



ビッグデータ活用基盤リファレンスアーキテクチャ 利用ガイド

株式会社NTTデータ
技術革新統括本部 システム技術本部

利用規約

ビッグデータ活用基盤リファレンスアーキテクチャに関する一切の資料（以下：本資料）の利用を希望されるお客様は、以下の利用規約（以下、本規約）にご同意いただくことを条件として、本資料を利用いただくことが可能です。

なお、お客様が本資料のダウンロードを行った時点で、本規約にご同意いただいたものとさせていただきます。

1. 本資料の著作権は、株式会社NTTデータ（以下：NTTデータ）が保有しています。
2. 本資料の利用者は、次項に定める条件を満たす場合に限り、本資料の全部または一部を、営利目的、非営利目的に関わらず、NTTデータの許諾なく複製、改変、頒布、公衆送信、販売、出版、翻訳/翻案することができるものとします。
3. 利用許諾の条件は次の通りとします。
 - (1) 本資料に次の著作権表示を必ず付し、また、付された著作権表示を削除しないこと。
著作権表示：© 2021 NTT DATA Corporation
 - (2) 本資料の複製物（本資料を改変したものを含む）を頒布、公衆送信をする場合は、本規約に定める利用条件を配布先に遵守させること。
 - (3) 本資料を改変する場合は改変箇所を明示して利用すること。
4. NTTデータは、本資料が第三者の著作権、特許権、実用新案権等の知的財産権に抵触しないことを一切保証するものではなく、また、本資料が第三者の権利を侵害する場合又は本資料の内容に誤りがあった場合でも一切責任を負いかねます。
5. NTTデータは、本規約による利用許諾を除き、NTTデータまたは第三者の著作権、特許権、実用新案権等の知的財産権に基づきいかなる権利を許諾するものではありません。
6. NTTデータは、本資料自体（改変された箇所も含む）、本資料のシステム開発への利用、開発されたシステムの使用、及び当該システムの使用不能等により生じるいかなる損害についても、なんら責任を負うものではありません。
7. 本資料を海外へ持ち出す場合及び非居住者に提供する場合には、「外国為替及び外国貿易法」の規制及び米国輸出管理規則等外国の輸出関連法規を確認のうえ、必要な手続きを行って下さい。
8. 本資料へのお問い合わせについては、NTTデータ(hadoop@kits.nttdata.co.jp)までご連絡下さい。
9. 本規約の成立、効力、解釈及び履行については、日本法に準拠するものとします。

はじめに

本書は、ビッグデータ活用基盤リファレンスアーキテクチャを作成した背景や概要および使い方等を解説することを目的としている。

なお、「ビッグデータ活用基盤リファレンスアーキテクチャ」とは、ビッグデータ活用基盤リファレンスアーキテクチャ利用ガイド(本書)、および以下の2つの資料の総称である。

- アーキテクチャ図
- 観点一覧

「ビッグデータ」というキーワードが登場して久しく、たくさんのデータを用いたサービス開発、業務改善、販売促進などのデータ活用プロジェクトは一般的に耳にするようになった。さらにデータの可視化/BI用途のように直接的にデータを利用するケースに加え、**AIや機械学習の活用、IoTの実現などデータを基礎とした技術の活用が以前にも増して求められる**ようになっている。また取り扱うデータ種類も増えており、データ活用プロジェクトを取り巻く状況変化は常に変化し続けている。

そういった状況下、データ活用基盤に活かせる便利な技術が次々登場している。実現環境もオンプレミスに加えてクラウドサービス、マネージドサービスの利用も進んでいる。NTTデータにおいても、**ストリームデータ活用基盤の運用性を向上させる技術、データレイクを高度化しアーキテクチャを刷新するための研究**が行われており、**旧来のアーキテクチャを一変させる可能性**も見出している。

しかしながら、選択肢が増えて自由になった反面、システムを作りサービスを実現するにあたって**「どこから考え始めたらよいかかわからない」、「どのように検討すればよいかかわからない」というケースが増えた**。また最新技術の活用の際には、**最初に定番のアーキテクチャを出発点とし、的確に議論を発展させる必要**がある。

そこで、NTTデータはこれまで手掛けてきた多くのデータ活用プロジェクトから、**共通する典型的なアーキテクチャの要素、特に重要な検討観点を抽出・体系化し、「ビッグデータ活用基盤リファレンスアーキテクチャ」**を作成した。

ビッグデータ活用基盤リファレンスアーキテクチャとは データ活用プロジェクトにおける基盤開発上流工程の参考文書

本資料の概要は以下の通りである。

- 利用ガイド(本書)
 - ビッグデータ活用基盤リファレンスアーキテクチャを解説した資料
- アーキテクチャ図
 - データ活用基盤における主要な構成要素を取り揃えた図
- 観点一覧
 - データ活用プロジェクトにおける基盤のランドデザインで検討すべき重要項目を整理した表

主な対象システム

- 幅広い利用者が多種多様なデータを自由に活用できるシステム
- スケーラブルなデータストアを有するシステム
- 以下に代表されるデジタル分野の要素が含まれるシステム
 - ◆ データレイク
 - ◆ データハブ
 - ◆ 可視化・BI基盤
 - ◆ AI/機械学習基盤
- データ基盤の変遷として、データウェアハウス中心としたデータ基盤から、データレイク中心にシフトし、近年ではデータレイクハウスが登場してきた。本資料では、これまでのNTTデータが手掛けてきた実績に基づき、検討の出発点としてデータレイク中心のシステムを取り扱う。

主な対象利用者

- ユーザー企業におけるシステム担当者
- ベンダー企業におけるアーキテクト、エンジニア
- これからデータ基盤を作りたい、データ基盤を作ろうとしたものの検討にお悩みの方

リファレンスアーキテクチャの特長

- アーキテクチャ図を見ながら、要件に応じて必要な構成要素を取捨選択することで、多くのプロジェクトにおいて必要とされるアーキテクチャの全体像を描くことが可能
- 基盤のグランドデザインで検討すべき重要項目を参照しながら議論することで、多くのプロジェクトで共同的に課題となる重要事項を踏まえ、着実にアーキテクチャを決めていくことが可能
- プロジェクト固有の内容を加筆しながら、検討すべき項目を育てていくことが可能
- 特定の製品、オープンソースソフトウェアに依存しない内容であり、多くのデータ活用プロジェクトで参考にすることが可能
- データ収集からデータ活用まで記載しており、データ活用基盤の全体を俯瞰した議論が可能

資料の説明 ～アーキテクチャ図～

データ活用基盤における主要な構成要素を取り揃えた図である。
構成は以下の通り。

- ビッグデータ活用基盤リファレンスアーキテクチャ図
 - データ活用基盤におけるデータパイプラインを実現するための以下の機能の構成要素およびデータの流れを記載している。
 - ◆ データ収集
 - ◆ データ蓄積
 - ◆ データ加工
 - ◆ データ活用
 - ◆ ストリーム処理

- ビッグデータ活用基盤リファレンスアーキテクチャ 達成できる要素
 - データパイプラインを実現するための各機能の説明を記載している。

資料の説明 ～観点一覧～

ビッグデータ活用案件における基盤グランドデザインで検討・考慮すべき内容を整理した表である。構成は以下の通り。

- 検討項目カテゴリの樹形図
 - 検討項目の大カテゴリ～中カテゴリの樹形図を記載している。

- プロジェクト共通項目
 - データ基盤の要件やシステム構成にかかわらず、どのプロジェクトでも共通的に検討すべき項目を記載している。
 - **プロジェクトや体制に関する内容がメインである。**
 - プロジェクト共通項目のみ「検討における注意点、補足事項」を公開している。

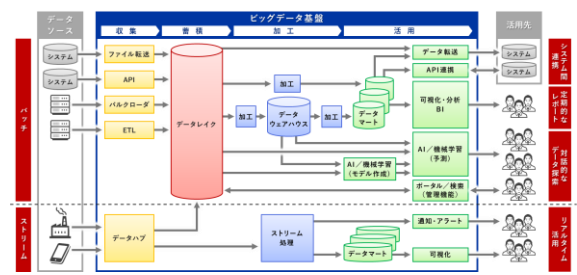
- システム化個別項目
 - データ基盤の要件やシステム構成により、機能・サービスを必要に応じて取捨選択し、検討する項目を記載している。
 - **システム構成や要件定義についての内容がメインである。**
 - システム化個別項目における「検討における注意点、補足事項」は非公開としている。ご相談は、NTTデータ(hadoop@kits.nttdata.co.jp)まで。

使い方 ~利用イメージ~

フェーズ



...



② 要件定義時にシステムの
スコープを定め、ステークホル
ダーと合意するために利用

アーキテクチャ図

① プロジェクト始まり時にどんな
ことを実現するか考えるために
システムイメージを利用

システム化個別項目				
項目	大カテゴリ	中カテゴリ	小カテゴリ	小カテゴリの説明
B.1.1.1	データ収集	収集対象データ種別	データソースからデータ庫内部に取り入れるデータの種別	構造化データ (Structured Data) RDBMSのテーブルやCSVなど、テーブル構造で表すことが可能なデータ 例: -CSV (Comma Separated Values) -TSV (Tab Separated Values) -システムログ
B.1.1.2			半構造化データ (Semi-Structured Data)	XMLやJSONなど、内部で多次元の構造を保持可能なデータ。厳密な定義はないが、ここでは以下の性質をもつものとする (不可能ではないが)SQLでの扱いが難しい -後方互換性を保たなければならないキーの変更や、前後スキーマの存在が可能 -情報を失わずにテキストで表すことが可能
B.1.1.3			非構造化データ (Unstructured Data)	テキストファイル (*.txt)、画像 (JPG, PNG, BMP, GIF) や音声 (WAV, MP3, AAC)、動画 (AVI, MP4, etc...) など センサーなどのIoTデバイスにみられる、独自フォーマットによりシリアライズされたフォーマット、シリアライズ/デシリアライズのためビットアサイン表など別途フォーマットが必要
B.1.1.4			特定のアプリケーション向けフォーマット	Word (*.docx) や Excel (*.xlsx)、PDF など用途が特定のアプリケーションで使われる形式

観点一覧

③ 要件定義時にシステム
のあるべき姿を決めるために
利用



NTT DATA

Trusted Global Innovator