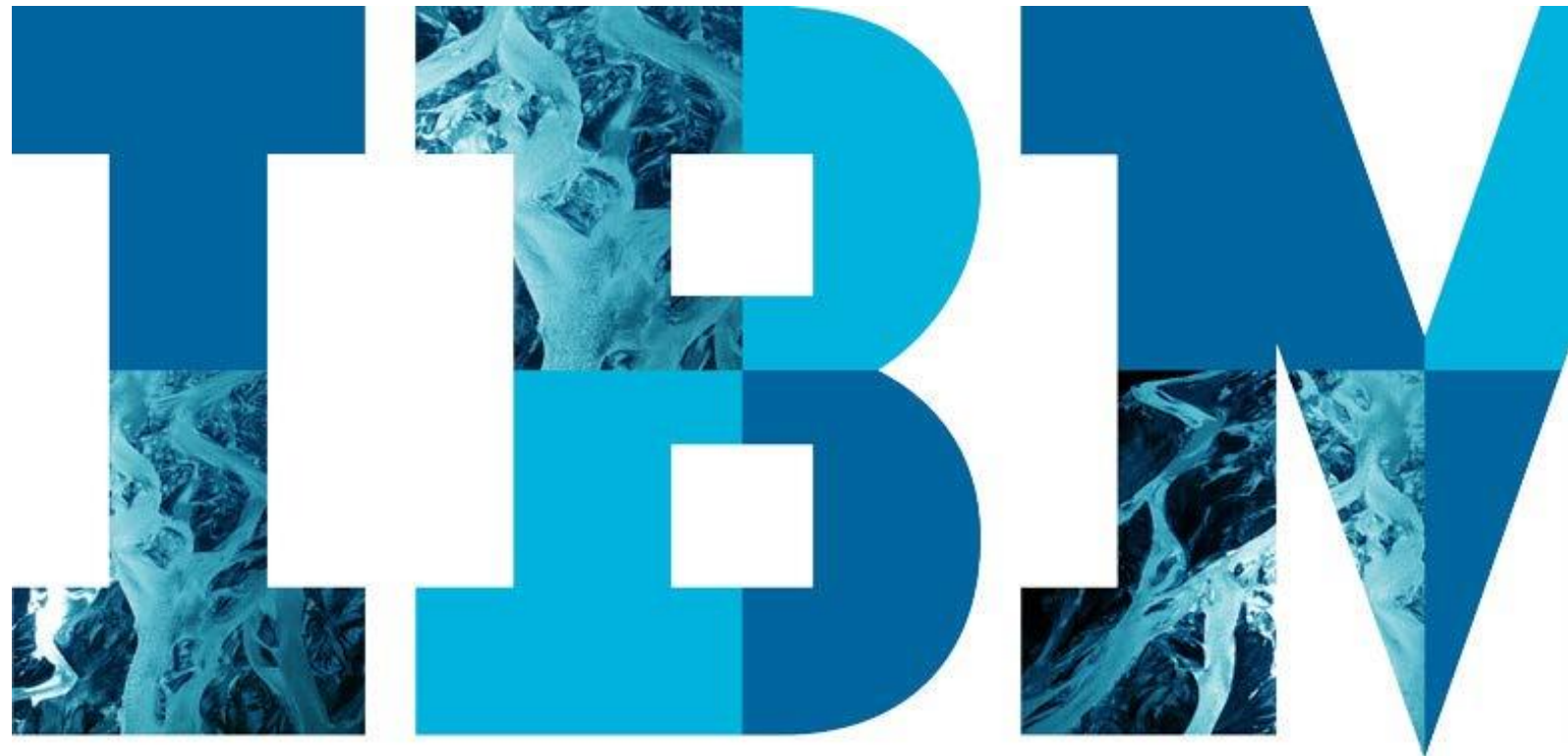


流動的データ層: これからのデータ・マネジメント

ハイブリッド・クラウドのビジネス環境において、
柔軟性と適応性がデータ・アーキテクトにとって不可欠な理由



1

データ・アーキテク
トにとっての
新たなビジネス
環境

2

流動的データ層
が不可欠な理由

3

必須要件

4

実世界における
流動的データ層

5

ハイブリッド型の
将来がもたらす
メリット

6

リソース

データ・アーキテクトにとっての新たなビジネス環境

データ・アーキテクトのビジネス環境は複雑さを増しています。データベースのクラウド化、クラウド・プラットフォーム、モバイル・データの同期などの新しいテクノロジーが台頭してきています。データ・アーキテクトは自分の管理下にあるデータを管理する方法を、かつてない程、変革することを求められているのです。

まず、データは自由に移動できる必要があります。データをユーザーが必要とするタイミングで必要とする場所に移動できなければなりません。リアルタイムでの移動が必要になることさえあります。

さらに、データの存在する場所にも自由度が必要です。データは、最も経済的かつ合理的に管理、保管、分析が行える場所に存在する必要があります。データ・センター、プライベート・クラウド、パブリック・クラウド、オンプレミスのストレージ、ユーザーのデバイス上など、さまざまな場所に、しかも複数の場所に存在することが求められます。

最後に、データを自由に表現でき、組み合わせることができる必要があります。様々なソースのデータを組み合わせることで、より詳細な洞察や分析が可能になりますが、そのためには、データがアクセス可能になっていて、かつ、組み合わせが可能になっていることが必要です。





しかし、このビジョンを実現するには、これまでの考え方を捨て、データ・マネジメントの構造を変えなければなりません。データ型やロケーションが重要事項だった時代は終わりました。より大切なのは、データがアクセスできることであり、データの反応性であり、データの正確性であり、アナリティクス・パワーがあるかどうかなのです。

この eBook は、データが最も必要となるタイミングと場所でデータを提供し、データをアクセスできるようにするための最新のアプローチである「流動的データ層」について解説します。

流動的データ層 (fluid data layer)を構築すれば、ビジネスにもかつてない程の柔軟性をのぞむことができます。ハイブリッド・クラウドの環境にデータ加工、統合、管理などの機能を併せ持つことで、データ・アーキテクトはデータをロケーションの制約から解放し、データがもたらす価値をフル活用できます。

流動的データ層が不可欠な理由

流動的データ層は、データ・アーキテクトにさまざまなメリットをもたらします。

- ハイブリッド・クラウドに迅速に処理をオフロードすることで、アップタイムを改善できます
- 季節変動や日々の需要のピークに備えた設備の過剰なプロビジョニングが不要になります
- データ・ソースからユーザーにたどり着く最も短いパスを見極めることでデータ移動のレイテンシーを最小化し、より迅速なアクセスを実現できます
- 大量のデータ移動なしに データ・アーキテクトがシステムを迅速に組み合わせることができるため、各部門は以前であれば実施にあたって多大なコストを必要とした短期的なプロジェクトを試すことができます
- ビジネス・ユーザーが新規アプリケーションのアイデアを思いついたときに、データ・アーキテクトはプロトタイプを実行して、そのアイデアの効果をすぐに検証できるようになります。



データ層とは何か？

現代のクラウドやモバイル環境においては、複数のデータ・センターやモバイル・デバイスにまたがってデータ・マネジメントを分散処理するケースが多くなっています。アプリケーション中の「データ層」と名付けられている最新のデータ・マネジメント・コンポーネントは、データベース部分だけを指すのではなく、データベース管理システムと当該システムの実行やスケールアップ（ダウン）、そしてセキュリティ面も保つ、分散インフラ全体のことを指しています。

流動的データ層により、前述のような非常に大きなビジネス上のメリットを提供しながら、組織上のメリットとしてデータ・アーキテクトがデータウェアハウス・インフラ自体をユーザーにさわらせないように保つことができます(図1を参照)。またデータ・アーキテクトは、データだけでなくデータベースそのものも、より簡単に移動できるようになります。例えば、アメリカの東海岸にデータ・センターを持つ音楽共有サービスの人気アジアで突然高まったとします。同社のデータ・アーキテクトはアジアにクラウド・ベースのデータ・センターを立ち上げることで、パフォーマンスを改善し、レイテンシーを短縮することができます。いずれかのデータ・センターがダウンした場合は、もう一方のデータ・センターがバックアップとしての役割を果たします。

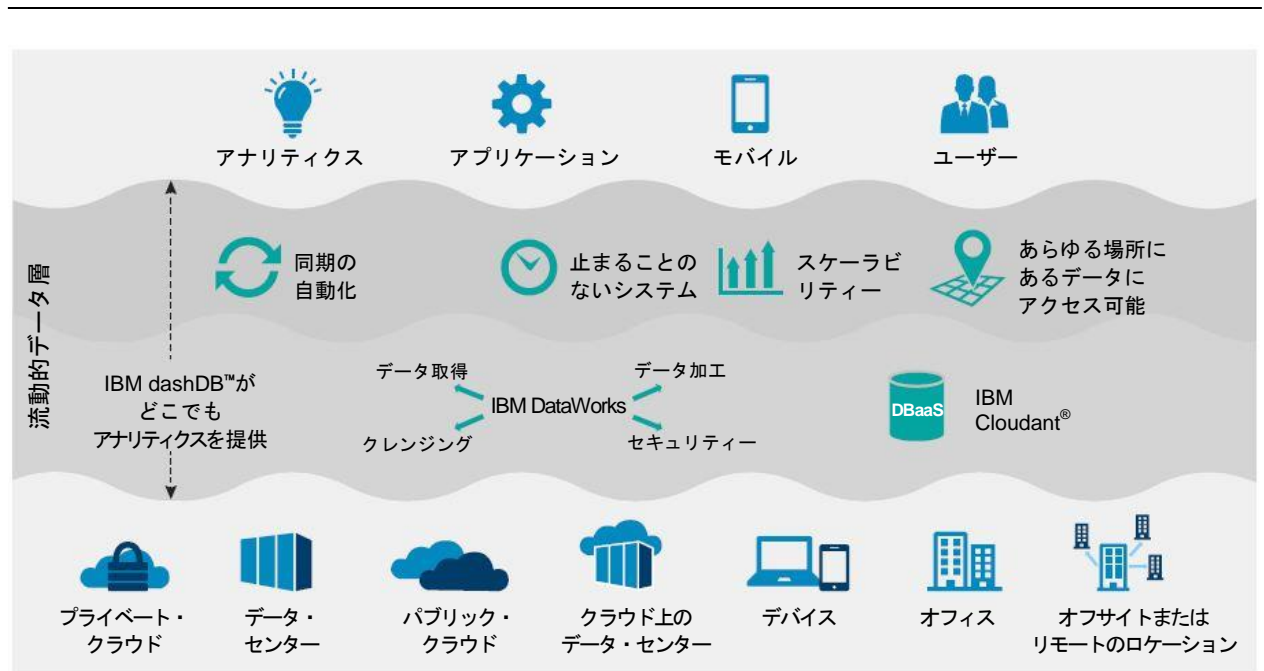


図1. 流動的データ層は複数のデータ・ソースにデータ・マネジメントを分散し、データのアクセスと移動を簡略化します。

1 データ・アーキテクトにとっての新たなビジネス環境

2 流動的データ層が不可欠な理由

3 必須要件

4 実世界における流動的データ層

5 ハイブリッド型の将来がもたらすメリット

6 リソース



この柔軟性は他の領域にもみられます。一般に企業が X-as-a-Service オファリングに関して持つ大きな懸念の1つとして、ロックインされてしまうことと柔軟性の欠如が挙げられます。しかし流動的データ層のおかげで、あるクラウド・ベースのシステムから別のシステムに移行する必要が生じて、コーディングのやり直しやデータ・アクセスの中断を行うことなくデータ・アーキテクトがアーキテクチャーを変更できるようになり、そのような懸念は払拭されます。

データ・ソースという点でもより柔軟性が高まります。データはサーバーのデータベース上にあれば管理が簡単なのですが、いつもそこに存在しているとは限りません。実際には、今日では、データはユーザーのモバイル・デバイス内にも存在しています。動画や、通話のログや、取引の詳細、ソーシャル・ネットワークへの投稿内に存在しているのです。あるいはインターネットに接続したクルマや組み込み型のセンサーからは、ストリーミング・データも発生しています。流動的データ層があれば、データ・アーキテクトは、セキュリティやアクセシビリティを保ちつつ、データを存在場所から取得し、必要とされる場所にプッシュすることができるのです。

必須要件



流動的データ層を構築するには、ビジネス・ニーズを満たすために組み合わせで活用でき、ニーズの変化に応じて簡単に統合と再利用ができるようなマネジメント機能がまとまって必要となります。

すなわち、以下の要件が必要となります。

- 社内データ・センターとクラウド・ベースのシステム間でデータを移動できること
- 構造化データと非構造化データを取り扱うことが可能なこと

- トラフィックや曜日、その他のパラメーターに応じて、スケールアップやスケールダウンが可能であること
- 既存のツール、システム、データ・ソースをクラウドのデータ層に統合できること

流動的データ層の連携・管理・分析・再設定をデータ・アーキテクトが行えるようにするためには、すべてのアプリケーションは相互に連携する必要があります。**この要件は、次に述べるデータ・マネジメント、データ加工、データ分析という3つの基本領域に分類できます。**

データ・マネジメント

データ層の中心となるのは、データベース管理システムとこれをサポートする分散インフラです。データ・アーキテクトは、この共通基盤を中心に、データ層に含まれる他の技術もいかしながら、データ・ソーシングやターゲットを行います。しかし、それだけでなく、パイロット・プロジェクトや季節ごとのニーズに対応する短期的なアプリケーション用に迅速にデータベースを立ち上げられることが求められます。それにはデータ層がクラウド上にデータベースを作成する能力を備えていることも重要となります。

クラウド・データベースを活用すると、開発者は環境のプロビジョニングを迅速に行うことができるため、コスト削減につながり、迅速なプロジェクトのテストにより俊敏性も向上します。さらにデータ・アーキテクトにとって重要なのは、データを収集する場所からデータを分析する場所までをカバーする「ポータブルなアナリティクス」を実現する方法があることです。その際インメモリー・テクノロジーとカラムナー・テクノロジーが加われば、よりスピーディーに洞察を提供することができるようでしょう。またデータ層と現在利用中のツールやシステム間とが統合できれば、再利用のチャンスが高まり、新たなテクノロジーの習得にかかる時間が必要なくなります。

データ加工

データ・アーキテクトは迅速に抽出・加工・ロード (ETL) を処理し、データのソースとターゲットを特定し、両者間のデータ交換を自動化する必要があります。データ・アーキテクトが社内外で一貫性のあるデータのアクセスと使用を保証できるよう、データ層にはデータの分類・プロファイリング・クレンジング・検証を行う機能が必要です。さらに、最近こそクラウド上でデータを移動する際のセキュリティについては不安がなくなりつつありますが、機密データの保護が引き続き重要であることには変わりはありませんので、流動的データ層は、保管中のデータと移動中のデータの両方を自動的に暗号化する必要があります。



データ分析

データ・アーキテクトは、遅延を最小限に抑えた上でストリーミング・データ (センサー・データ、テレマティクス、GPS の情報、地域別の売上データなど) を分析する必要があります。ソースの数に関係なく、データがシステムに入った時点でほぼリアルタイムでアナリティクスを実行できるようなシステムを設計できることが理想です。複数のソースから提供されるデータを分析するほうが、洞察にコンテキストが追加されて、理解を深めることができるからです。

非構造化データに関しても、同様にスピーディーに分析できることが必要です。流動的データ層を活用すれば、非構造化データベース (Apache Hadoop システムからインメモリー・データベースに提供される NoSQL データや JSON データまでを含む) から得られたアウトプットもデータ・アーキテクトが組み合わせることができるようになります。その際、適切でないスキーマに無理にデータを合わせる必要がないため、より俊敏な開発を行うことができます。さらに、データ・アーキテクトが非構造化データベースについても SQL でクエリーをかけ、高レベルな統計分析を行えるように、ツール上も制約がないことが必要です。

クラウド上のデータ・マネジメントに IBM ソリューションを選択する理由

① IBM は、クラウド上で利用できるようになったデータ・マネジメントに関する卓越した知的財産を保有しています。

② IBM は自社のクラウド・データ・サービスを連携させることで、より使いやすいソリューションを実現しています。他の統合されていないサービスでは必要となる統合とプロビジョニングの手間が不要になります。

③ IBM クラウド・データ・サービスは、信頼性とパフォーマンスを実現し、グローバルでの柔軟なフットプリントをもつ IBM SoftLayer® のベアメタル・プラットフォーム上で稼働します。

④ IBM はパブリック、プライベート、ハイブリッド・クラウドのどれにでもお客様のクラウド・データをデプロイできる柔軟なオプションを提供するため、コスト効率、制御、パフォーマンスを最大限に高めることができます。

IBM のメリット: 統合されたデータ・サービス

IBM はさまざまな種類の強力なデータ・サービスを提供しており、その多くは IBM Bluemix® を通じて開発者とデータ・アーキテクトに提供されています。このようなサービスが統合されて、あらゆるアプリケーションの開発と設定をシンプルにしてくれる流動的データ層が実現されます。開発に最大限の柔軟性をもたらし、ロックインの影響を最小化するために、IBM は Hadoop、Spark、CouchDB などのオープン・ソース・テクノロジーを使用した多くのオファリングを開発しています。IBM のデータ・サービスのポートフォリオの詳細情報を確認するには、このリソースを参照してください。

実世界における流動的データ層

流動的データ層のメリットを享受できるのは特別な業界に限ったことではなく、洞察や効率を追い求めデータ活用したいあらゆる業界です。さらに重要なポイントとして、本アーキテクチャーにより、大量のデータをもっているのに共有と分析ができていなかった業界の障壁を解消することも可能です。データがどこにありと問題でなくなるわけですから。

IBM のお客様である Physion は同社のクラウド・ベースの Ovation というアプリケーションのために流動的データ層を構築しました¹。これにより、研究者は実験データをより簡単に保存・整理・管理・共有できるようになり、科学的なコラボレーションの障壁が解消され、個々の研究者のデスクトップや社内のサイロに取り残されていたデータの価値を最大化することが可能となりました。

ある小売企業のデータ・サイエンティストは在庫データなどの構造化データをソーシャル・メディアや気象サービスが提供する非構造化データと組み合わせるクラウド・ベースのデータウェアハウスを構築することで、今後発生する悪天候の条件を予測できるようになりました。この結果、在庫の補充が必要な店舗に道具や食料品を出荷できるようになりました。

医療業界では、複数のシステムのデータを取り込む機能を活用することで、医療機関と公衆衛生機関が疾病の大流行に関する情報を共有できるようになりました。ソーシャル・メディアを分析して不正確な医療情報を特定することで、情報提供の効果を最大限に高めるために、どのようにしてどこに正確な情報を配信すべきか見極められるようになりました。

次に、クルマの相乗りサービスの例でどんなメリットが実現できるかみてみましょう。さまざまなデータ・ソースのデータを集約することで、カスタマー・エクスペリエンスを改善し、ドライバーがタイミングよく、ニーズのある場所にいる ようにできます。データ・アーキテクトはサービスの需要が最も高いタイミングを示す履歴データを活用して、現在の需要に関するリアルタイムの情報と組み合わせることができます。さらに、追加の天候データと交通データをロードすることにより、より精度の高い分析を提供できます。ほぼリアルタイムでこの情報を提供することによって、クルマの相乗りサービス会社はドライバーが乗客を乗せるための最適な場所を示し、乗客により正確な所要時間を示せるようになります。乗客は待ち時間を短縮でき、ドライバーはビジネス・ボリュームを増大できます。

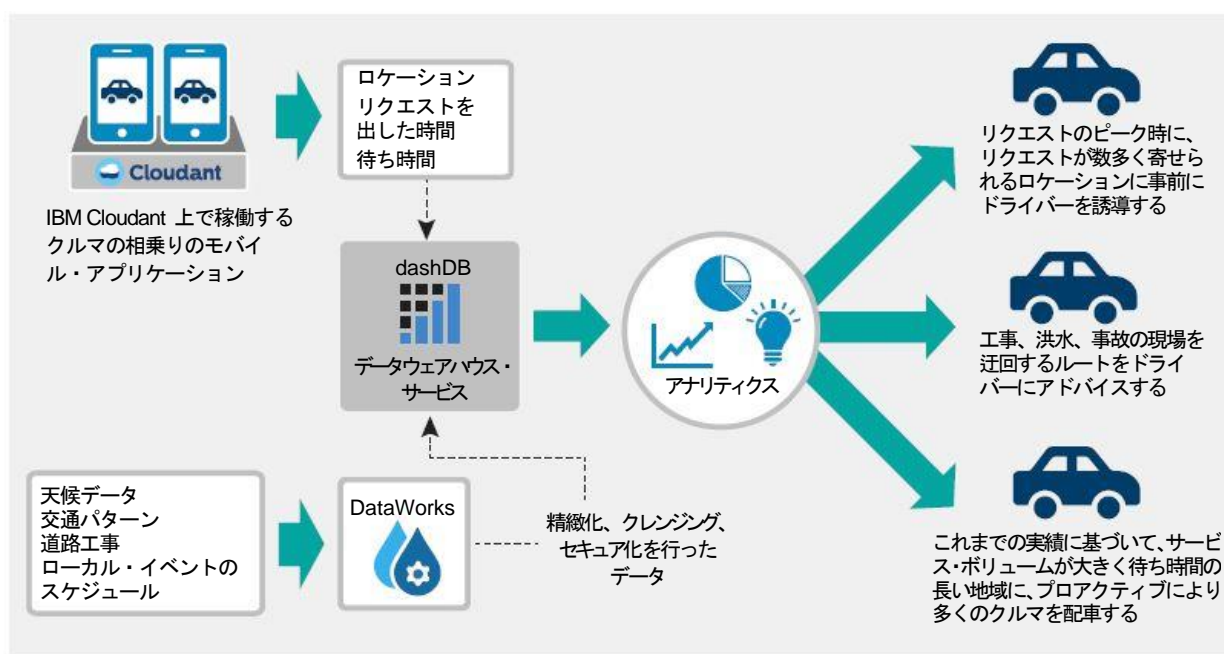


図 2. 複数のデータ・ソースとデータ型を組み合わせ、すべてのデータを分析し、ほぼリアルタイムでデータを提供することにより、クルマの相乗りサービスは乗客の待ち時間を短縮し、ドライバーの稼働時間を改善することができます。

ハイブリッド型の将来がもたらすメリット



流動的データ層は、アクセシビリティや俊敏性、適応性の基礎となります。信頼性のあるデータにユーザーがセルフサービス形式でアクセスすることをデータ・アーキテクトが実現できるようになります。開発者には新規アプリケーションにデータ・サービスを組み込む機能を提供し、ビジネス・アナリストには分析用のデータを発見・活用・提供する機能を提供できます。複数の場所(モバイル、オンプレミス、クラウド等)でデータを多重に保持し、全ユーザーが中断なくデータにアクセスできるようにします。

需要に応じてデータ層を拡張できれば、データ・サイエンティスト向けに分析用の情報をより大量に保存できます。さらに重要な(もしかしたら最重要な)ポイントとして、データ層が本来持つスケラビリティにより必要に応じてデータ層の拡張や再配置が行えるので、将来さまざまなニーズに対応しようとしたときに大規模な改修やアップグレードをしなくてもすむことです。

広範なデータ・マネジメント、データ加工、アナリティクスの機能を提供する IBM の製品とサービスのポートフォリオにより、データ・アーキテクトは、さまざまな機能を組み合わせて、よりスピーディーにより先進的なシステム環境を構築できるようになります。ポートフォリオに含まれるすべての製品を使用しなければならないということではありません。アーキテクトは 1 つのプロジェクトから始めて、少しずつプロジェクトを拡張し、必要に応じて機能の追加と変更を行うことができます。この柔軟性のおかげで、クラウド、オンプレミス、ハイブリッドで稼働するシンプルで使いやすくセキュアなサービスが提供でき、アプリ構築に必要なシームレスなデータ統合やデータ・ガバナンスが実現します。これこそ、ますます複雑化し、競争が激化する今日のビジネス環境でデータ・アーキテクトが必要とするものなのです。

リソース

流動的データ層を構築・強化するために役立つ IBM サービスの詳細については、以下の Web サイトにアクセスしてください。

- Bluemix: ibm.com/software/bluemix
- IBM Cloud Marketplace: ibm.com/marketplace/cloud/jp/ja-jp
- IBM クラウド・データ・サービス: ibm.com/analytics/jp/ja/technology/cloud-data-services/
- IBM Cloudant: ibm.com/software/jp/info/cloudant
- IBM DataWorks: ibm.com/software/products/ja/dataworks/
- IBM dashDB: ibm.com/software/jp/info/dashdb/



© Copyright IBM Corporation 2015
日本アイ・ビー・エム株式会社
IBM Analytics

〒103-8510
東京都中央区日本橋箱崎町 19 番 21 号

Produced in Japan
2015 年 5 月

IBM、IBM ロゴ、ibm.com、Bluemix、Cloudant、dashDB および DataWorks は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

SoftLayer は、IBM のグループ企業である SoftLayer, Inc. の商標または登録商標です。

Java およびすべての Java 関連の商標およびロゴは Oracle やその関連会社の米国およびその他の国における商標または登録商標です。

本書の情報は最初の発行日の時点で得られるものであり、予告なしに変更される場合があります。すべての製品が、IBM が営業を行っているすべての国において利用可能なものではありません。

記載されているお客様事例は、例として示す目的でのみ提供されています。実際の結果は特定の構成や稼働条件によって異なります。本書に掲載されている情報は特定物として現存するままの状態を提供され、第三者の権利の不侵害の保証、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任なしで提供されています。IBM 製品は、IBM 所定の契約書の条項に基づき保証されます。

¹ Physion のライフサイエンス・アプリ:IBM Cloudant のお客様事例 <https://cloudant.com/resources/case-studies>



Please Recycle