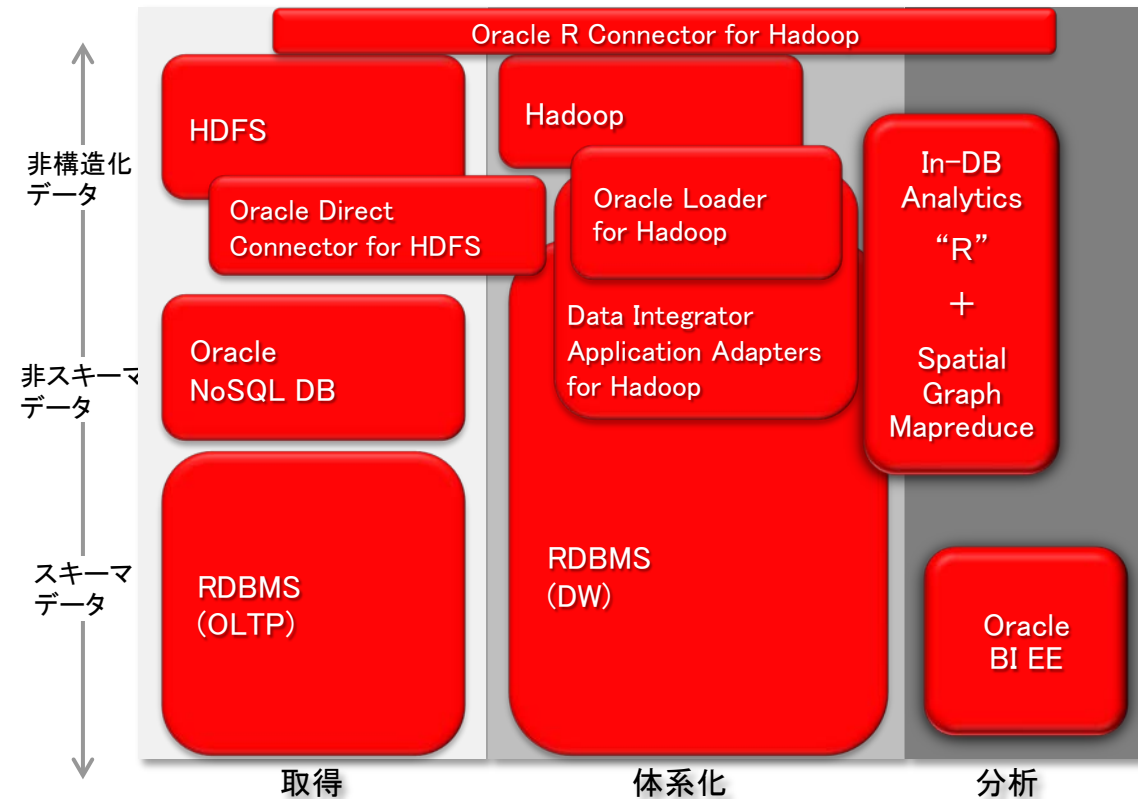
新たに広まりつつある、“良いテクノロジー”の利活用

- ✓ 非構造化分散ファイルシステム
- ✓ Key Value型データベース
- ✓ 統計解析

扱う情報の種類により求められる技術、スキルが異なる



あらゆるデータを網羅的に活用するための統合情報基盤



Hadoop、HDFS

- ✓ 非構造化データの蓄積、処理
- ✓ 開発・運用・管理に実績あるCloudera

Oracle NoSQL DB

- ✓ 準構造化データの蓄積
- ✓ 高速処理、高拡張性

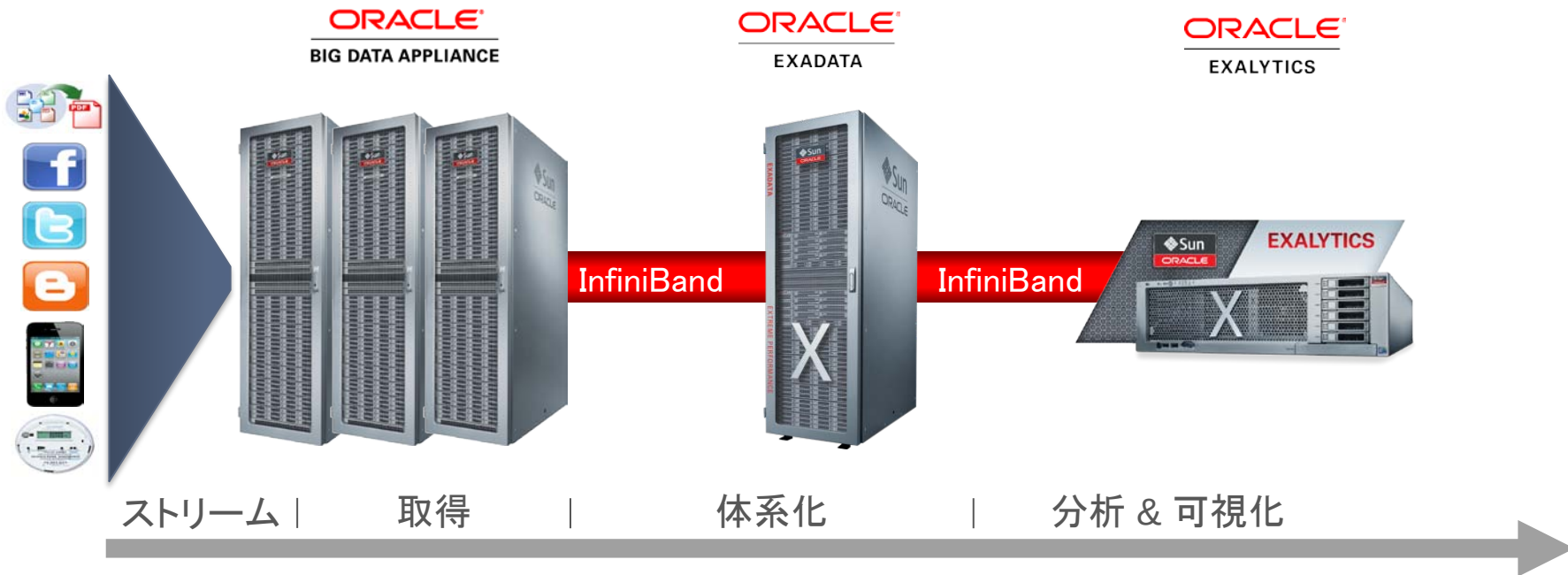
Oracle Big Data Connectors

- ✓ 非構造化データからデータベースへの連携
- ✓ 高速データローディング
- ✓ データベースから非構造化データへのアクセス

In-DB Analytics

- ✓ オープン言語“R”の採用
- ✓ 大量データを用いた統計解析、マイニング

あらゆるデータを網羅的に活用するための統合情報基盤

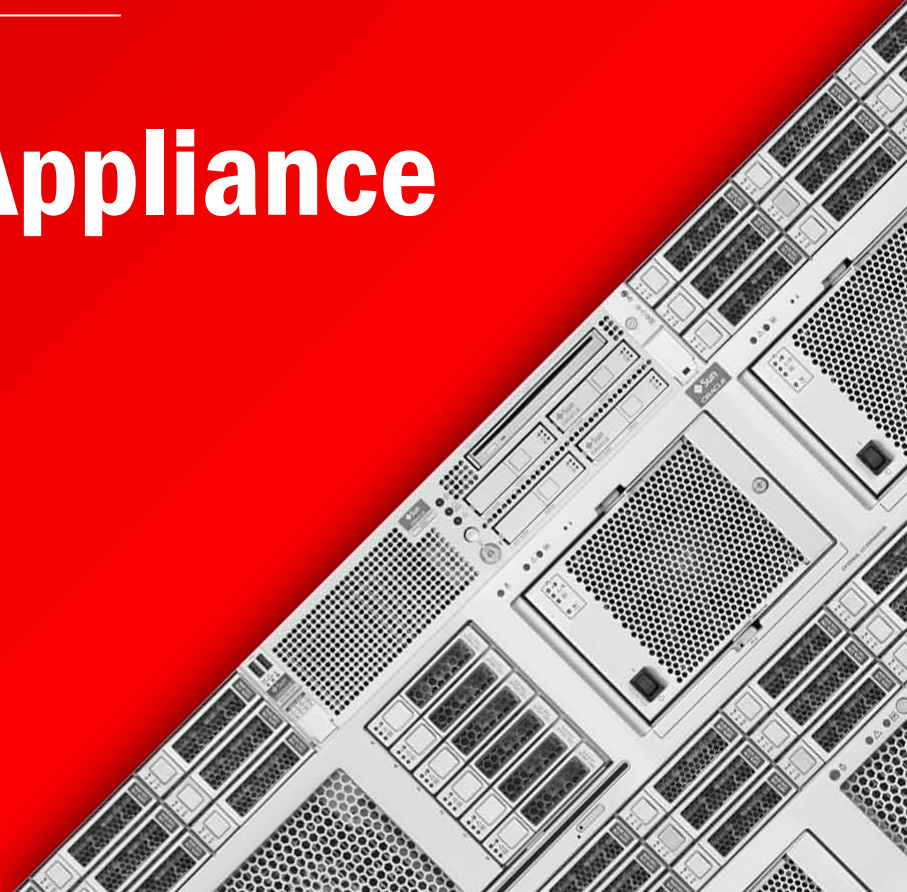


例えば、従来の方法でHadoop システムを構築するとすると



- ラックを組み上げる時間は？
- ソフトウェアを最適に構成するのにかかる時間は？
- コストと難易度はどの位のものなのか？

Oracle Big Data Appliance (BDA)



Oracle Big Data Appliance ハードウェア

- 18ノードのSun X4270 M2 Servers
 - 864 GB メモリ
 - 216 コア
 - 648 TB ストレージ
- 40 Gb/s InfiniBand Fabric
 - ラック間／ノード間
- 10 Gb/s Ethernet
 - データセンターのネットワークへ接続



ノード – Sun Fire X4270 M2



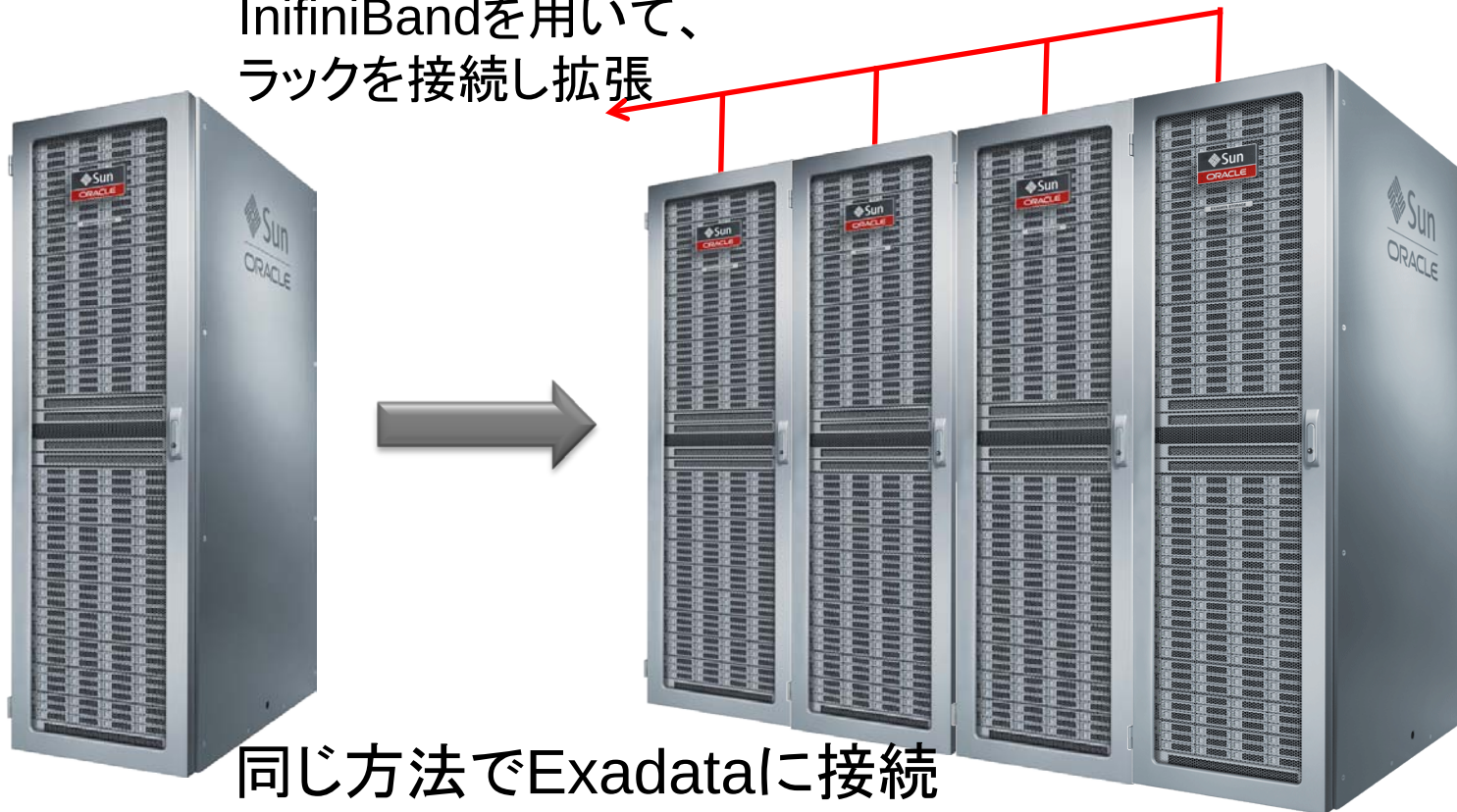
プロセッサ	2 Six-Core Intel® Xeon® X5675 プロセッサ (3.06 GHz)
メモリ	48GB (6 * 8GB), 96 GB または 144GBに拡張可能
ディスク	12 x 3 TB 7.2K RPM 高容量 SAS (hot-swap)
ディスク・コントローラ	512MBバッテリー・バックド・キャッシュ, HBA
ネットワーク	2 InfiniBand 4xQDR (40Gb/s) ポート(1 Dual-port PCIe 2.0HCA) 4 Gigabit Ethernet Ports内臓
リモート管理	1 Gigabit Ethernet portを備えたILOM
電源	冗長構成のホット・スワップ可能な電源とファン

メモリ拡張キット

- 特定のノードがより多くのメモリを必要とする場合
 - 48GBから96Gb (12 x 8GB)または144GB (18 x 8GB) に変更可能
- 例
 - クラスター内の全ノード
 - 一部のノード(Name Node, HBase master etc.)
- 注意点
 - 144GBへのアップグレードは、メモリ帯域削減のため、若干のパフォーマンス減退につながる可能性あり(クロック: 1333MHz→800MHz)

水平型拡張モデル(スケールアウト・モデル)

InfiniBandを用いて、
ラックを接続し拡張



同じ方法でExadataに接続

ORACLE

Oracle Big Data Appliance ソフトウェア

最高のパフォーマンスを出せるように
調整済みのソフトウェアがインストール済み

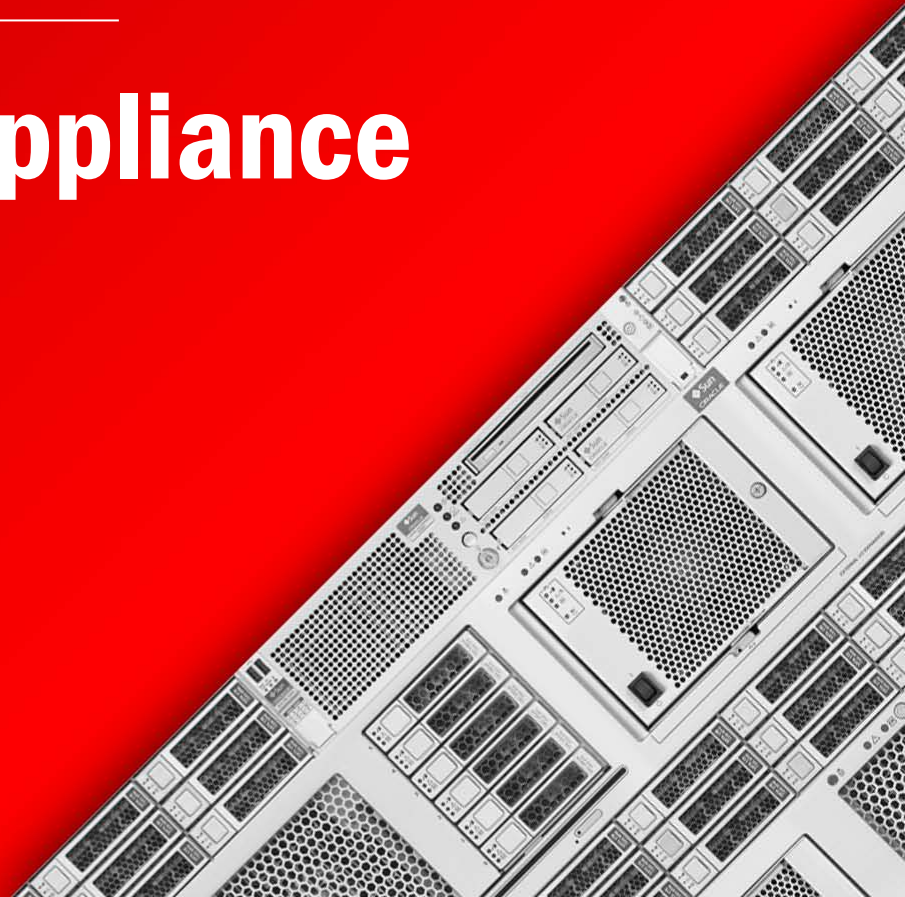
- Oracle Linux 5.6
- Java Hotspot VM
- Cloudera CDH
- Cloudera Manager
- Open Source R Distribution
- Oracle NoSQL Database
- Oracle Big Data Connectors



(注意)MySQL Standard Edition

- 管理用途のみの利用
 - Cloudera Manager
 - Hive のメタストア
 - ODI Agent

Hadoop@Big Data Appliance



HadoopはClouderaディストリビューションを採用

Cloudera's Distribution Including Apache Hadoop (CDH)

Oracle Selects Cloudera to Provide Apache Hadoop Distribution and Tools for Oracle Big Data Appliance

REDWOOD SHORES, Calif. and PALO ALTO, Calif. - January 10, 2012

News Facts

- Oracle and Cloudera, a leader in Apache Hadoop-based software and services for the enterprise, today announced a joint agreement to provide an Apache Hadoop distribution and tools for Oracle Big Data Appliance.
- Oracle has integrated Cloudera's Distribution Including Apache Hadoop (CDH) and Cloudera Manager software into Oracle Big Data Appliance, an engineered system designed to provide high performance and scalable data processing environment for Big Data.

Cloudera ディストリビューションを選択した理由

- 重要機能を迅速に進化
 - コミュニティにおけるHadoop専門家が構築
 - 難解ではなく、実践的
 - 大型クラスターで必要なものにフォーカス
- 大規模環境で実証済み
 - Hadoopを使用する大規模な本番環境で使用
 - これらの環境において、極めて安定的に稼働
- Clouderaで管理およびテスト
 - オープン・ソースでありながらコンポーネントをきちんと管理
 - リッチな管理GUIツールを包含

CDHに梱包されるコンポーネント

ディストリビューションの詳細

- Apache Hadoop
- Apache Hive
- Apache Pig
- Apache HBase
- Apache Zookeeper
- Apache Flume
- Apache Sqoop
- Apache Mahout
- Apache Whirr
- Apache Oozie
- Fuse-DFS
- Hue

最新の詳細はこちらをご参照ください: <http://www.cloudera.com/hadoop-details/>

CDH のバージョン

- CDH3

- 最新のものインストールされる
- BDA出荷開始時点(2012/1/10)ではupdate 2
- 現時点(2012/4/6)では、CDH3 update 3
 - HDFSパフォーマンスの大幅改善
 - MapReduce TaskTrackerのDisk障害に対する耐性の向上

- CDH4

- 現在ベータ版
- BDAへの適用は現時点では時期未定

Hadoop ソフトウェア・レイアウト (標準)



- ノード1:
 - Name Node, Balancer , HBase Master*
- ノード2:
 - Secondary Name Node, Cloudera Manager, Zookeeper*, NoSQL Database Administration
- ノード3:
 - JobTracker, HUE, MySQL Master, ODI Agent, Hive
- その他のノードもしくは追加BDA
 - HDFS Data Node, NoSQL DB Storage Node*
 - (ノード4)
 - Name Node data backup

*構成上のオプション

Cloudera Manager の管理画面

cloudera manager Services ▾ Hosts Activities Logs Events Reports Search Welcome admin ⚙ Help ▾ Logout

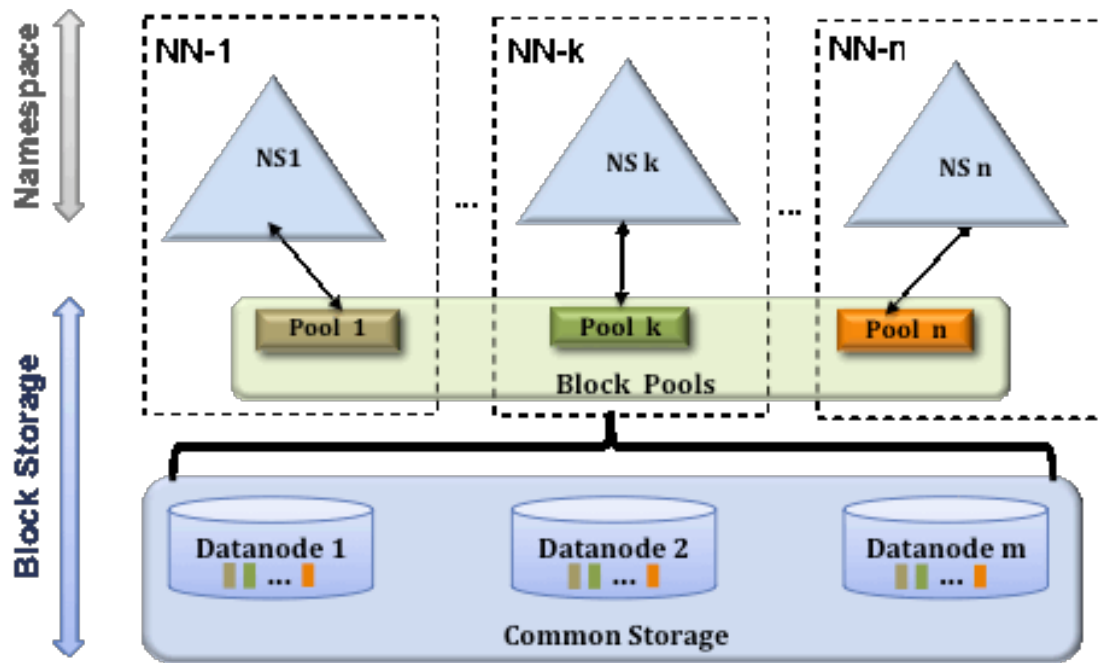
Jan 27 11:50 AM Jan 27 12:00 PM Jan 27 12:10 PM Jan 27 12:20 PM Jan 27 12:30 PM Jan 27 12:40 PM Jan 27 12:50 PM Jan 27 1:00 PM

Services (Current) ◆ Switch to Current

Add a Service Generate Client Configuration

Name	Type	Status	Health	Role Counts	
hdfs1	HDFS	✓ Started	✓ Good	1 Secondary NameNode , 1 NameNode , 1 Balancer , 7 DataNodes	Actions ▾
hue1	Hue	✓ Started	✓ Good	1 Beeswax Server , 1 Hue Server , 1 Job Designer	Actions ▾
mapreduce1	MapReduce	✓ Started	✓ Good	1 JobTracker , 4 TaskTrackers	Actions ▾
mgmt1	Cloudera Management Services	✓ Started	✓ Good	1 Event Server , 1 Activity Monitor , 1 Alert Publisher , 1 Service Monitor , 1 Resource Manager	Actions ▾

次世代HDFS:HDFS Federation とデータブロックのプール化



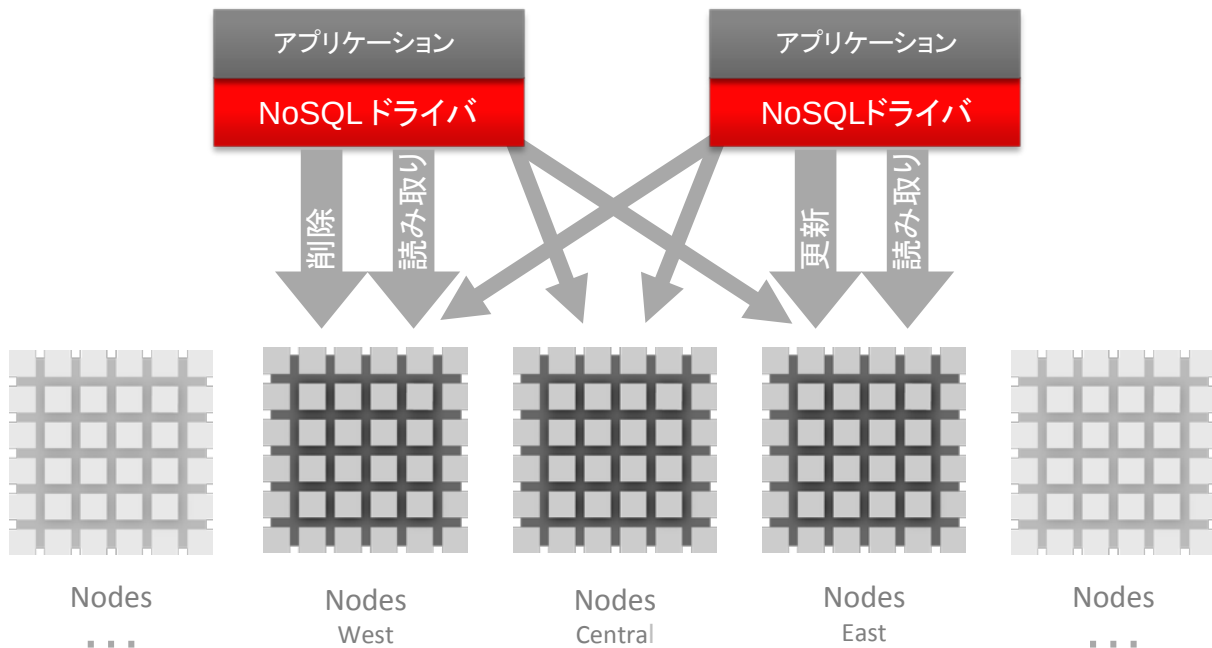
広帯域のネットワークが益々必要

<http://hadoop.apache.org/common/docs/r0.23.0/hadoop-yarn/hadoop-yarn-site/Federation.html>

Oracle NoSQL Database



Oracle NoSQL Database



キーバリュ型データベース

データモデルの容易な変更

高い拡張性と可用性

透過的な負荷分散

BerkeleyDBを元に開発

Oracle NoSQL DBのスケラビリティ

処理に使用するノードの数

24 Node Cluster

48 Node Cluster

96 Node Cluster

処理スピード＝一定
メモリ: 16ms

処理に使用するノードの数
15:28:55 (ほぼ、1:2:4)

Scalability Factor:

94%

92%

ノードの数に応じて大量処理が可能

管理ツール及びログファイル

- GUI/CUIでの管理ツール
- 各種APIでの管理／統計情報の取得
- 各種ログファイル: ログレベル指定可
- 様々なチューニング・パラメータ

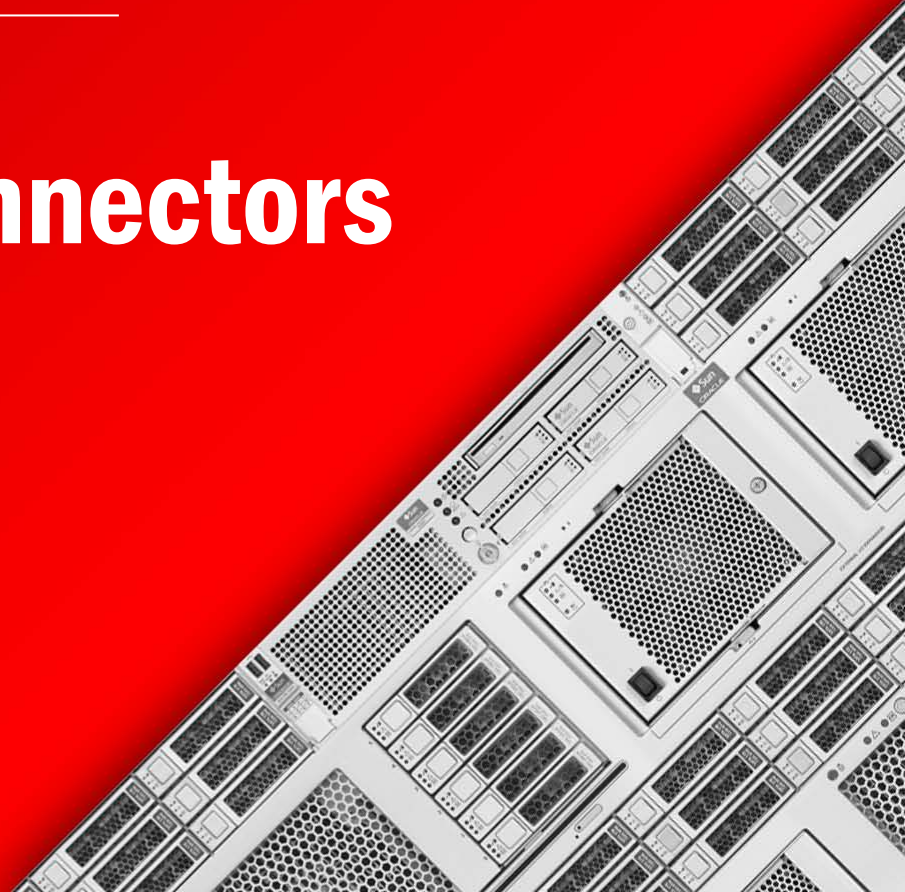
The screenshot shows the Oracle NoSQL Database management interface. The left pane displays the 'Topology Browser' with a tree view of the database topology, including nodes like 'SN 1 kinichi-VirtualBox:5000' and 'SN 2 kinichi-VirtualBox:5000'. The right pane shows the 'RepNode group 1, node 1' configuration details, including properties like 'storageNodeid', 'config/properties', 'nodeHostPort', 'helperHosts', 'cacheSize', 'javaMiscParams', 'loggingConfigProps', 'requestQuiesceTime', 'mMaxActiveRequests', 'mNodeLimitPercent', 'mRequestThresholdPercent', 'mEnvConsistencyLag', 'maxTrackedLatency', 'statsInterval', 'threadDumpInterval', and 'activeThreshold'.

Resource	Time	Op Type	Interval							Cumulative						
			TotalOps	PerSec	Min	Max	Avg	95th	99th	TotalOps	PerSec	Min	Max	Avg	95th	99th
rg1-rn1	11/17/11 14:19:50	single	932	93	3	75	6.0	12	18	932	20	3	75	6.0	12	18
rg1-rn1	11/17/11 14:20:00	single	2112	211	2	51	3.8	7	13	3049	54	2	75	4.5	9	18
rg1-rn1	11/17/11 14:20:10	single	2146	214	1	41	3.9	7	16	5195	78	1	75	4.3	9	16
rg1-rn1	11/17/11 14:20:20	single	2105	210	1	41	4.0	7	16	7300	96	1	75	4.2	8	16
rg1-rn1	11/17/11 14:20:30	single	2593	259	2	19	3.2	4	7	9893	115	1	75	3.9	8	14
rg1-rn1	11/17/11 14:20:40	single	2556	255	1	27	3.3	5	8	12449	129	1	75	3.8	7	13
rg1-rn1	11/17/11 14:20:50	single	2638	263	1	17	3.2	4	7	15087	142	1	75	3.7	7	12
rg1-rn1	11/17/11 14:21:00	single	2515	251	1	15	3.3	5	7	17602	151	1	75	3.6	6	11
rg1-rn1	11/17/11 14:21:10	single	2559	255	1	16	3.2	5	7	20161	160	1	75	3.6	6	11
rg1-rn1	11/17/11 14:21:20	single	2128	212	1	25	3.9	7	12	22289	163	1	75	3.6	6	11
rg1-rn1	11/17/11 14:21:30	single	2612	261	2	14	3.2	4	7	24901	170	1	75	3.6	6	11
rg1-rn1	11/17/11 14:21:40	single	2640	264	1	26	3.2	4	7	27541	176	1	75	3.5	6	10
rg1-rn1	11/17/11 14:21:50	single	2653	265	2	14	3.1	4	7	30194	181	1	75	3.5	6	10
rg1-rn1	11/17/11 14:22:00	single	2669	266	1	17	3.1	4	7	32863	186	1	75	3.5	6	10
rg1-rn1	11/17/11 14:22:10	single	2604	260	2	28	3.2	4	7	35467	190	1	75	3.4	6	10

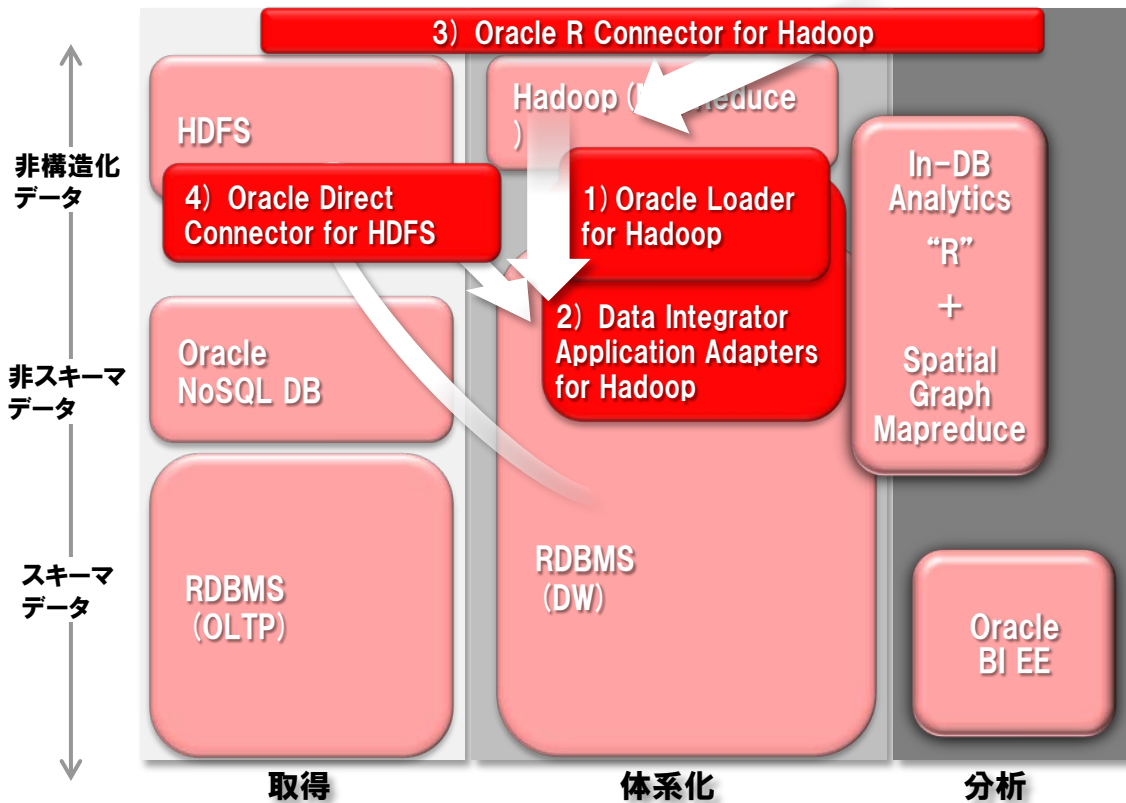
HDFSとOracle NoSQL Databaseの比較

Hadoop Distributed File System (HDFS)	Oracle NoSQL Database
ファイルシステム	データベース
並列検索	インデックス
特定の構造なし	シンプルなデータ構造
大量の書き込み	大量のランダムな読み取りと書き込み
バッチ指向	リアルタイム

Oracle Big Data Connectors



Oracle Big Data Connectors



1) Oracle Loader for Hadoop

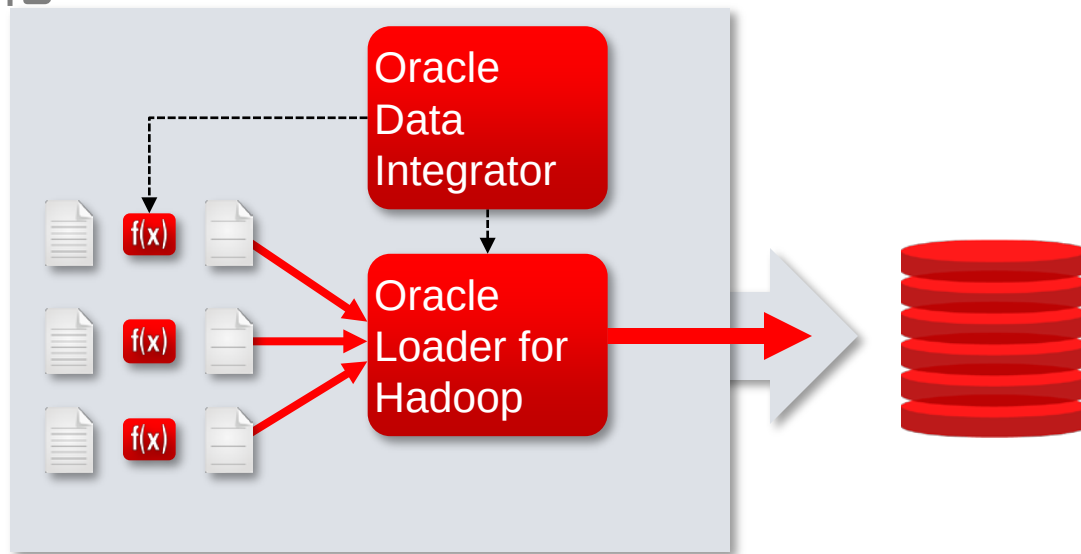
2) Oracle Data Integrator Application Adapters for Hadoop

3) Oracle R Connector for Hadoop

4) Oracle Direct Connector for HDFS

Oracle Data Integrator

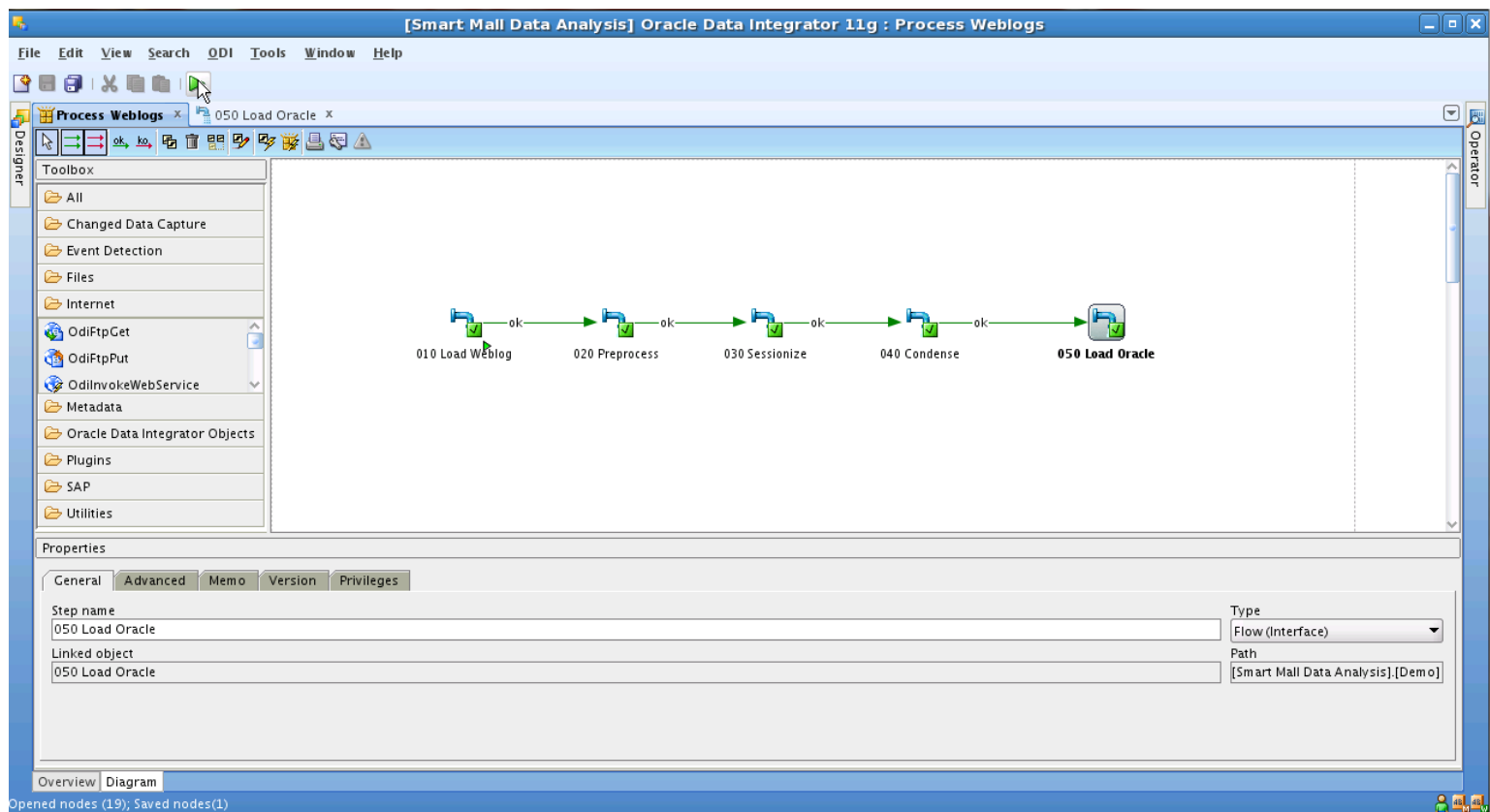
Oracle Loader for Hadoopと一体になりMapReduceコードを簡素化



MapReduceコードを自動生成

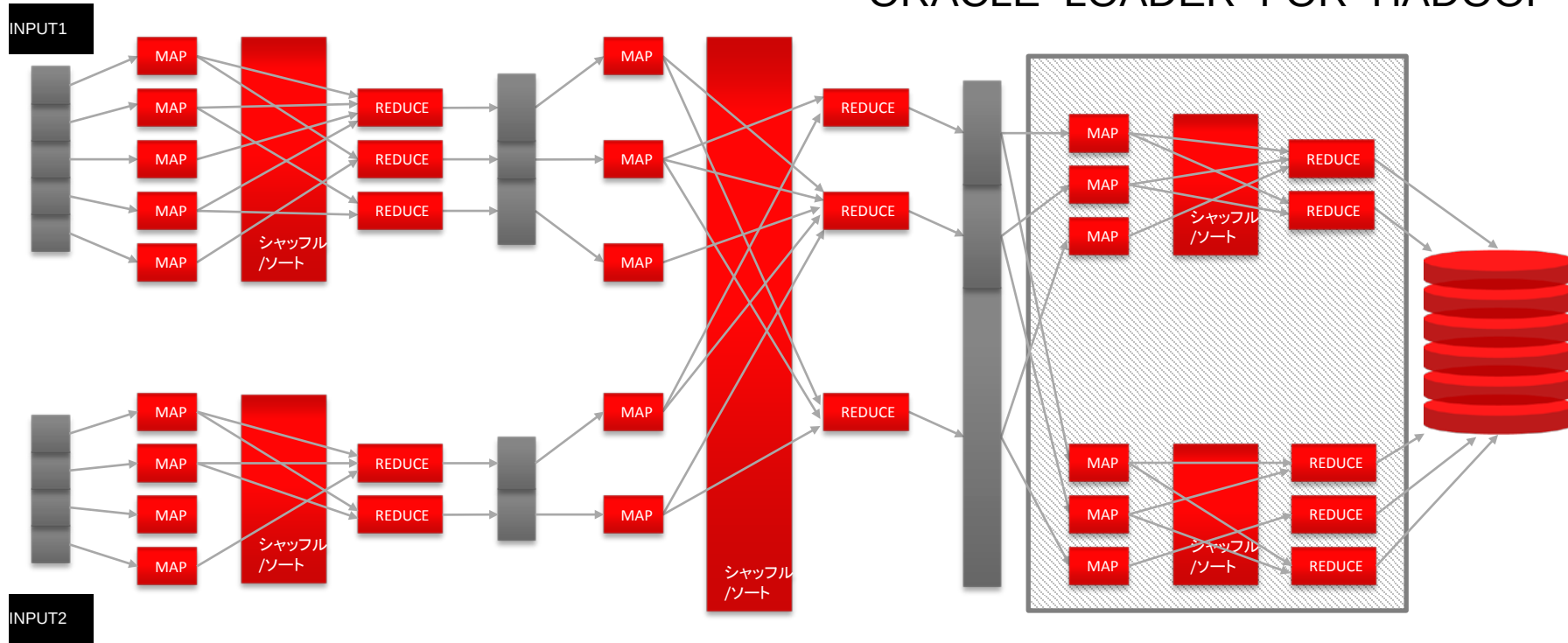
プロセスを管理

データウェアハウスへロード



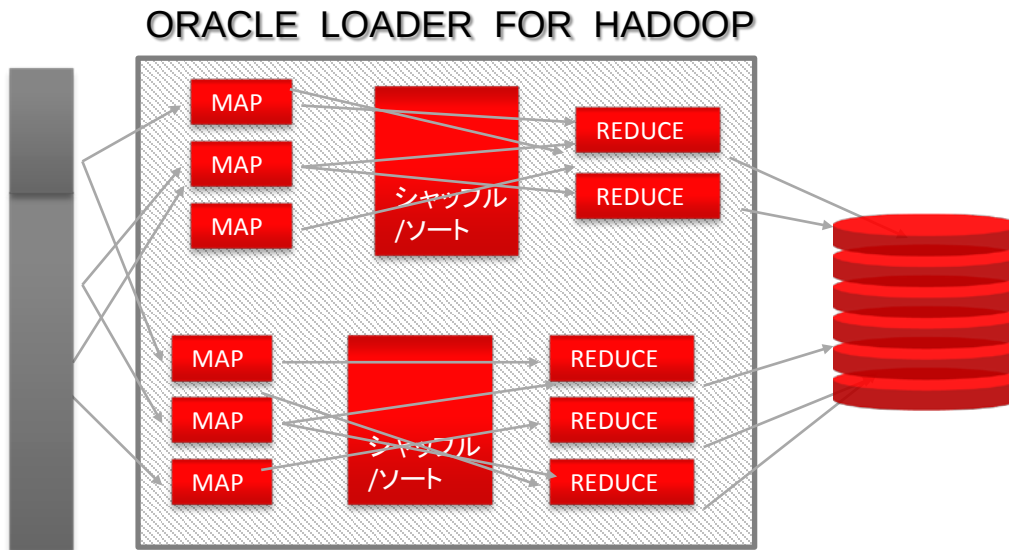
Oracle Loader for Hadoop

ORACLE LOADER FOR HADOOP



Oracle Loader for Hadoop

Hadoopクラスターを使用しOracle DBへの高速ロードを実現



MapReduce ワークフローの最終段階として利用

パーティション・テーブルもサポート

オンラインとオフラインのロード

Oracle Loader for Hadoopの優位点

- 入出力および中間処理が高速化
 - 最新のオープンソースとOracleのknow-howの結晶
- Oracle Database に最適化
 - Oracle Database の機能を最高に生かす仕組み
- MapReduce ジョブとして独自に実装可能
 - APIを提供

優位点の詳細:ロード・バランシング

- 全てのReducer に対する負荷をほぼ均等
 - MapReduce パーティショニング・スキームを作成
 - Oracle Databaseに最適なサンプリングのノウハウを反映
- 同時にロードも調整
 - パーティション間で分布が均一でない場合
- ロードの速度低下を防ぐ
 - Reducer へのバランスが悪いロードの場合
- Apache Commons Math ライブラリを使用

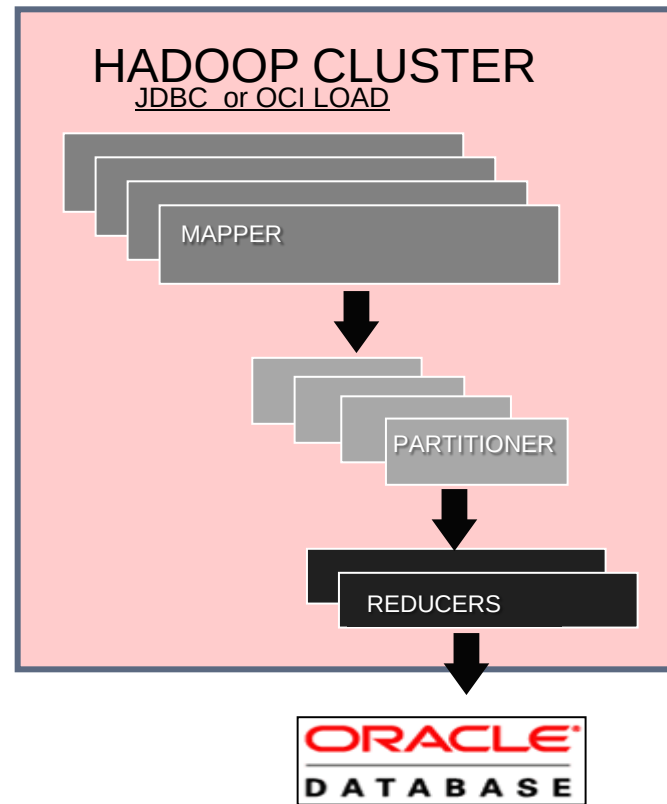
優位点の詳細: 入力処理

- 様々な入力フォーマット
 - CSV等のDelimited text
 - Hive tables
 - APIを利用した独自実装
- 高速な入力処理が可能
 - バイナリ・フォーマットはAvro record files



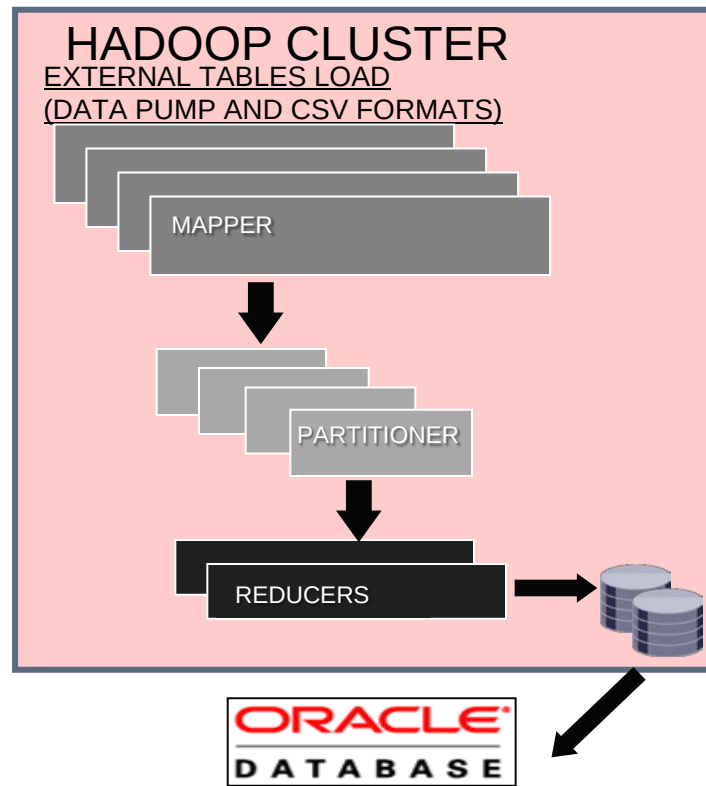
優位点の詳細: 出力処理(オンライン)

- 2つの方式をサポート
 - JDBC
 - OCI Direct Path
- 複数のOracle Database パーティション表への高速な転送
- Oracle Wallet サポート
 - 実行時にoraclepki.jarが必要



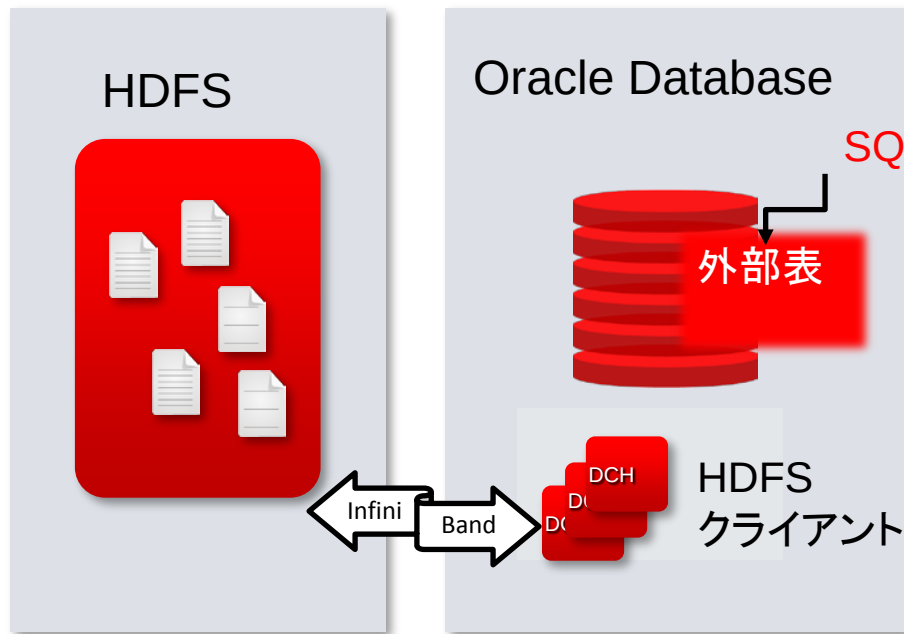
優位点の詳細: 出力処理(オフライン)

- Data Pump フォーマット
 - 外部表のためのバイナリファイル
 - 外部表のためのSQL
 - Direct Path の並行処理をサポート
 - 外部表に対する最速のオプション
- CSV, delimited text
 - SQL*Loader 等でロード



Oracle Direct Connector for HDFS

Oracle Databaseから直接アクセス



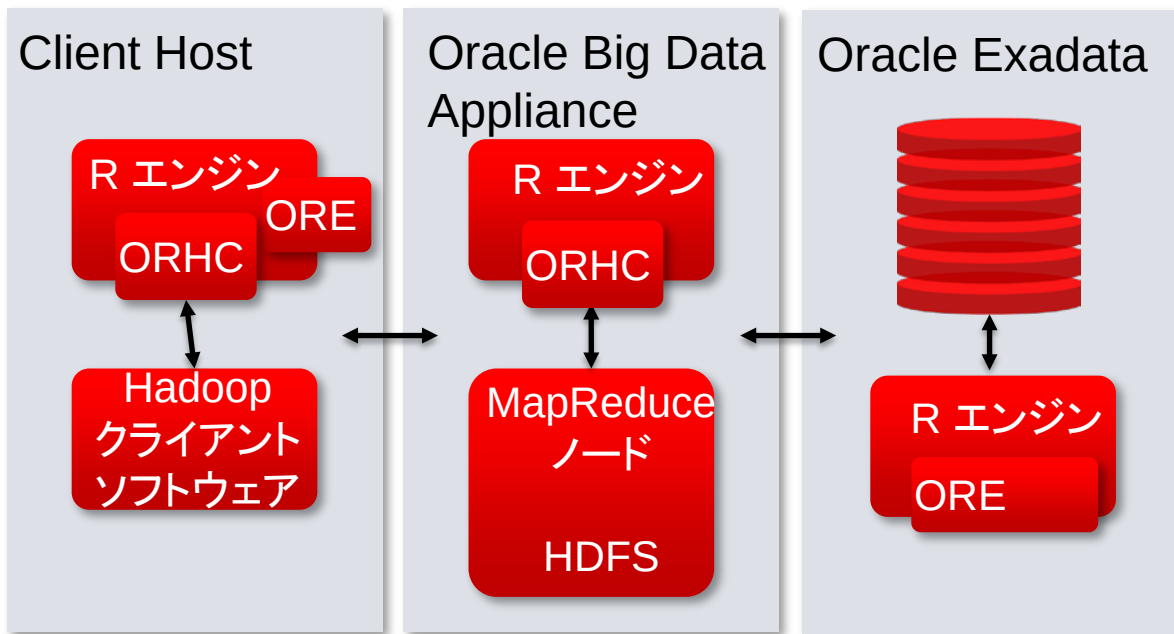
HDFSへのSQLアクセス

外部テーブルビュー

データ・クエリまたはインポート

Oracle R Hadoop Connector

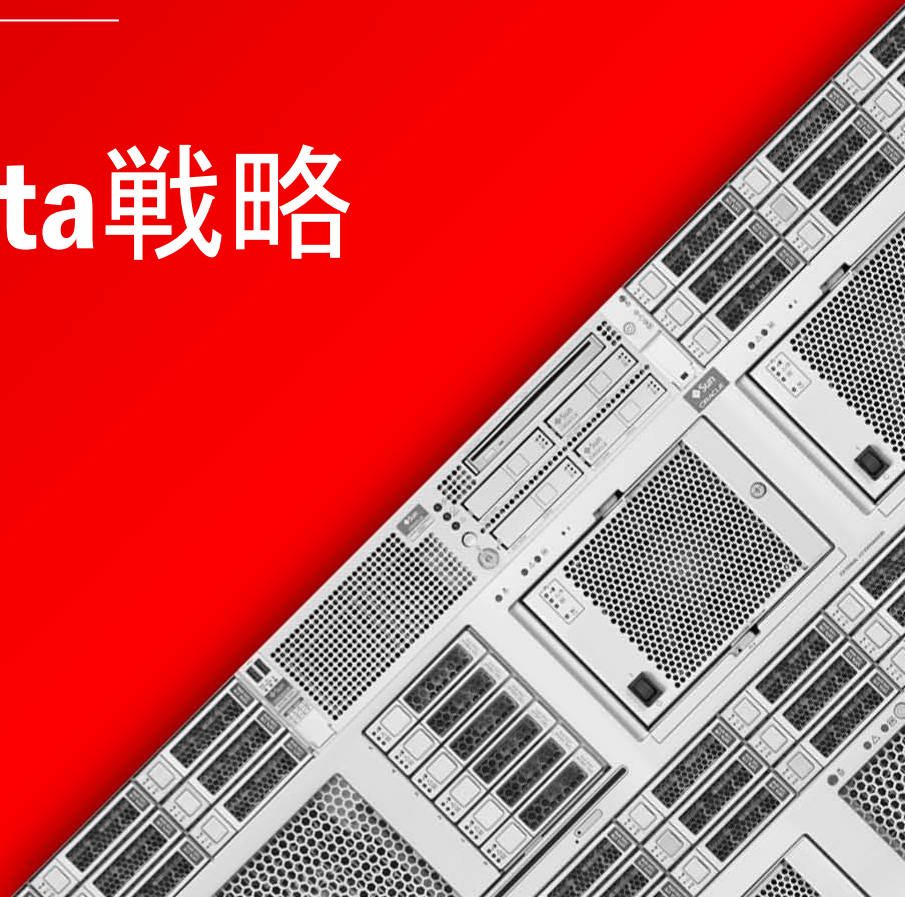
HadoopへのネイティブRアクセス



ネイティブ R MapReduce

ネイティブ R HDFS アクセス

Oracle のBig Data戦略 皆様と共に



Big Data利活用の3ステップ

構造化データの利活用

- ✓ 既存の業務系・情報系データベースに蓄積されたデータを、事業におけるコスト削減と売上向上に貢献する仕組みとして確立

非構造化・構造化データを融合した利活用

- ✓ 非構造化データの取得・体系化・分析に先進的に取り組んできた企業からの要望は、既存の構造化データと非構造化データを融合したデータ利活用

ビジネス・インテリジェンス、ビジネス・インサイトとしての利活用

- ✓ データマイニングや統計解析などの高度なデータ分析により、それぞれの企業や事業にとって価値がある洞察を掘り出すことを実現

Oracle and Big Data : Big Data for Enterprise

The screenshot shows the Oracle Big Data website as it appeared in early 2012. The browser window displays the URL <http://www.oracle.com/us/technologies/big-data/index.html>. The Oracle logo is prominently displayed at the top left. A navigation bar includes links for Products and Services, Solutions, Downloads, Store, Support, Training, Partners, and About, along with a dropdown for the Oracle Technology Network. The main content area is titled "Oracle and Big Data" with the subtitle "Big Data for the Enterprise". It features a large image of server racks and a detailed paragraph explaining the importance of big data in business. Below this, several links to white papers and a press release are provided. On the right side, there are promotional banners for a webcast and an online forum, both with "REGISTER NOW" buttons. The footer of the page includes a link to "Products and Solutions for Big Data".

Oracle Big Data | Oracle Products

http://www.oracle.com/us/technologies/big-data/index.html

ORACLE

(Sign In/Register for Account | Help) United States Communities I am a... I want to... Secure Search

Products and Services Solutions Downloads Store Support Training Partners About Oracle Technology Network

Technologies > Big Data

Oracle and Big Data

Big Data for the Enterprise

The term big data draws a lot of attention, but behind the hype there's a simple story. For decades, companies have been making business decisions based on transactional data stored in relational databases. Beyond that critical data, however, is a potential treasure trove of less structured data: weblogs, social media, email, sensors, and photographs that can be mined for useful information.

Oracle offers the broadest and most integrated portfolio of products to help you acquire and organize these diverse data sources and analyze them alongside your existing data to find new insights and capitalize on hidden relationships. Learn how Oracle helps you acquire, organize, and analyze your big data:

- White Paper: Big Data Strategy Guide (PDF)
- Analyst White Paper: Winter Corporation, Big Data: Business Opportunities, Requirements, and Oracle's Approach (PDF)
- Press Release: Oracle Selects Cloudera for Apache Hadoop Integration Into Oracle Big Data Appliance
- Analyst White Paper: IDC, Oracle's All-Out Assault on the Big Data Market: Offering Hadoop, R, Cubes, and Scalable IMDB in Familiar Packages (PDF)
- White Paper: Oracle: Big Data for the Enterprise (PDF)
- Podcast: Big Data Panel Discussion (28:36)

Oracle 1-800-633-0738

Have Oracle call you

Global contacts

Webcast and Live Chat, April 4, 2012

Introducing Oracle's Latest Business Analytics Strategy

ORACLE BUSINESS ANALYTICS
See more. Act faster.

REGISTER NOW

Online Forum

Big Data Essentials

REGISTER NOW

Products and Solutions for Big Data

OTN(Oracle Technology Network)での情報交換

OTN Discussion Forums : Big Data

https://forums.oracle.com/forums/category.jspa?categoryID=512

ORACLE

Forum Home » Big Data

Category: Big Data

Up one category Back to main category

Forum / Category	Threads / Messages	Last Post
Big Data and Big Data Appliance	3 / 7	2012/03/09 9:43 Last Post By: Jean-Pierre Dij... »
Big Data Connectors	4 / 10	2012/03/23 13:05 Last Post By: atiru »
NoSQL Database	81 / 386	2012/03/26 6:15 Last Post By: Kun MI »

Recent threads in this category:

Thread	Author	Forum	Replies	Last Post
Can not deploy a remote storage node	Kun MI	NoSQL Database	5	2012/03/26 6:15 Last Post By: Kun MI »
ODI Application Adapter for Hadoop	user570138	Big Data Connect...	2	2012/03/23 13:05 Last Post By: atiru »
Can I open *.jdb file ?	foreversunyao	NoSQL Database	4	2012/03/22 20:00 Last Post By: foreversunyao »
Service start stop scripts and Nagios	jcsuedelvalle	NoSQL Database	4	2012/03/19 17:36

Welcome, Guest

- [Sign In / Register](#)
- [Guest Settings](#)
- [Search](#)
- [FAQ](#)

Search Category

Last Post

[Re: Can not deploy a remote...](#)
Last Post By: [Kun MI](#)
Last Post At: 2012/03/26 6:15

Popular Tags

11g apex api bdb coherence
database dbd deploy hadoop hostname
java **nosql** nosqldb oracle plsql
[View all](#)

Q&A

ご質問・ご相談はOpenWorld終了後もお受けしております

あなたにいちばん近いオラクル

Oracle Direct

0120-155-096

(平日9:00-12:00 / 13:00-18:00)

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

Oracle Direct	検索
---------------	----

各種無償支援サービスもございます。

ORACLE

Hardware and Software

ORACLE®

Engineered to Work Together

ORACLE®

ORACLE®