

ビッグデータ分析の導入工程の定型化とクラウドデザインパターンの適用

NTTコミュニケーションズ株式会社 柏村 洋平 yohei.kashimura@ntt.com

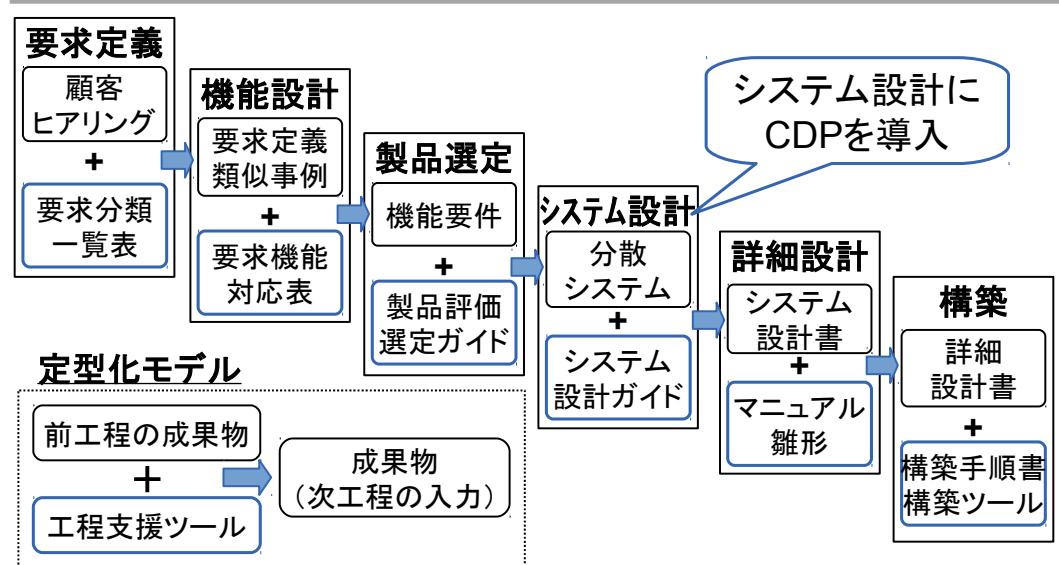
開発における問題点

ビッグデータ分析の要求、および、要求を実現するシステムは多様であり、システムを導入するためには、検討事項が多く、効率が悪い。
また、ビッグデータ分析では、システムのサイジングの難しさや、導入効果の事前見積もりが難しいことにより、投資判断が難しく、低コスト、かつ柔軟なリソースの確保が要求される。

手法・ツールの適用による解決

ビッグデータ分析の導入工程を定型化し、作業の明確化と、工程支援ツールの整備を行う。
クラウドを利用し、リソース確保の柔軟性を実現する。ビッグデータ分析で使用する分散システムを、クラウド上に構築するための設計指針として、クラウドデザインパターン(CDP)を適用する。

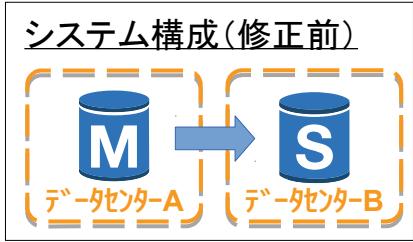
ビッグデータ分析の導入工程の定型化



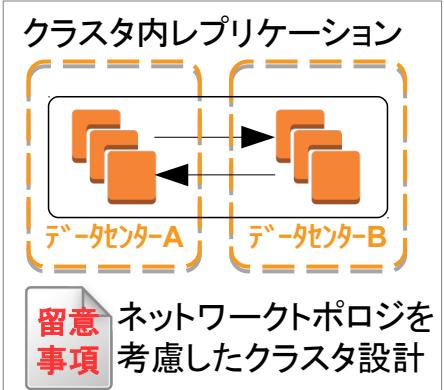
- ① 工程の定型化
 - i. 工程で実施すべき作業を定義
 - ii. 作業を定型化する支援ツールを整備
- ② クラウドデザインパターン(CDP)の導入
 - i. 機能に対するCDP適用可否の判定
 - ii. システム構成、留意事項を修正・追記
→ ビッグデータ分析に特化した修正を加え、パターンを適用可能にする。
 - iii. CDP適用による品質影響の評価
→ CDP適用マップを作成し、適用可否と効果を可視化することで、パターン選択を支援。

CDP適用例

DB-Replication 適用例
DB-Replicationパターンに、分散システムに適用するためのシステム構成の修正、留意事項の追記を行った。



システム構成 以下の2つの構成から、選定した製品の特徴、各構成の留意点を考慮して選択する。



CDP適用結果

適用結果
評価を行った49パターンのうち29パターンを適用可と判定し、そのうち8パターンについて追記、修正を加えた。
全ての機能に対し、効率性に寄与するパターンが存在し、クラウドの柔軟性を活用できることを確認した。

表: CDP適用マップ(一部抜粋)

機能	収集			データストア			加工		分析		
	収集	巡回	バッファ・キュー	RDBMS	NoSQL	分散FS	抽出・変換	集約・突合	リアルタイム	オンライン	バッチ
パターン											
Stamp	UM	UM	UM	UM	UM	UM	UM	UM	UM	UM	UM
Scale Out	UE	UE	UE		UE	UE	UE	UE	UE	UE	UE
Write Proxy				E	E	E					
DB-Replication				R	R	R					
Job Observer		E	E				E	E	E		E

【品質特性】 F: 機能性, R: 信頼性, U: 使用性, E: 効率性, M: 保守性, P: 移植性