



2014 Asakusa Framework Day

Hadoop活用への技術的取組と Asakusa Frameworkの適用評価について

創造・信頼・成長

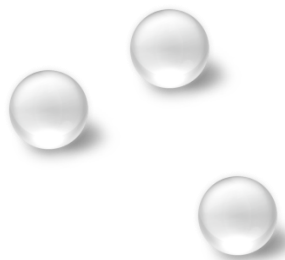


www.ns-sol.co.jp

2014年11月28日

産業・流通ソリューション事業本部

折原 耕路



新日鉄住金ソリューションズ株式会社

はじめに

1. Hadoopへの取組

2. Asakusa Frameworkへの取組

2-1. Amazon Web Service環境とオンプレミス環境

2-2. Asakusa FrameworkとHive

3. ソリューション・サービス

社名	新日鉄住金ソリューションズ株式会社		
設立	1980年10月 2012年「新日鉄住金ソリューションズ」に社名変更 2001年4月 新日鉄情報通信システム(ENICOM)と新日鉄EI事業部が新日鉄ソリューションズとして事業統合		
事業内容	<p>経営及びシステムに関するコンサルテーション</p> <p>情報システムに関する企画・設計・開発・構築・運用・保守及び管理</p> <p>情報システムに関するソフトウェア及びハードウェアの開発・製造並びに販売及び賃貸</p> <p>ITを用いたアウトソーシングサービスその他各種サービス</p>		
資本金	129億5,276万3,000円		
売上高	1,800億円 (連結)	1,587億円 (単独)	2014年3月期
従業員数	5,052名 (連結)	2,383名 (単独)	2014年3月末時点
連結子会社	<p><u>ITサービス</u> NSSLCサービス株式会社</p> <p><u>コンサルティング</u> NSフィナンシャルマネジメントコンサルティング株式会社 株式会社金融エンジニアリング・グループ</p> <p><u>海外事業</u> 新日鉄住金軟件(上海)有限公司 NS Solutions USA Corp NS Solutions Asia Pacific Pte.Ltd NS Solutions IT Consulting Europe Ltd</p> <p>Thai NS Solutions Co., Ltd. PT.NSSOL SYSTEMS INDONESIA</p> <p><u>地域事業</u> 北海道NSソリューションズ株式会社, 東北NSソリューションズ株式会社, 株式会社NSソリューションズ東京, 株式会社NSソリューションズ関西, 株式会社NSソリューションズ中部, 株式会社NSソリューションズ西日本, 大分NSソリューションズ株式会社</p> <p><u>合併会社</u> エヌシーアイ総合システム株式会社 日鉄日立システムエンジニアリング株式会社</p>		

コンサルティング

企画・設計

開発

運用・保守

業務ソリューション

お客様の業種・業務に関する豊富な知識と経験に基づくソリューションの提供

- 製造・消費財・流通・サービス分野
 - SCM(Supply Chain Management)
 - PLM(Product Lifecycle Management)
 - ERP(Enterprise Resource Planning)
 - SCE(Supply Chain Execution) 等
- 金融分野
 - 市場系業務・ディーリングシステム
 - ALM・原価計算、統合収益、リスク管理
 - カードプロセッシングASP、外為インターネットサービス、与信判定等サービス等
- 社会・公共・公益分野
 - 官公庁等LAN/WAN等基盤OAシステム
 - 調査統計処理、電子申請
 - 学事業務支援
 - 高度科学技術解析支援 等

基盤ソリューション

お客様の重要な業務を支える緻密で信頼性の高いシステム基盤を設計・構築

- クラウドサービス
 - クラウドコンピューティングITインフラ “absonne”
 - プライベートクラウド “absonne Enterprise Cloud Framework”
 - 開発環境サービス “NSSDC”
 - デスクトップ仮想化サービス “M³Daas@absonne”
- システムライフサイクルサポート
 - システム基盤の提案、企画、設計、調達、構築・実装、運用・保守 等
- 大規模検証サービス
 - 製品組み合わせ技術検証
 - 性能評価、性能管理技術 等

ビジネスサービス

製鉄システムの経験に基づく運用・保守等のアウトソーシングサービスの提供

- ITインフラアウトソーシングサービス
 - 弊社フレームワークに基づき標準化したインフラ運用サービス “NSFITOS”
- 運用・保守サービス
 - 店舗システムのリモート集中監視
 - 通信会社のネットワーク設計・構築・運用
 - ECサイトのサーバーハウジング 等
- 図面・文書ASP/BPOサービス
 - 金融機関向け文書管理、取引報告書電子配信
 - オンデマンドアーカイブ
 - CAD/図面情報の交換/共有
 - 法定電子公告 等

コンサルティング

研究開発・技術サポート

1. Hadoopへの取組

大量のデータの蓄積が可能

Volume

リアルタイムに情報を収集する基盤

Velocity

技術の発展
(Hadoop、MapReduce、DWHアプライア、Variety

技術コストの低減



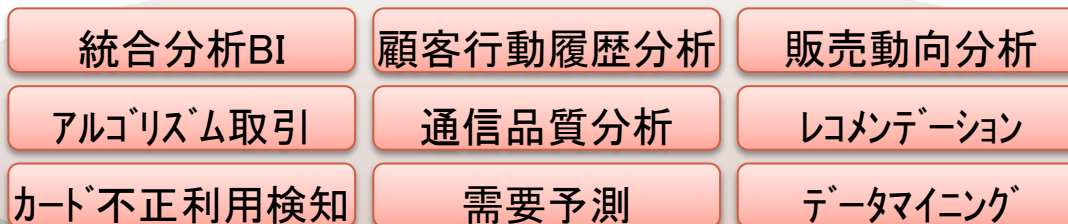
分散コンピューティングが身近になり、
今まで顧みられなかったデータの活用が可能に

- **巨大企業**(スーパーコンピューターを保持など)や**国家レベル**しかできなかった情報活用が、一般企業でも可能な時代に



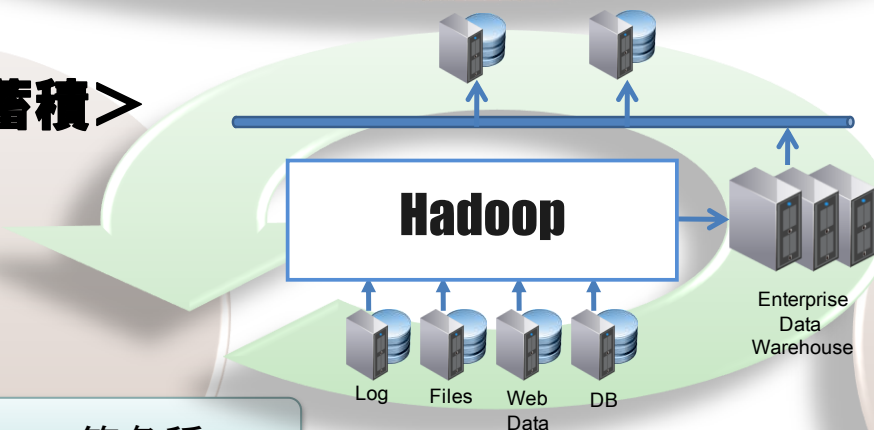
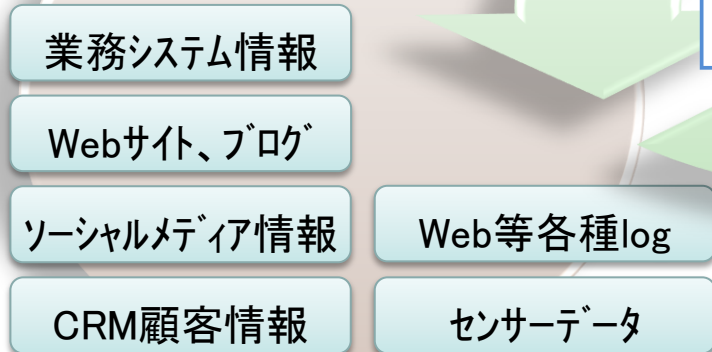
収集・蓄積、共有、分析・活用のサイクルを廻すこと

<データ分析・活用>

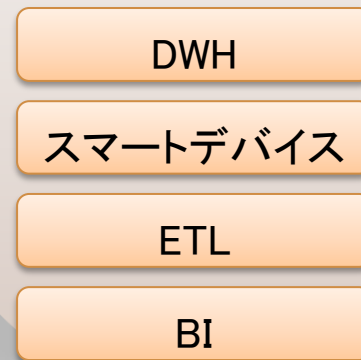


分析システム ECサイト

<データ収集・蓄積>

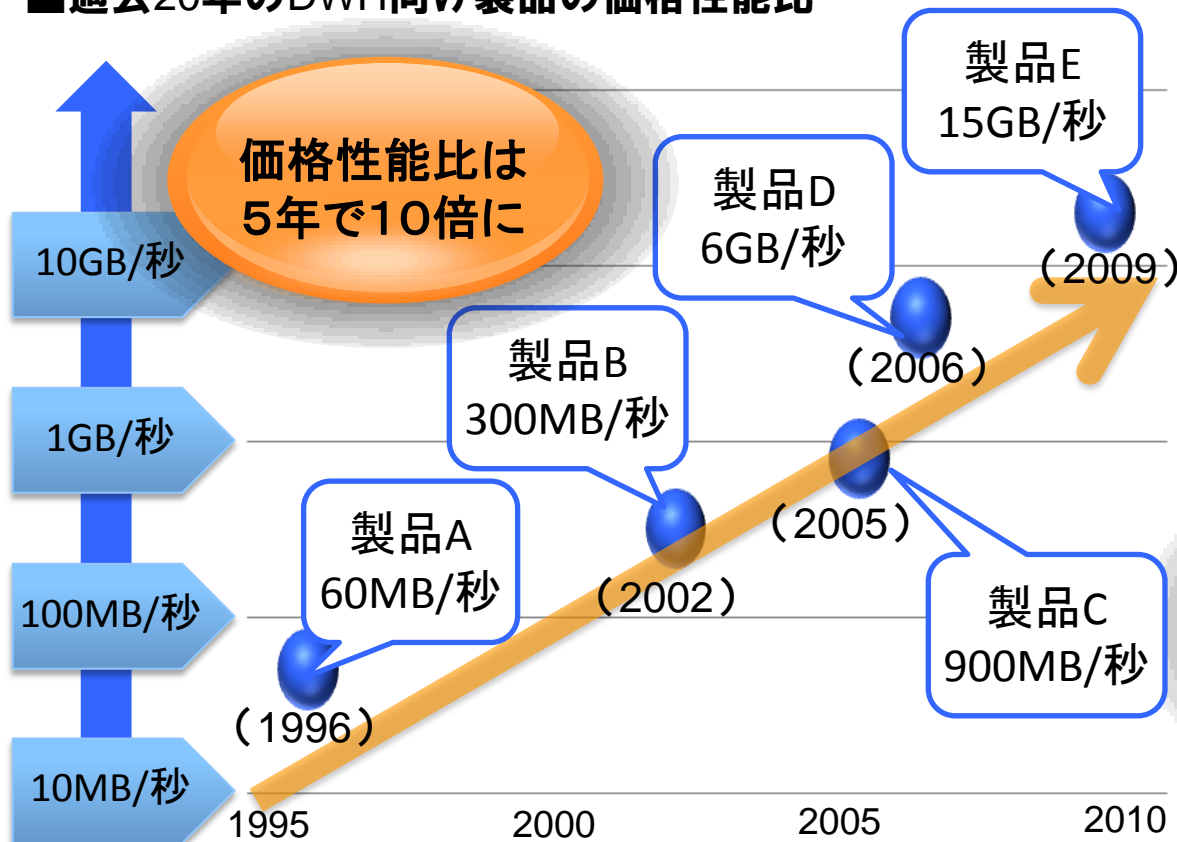


<データ共有>



テクノロジーの進化で、扱えるデータ量と種類の幅が広がり リアルタイムな活用が可能に

■過去20年のDWH向け製品の価格性能比

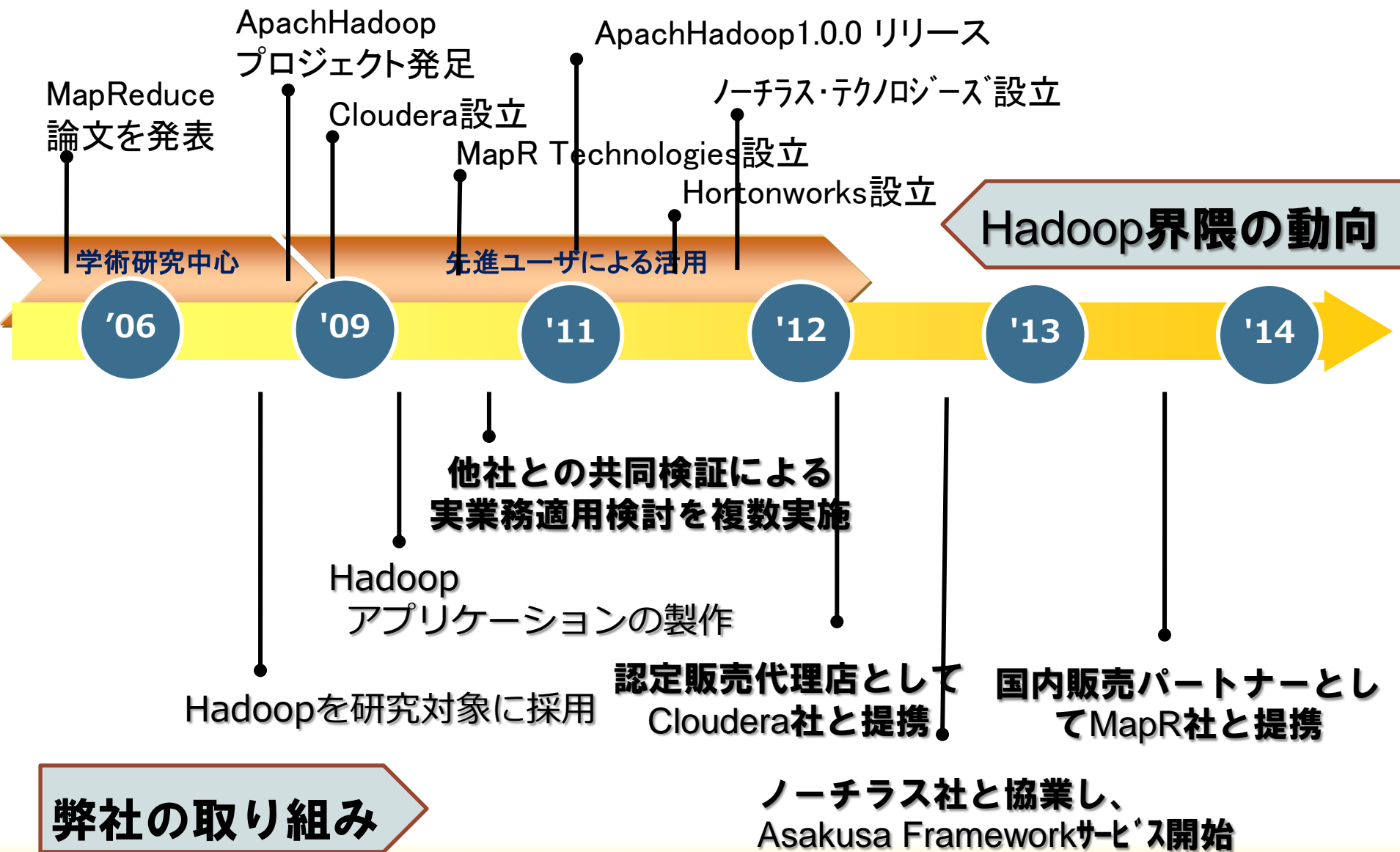


新しいテクノロジー・利用形態

- Hadoopの普及
- クラウドの活用
- 高性能アプライアンスの普及
- 位置情報/画像情報の利用
- テキストマイニングの発達

あきらめていたことが
実現できる時代に

弊社におけるHadoopの取り組み



Hadoopへの期待と課題

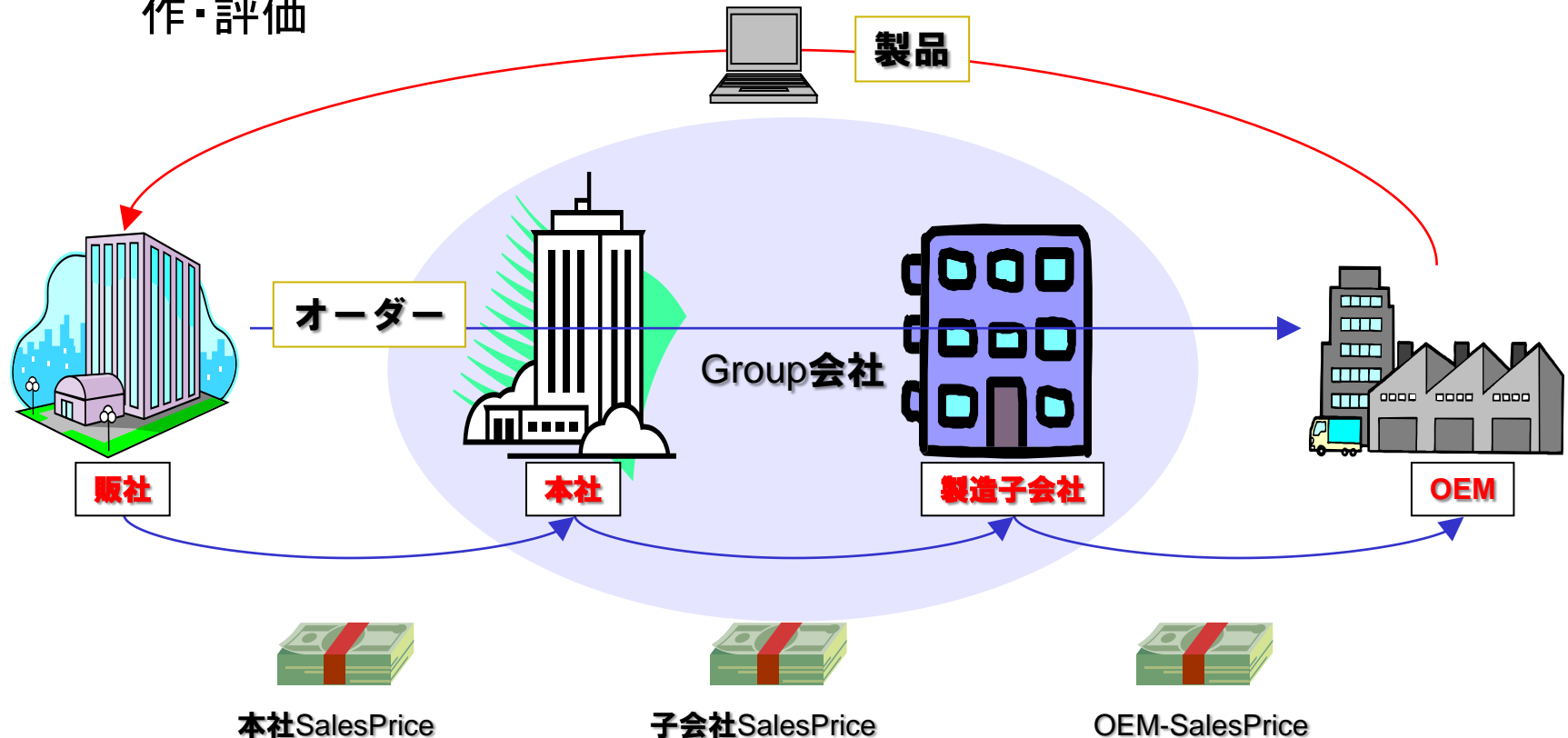
- ◆ 背景: データ量の増大と大量データの高速度処理
 - ログデータ、SNSのデータなど、業務系データよりはるかに多い量のデータの高度な活用が求められている
 - データが構造化されているとは限らない
- ◆ 生じる課題
 - すべてのデータを構造化してRDBMSへ格納するにはコストが高い
 - 既存テクノロジーでは有効な時間内に処理が完了しない場合
- ◆ 課題を解決すると期待される新しいパラダイム
 - MapReduceアルゴリズム
 - オープンソースの並列処理プラットフォーム
 - これらを実現している製品の 하나가 Hadoop
- ◆ DWH及び大量データ分析用途におけるHadoopへの期待
 - 超大規模な並列実行環境を安価に構成できる可能性がある
 - SQLでは実現が難しい複雑な分析、機械学習などによるデータ活用
- ◆ Hadoopを実適用するにあたっての課題
 - エンタープライズ向け機能(セキュリティ/運用等)のカバー方法
 - アプリケーションの開発
 - MapReduceを活用したデータ活用のためのノウハウが必要
 - SQLライクな言語やGUIでHadoop上のデータを利用できるプロダクトも存在するが、まだ機能面で制約が多い

2. Asakusa Frameworkへの取組

2-1. AWS vs オンプレミス

2-2. Asakusa Framework vs Hive

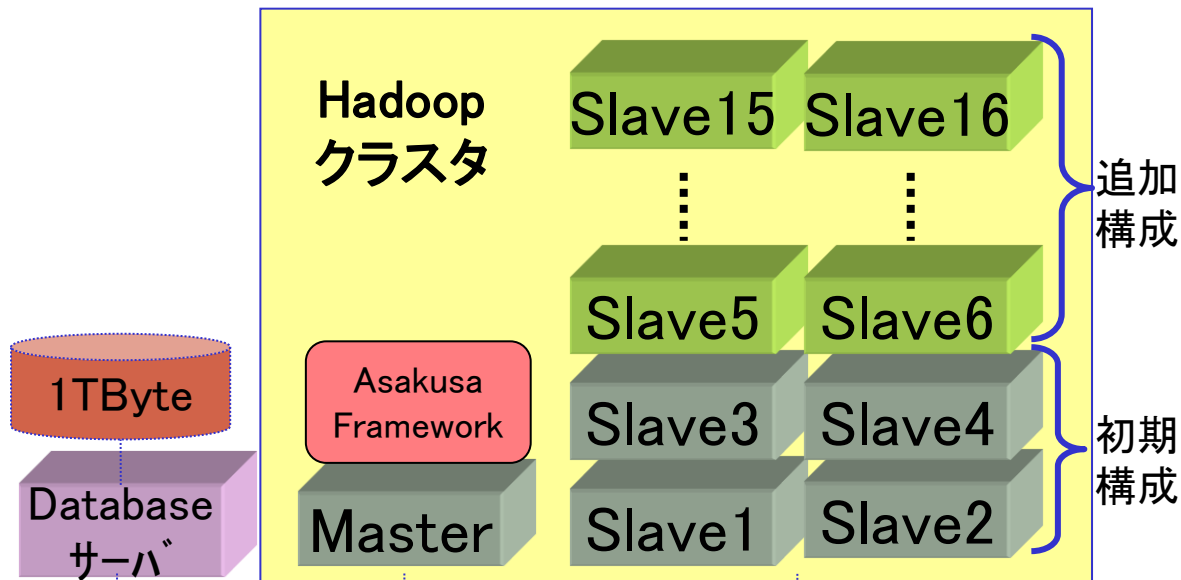
- ◆ OEM生産を想定した評価モデル
 - 各パーツ価格や構成変動に伴う販売価格の積み上げバッチ処理を製作・評価



・製品約10万個、スペック約100万件

■ 検証環境

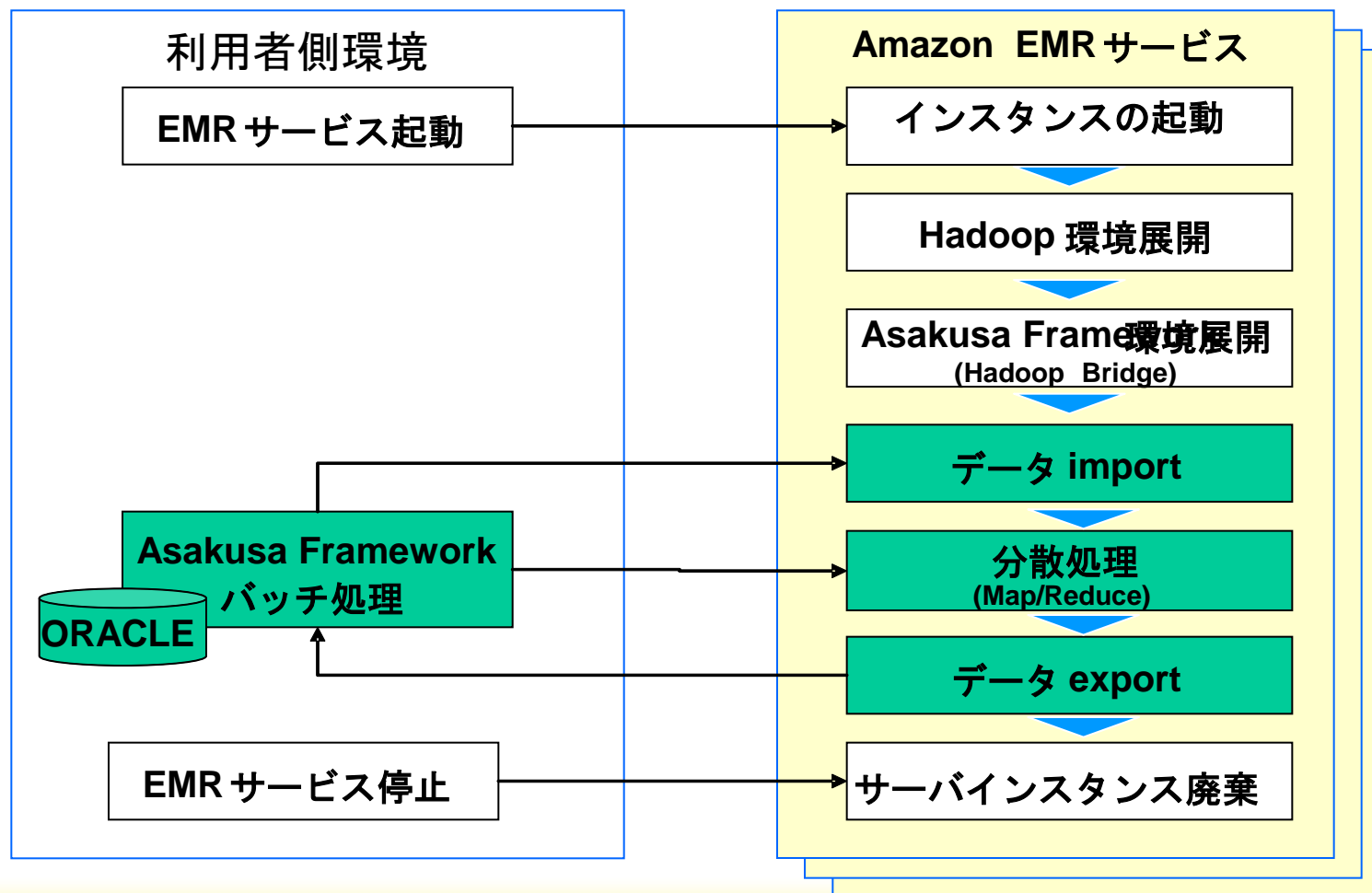
自社検証環境



機器	IBM BladeCenter
仕様	<ul style="list-style-type: none"> ・Intel Xeon 2GHz 2CPU × 2 core ・RAM8GB
OS等	<ul style="list-style-type: none"> ・CentOS 6 (x86 64bit版) ・Hadoop (CDH)

■ 検証環境

Amazon Web Serviceの提供するElastic MapReduce(EMR)サービスの利用



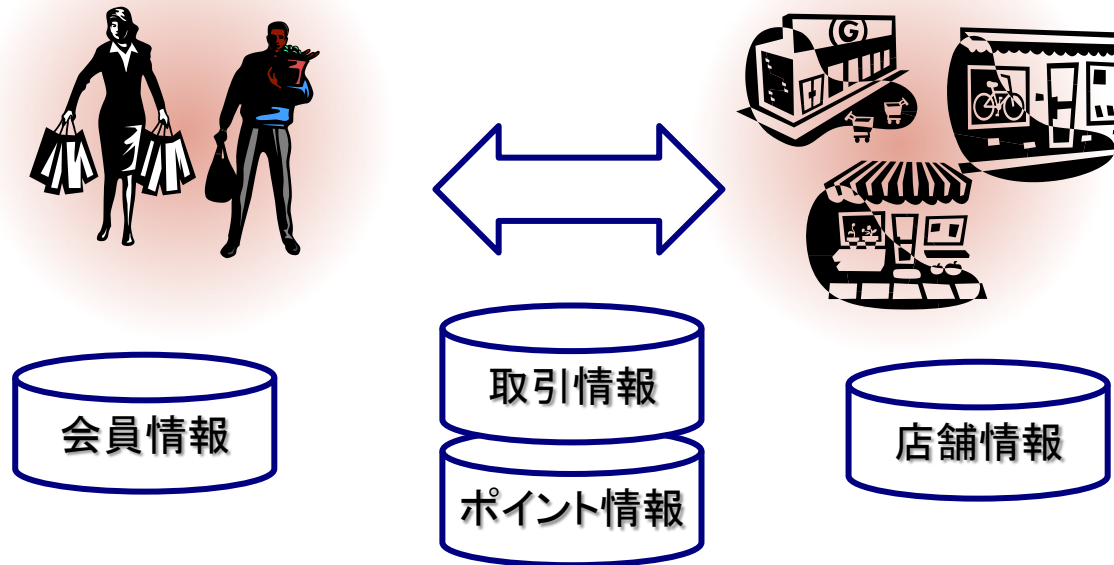
計算結果件数 : 4,021,574 (約 400万)
積上げ総レコード数 : 182,086,465 (約 1.8億)

	32node
AWS (EMR)	40分
オンプレミス	-
PL/SQL (参考)	-

Screen only
(講演にてご覧ください)

- ◆ PL/SQL

◆ コンシューマ企業における購買実績の分析



購入者	延べ800万人
商品	150万点
購買明細	20億件以上

店舗別、会員属性別、時間帯別等の購買額の計算

		多ノード環境	高スペック環境
サーバ構成	Master:	1台	1台
	Slave:	16台	4台
Slave	CPU:	Intel Xeon®2.00GHz コア4(2CPU × 2コア)	Intel Xeon®2.26 GHz コア6(スレッド12)
	メモリ:	8GB	48GB

HadoopはCDH

参考Benchmark	多ノード環境	高スペック環境
Hadoopテストツール:hadoop-*-test.jar 分散処理環境の性能(時間の逆比)	1	0.2~0.6
円周率の計算:Tachus PI サーバ1台あたりの性能(時間の逆比)	1	1.7~3.6
UNIXベンチマークツール:Unix Bench サーバ1台あたりの性能(指標)	1	1.6

*多ノード環境を”1”とした場合の性能比に変換.数値が大きい程性能が高い

多ノード
環境

高スペック
環境

1支店

多ノード
環境

会員・所属

時間帯

会員属

Screen only
(講演にてご覧ください)

分

分

分

- Hive、Asakusa Frameworkでほぼ同程度の処理時間オーダー
 - 圧縮形式の違いや細かいチューニングによる改善差異の可能性はあり。
 - 実行処理の構造による処理時間差異あり。
- Hive(HiveQL)の場合、SQLから比較的簡単に置き換えが可能である。
- Asakusa Frameworkの場合でも、SQLからの置き換えは可能であるが、ではSQL実行計画を独自に作り込む必要がある。
- より複雑な操作(中間データの再利用 / 出力の分割 / 中間データに対するデータ操作)であれば、Asakusa Framework DataFlowの強みがでてくる。
- 単純なSQLで可能な処理はHiveが適していると思われる。

まとめ

- ◆ Hive、Asakusa Frameworkのどちらでも対応は可能であり、要件に合わせた選択をするべきである。
- ◆ 単純なBI向け処理の場合、Asakusa FrameworkよりもHiveによる対応が適しているが、複雑なバッチ処理になるほど、中間処理結果を後続処理で利用するなどAsakusa Frameworkの方が効率の良い設計が可能と思われる。

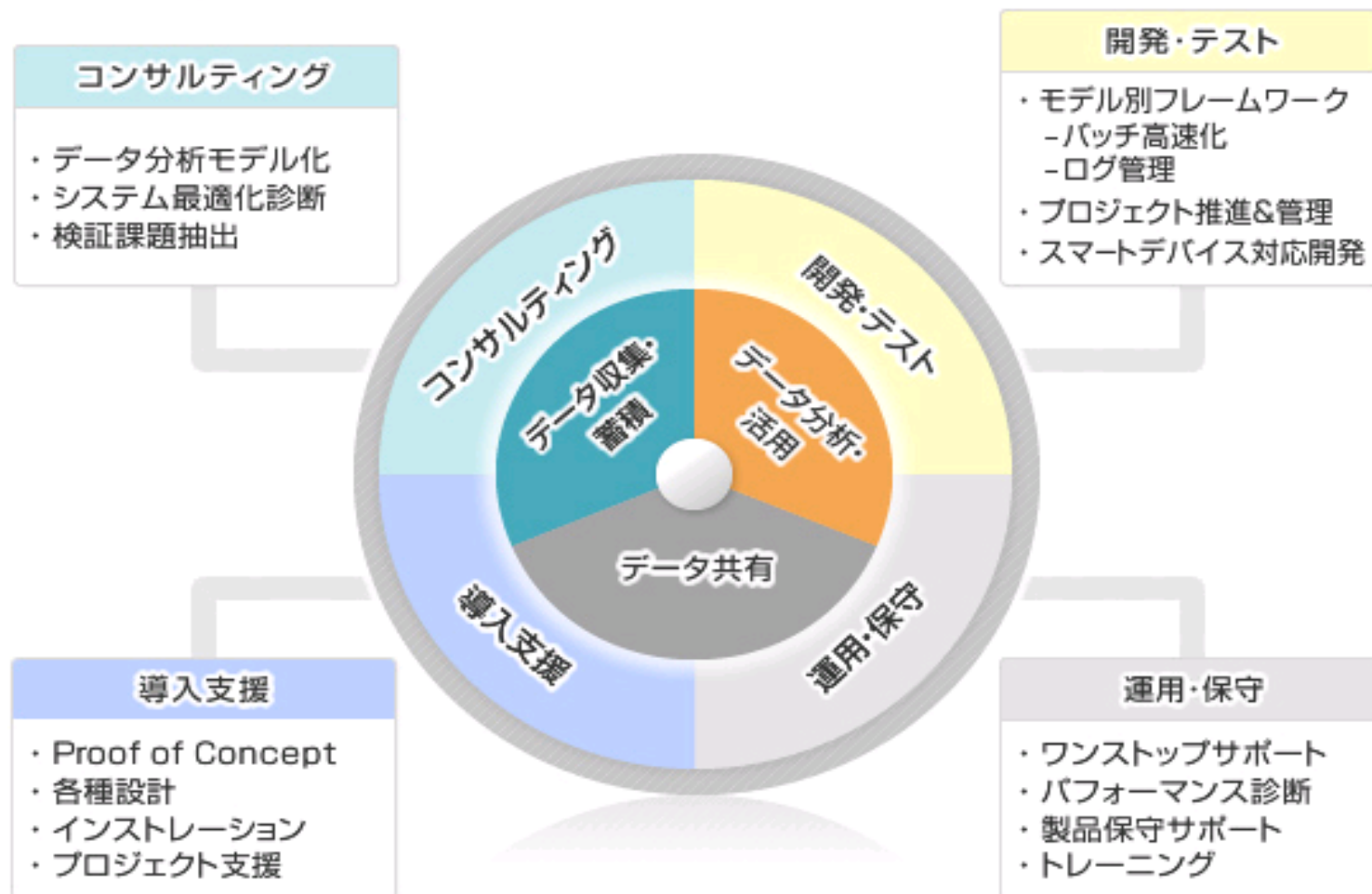
エンタープライズ開発向き

- ◆複雑な基幹バッチのビジネスロジックでは、DFP(データフロープログラミング)手法がパフォーマンスに有利
- ◆DFPはCOBOL設計者にも受入れやすい、データフローを設計に利用できる
- ◆OSSであり、ベンダーロックインしない
- ◆エンタープライズ開発向けのツールが揃っている

3. ソリューション・サービス

ビッグデータソリューション

新日鉄ソリューションズはHadoopをコアとしたデータ活用に向けトータルソリューションを提供しています。



ビッグデータソリューション = 並列処理テクノロジーの活用

安定したHadoopを活用するシステム構築



- ・Hadoopを適切に活用することで課題解決への期待は大きい
- ・一方、従来とは異なる新しい技術のため、設計・開発・運用における対応が重要となる
- ・システムライフサイクルを踏まえてサービスを整備

適用処理

[大量データの業務への取り込み]

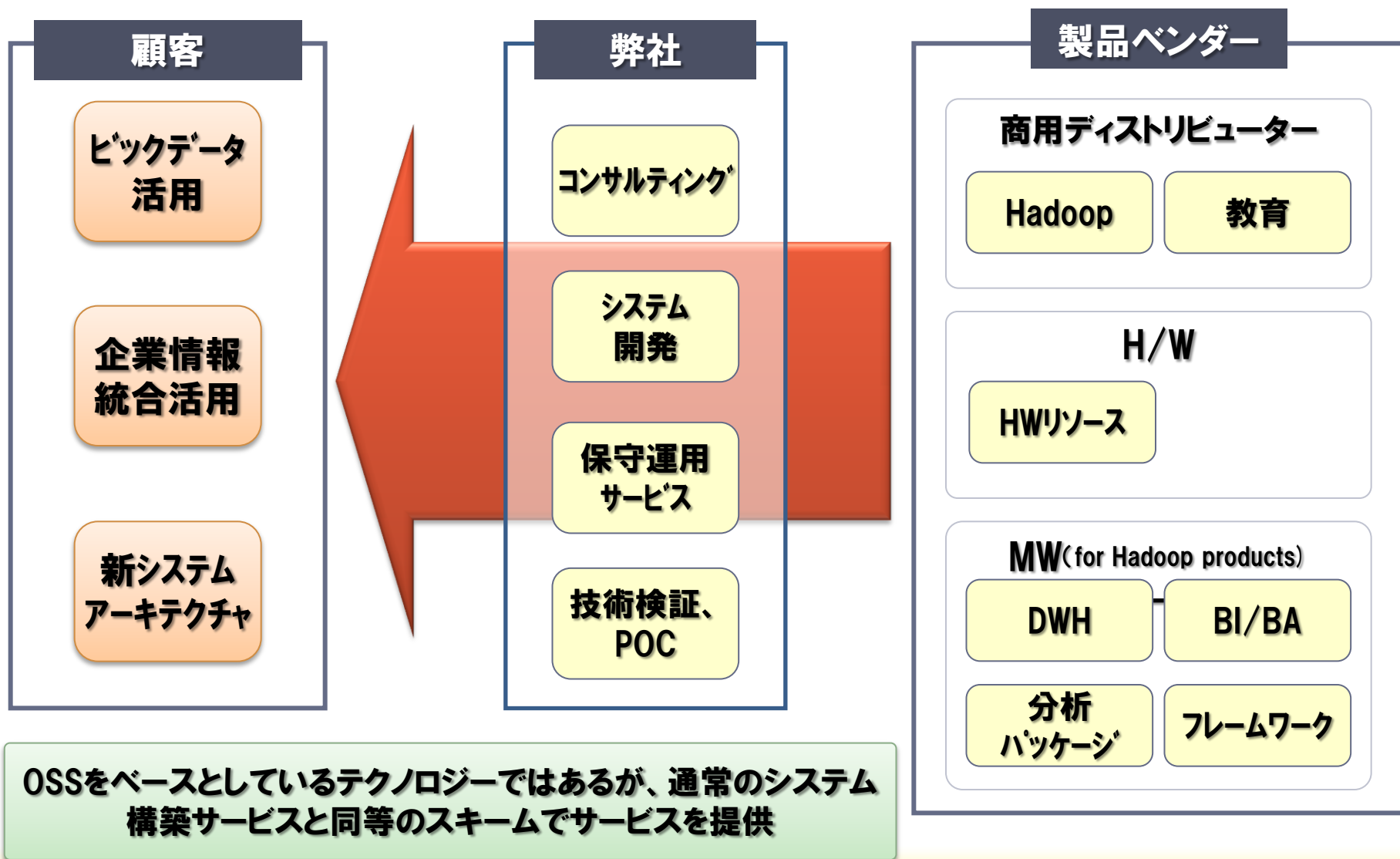
- 1) 多量データの現実的な処理時間で検索(並列実行)
- 2) 非構造化情報の構造化処理(SNSやログ情報の分析)
- 3) ペタバイト規模のデータをコストパフォーマンスの高い環境へ格納

[バッチ処理の高速化]

- 1) 数時間かかる処理を、並列化機構のハードリソース増加による処理性能のスケールビリティを活用

<マイニング処理>

- ・保存期間のロング化による、分析精度の向上
- ・生データの長期保管による、将来への布石
- ・リアルタイム化による、早期のトレンド分析





<http://www.nssol.nssmc.com/>

- ・NS(ロゴ)、NS Solutions、NSSOL、absonne、absonne Enterprise Cloud、Framework、NSSDC、M3Daas、NSFITOSは、新日鉄住金ソリューションズ株式会社の登録商標です。
- ・ORACLEは米国ORACLE Corp.の登録商標です。
- ・その他本文記載の会社名及び製品名は、それぞれ各社の商標又は登録商標です。