SAP HANA Vora™ と MapR 連携の Lenovo レファレンス・アーキテクチャ

SAP HANA® と MapR の連携を実証した Lenovo Sysytem x における検証レポート

SAPホワイト・ペーパー

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

【目次】

はじめに……1
ソリューション概要……2
　ビッグデータを統合する SAP HANA Vora……2
　ビッグデータ基盤 MapR……3
ハードウェア概要……4
　Lenovo System x Solutions for SAP HANA とは……4
　Lenovo System x Solution for SAP HANA アプライアンス・モデル の特長……6
　MapR on Lenovo System x for SAP HANA Vora とは……7
　MapR on Lenovo System x for SAP HANA Vora の特長……8
ソフトウェア概要……9
　SAP HANA Vora の主な機能……9
　MapR に統合されている SAP HANA Vora……10
MapR の概要とコンポーネント……11
　MapR 概要……11
SAP HANA Vora ユースケース……13
　SAP HANA Vora 活用の 3 つのシナリオ……13
データ階層管理……14
　データ階層管理とは……14
　ダイナミックティアリングによるウォーム・データの階層管理……14
　ビッグデータでの階層管理……15
　階層に分かれたデータを SAP HANA スマートデータアクセスで統合……15
アナリティクス……16
　SAP BusinessObjects Lumira……16
　誰もが使えるセルフサービス方式のデータビジュアル化環境……16
ソリューション・コンポーネント……17
ハードウェア・コンポーネント……17
　Lenovo System x3850 X6 のハードウェアの特長……18
　MapR on Lenovo System x Solution for SAP HANA Vora 推奨構成……18
　Lenovo System x3650 M5 のハードウェアの特長……19
　Lenovo System x3650 M5 運用管理の特長……20
　SAP HANA と MapR for SAP HANA Vora を連携するネットワーク……20
ソリューション構成例……22
検証済み構成とサイジングのガイド……24
SAP HANA Vora と MapR 連携検証について……26
　検証構成……26
　データロードとテーブル作成……27
　ビューの開発……28
　Vora Tools と Lumira によるデータアクセス……28
　SAP……30
　HANA Studio による仮想テーブルの構成……30
まとめ……31
SAP ジャパンについて……32
マップアール・テクノロジーズについて……32
参考資料：System x サーバ構成詳細……33
　Lenovo System x Solution for SAP HANA アプライアンス・モデル……33
　MapR on Lenovo System x Solution for SAP HANA Vora　推奨構成 (約 50TB)……33

はじめに

様々なテクノロジートレンドが変化し、デジタル変革に取り組む企業にとっては、ビッグデータの活用が急務となっています。これまで企業が投資してきた、ビジネスに欠かせない「守りのシステム」である基幹系データのデータを活用しつつも、来たるビッグデータや IoT 時代に備えて「攻めのシステム」をビジネススピードに沿うように構築するには、どのようなシステムアーキテクチャで設計すべきかベストプラクティスが求められています。

こういった課題意識を持ったお客様のため、インメモリ・プラットフォームを提供する SAP、ビッグデータのコンバージドプラットフォームを提供する MapR、ハードウェアに強みをもつ Lenovo、これらを統合してシステム構築を手がける伊藤忠テクノソリューションズの4社が共同で検証を行い、レファレンス・アーキテクチャを作成しました。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

ソリューション概要

ビッグデータを統合する SAP HANA Vora

SAP HANA Vora は、エンタープライズデータとビッグデータを一つのプラットフォームに統合します。
ビッグデータの活用が叫ばれる昨今、企業がこれまで整備してきた基幹系 ERP システムに保存されている売上情報や顧客情報を更に活用したり、販売計画や需要予測、製産計画、といった計画業務と組み合わせて分析することで、これまで得ることができなかったビジネス上の洞察を得ることが可能となり、さらなる競争優位の源泉とすることができます。

SAP HANA Vora は、以下の3つの機能を提供します。

- ●ビジネスデータとビッグデータの横断的な分析
- ●ビジネスデータ分析のための階層 /OLAP 分析
- ●データ高度活用のための分析プラットフォーム

SAP ERP に代表される基幹系業務データと、ビッグデータ分散フレームワーク Hadoop のデータを連携したいというニーズをどのように満たすことができるでしょうか。ERP のデータはデータベースに格納されているため SQL を使ってアクセスする必要があり、一方で、Hadoop で分析を行っているデータサイエンティストは Scala や R といったオープンソース言語を日々扱っています。両者がデータのやり取りを行おうと思うと、ERP 管轄部門にデータ抽出を依頼するのに時間がかかったり、基幹系システムのデータ構造が複雑でデータを読み解くのが困難であったり、一筋縄ではいかないのが現実です。そこで、SAP は、この基幹系ビジネスデータとビッグデータを結びつけるべく、SAP HANA Vora をリリースしました。
加えて、SAP HANA Vora は、ビジネス分析で求められる OLAP 分析の機能をビッグデータの分散環境に提供します。ビジネスデータは、企業の組織構造や扱う製品のカテゴリなど、階層構造で格納されているのが一般的です。ビジネスデータの分析には、データ階層の定義、OLAP 機能が求められます。

ビッグデータ時代になると、データの移動コストを考慮した場合、分析をデータが保存されている場所で直接行うのが一般的です。Hadoop エコシステムでは、Spark MLlib といった機械学習やストリーム分析のプロジェクトも存在しています。
しかし、目的別に機能が提供されているため、新しい分析のために新しいコンポーネントや技術を検討する必要があり、すぐに分析をはじめられる環境ではありません。こうした問題を解決するため、

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

SAP HANA Vora 1.3では、単にビジネスデータとビッグデータを結合するプラットフォームを提供するにとどまらず、時系列分析やグラフデータ分析といった高度な分析機能を提供することにより、ビッグデータの保存場所とするだけでなくさらに深い洞察を得るためのプラットフォームとして活用することができます。

ビッグデータ基盤 MapR

ビジネスニーズが変わったことにより、より量も種類も多くのデータを活用する時代になり、システムに対する要求も以前とは変わってきました。従来の RDBMS ベースのみでは活用しきれないデータを活用するために、新しいデータ基盤として Hadoop が登場しました。MapR はその Hadoop をさらに進化させ、オープン API を備え、データ保護の機能を追加する一方で単一障害点やボトルネックを排除した、真のエンタープライズグレードのビッグデータ基盤を提供します。
SAP HANA や SAP HANA　Vora のデータ基盤として考える際には、以下の機能が必要となります。

- データを長期間蓄積する高速なストレージ機能
- ビッグデータの処理基盤である Spark の機能
- データ量あたりのコスト効率

さらに、昨今のデータ活用系のシステムの利用用途はよりミッションクリティカルに近づいているため、以下の要素が求められます。

- 単一障害点が無いこと
- 内部処理上もデータロード上にもボトルネックが無いこと
- 重要なデータを保護する機能

つまり、単なるストレージでは Spark の機能がないことによるデータの活用性に問題があります。Spark の機能を有する MapR 以外の Hadoop では、単一障害点の存在やパフォーマンスボトルネックの問題が残ります。MapR は業界で唯一これらの問題点を根本から解決してあるため、SAP HANA や SAP HANA Vora のデータ基盤として、コストやパフォーマンスはもちろんのこと、ミッションクリティカルな利用にも対応でき、安心して活用できる唯一のビッグデータ基盤となります。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

ハードウェア概要

本ソリューションは、Lenovo System x Solutions for SAP HANA のアプライアンス・モデルと、MapR on Lenovo System x for SAP HANA Vora　ソリューションを組み合わせた、2 つのソリューションから構成されています。

業界標準ともいえるインメモリ統合プラットフォームの SAP HANA と、MapR Hadoop ソリューションを透過的に利用できる新世代のビッグデータ・アナリスティックを実現するソリューションおよびその特長は以下の通りです。

Lenovo System x Solutions for SAP HANA とは

「Lenovo System x Solutions for SAP HANA」は、次世代 x86 サーバのアーキテクチャ「第 6 世代 Enterprise X-Architecture(X6)」に準拠し、大容量メモリを搭載可能な「Lenovo System x3850 X6」や「Lenovo System x3950 X6」に、SAP のインメモリ・ソフトウェアをあらかじめ搭載した SAP HANA 専用アプライアンスです。いずれのサーバも、リアルタイムでの検索やモデル化、分析のワークロードに最適化されたサーバ機種、プロセッサー数、メモリ容量などで構成されています。

「Lenovo System x Solutions for SAP HANA」は、データ容量に応じた 5 種類のモデルからなり、x86 サーバの OS 上に、分散ファイルシステムを搭載し、複数台のサーバを並列に接続して SAP HANA 全体の処理性能を容易に向上することが可能です。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

「Lenovo System x Solution for SAP HANA」は、Lenovo の X6 アーキテクチャをベースとした SAP HANA のワークロードに特化したソリューションです。本アプライアンスは、ワークロードの特性に合わせた構成やテクノロジーを採用しているため、優れた性能を発揮することが可能になります。

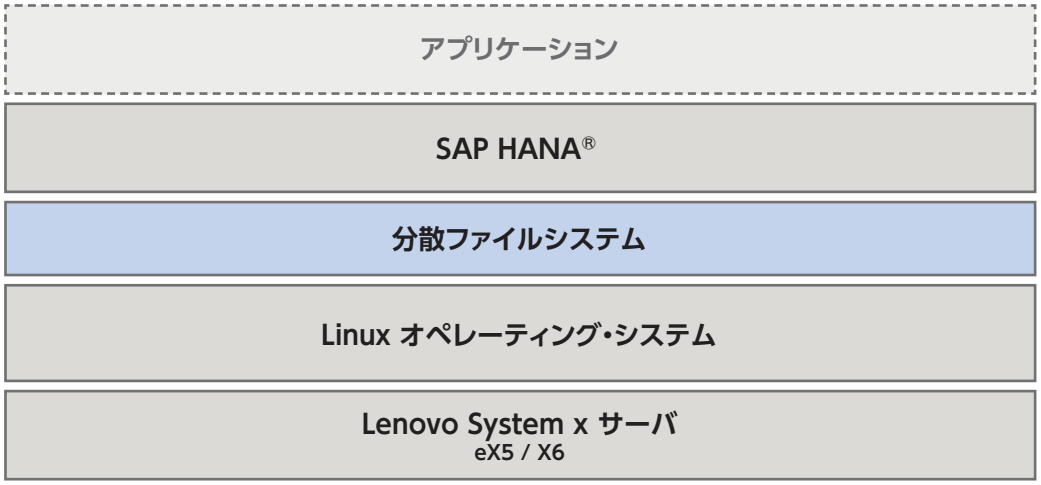


図 1　Lenovo System x Solution for SAP HANA アプライアンスの構成

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

Lenovo System x Solution for SAP HANA アプライアンス・モデル の特長

- Lenovo System x ハイパフォーマンス製品をベースに、シングルノード構成またはスケールアウト構成をラインアップ
- メモリ容量とデータ容量に応じたモデルから、お客様のニーズに応じてシンプルに選択可能
- 各モデルのログ領域に SSD (フラッシュ・ドライブ) を採用し、高パフォーマンスを実現
- S 構成から M 構成に、M 構成から L 構成に、L 構成から XL 構成に、容易にスケールアップ可能
- 複数の M, L, XL 構成 (512GB 構成以上) によるスケールアウト構成が可能
- 外部共有ディスクを用いずに高可用性 (HA) 構成やスケールアウト構成が組める分散ファイルシステム搭載
- 高いディスク I/O 要件やスケールアウト構成に対応

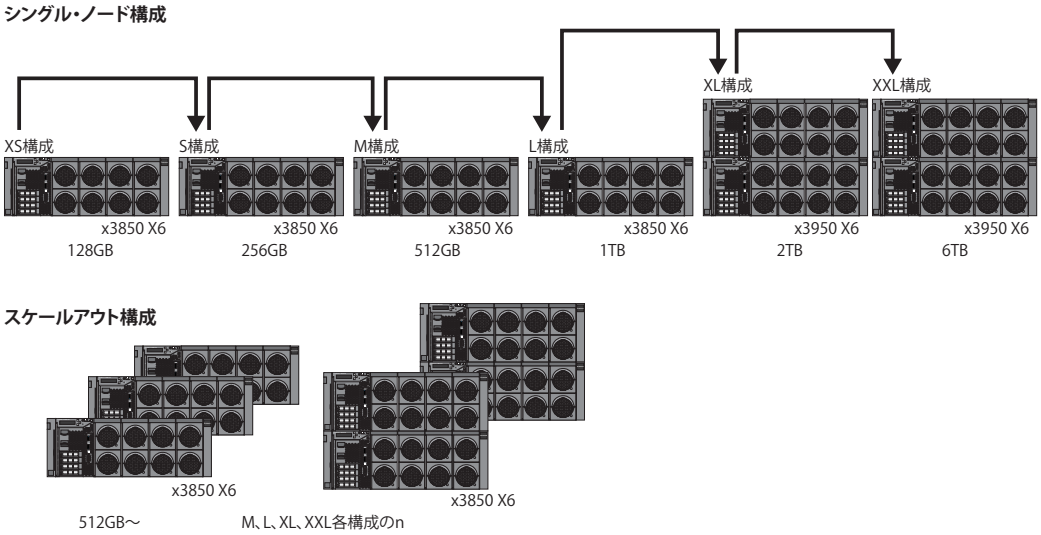


図 2　Lenovo System x Solution for SAP HANA アプライアンス・モデル

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

MapR on Lenovo System x for SAP HANA Vora とは

「MapR on Lenovo System x for SAP HANA Vora」は、ビッグデータ時代のエンタープライズ企業向けの、これまでにないデータ分析用「データレイク」ソリューションです。 Lenovo System x サーバに MapR を搭載する Software-Defined Storage (ソフトウェア・デファインド・ストレージ: SDS) で、さまざまな情報のインプットと分析を可能にします。安価な x86 サーバをベースにした本ソリューションは、SAP HANA や Netezza 、Exadata などの従来のデータウェアハウス (DWH) との連携はもちろん、既存 DWH から分析系基盤への移行や、ETL (Extract、Transform、Load) 処理のオフロード化を実現します。

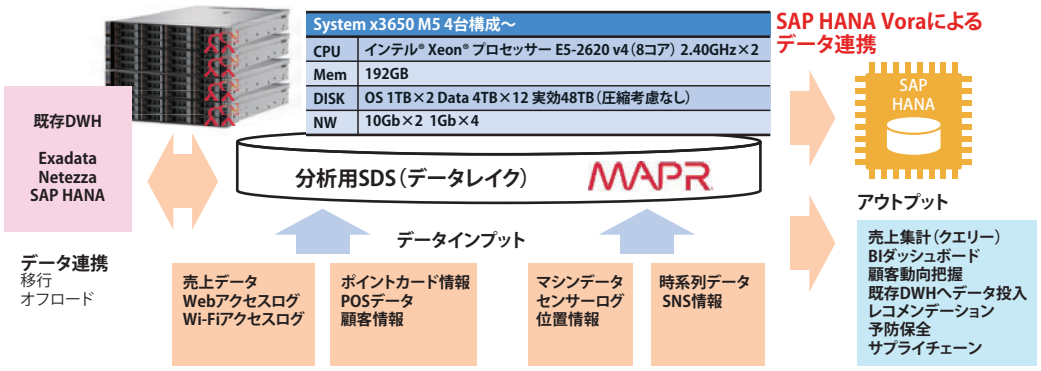


図 3　MapR on Lenovo System x for SAP HANA Vora

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

MapR on Lenovo System x for SAP HANA Vora の特長

- MapR コンバージド・データ・プラットフォームが稼働する Lenovo System x は最小 4 台からデータレイクのニーズに応じて 1 台ずつスケールアウト
- 基幹系、業務系の構造化データと、画像や音声といった非構造化データを一元管理可能な大規模なデータレイクをサポート
- 高速なデータ圧縮による効率化と高速なデータ読み込みを可能にする高いパフォーマンス
- データ抽出 (Extract)、データ変換 (Transform)、データロード (Load) のワークフロー統合を実現し、データのサイロ化の解消
- データバックアップやスナップショットも容易に対応

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

ソフトウェア概要

SAP HANA Vora の主な機能

SAP HANA Vora は単なるビッグデータの保存先のみならず、ビジネスデータと組み合わせて高度な分析を行うプラットフォームとしての様々な機能を提供します。代表的な機能は以下のとおりです。

- SAP HANA および Hadoop 間のデータ統合
- ビジネスデータ分析のための階層 /OLAP 分析
  - Hadoop データに対する階層機能
  - 通貨換算機能
  - データモデラーによる GUI ベースの OLAP データモデリング機能
- データ高度活用のための分析プラットフォーム
  - 時系列分析
  - グラフデータ分析
  - JSON ドキュメントストア
  - ディスク・ストレージ

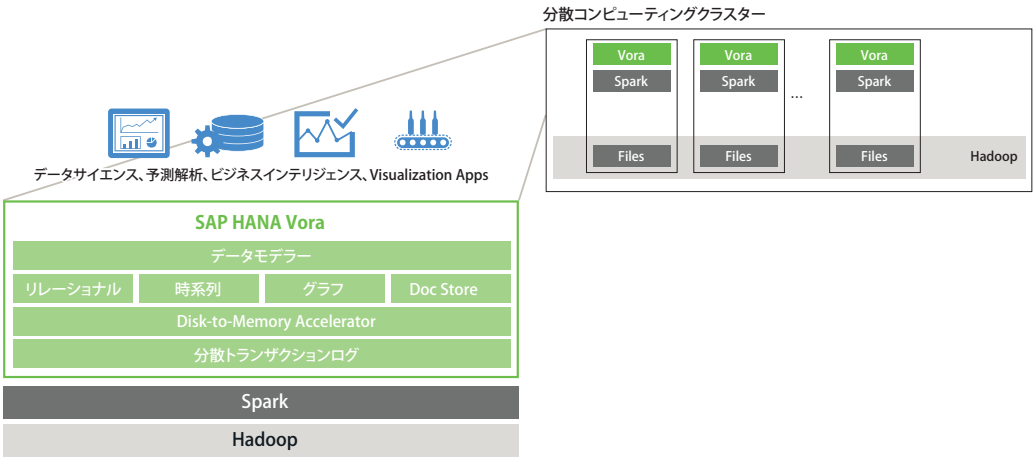


図 4 SAP HANA Vora の主な機能

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

エンタープライズ向け機能

企業データは階層構造で格納されているのが一般的です。したがって、企業データの分析には、階層構造を扱えること、また、グローバル企業にとっては、通貨換算は必須の機能です。分析を行うために、テーブルをモデリングして、スタースキーマを設計する必要があります。SAP HANA Vora は、こうしたエンタープライズ向け機能を提供します。

通貨換算では、Spark の通貨換算関数の他、SAP ERP の通貨換算テーブル基準の通貨換算機能が利用できます。

データモデリングは、データモデラーを利用して、GUI によって直感的に定義することができます。また、ビジネスユーザーは、SAP BusinessObjects Lumira のような対話型のアナリティクス分析ツールを利用して SAP HANA Vora にアクセスすることで、視覚的な分析を行うことができます。

データ高度活用のための分析プラットフォーム

SAP HANA Vora は、格納されたデータに対して、時系列予測やグラフ分析といった高度な分析機能を提供しています。

時系列テーブルとして定義すると、時系列データに適したデータ圧縮が適用されます。また、トレンド予測を行うことができます。

MapR に統合されている SAP HANA Vora

SAP HANA Vora は、MapR Converged Data Platform に統合されています。

インストーラも MapR に対応しており、既存の MapR クラスタに追加してインスールが可能です。管理ツールである MapR Control System と統合されています。

SAP HANA Vora は、Spark のプラグインとして動作します。必ずしも、SAP HANA を必要としません。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

MapR の概要とコンポーネント

MapR 概要

MapR Converged　Data Platform は、ビッグデータや IoT 時代においてデータ活用に必要な各種の機能を 1 つの製品に統合したデータ基盤ソフトウェアです。MapR に包含する機能は、データを格納し保護するストレージ機能、バッチ活用の機能である Hadoop、リアルタイム活用の機能である Spark、さらに NoSQL データベースの MapR-DB、そしてストリーミングデータを収集するイベントストリーミング機能の MapR Streams が含まれています。これにより、お客様は例えば、データ活用ニーズがバッチからリアルタイムに変わったとしてもシステムを作り直したり、追加システムを構築することなく、機能をアクティベートするだけで対応できます。そして MapR にはオープン API が備えられ、SAP HANA や HANA Vora を始めとした商用のデータ分析、データ活用のソフトウェアと容易に連携ができ、お客様のシステムをロックオンしません。さらに、MapR は昨今注目されている x86 サーバ上で稼働するソフトウェアデファインド型で、データ量や処理に応じて容易にスケールアウトが可能です。

SAP HANA のデータ基盤を考えた際に、単なるストレージだとコストおよび、Spark や Hadoop の機能がないことによるデータの活用性に問題があり、それらを考慮して MapR 以外の Hadoop を活用すると、単一障害点の存在やパフォーマンスボトルネックの問題が残ります。MapR は業界で唯一これらの問題点を根本から解決してあるため、SAP HANA のデータ基盤として、コストやパフォーマンスはもちろんのこと、ミッションクリティカル性にも対応でき、安心して活用できる唯一のデータ基盤となっています。

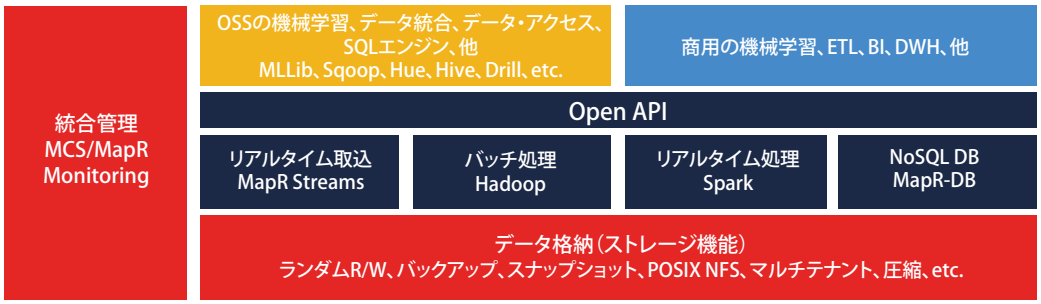


図 5 MapR の概要

SAP HANA および SAP HANA Vora のデータ基盤として MapR を利用する際には、MapR の以下のコンポーネントが必要となります。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

SAP HANA の基盤として	MapR Converged Data Platform 5.2 Enterprise Edition Base Module-Hadoop POSIX Client
SAP HANA Vora の基盤として	MapR Converged Data Platform 5.2 Enterprise Edition Base Module-Hadoop POSIX Client Apache Spark
Linux OS の要件	http://maprdocs.mapr.com/home/InteropMatrix/r_os_matrix.html

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

SAP HANA Vora ユースケース

SAP HANA Vora 活用の 3 つのシナリオ

SAP HANA Vora は、単独で活用することができます。加えて、インメモリ統合プラットフォーム SAP HANA と連携したり、SAP S/4HANA のような基幹系システムと相互に データ活用を促進することができます。

活用シナリオには以下の3つが考えられます。

タイプ 1: 新しいインサイトの獲得
タイプ 2: ビジネスコンテキストを追加
タイプ 3: データドリブンアプリケーションの構築

タイプ 1: 新しいインサイトの獲得

タイプ 1 は、SAP HANA Vora の単独利用です。SAP HANA Vora は、分析機能を提供しているため、別途 Hadoop エコシステムの様々な製品を追加して導入することなく、即座に高度な分析が可能です。

タイプ 2: ビジネスコンテキストを追加

タイプ2は、エンタープライズデータとビッグデータのマッシュアップです。基幹システムである SAP S/4HANA に記録された売上等の経年データと、SAP HANA Vora に格納されたソーシャルデータをマッシュアップして分析の精緻化を行うことができます。
例えば、ファッションアイテム等のトレンドに左右される商品に関しては、新製品の発売後、ソーシャル・ネットワークで話題にのぼる傾向にあるものは、売上が期待できます。ソーシャルでの出現頻度と売上実績を比較することで、今後の販売計画の予測をより確かなものにすることができます。

タイプ 3: データドリブンアプリケーションの構築

タイプ3は、ビジネスデータの意思決定を Hadoop からのデータとマッシュアップして精緻化します。ホット・データはインメモリ・プラットフォーム SAP HANA に格納しておき、過去のデータはビッグデータのコンバージドプラットフォームである MapR の Hadoop に保存しておきます。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

データ階層管理

データ階層管理とは

SAP HANA は統合インメモリ・プラットフォームです。データ管理についても高度な機能を提供しています。最新データや頻繁に参照されるデータを「ホット・データ」としてインメモリ上に展開しつつ、過去の売上データなど履歴データを「ウォーム・データ」としてディスク・ストレージに保存したほうがコスト効率が高い場合があります。SAP HANA は、ダイナミックティアリング機能によりマルチストアテーブルを提供しています。これにより、メモリ・ディスク間でストレージを動的に管理することができます。また、センサーデータなどの大量データは MapR ファルシステムのような分散ストレージに「ビッグデータ」として格納するほうが好ましい場合もあります。このように、データを階層で管理することで、コスト効率の高いストレージ環境を構築することができます。

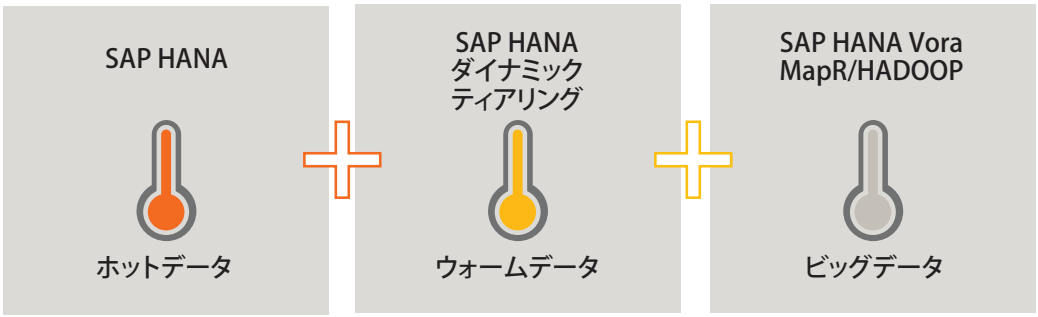


図 6　データ階層管理

ダイナミックティアリングによるウォーム・データの階層管理

SAP HANA において、ダイナミックティアリング機能を利用することで、テーブルの保存先としてメモリに最適化されたパーティションに加えて、従来型のディスクを選択することが可能になりました。SAP は Sybase の持つカラム型の特許を多数保有しており、これを活用してディスクに最適化されたパーティションも高速に扱うことができます。マルチストアテーブルとして定義されるため、データベースのカタログオブジェクトは単一のテーブルとして認識され、ユーザーにデータがメモリにあるのかディスクにあるのかを全く意識させません。直近の頻繁に参照するデータをメモリ・パーティションに常に展開しておき、履歴データはディスク・パーティションに保存しておくことで、効率的なデータ格納が行えます。また、メモリ・ディスク間の保存先の変更は必要に応じて双方向にいつでも行うことができます。マルチストアテーブルであっても、バックアップ（フル、増分、差分）、データの暗号化をサポートしており、信頼性が求められるエンタープライズの要求に応えます。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

ビッグデータでの階層管理

SAP HANA Spark Controller を利用することで、SAP HANA から、SAP HANA Vora に対して、透過的なデータアクセスが可能になります。SAP HANA は、Smart Data Access と呼ばれる、データソースに対して仮想的なデータアクセス機能を提供しています。SAP HANA からは、外部データソースのテーブルを Virtual Table として定義することで、同様にデータ操作を行うことができます。ユーザーはデータの場所を気にせず、複数のデータソースからのデータを組み合わせて分析を行うことができます。

これにより、MapR のファイルシステム上に展開されたテーブルと、SAP HANA 上のテーブルを結合してクエリ実行が可能となり、ビジネスデータとビッグデータの横断的な分析が行えます。

階層に分かれたデータを SAP HANA スマートデータアクセスで統合

データ階層管理を行った場合でも、ユーザー視点で、データのアクセス先を考慮する必要はありません。

SAP HANA のインメモリ・テーブルは、マルチストアテーブルの構造を採ることで、ユーザーがインメモリかディスクかを考慮することを不要にしました。ビッグデータに保存されたテーブルは、スマートデータアクセス機能により、ヴァーチャル・テーブルとして定義可能です。SAP HANA Data Warehouse Foundation 機能を利用することで、インメモリ・テーブルとヴァーチャル・テーブルをユニオン・ビューとしてまとめることができます。ビジネスユーザーは、アナリティクスツールから、このユニオン・ビューに対してアクセスすることで、透過的に横断的な分析が可能になります。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

アナリティクス

SAP BusinessObjects Lumira

誰もが使えるセルフサービス方式のデータビジュアル化環境

SAP BusinessObjects Lumira 概要

SAP BusinessObjects Lumira でデータをフルに使ってビジネスに活かしましょう。データビジュアル化ソフトウェアを活用すると、見栄えがよい対話操作型のマップ、チャート、インフォグラフィックを簡単に作成できます。この環境では、Excel をはじめとする多様なソースからデータを取り込み、直感的なダッシュボードでビジュアル BI 分析を実行し、得られた洞察をストーリーに沿ってまとめ、安全な環境で共有することが可能です。

SAP BusinessObjects Lumira, server edition なら、セルフサービス方式で利用できるビッグデータのビジュアル化機能を、組織全体に 1 つのプラットフォームで提供できます。ユーザーとアクセス権限を管理するためのツールも完備しているため、安全に洞察を共有することが可能です。

SAP BusinessObjects Lumira チーム用サーババージョンは、中堅・中小企業の全社規模または大企業の部門単位での利用に適した、お求めやすい価格でパワフルな機能を提供するソリューションです。動作の軽いインメモリサーバが採用されており、データの分析、ビジュアル化、共有がいずれも高速に処理されるため、より短時間で洞察を獲得することができます。

SAP BusinessObjects Lumira の無料パーソナル版をダウンロードして、今すぐビジュアルデータ分析を始めましょう。この環境では、さまざまなデータベースに簡単に接続し、説得力あるデータビジュアル表現を作成し、得られた知見をほかの人と簡単に共有することができます。

SAP BussinessObjects Lumira の対応するデータソース

SAP BussinessObjects Lumira は多様なデータソースに対応しており、SAP HANA や SAP HANA Vora に対して直接アクセスし、対話型分析を行うことができます。SAP HANA のスマートデータアクセス機能と SAP HANA Vora を利用してデータソースを仮想統合しておけば、分析ユーザーはデータの保存場所を意識せずに分析が可能です。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

ソリューション・コンポーネント

本検証で評価を行った Lenovo System x Solutions for SAP HANA のアプライアンスと、MapR on Lenovo System x for SAP HANA Vora　を透過的に連携するソリューションのダイアグラムは下記の通りです。

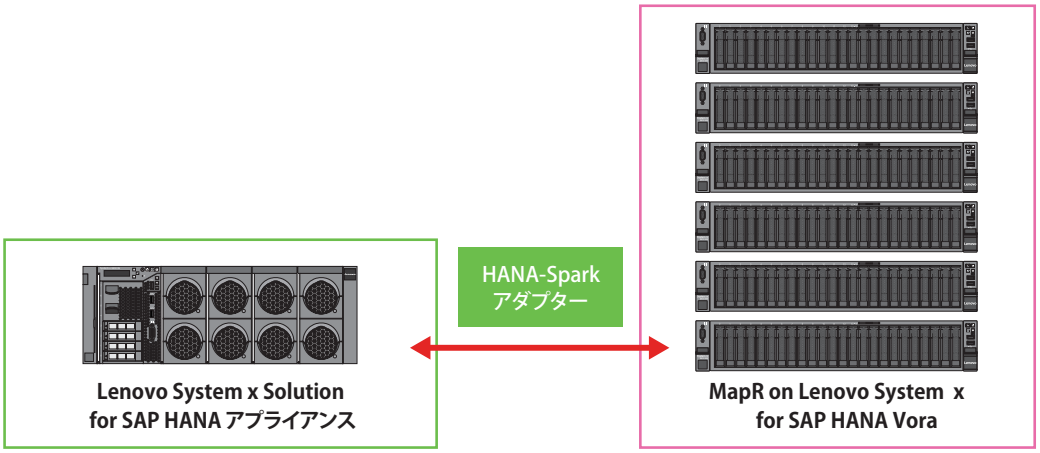
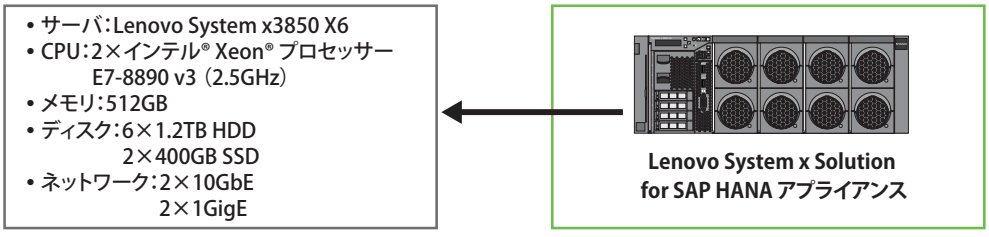


図7:　ソリューション・コンポーネント

ハードウェア・コンポーネント

Lenovo System x Solution for SAP HANA アプライアンス



※構成詳細は参考資料に添付いたします

図8:　ハードウェア・コンポーネント

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

Lenovo System x3850 X6　のハードウェアの特長

Lenovo System x3850 X6 はユニークな「ブック型」モジュラーデザインを持ち、最新のプロセッサーとメモリ技術をサポートした SAP HANA アプライアンスとして認定された、シームレスな拡張性と最大のパフォーマンスを実現するシステムです。

- 卓越した計算パフォーマンス
- メモリ可用性を最大化 (1 サーバあたり最大 12TB)
- 優れたパフォーマンスの大容量ストレージ
- 増加するシステム需要に対応できる拡張性
- 同一シャーシでの複数世代のテクノロジーを利用可能
- 自己回復力のあるプラットフォームによりアップタイムを最大化

MapR on Lenovo System x Solution for SAP HANA Vora　推奨構成

当構成はサーバ 4 台での推奨構成です。約 50TB のデータ量に対応いたします。

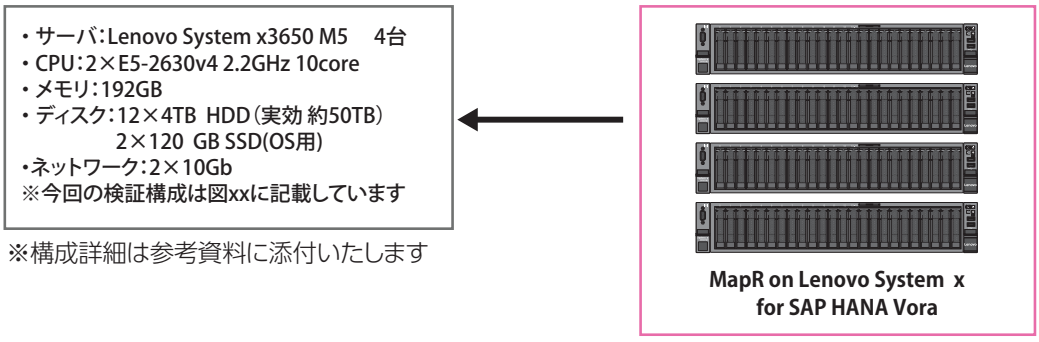


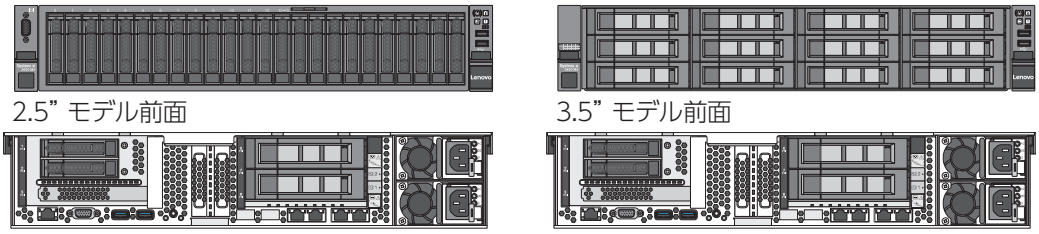
図9: MapR on Lenovo System x Solution for SAP HANA Vora

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

Lenovo System x3650 M5　のハードウェアの特長

Lenovo System x3650 M5 は、パワフルで優れた汎用性を誇る 2 ソケットの 2U ラック・サーバです。x3650 M5 は、最大 2 個のインテル® Xeon® プロセッサー E5-2600 v4 製品ファミリーと、2 ソケット・サーバでは屈指の大容量ストレージ（最大 148 TB）、および HDD/SSD/NVMe や 2.5/3.5 インチ等の豊富なストレージ・プラットフォームの選択肢により、クラウドからビッグデータまで、あらゆるワークロードを最適化し、ビジネスを加速させることが可能です。

大容量のディスク構成をとることができるのが最大の特徴のサーバで、最大で前面に 2.5”ドライブを 24 本、背面に 2.5i”ドライブを 2 本と 3.5”ドライブを 2 本まで搭載可能な 2.5”タイプのモデルと、最大で前面に 3.5”のドライブを 12 本と、背面に 2.5“ドライブを 2 本と 3.5”ドライブを 2 本まで搭載可能な 3.5”タイプのモデルがあります。



2.5”モデル前面　3.5”モデル前面

2.5”モデル背面　3.5”モデル背面

図10: Lenovo System x3650 M5 のハードウェア外観

さらに、搭載するドライブの種類もさまざまなタイプを搭載することができ、サポートされるドライブは下記の通りです。

- 400GB ～ 2TB NVMe（PCIe/2.5”Bay)
- 120GB ～ 3.8TB　SSD(2.5”/3.5”Bay)
- 300GB ～ 2TB SAS/SATA（2.5”Bay)
- 1TB ～ 10TB（3.5”Bay)

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

Lenovo System x3650 M5　運用管理の特長

サーバに搭載するドライブの設定を、専用の管理チップ Integrated Management Module (IMM) の GUI 画面を通じて容易に操作できます。万一の障害時にも、IMM ポートから、システムログ (syslog) や SNMP トラップ、メールでの通知が可能なため、柔軟な運用管理できるとともに、高い耐障害性を実現可能です。

統合管理ツールである、Lenovo XClarity を使用することで複数台のサーバの統合管理や Firmware の更新を行うことも可能です。

また、一般的な Hadoop 環境の場合、管理専用のサーバとして、Management Node が必要になります。本評価で採用した MapR on Lenovo System x Solution for SAP HANA Vora ソリューションなら、すべてのノードをリソースとして使用できるため管理専用サーバが不要となり、システム全体として、サーバ台数の節約が可能です。

SAP HANA と MapR for SAP HANA Vora を連携するネットワーク

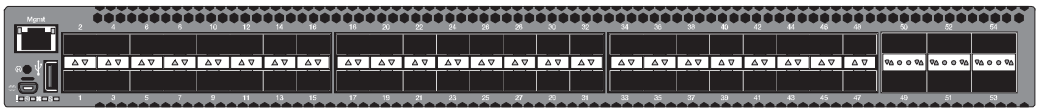


図11: Lenovo RackSwitch G8272 の外観

Lenovo ネットワークスイッチのハードウェアの特長

機能豊富な Lenovo® RackSwitch™ G8272 は、10 Gbps Ethernet (SFP+) × 48 ポートと 40 Gbps Ethernet (QSFP+) × 6 ポートを柔軟に構成可能な 1U スイッチです。SFP+ ポートは必要に応じて 1 Gbps Ethernet (SFP) 接続をサポートします。消費電力がわずか 123 W の G8272 は、同等のスイッチ 1 よりも電力を節約できます。また、標準装備の冗長電源と冗長冷却ファンにより、信頼性と効率性を高められます。600 ナノ秒未満という低遅延と優れた 3 年間保証により、G8272 は、低コストでありながら理想的なパフォーマンスを実現する選択肢となります。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

また、L2 レベルでの冗長化はスパニングツリーはもちろん、VLAG 機能によりスイッチをまたいだリンクアグリゲーションが可能ですので、サーバとスイッチ間の帯域を有効活用することができます。最新の Network 機能にも対応しており、OpenFlow、VXLAN Gateway、NSX、FCoE、ONIE 対応などの機能に対応しており柔軟で安価にお客様環境に対応したネットワークを構築することが可能です。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

ソリューション構成例

本セクションでは、SAP HANA Vora の各種サービスと、ランドスケープ内にどのように適切に配置すべきかを紹介します。

Vora Manager により、Vora 各サービス構成の管理を行います。

サービス名	機　能
Consul（HashiCorp）	Consulはディスカバリーサービスに使用され、各サービスのハートビートチェック等を行います。
Nomad（HashiCorp）	Nomadはスケジュールとリソース管理を行います。
SAP HANA Vora UI Manager	SAP HANA Vora UI Managerは、各種Voraサービスのステータス状況を表示、サービスの開始・停止、設定等を行うことができます。

SAP HANA Vora は以下のようなサービス群から構成されています。

サービス名	機　能
Vora Distributed Log	SAP HANA Voraカタログ用のパーシスタンスを提供するDistributed Log用のマネージャーサービス
Vora Catalog	SAP HANA Voraのメタデータストア
Vora Disk	データをディスクに保存するエンジン
Vora Document Store	ドキュメントを処理するエンジン
Vora Graph	グラフデータを処理するエンジン
Vora In-memory Engine Server	SAP HANA Voraのリレーショナルインメモリエンジン
Vora Landscape Server	データのパーティショニングやデータベースエンジンの配置を管理するサービス
Vora Thriftserver	HIVE JDBC ドライバーへのゲートウェイサービス
Vora Time Series	時系列データを処理するエンジン
Vora Transaction Manager	ユーザートランザクションを管理
Vora Transaction Coordinator	(メタ)データの一貫性を担保するサービス
Vora Transaction Lock Manager	ユーザーセッションにおけるデータベースエンジンのクエリ実行のドライバー

ノードの種類	説　明
管理ノード	クラスタプロビジョニングツール用のノード(MapR Control System)
マスターノード	クラスタのマスターノード、CLDB(ネームノード相当)やZooKeeperが稼働する
ワーカーノード	クラスタの計算ユニットが稼働するノード、データノード、ノードマネージャ
ジャンプボックス	クライアントツールのみが稼働するノード

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

サービスは、以下の情報に基づき、ランドスケープ内に適切に配置します。

サービス名	稼働ノード
Vora Manager	クラスター内のすべてのノードに配置 □マスターノード（可用性のため複数マスターノード） □ワーカー（マスター以外のすべてのワークノード） □クライアント（すべてのノードに配置）
Vora Distributed Log	マスターノードに配置　本稼働環境では5以上を推奨、可能な限りすべてのノード、上限なし
Vora Catalog	単一ノードに配置（マスターノード、Distributed Logの稼働するノード）
Vora Disk	1ノード以上に配置
Vora Document Store	1ノード以上に配置
Vora In-memory Engine Server	各Workノードに配置
Vora Landscape Server	1ノード以上に配置
Vora Thriftserver	1ノード以上に配置、ジャンプボックス
Vora Time Series	1ノード以上に配置
Vora Tools	1ノード以上に配置、ジャンプボックス

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

検証済み構成とサイジングのガイド

SAP HANA Vora を利用する場合、クラスターノードはそれぞれ以下の最低限の必要条件を満たしている必要があります。

- 4 コア 以上
- RAM 8 GB 以上
- 20 GB 以上のストレージ容量 (MapR-FS 用)

今回の検証環境

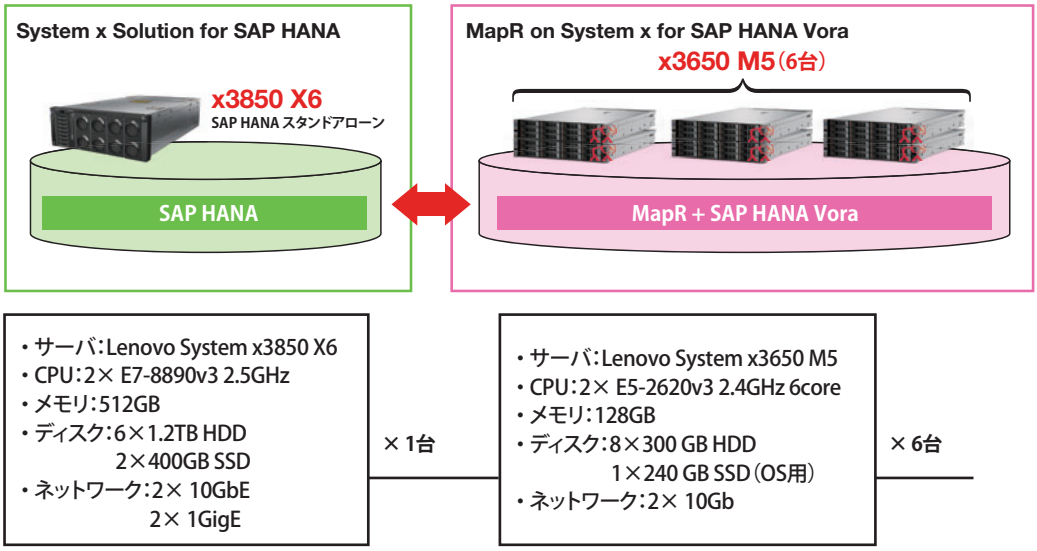


図12: 今回の検証環境

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

SAP HANA Vora 用 MapR 推奨構成

- 12 コア以上
- RAM 128GB 以上（Spark 利用）
- 内蔵ディスク 10 本以上（容量に応じたトータルディスク容量）

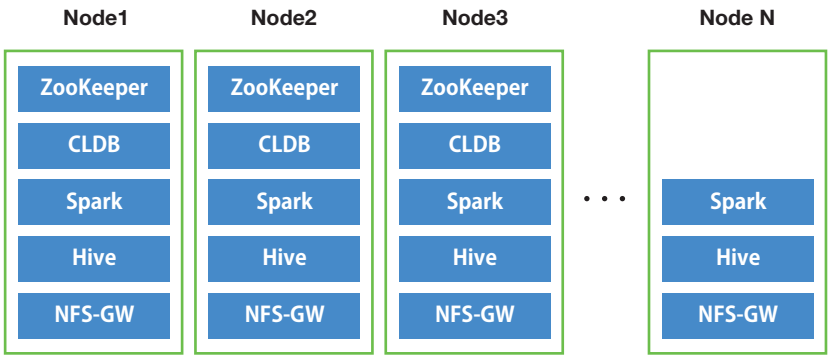


図13: HANA Vora用 MapR 構成例

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

SAP HANA Vora と MapR 連携検証について

検証構成

SAP HANA Vora の検証構成として Lenovo 社製の 6 台のサーバに検証実施時点で最新版の SAP HANA Vora 1.2 を導入しています。今回は検証用途のため 6 台のサーバの役割として下図のようにノード 1 についてはジャンプボックス、管理ノード、マスターノード及びワーカーノードの全ての機能を割り当て、ノード 2 ～ 6 にはマスターノード及びワーカーノードの機能を割り当てています。また、SAP HANA から SAP HANA Spark Adapter を介した SAP HANA Vora との連携を行うために、SAP HANA Vora の全てのノードに SAP HANA Spark Controller 1.6 Patch Level 1 を導入しています。Windows サーバには SAP BussinessObjects Lumira と SAP HANA Studio を導入し SAP HANA 及び SAP HANA Vora のクライアントとして活用しています。Vora Tools や MapR Control System（MCS）に Web UI を用いてアクセスする際にもこのクライアントを活用しています。

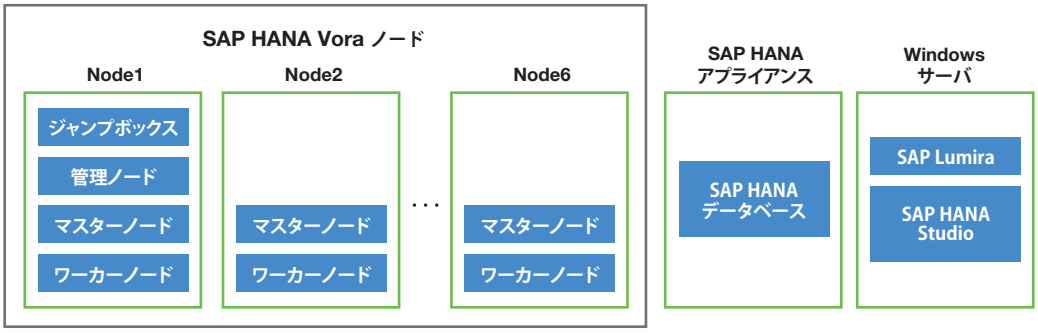


図14: SAP HANA Vora と MapR の検証構成

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

データロードとテーブル作成

本検証では SAP HANA 上のテーブルから CSV ファイルとしてエクスポートしたデータを利用しています。その際に MapR FS の NFS 機能を用いて Hadoop 上の大容量のディスク領域内に数百 GB 単位の CSV ファイルを直接的に配置し、必要に応じてデータ加工を行い SAP HANA Vora のデータソースとして利用しています。これらのファイルに対して Vora Tools の SQL Editor 機能を用いて、Create Table ステートメントを実行することで、SAP HANA Spark sqlContext 及び SAP HANA Vora エンジン内にテーブルが作成されます。その際に適切なファイルパス、ファイル名、テーブル名が構成オプションとして必要になります。

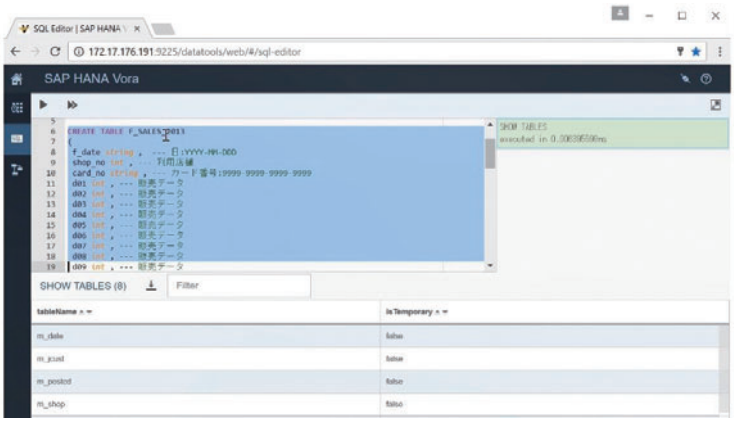


図15: データロードとテーブル作成

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

ビューの開発

Vora Tools の Web UI から Modeler の機能を使って SAP HANA Vora 上に作成済みのテーブルを選択して GUI ベースによる OLAP のモデリングが可能になっています。SQL Editor 上で SQL ステートメントを記述してビューを作成することも可能ですが、SAP HANA Studio のモデリング環境や各種 BI ツールの GUI ベースの開発環境に馴れ親しんでいる開発者にとっては、Modeler による開発支援環境は大きなメリットがあると考えられます。

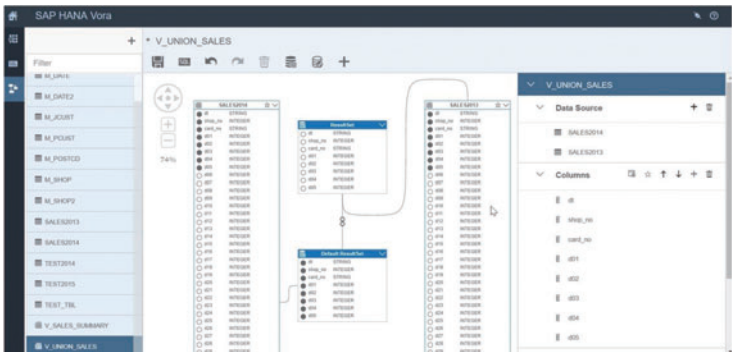


図16: ビューの開発

Vora Tools と Lumira によるデータアクセス

SAP HANA Vora で標準的に利用可能なデータ参照用のツールとして Vora Tools と SAP 社のデータ ディスカバリー ツールである Lumira があります。Vora Tools の SQL Editor 画面からは SQL ステートメントを用いたクエリが実行できます。Lumira を用いる場合は Simba technologies 社が提供する JDBC ドライバーをセットアップして Generic の JDBC ドライバーとして Vora にアクセスすることができます。

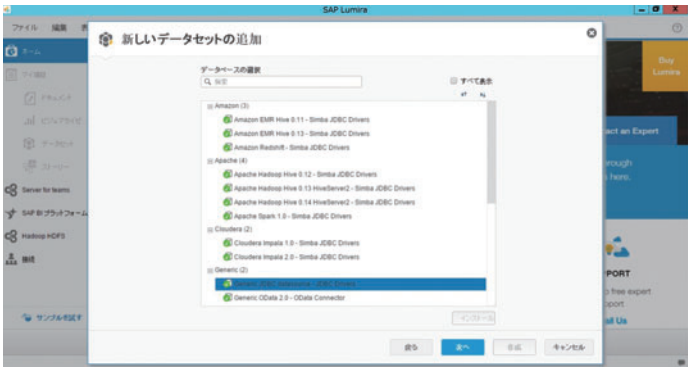


図17: SAP Lumira

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

これにより、SAP HANA や他の DWH などと同様に従来より BI ツールを用いてデータ分析を行っていた環境と同様に視覚的なデータ分析が可能になります。また、クラウド ベースの DWH にアクセスする際に用いるようなフリーの SQL クライアント ツールを用いた SAP HANA Vora へのアクセスもできるので、必要に応じたツールの選択が可能です。

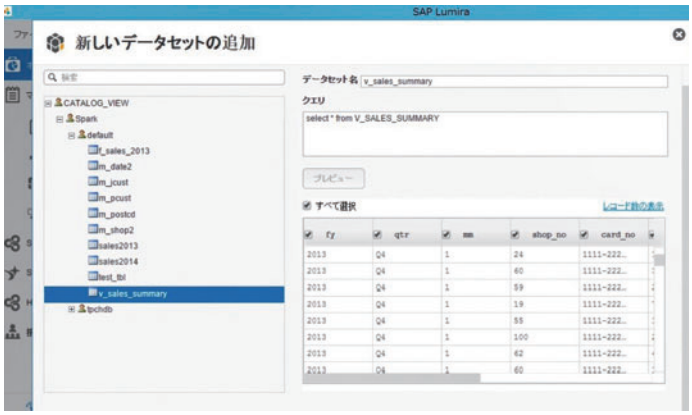


図18: SAP Lumira によるデータアクセス

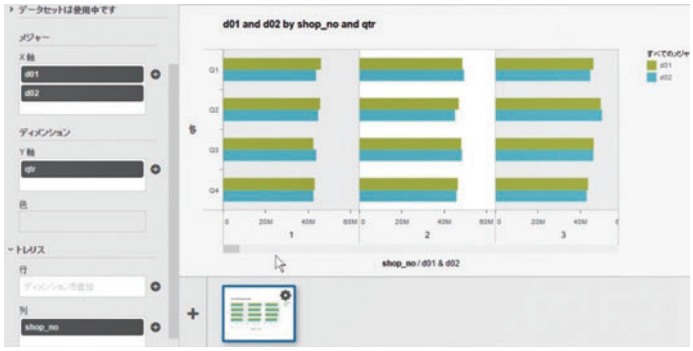


図19: SAP Lumira によるデータアクセス

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

SAP HANA Studio による仮想テーブルの構成

SAP HANA のスマート データ アクセス機能により SAP HANA Spark Controller を用いた SAP HANA と SAP HANA Vora 間の連携が実現され、SAP HANA Vora 上のテーブルを SAP HANA の仮想テーブルとして定義できようになります。実体となるデータは SAP HANA Vora 上に存在し、クライアント ユーザーからは SAP HANA を介して透過的に SAP HANA Vora 上のデータにアクセスすることができるようになるため、複数のデータソースに対して SAP HANA Studio から単一データソースのイメージでデータモデルを管理することができるようになります。

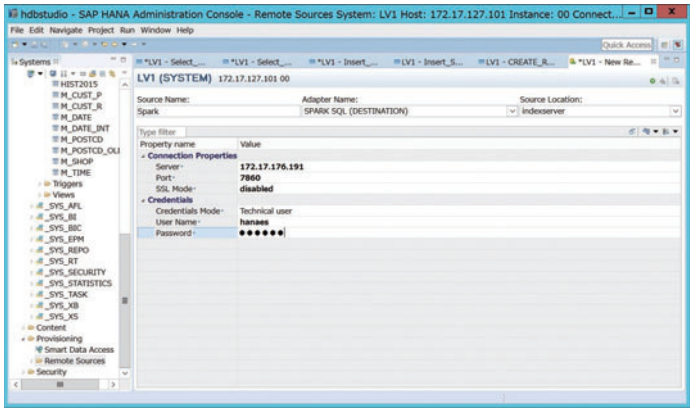


図20: SAP HANA Studio

この仕組みを用いて SAP HANA 上には利用頻度の高いホット・データを配置し、SAP HANA Vora 上には利用頻度の低いコールドデータをアーカイブとして維持するようなデータの階層化を行うことで SAP HANA のメモリ リソースを有効活用しつつビッグデータ分析基盤を効率的に維持管理することができるようになります。

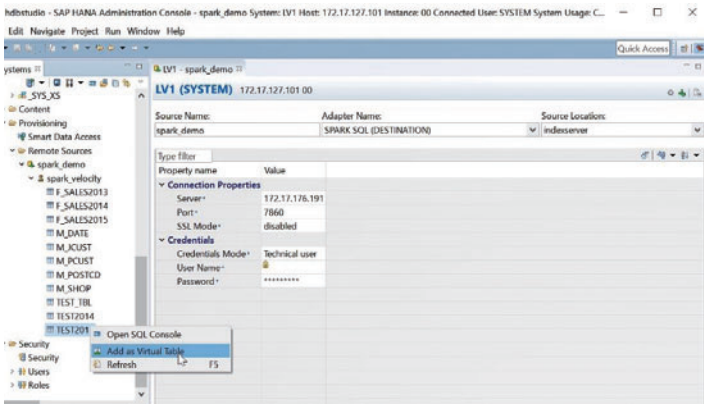


図21: SAP HANA Studio

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

まとめ

ビッグデータ分析や IoT への対応をはかる際に、データの温度は 1 つの重要なクライテリアであるが、温度毎にデータの保存場所が異なるとコストの上では効果的だが、活用の上では効率が落ちます。リアルタイムにアジャイル分析を行う対象となる最も温度の高いホット・データを SAP HANA に、量や種類が多く長期間のデータや保存義務のあるデータ、さらに生データといったコールドデータを MapR に蓄積します。当検証では、SAP HANA Vora を活用することで、データがそれらのシステムに分かれていても、透過的かつ統合的にデータへのアクセス、つまり分析ができることが確認できました。そして、この統合分析プラットフォームを検討する企業が、稼働検証をせずとも迅速に導入プロジェクトを開始できるよう検証済のレファンレンス・アーキテクチャを作成しました。

データ活用を検討する企業は、データの温度によってシステムを変えることでコスト効率を上げながら、あらゆるデータを透過的かつ統合的に活用できるこの統合分析プラットフォームによって、大きなビジネスメリットを得られると確信しています。さらに、レファンレンス・アーキテクチャ化により、導入期間も短縮することができることを期待しています。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

SAP ジャパンについて

SAP ジャパンは、エンタープライズ・アプリケーション・ソフトウェアにおけるマーケットリーダーとしてあらゆる業種におけるあらゆる規模の企業を支援している SAP SE の日本法人として、1992 年に設立されました。SAP は、企業が市場での優位性を保持するため、バックオフィスや役員会議室、倉庫や店頭で、さらにデスクトップ環境やモバイル環境などにおいて、企業がより効率的に協業を行い、より的確なビジネス判断を行うための様々なソリューションを提供します。企業が継続的な収益性の高い事業を実現することに貢献する SAP のアプリケーションやサービスは、世界各国 296,000 社以上の顧客企業に利用されています。国内でも日本企業の情報化の推進、国際競争力および企業価値の向上に貢献しています。

マップアール・テクノロジーズについて

MapR は、トップランクの Hadoop および Spark をグローバルイベントストリーム、リアルタイムデータベース、エンタープライズストレージに統合した業界唯一のコンバージド・データ・プラットフォームを提供し、顧客がデータの持つ巨大なパワーを活用できるように支援しています。ミリ秒単位の応答が求められる不正検知、安全で高可用性が求められるヘルスケアにおけるデータ主導の洞察、侵入検知のためのペタバイト規模の分析、カスタマーエクスペリエンス向上のための業務処理と分析処理の統合など、厳しい要件を伴う製品ニーズを持つ企業が MapR を利用しています。ほとんどの顧客は 12 ヶ月以内に投資の回収を達成し、5 倍以上の ROI を実現しています。MapR は、ワールドクラスのプロフェッショナルサービスと、4 万 5 千人の開発者、データアナリスト、システム管理者がビッグデータスキルの習得に利用した無料オンデマンドトレーニングを通して、顧客の成功を確実なものにします。MapR は米国カリフォルニア州に本社を置き、東京都千代田区に日本法人を構えています。

■ お問い合せ先
マップアール・テクノロジーズ株式会社 営業部　Tel：03-6386-0920　Email：sales-jp@mapr.com

作成者：
SAP ジャパン株式会社

発行元：
SAP ジャパン株式会社　　東京都千代田区麹町 1-6-4 SAP ジャパンビル
電話番号：03-6737-3000（代表）・電子メール：info.japan@sap.com・Web：http://www.sap.com/japan/
2017 年 1 月

Copyright © 2017 SAP Japan. All rights reserved.
SAP、SAP ロゴ、記載されているすべての SAP 製品およびサービス名はドイツにある SAP SE やその他世界各国における登録商標または商標です。
またその他記載された会社名およびロゴ、製品名などは該当する各社の登録商標または商標です。

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

参考資料：System x サーバ構成詳細

Lenovo System x Solution for SAP HANA アプライアンス・モデル

型番	製品名	数量
6241HXJ	Systemx3850X6/XeonE7-8880v3(18)2.30GHz-1866MHz×2/PC3L-12800 512.0GB(32×16)(Chipkill)/HSSAS 10K7.2TB(1.2×6)+800GB(400×2)/RAID-M5210/RAID-5200-2GF/MLNX-CX-3/POW(1400W×4)/OSなし/SAP HANA SW (ライセンスは別途入手ください) /3年保証24×7(CRU)/SS90	1

MapR on Lenovo System x Solution for SAP HANA Vora 推奨構成 (約50TB)

型番	製品名	数量
8871D4J	System x3650 M5(HS 3.5)/XeonE5-2630v4(10) 2.20GHz-2133MHz×1/PC4-19200 16.0GB(16×1)(Chipkill)/RAID-M5210/RAID-5200-1GF/POW(750W×1)/OSなし/3年保証24×7(CRU)/SS90	4
00YJ198	XeonE5-2630v4(10) 2.20GHz-2133MHz	4
46W0825	8GB 2R PC4-19200 RDIMM	32
46W0829	16GB 2R PC4-19200 RDIMM CK	28
00FN143	4TB 7.2K 6Gbps NL SATA 3.5型 Gen2 HS 512e HDD	48
00WG620	S3510 120GB SATA 2.5型 Gen3 HS MLC Enterprise Entry SSD	8
00KA498	PCIe ライザー・カード (2x8 FH/FL + 1x8 FH/HL)	4
00FK658	追加2HDDキット(2.5型/背面ベイ)	4
46C9114	ServeRAID M1215 SAS/SATA コントローラー	4
00AG570	Emulex VFA5.2 2x10 GbE SFP+ PCIe アダプター	4
00FK932	高効率 750W リダンダント電源機構	4
46M2593	NEMA5-15P to IEC C13 電源ケーブル (2.8m)	8

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載

Lenovo SAP HANA アプライアンスについてのお問い合わせ・資料請求

レノボ・ジャパン株式会社

〒101-0021　東京都千代田区外神田四丁目14番1号　秋葉原UDX

0120-68-6200　受付時間：月曜日～金曜日 9:00～17:30
（土、日、祝日、5月1日、12月30日～1月3日を除く）

hojin_jp@lenovo.com

http://www.lenovojp.com/business

Lenovo、レノボ、レノボロゴ、ThinkCentre、ThinkPad、ThinkStation、ThinkServer、New World New Thinking、ThinkVantage、ThinkVision、ThinkPlus、TrackPoint、Rescue and Recovery、UltraNavは、Lenovo Corporationの商標。
Intel、インテル、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Intel Atom、Intel Atom Inside、Intel Core、Core Inside、Intel vPro、vPro Inside、Celeron、Celeron Inside、Itanium、Itanium Inside、Pentium、Pentium Inside、Xeon、Xeon Inside、Ultrabook は、アメリカ合衆国および/またはその他の国における Intel Corporation の商標です。他の会社名、製品名、サービス名等は、それぞれ各社の商標または登録商標。

Lenovo.

レノボ・ジャパン株式会社

〒101-0021 東京都千代田区外神田四丁目14番1号 秋葉原UDX

http://www.lenovojp.com/business/

Intel Inside®
パワフルなデータセンターを
インテル® Xeon®
プロセッサー搭載