

～ データ経営の鍵は「データ統合管理」 ～

# データ統合に必要なツールとデータ統合のステップとは



# データ統合管理はできていますか？

社会全体のデジタル化が進展する中、自社が保有する「データ」の質や量が企業の競争力の一つとして注目されるようになりました。しかし、日本企業によくあるのが、自社内に多数のデータを保持しているにもかかわらず、そのデータが分散していたり、重複・欠損していたりするなど、データの活用環境が整っていないという課題です。

そこで本ホワイトペーパーでは、

- 「どのようにデータを統合管理すればよいのか分からない」
- 「データ統合管理にはどういったツールが必要になるのだろうか」
- 「データ統合はどのような手順を進めればよいのだろうか？」

といったお悩みを持つ情報システム部門に対し、事例を交えながらデータ収集や分析の重要性、データ統合管理の概要と目的、データ統合管理に必要なツール、データ統合のステップについて解説します。

## 目次

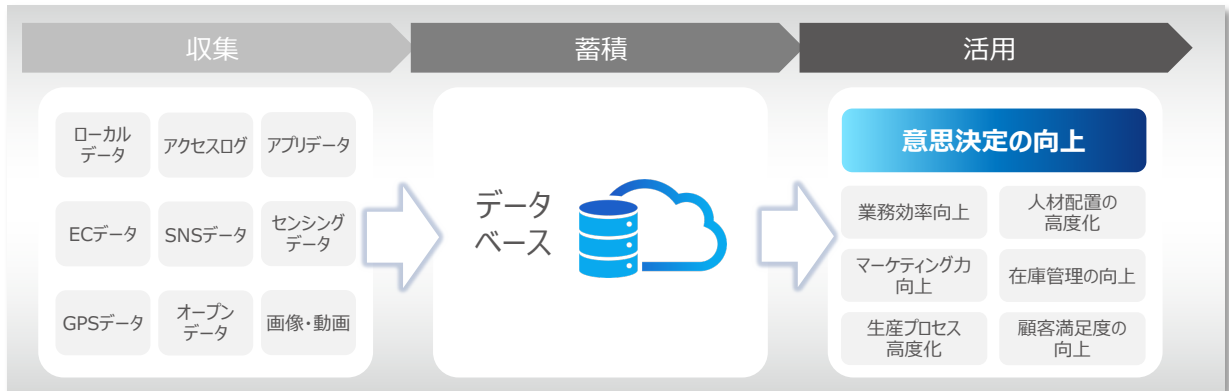
1. データ収集・分析の重要性～データに基づく意思決定の浸透～
2. データ経営の鍵は「データ統合管理」～データ統合管理の概要と目的～
  - 2-1. データ統合管理とは
  - 2-2. データ統合管理の目的
3. データ統合管理に必要なツール
  - 3-1. データ蓄積ツール（DWHなど）
  - 3-2. データ収集ツール（iPaaS）
  - 3-3. データ分析ツール（BIツール）
4. データ統合のステップ
  - 4-1. 統合対象の決定
  - 4-2. 対象データの加工
  - 4-3. データの集積
5. データ活用事例3選
  - 5-1. 事例1：IoTデータから既存顧客のニーズを予測することで営業効率を高め、縮小する市場においても収益を維持
  - 5-2. 事例2：気象データとSNSデータを活用して季節商品の需要を予測し、在庫を大幅削減
  - 5-3. 事例3：顧客データを分析し、取扱ブランドの評価をリアルタイム把握。仕入れやマーケティング施策に活かし、リピーターを増加
6. まとめ





# 1. データ収集・分析の重要性

## ～データに基づく意思決定の浸透～



データを活用した経営や商品開発・サービス提供は、現代的な企業運営のキーポイントとなっています。特に近年においてはデータ活用の注目度が高まっているといえるでしょう。この背景には、以下の3つのポイントがあります。

### データに基づく施策決定の浸透

IT技術の発達とともに、企業においてもITの活用が進んでいます。ITで業務を遂行することにより、顧客・製品・製造・流通・販売などあらゆる分野のデータがデジタル化され、再利用可能な状態となりました。

特に2010年代では、ビッグデータと呼ばれる大量のデータを解析して知見を見出す取り組みが普及し、データサイエンスの重要性が認知されるようになりました。企業経営においても、ERP（統合基幹業務システム）の普及などにより、データに基づく意思決定が浸透したといえるでしょう。このように、技術面・経験面などでデータ活用の素地が形成されたことが、近年のデータ活用の一般化につながっていると考えられます。

### DXの推進

経済産業省が2018年に公表した「DXレポート」を契機に、日本においてデジタルトランスフォーメーションが注目されるようになりました。DXレポートでは、「デジタル技術やデータの活用ができなければ、今後企業が構想力を維持できなくなる」と警鐘を鳴らしており、これにより日本企業にデータ活用推進の意識づけがされたといえます。

### AIの活用

近年の大きなトピックスとして、ディープラーニングをはじめとしたAI技術の進展が挙げられます。現在では、画像認識や音声認識の精度はすでに人間を超えている状況であり、今後もAI技術の適用範囲は広がっていくでしょう。

AIを活用するためには、AIに読み込ませる学習用データが必要ですが、データの質・量は企業における重要な差別化要因となります。よって、AIを有効活用していくためには、「自社が保有するデータを如何に収集できるか」がポイントとなります。AI活用の文脈の中でも、データの重要性は高いといえるでしょう。

## 2. データ経営の鍵は「データ統合管理」 ～データ統合管理の概要と目的～

### 2-1. データ統合管理とは

データ統合管理とは、文字通り「データ統合」と「データ管理」を併せて実施することです。データ統合とは、部門ごと・業務領域ごとに分散したデータを一カ所にとりまとめ、誰もがデータの在りかを把握できるような状態にすることを指します。また、データ管理では、データをアクセスしやすい場所に保管したり、使いやすい形式を維持したりしながら、鮮度や正確性を向上させていきます。

つまり、「社内に散らばるデータを一元化して管理し、分析することで業務効率化や新たな付加価値の創造に役立てるためのアクション」をデータ統合管理と呼ぶことが多いでしょう。データ統合管理の方法はさまざまですが、ここでは代表的なシステム・ツールである「データレイク」と「データカタログ」について以下で整理しておきます。

#### データレイク

データレイクとは、目的・用途が明らかではないビッグデータを保管する湖のような場所です。複数のルートから流れ込むデータを、発生したままの姿で蓄積しておきます。データレイクを使用することで、社内のあらゆる場所から誰もが簡単に、かつ迅速に自分の欲するデータを探せるような状態を維持するわけです。データレイクに集積されるデータは未処理であることが多く、用途や目的に応じて加工・編集されたのちに使用されます。一般的には非構造化データの分析など、主にデータサイエンス領域で活用されることが多いでしょう。

#### データカタログ

データカタログは、データの場所を特定するためのツールです。データレイクへ蓄積されるデータは、日々増大し、多様性を増していきます。そのため、「どこかに保存されているはずだが、見つけれない」「目的や意味がわからないデータが増えて管理しきれない」といった課題が発生しがちです。こうした課題を解決するのがデータカタログです。データカタログは、データの名称・長さ・型・説明などを含む「メタデータ」を管理することで、データ所在地の特定や意味の理解を促進します。データカタログは、メタデータの管理・育成次第で使い勝手が大きく変化するという特徴を持っています。

### 2-2. データ統合管理の目的

データ統合管理の最終的な目的は、「DXを実現するための基盤を整えること」です。また、短期的な目標としては、「データドリブン経営の推進」「データ民主化の実現」などが挙げられるでしょう。データドリブン経営では、非構造化データを活用した戦略的な意思決定を行います。また、データ民主化では、社内のだれもが自由にかつ柔軟にデータへアクセスし、担当業務に活用できる環境を構築します。データドリブン経営とデータ民主化は、社内のだれもが定量化されたデータを参照でき、新たな知見を得ることでビジネスの成長を促すものです。

また、データ統合管理には「情報のサイロ化を防ぐ」という目的もあります。業務システムの数や種類が増えるにつれ、情報の散逸が進んで相互連携が難しくなる「情報のサイロ化」が課題となるケースが増えています。情報のサイロ化が起こると、意思決定スピードが低下したり業務品質が悪化したりと、企業経営に甚大な悪影響をもたらすことも少なくありません。また、時間の経過とともに解決が難しくなるため、早急な対策が望まれます。

## 3. データ統合管理に必要なツール

データ統合管理には、データ蓄積、データ収集、データ分析のためのツールがそれぞれ必要です。本章では、それらの必要なツールについて解説します。

### 3-1. データ蓄積ツール（DWHなど）

データ蓄積ツールには、主にDWH・DMP・CDPが挙げられます。それぞれ以下で解説します。

#### DWH

DWHとは「Data Ware House」の略称であり、様々なシステムからデータを集めて格納するデータベースを意味する言葉です。多様なデータを収集することで、より高度な分析を実現することができます。例えば、販売システムで保持している売上データと製造システムで保持している生産データをクロスで分析して、より高度な生産計画を立案するケースなどが考えられるでしょう。

一般的に、自社システムの多様なデータは、システムにより重複があったり形式が異なったりするため、自社データを集約して活用するためにはデータを整える作業が必要となります。そこで、DWHにデータを格納する際には、データをクレンジングしたうえで、多様なデータを構造化して利用しやすいような状態にします。この作業は大変ではありますが、一度構造化されたデータはビジネスユーザーなどITに関する知見がないユーザーでも利用しやすいものとなります。

#### DMP

DMPとは「Data Management Platform」の略称であり、マーケティングに必要となるデータを集約して格納したデータベースのことです。顧客に関する様々なデータを収集し、一元的に保持することで、高度なマーケティングに用いることができます。

DMPは、自社で保有するデータ（いわゆる1stパーティデータ）を中心に自社用に用意されるプライベートDMPと、第三者が保有する外部データ（いわゆる3rdパーティデータ）を提供するパブリックDMPの2つに分かれます。

#### CDP

CDPとは「Customer Data Platform」の略称であり、自社が保有する複数の情報システムからデータを収集し、顧客をキーとしたデータベースを構築することを意味します。自社の顧客データを一元化することで、例えば事業Aの顧客データと事業Bの顧客データを組み合わせ、より高度なマーケティングを実現することができます。

CDPは前述したDMPのうち、プライベートDMPに近い概念であり、必要に応じて自社データだけではなくパブリックDMPからデータを取得し、自社データと組み合わせる利用することもあります。

## 3-2. データ収集ツール (iPaaS)

データ収集のために活用できる手段の一つとして、iPaaSが挙げられます。iPaaSを用いることで、システムからデータストレージへの連携処理を簡単に構築することができます。以下では、iPaaSの概要や導入効果について解説します。

### iPaaSの概要

iPaaSは、複数のシステム間連携を自動化するためのクラウドサービスです。iPaaSを用いることで、ノーコード・設定作業のみでシステム間の連携処理を実現することができます。iPaaS自体がクラウドサービスであるという特性上、iPaaSはSaaS等のクラウドサービス同士の連携に適していますが、サービスによってはオンプレミス環境のシステムとも連携可能です。

多くの企業でクラウドサービスの導入が進む中、クラウドサービス同士を連携させて効率的に業務に活用していきたいというニーズが生まれました。そのニーズに対応する形で、iPaaSの活用が徐々に進んでいます。

### iPaaSの導入効果

iPaaSを導入することにより、以下の効果が期待できます。

#### 連携処理の自動化によるコスト削減

通常であれば、クラウドサービス間のデータ連携処理は人手で実施しますが、iPaaSを用いることで省力化が実現します。これにより、人的リソースや人件費の節約につながります。

#### システム間連携処理機能の開発削減

システム間のデータ連携を行う際には、人手で行う以外に、連携処理機能を開発する方法もあります。しかし、一般的に連携処理機能の開発コストは高くなりがちであり、双方のシステム変更を起因とするメンテナンスの手間も発生します。iPaaSを導入することで、面倒になりがちな連携処理の開発範囲を削減することができます。

### 3-3. データ分析ツール（BIツール）

続いて、データ分析に有用であるBIツールについて解説します。

#### BIツールの概要

BIツールとは、自社が保有するデータを分析し、データの集計や抽出・グラフ化・レポート化などを実現するツールのことです。企業経営においてデータ活用の重要性が意識されるようになり、BIツールの活用が進んでいます。

自社のデータを元にデータベースを構築しさえすれば、BIツールによって自由にデータ分析が可能です。BIツールの利用にはコーディングは不要であるため、ユーザー自身が集計方法を指定することができます。

#### BIツールの導入効果

BIツールを導入することにより、以下の効果が期待できます。

##### 柔軟なレポート作成・データ分析を実現

従来であれば、レポート作成のためにはシステム開発を行う必要があり、項目や集計方法を変更するためにはプログラミングが必要でした。BIツールを導入すれば、ユーザー側でも設定を変更するだけでレポートやデータ分析方法を修正することができます。

##### データ分析の効率向上

Excelなどを利用して生データをグラフ化し、データ分析を行うこともできますが、一定の時間と手間が必要となります。特に、月次報告などで定型的な分析作業が必要な場合は、BIツールを用いることで作業の効率化が可能です。



## 4. データ統合のステップ

ここでは、データ統合の具体的なステップについて解説します。データ統合を成功させるには、「統合対象の決定」「対象データの加工」「データの集積」という3つのステップを適切に進めていく必要があります。以下は、各ステップの具体的な内容です。

### 4-1. 統合対象の決定

まず、社内システム・従業員個人のPCなどに存在するデータをリスト化し、精査したうえで統合対象を決定します。社内システムについては、基幹システムとサブシステムを含めて横断的に精査していきましょう。一般的には、下記のようなデータが統合対象になります。

- 業務アプリケーションのトランザクションデータ
- 製品、サービスのマスタデータ
- CRMなどに蓄積された顧客データ
- 外部IoTシステムから集められたセンサーデータ

### 4-2. 対象データの加工

統合対象が決定したら、対象データの加工を行います。データ加工では、次のようなタスクを進めていくことになるでしょう。

#### データのフォーマット統一

ファイル形式や文字コード、項目のデータ型などを決定し、統一されたフォーマットを作成します。複数の文字コード（Shift-JIS、UTF-8など）が混在する場合は、いずれかひとつに指定して文字化けが生じないようにしてください。また、日付や通貨については桁数や通貨単位を統一し、分析ツールで扱いやすい形に整えていきましょう。

#### 欠損値および異常値の検出と補正

何らかの事情でデータが正常に取得できていない「欠損値」や、本来取得すべきでないデータが取得されている「異常値」などに対しては、適宜補正を行います。欠損値や異常値は、統合後のデータ連携などにおいて不具合の温床になるため、早い段階で取り除くことが大切です。

#### 重複データの削除と名寄せ

対象データの中に重複がある場合は、削除するか、名寄せによって関連付けを行うかの判断が必要です。特に名寄せは、データの使い勝手や価値を向上させるために無くてはならない工程です。一般的に名寄せは、「共有キー」を設定したうえで、関連性のあるデータを紐づけていきます。また、紐づけ作業と並行してデータの調整も行いましょう。例えば「列構造調整」ではデータを管理する列の数を調整し、行列構造調整ではデータが開始される位置の調整を行います。



### 4-3. データの集積

データ加工のあとは、データの集積方法を決定し、集積を進めていきます。データの種類・活用の方向性・目的によって集積場所のタイプは異なるため、複数の候補から最適な方法を選択していきましょう。前章までで、データレイク、DWH、DMP、CDPについて紹介しました。以下では、データ統合で使用されるその他の集積方法として、PIM、DAM、データ仮想化について説明します。

#### PIM（プロダクト・インフォメーション・マネジメント）

PIMは、日本語で「商品情報管理」と翻訳されます。商品情報とは、製品の基礎情報（仕様や価格など）に、訴求力を高める情報（商品説明、ブランドロゴ、キャッチコピー）を加えたものです。PIMを使ったデータ統合では、主にブランディングやプロモーションの効率化・質の向上が期待できます。

#### DAM（デジタル・アセット・マネジメント）

DAMは、企業内に存在する情報資源を集積・管理するための仕組みです。デジタルコンテンツ（テキスト・動画・静止画やカタログに使用するデータ）を、統一のプラットフォームで管理します。DAMを使用することで、メタ情報が整理されてアクセス性が向上したり、使用するメディアに合わせたデータ変換作業を効率化できたりといった効果が期待できます。また、コンテンツの著作権管理も容易になるため、デジタルコンテンツの管理コスト削減にもつながるでしょう。

#### データ仮想化

データ仮想化は、複数のデータソースから取得したデータを仮想化し、「仮想レイヤー」と呼ばれる場所で一元的に管理する仕組みです。データ仮想化ではオリジナルデータの物理的な移動・保存が発生しないため、厳密に言えばデータの集積は行われません。ただし、仮想的にデータが一元化されることで、保管場所の容量を気にすることなく、迅速なデータ活用が可能になります。また、仮想化される前のオリジナルデータには一切アクセスが発生しないため、データ消失のリスクも低減させることができます。



## 5. データ活用事例3選

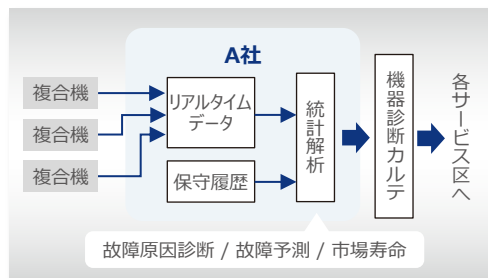
最後に、データを活用して経営成果を挙げた事例を3つ紹介します。

### 5-1: 事例1

#### IoTデータから既存顧客のニーズを予測することで 営業効率を高め、縮小する市場においても収益を維持

オフィス用機器メーカーのA社は、顧客企業の倒産や事業縮小に伴い、売上の低下に直面していました。新規顧客の獲得と並行して、既存顧客の維持が喫緊の課題でした。そこでA社は自社製品にIoTセンサーを取り付け、機器の稼働状況や故障タイミングなどのデータを収集し、それらIoTデータと過去の発注履歴や問い合わせ履歴などのマーケティングデータを基に、買い替え予測モデルを開発しました。

結果として、買い替えニーズの高い顧客に適切なタイミング・提案内容でアプローチでき、成約率を3倍以上にすることができました。顧客企業にとっても、機器故障によるストレスが軽減し、顧客の満足度向上にもつながっています。

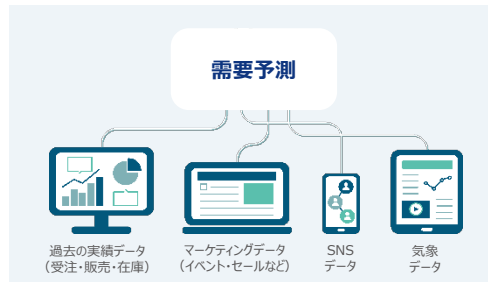


### 5-2: 事例2

#### 気象データとSNSデータを活用して季節商品の需要を予測し、 在庫を大幅削減

食品メーカーのB社は、季節商品の在庫管理に課題を抱えていました。売り逃がしを避けるために安全在庫を多めに確保していましたが、季節商品は販売期間が短いため、シーズン後には一定量の廃棄が出ている状況でした。そこでB社は、季節商品の需要予測精度を高めるために、体感指数を分析するサービスを取り入れました。

具体的には、外部の気象データとSNSのデータを組み合わせて地域ごとの体感指数を算出するものです。この予測結果を生産計画に反映させたところ、余剰在庫を35%削減することに成功しました。

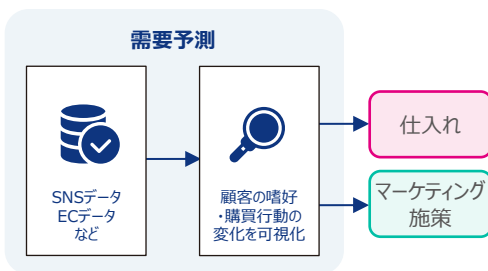


### 5-3: 事例3

#### 顧客データを分析し、取扱ブランドの評価をリアルタイム把握。 仕入れやマーケティング施策に活かし、リピーターを増加

アパレル系のECを運営するC社では、競争力強化のため、顧客のニーズを捉えた商品ラインナップやマーケティング施策が必要であると感じていました。そこでC社はSNSやEC上の行動データなどから、ブランドの認知率や、顧客の嗜好・購買傾向の変化をタイムリーに把握し、30日以内に売れる商品を90%の確率で予測できるように改善を図りました。

その結果、予測に基づいた仕入れやマーケティング施策を行うことで、売上増加と余剰在庫の削減を実現することができました。また、返品率の低下によって物流リソースに余裕ができたことで配送リードタイムが短縮するとともに、顧客の満足度が高まりリピート率向上につながっています。

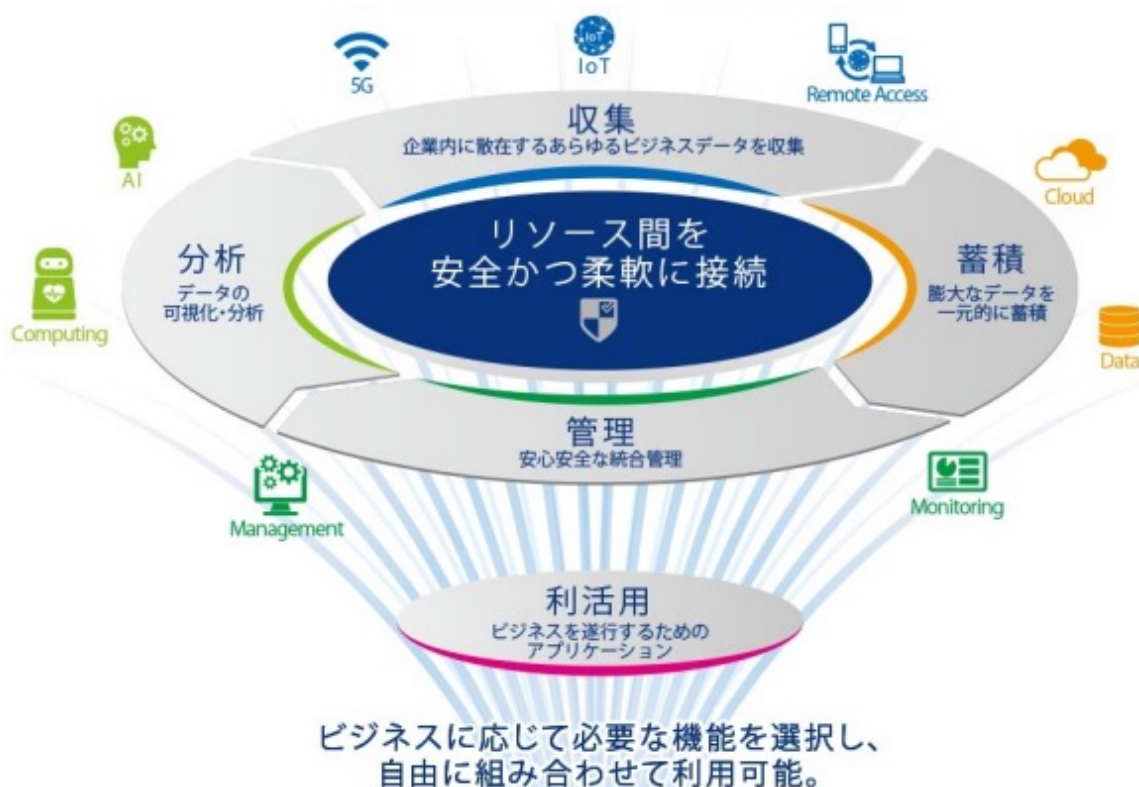


# まとめ

本資料では、データを活用した経営の重要性やデータ統合管理、必要なツール、データ統合のステップ、データ活用事例を解説しました。DXは一朝一夕に為し得るものではなく、中長期的なスパンで推進していく忍耐が問われる取り組みです。しかし、まずは市販の製品やサービスを活用して、ニューノーマルな働き方に対応していくことで、新しいデジタル技術を柔軟に取り入れる文化が根付いたり、データを蓄積できたりと、今後DXを成功させるための基盤となります。自社の状況と照らし合わせて最適な製品・サービスを選択することが重要になりますが、判断が難しい場合は、お気軽にお問い合わせください。

## データの収集・蓄積・活用方法の具体的な方法についてのご相談がある方は、NTTコミュニケーションズへお問合せください。

NTTコミュニケーションズでは、DXの実現に向けた取り組みをトータルでサポートしております。私たちは、企業に点在するデータの活用を容易にし、成長エンジンに変える事でお客様のDX実現までのプロセスをトータルでご支援させていただきます。



### 〈新たなビジネス価値の創出〉 意思決定の迅速化による競争力強化・新たなビジネス領域への進出

データ収集	データ蓄積	データ利活用
エッジデバイスやデータレイクなどからセキュアなネットワークを介してデータをスムーズに収集	大企業のビッグデータを統合的に蓄積するクラウド型のデータ蓄積基盤。オンプレミスからの移行やハイブリッドクラウド化にも対応	BI / BAツールからデータマイニングツール、専門知識不要のデータ分析ツールまで幅広いソリューションをラインナップ

NTTコミュニケーションズのDXの取り組みsdpfとは？ ▶

