

db tech showcase 2022 【E5】

IoT/ビッグデータ向けデータベースGridDB Deep Dive

TOSHIBA

東芝デジタルソリューションズ株式会社
ソフトウェアシステム技術開発センター
シニアフェロー 服部 雅一 (masakazu.hattori@toshiba.co.jp)

自己紹介

東芝デジタルソリューションズ株式会社 シニアフェロー

- 入社以降、AIシステムやデータベースの研究開発に従事
- 2010年、(社)情報処理学会より「喜安記念業績賞」受賞
- 2012年、スケールアウト型DB “GridDB” の研究立ち上げ
翌年、V1.0を上市。以降、チーフアーキテクトとして開発を主導

その他、2006年 (社)日本電気協会 関東電気協会 電気関係事業従業員功績者表彰(最優秀賞)

2008年(社)発明協会 関東地方発明表彰(日本弁理士会会長奨励賞)

2018年(社)発明協会 関東地方発明表彰、など表彰多数

- 2020年、日本データベース学会の第4回「業績賞」受賞



日本データベース学会の第4回「業績賞」受賞

DB技術と産業の発展に貢献との評価



Agenda

0 1 GridDBの概要

0 2 GridDBを使うメリット

0 3 自律的なDBクラスタ

効率良く複数サーバを利用してパフォーマンスアップ

0 4 イベント駆動アーキテクチャ

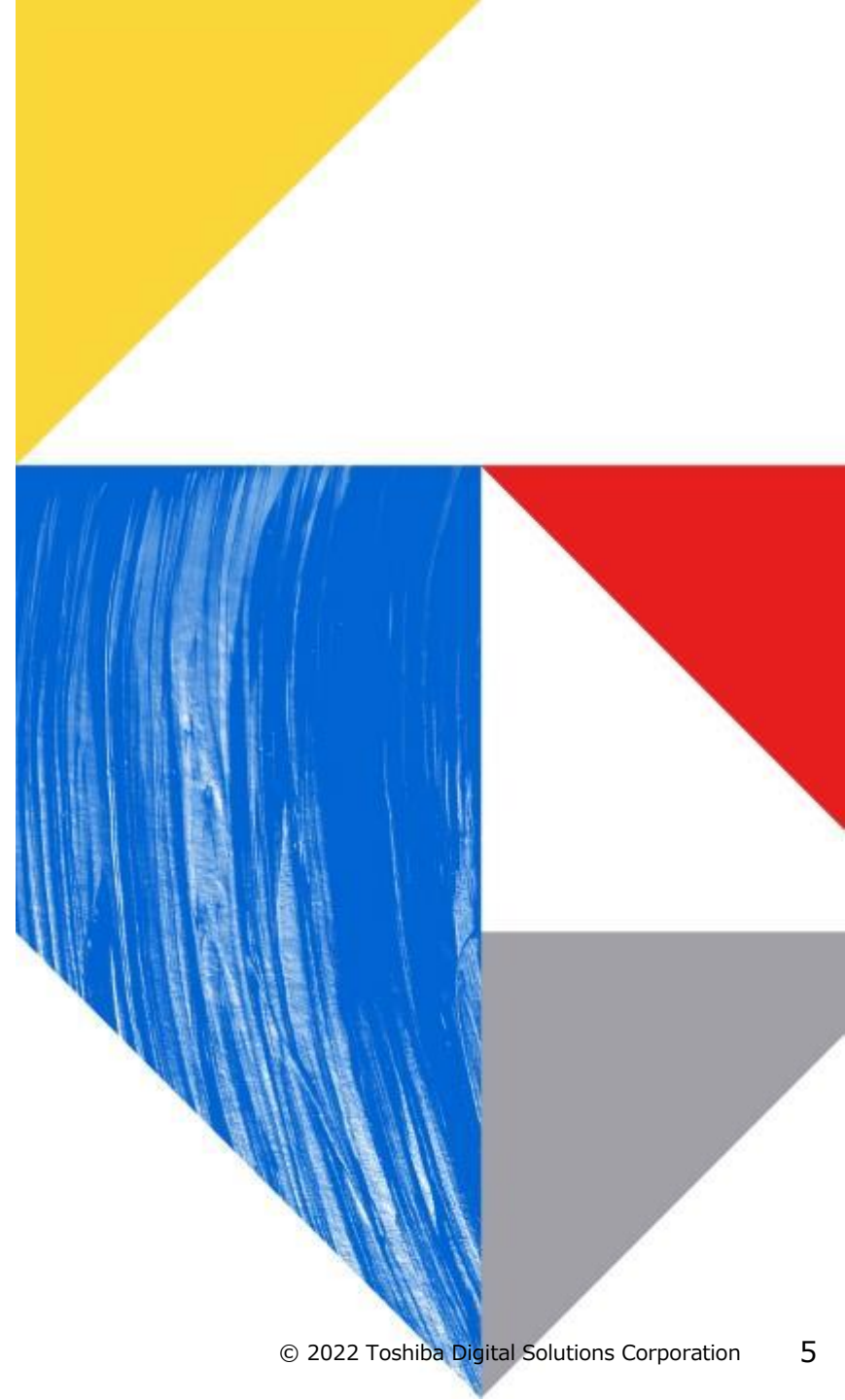
効率良く複数コア、メモリを利用してパフォーマンスアップ

0 5 他DBとのベンチマーク

0 6 おわりに

01

GridDBの概要



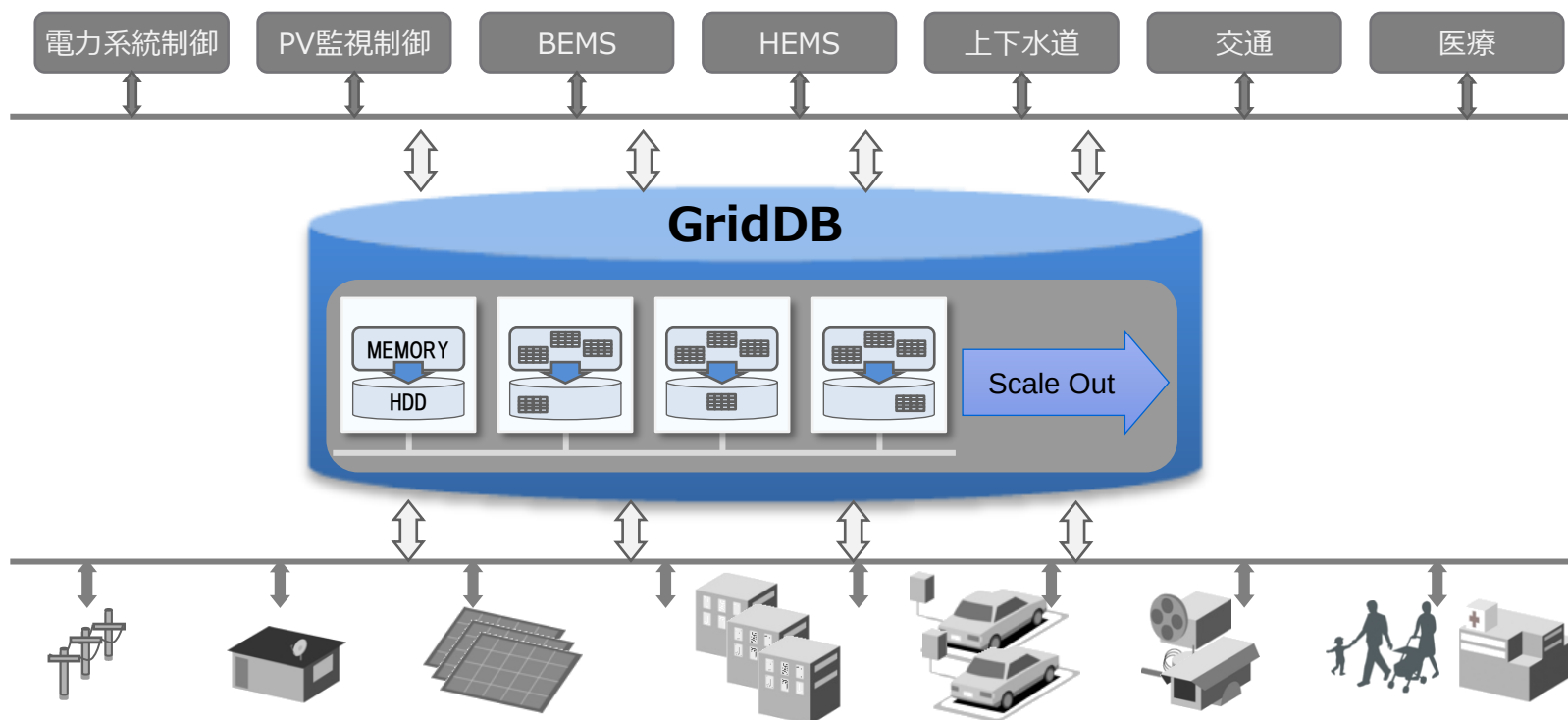
「×デジタル」により新たな価値を創造

今後、さまざまなモノのデータをサイバー空間に収集し、AI・デジタル技術で分析した結果を、フィジカル世界にフィードバックすることで価値を創造できるようになります。こういった世界が広がるなか、私たちは、モノづくりの幅広い知見・実績（フィジカル技術）と、AI・デジタル技術（サイバー技術）を掛け合わせ、さまざまな社会課題を解決していきます。



開発背景

- 社会インフラシステムの維持管理には稼働データを管理するデータベースが必要
- システム規模の増大 → ①RDBではその要件を満たせない、②NoSQLの信頼性に不安



GridDBを自製化

ペタバイト級IoTデータを高速に処理するスケールアウト型データベース GridDB

- 2022年 アーキテクチャ刷新、クラウド対応強化
- 2019年 1ノードあたりペタバイト級のデータ管理に対応
- 2017年 SQL並列分散処理化
- 2016年 OSS公開
- 2015年 GridDB/NewSQL (SQLインターフェイス機能)
- 2013年 GridDB (NoSQLのみ) 上市
- 2011年 研究開発スタート

単体製品、クラウドサービス、Meister DigitalTwin基盤などを通してビジネス展開

時系列 データ指向



高頻度で大規模な時系列データを効率よくリアルタイム処理する時系列データ指向

ペタバイト級の 高い処理能力



ペタバイト規模のデータを扱うためにさまざまな工夫を組み込み、高い処理能力を実現

高い信頼性と 柔軟な拡張性



障害の発生時やサーバ増設においてもノンストップ運用を実現する高い信頼性と柔軟な拡張性

開発の俊敏性と 使いやすさ



NoSQLインターフェースだけではなく、SQLインターフェースを用意し、開発の俊敏性と使いやすさを実現

時系列データ：時間とともに変化するデータ

3つのエディション



GridDB Community Edition

高頻度・大量に発生する時系列
データの蓄積とリアルタイムな活用を
スムーズに実現する次世代の

**オープンソース
データベース**



GridDB Enterprise Edition

高頻度・大量に発生する時系列
データの蓄積とリアルタイムな活用を
スムーズに実現し、
ビジネスを大きく成長させるために

**最適化された
次世代のデータベース**



GridDB Cloud

高頻度・大量に発生する時系列
データの蓄積とリアルタイムな活用を
スムーズに実現する

**クラウドデータベース
サービス**

社会インフラを中心に、高い信頼性・性能が求められるシステムで採用

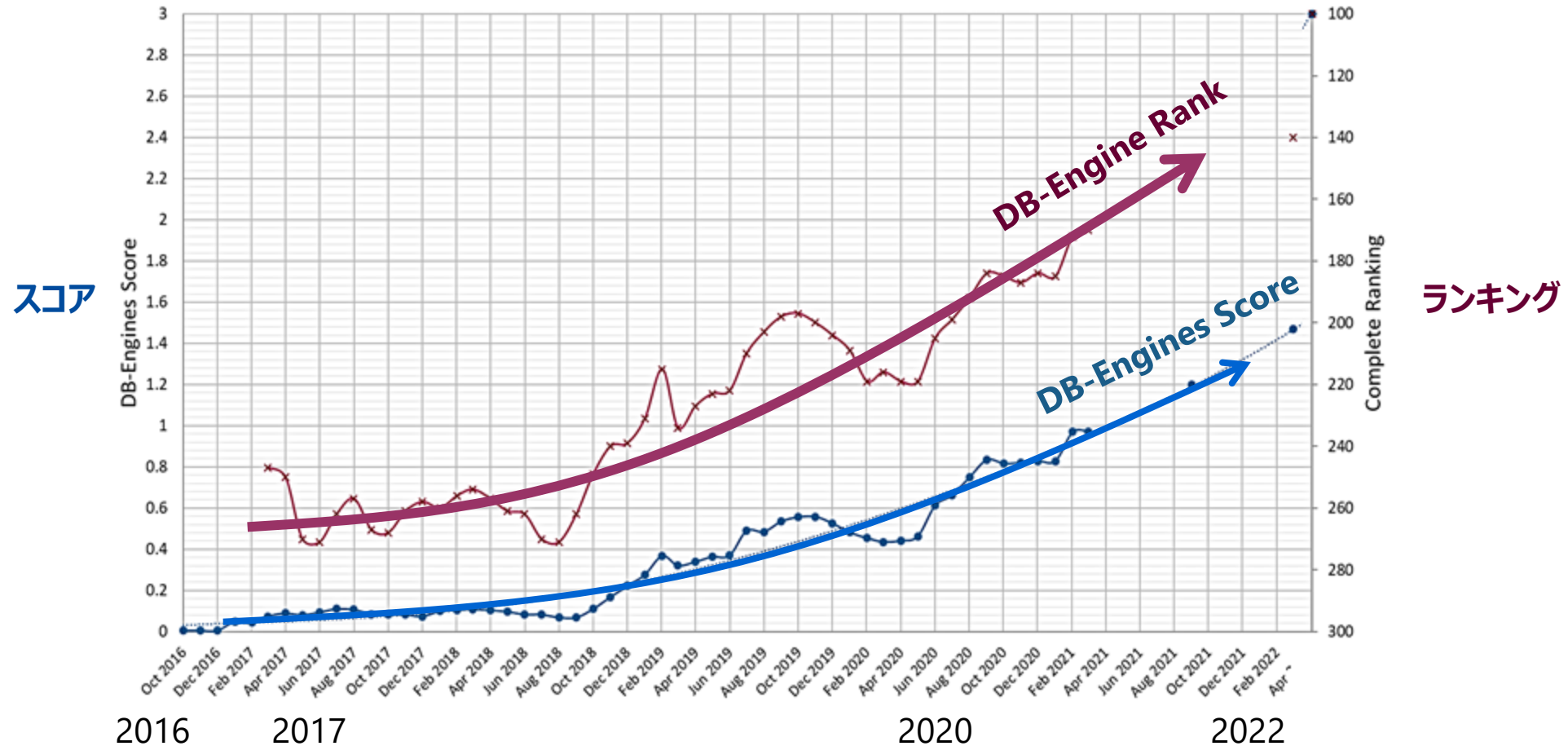
- 電力会社 低圧託送業務システム
スマートメータから収集される電力使用量を集計し、需要量と発電量のバランスを調整
- HDD製造会社の品質管理システム
製造装置のセンサーデータを長期にわたって蓄積・分析し、品質分析・改善に適用
- 半導体製造ラインのトレーサビリティ管理
製造履歴や品質履歴、材料データなどのデータを横串で分析し、製品の品質管理やトレーサビリティに適用
- 半導体製造ラインの故障検出システム
製造ラインのセンサーデータをリアルタイムにAIで分析し、製造ラインの異常を検出
- デンソー ファクトリーIoT
工場のDigital Twinを実現し、生産性を向上。
- DENSO International America 次世代車両管理システム
車両の各センサーデータを用いた車両管理システムのPoC

各エディションの違い

インターフェイスはほぼ同じ。クラスタ構成の有無の違い

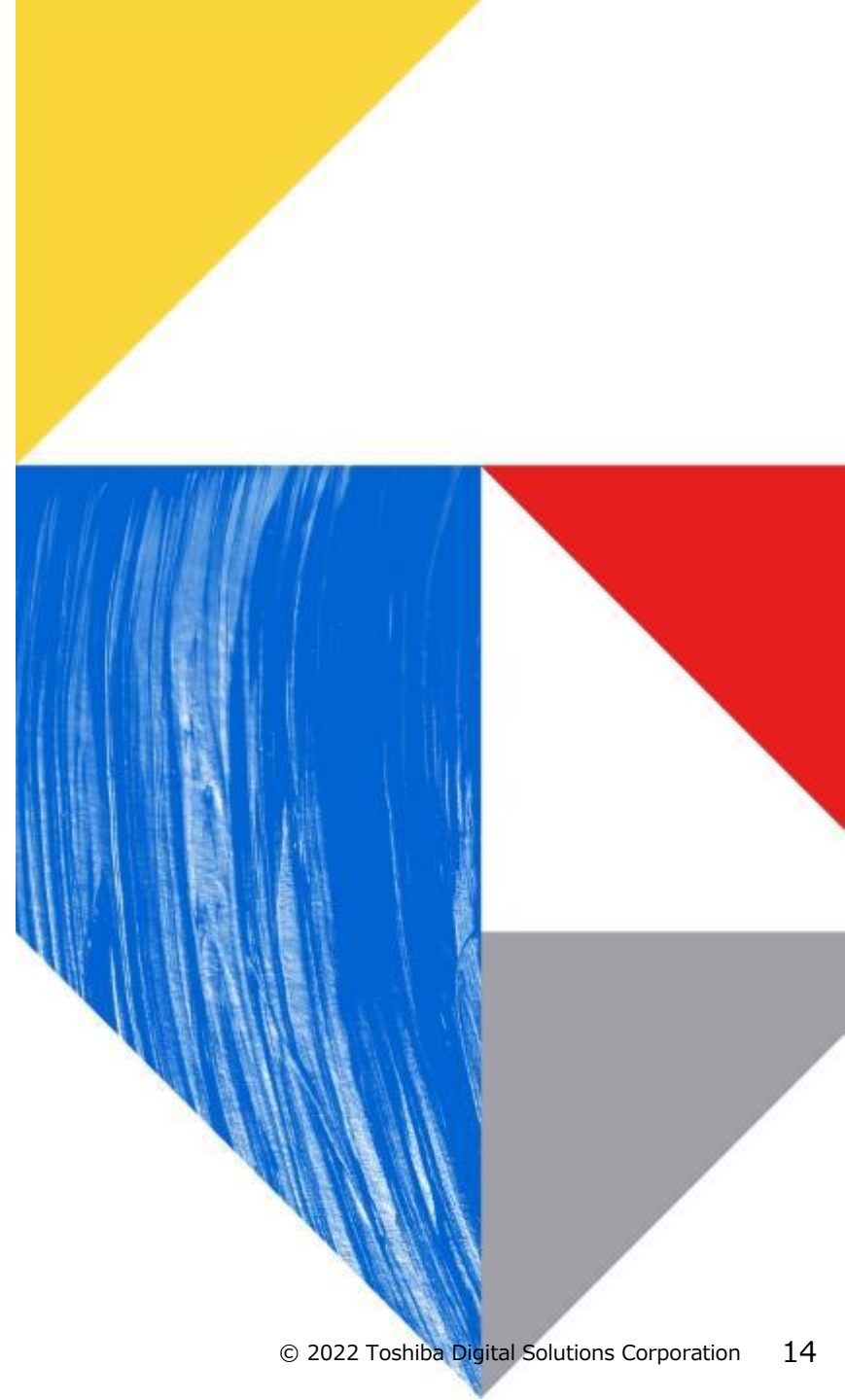
項目	機能	Community Edition	Enterprise Edition	Cloud
サポート			✓	✓
プロフェッショナルサービス			✓	✓
データ管理	時系列コンテナ	✓	✓	✓
	コレクションコンテナ	✓	✓	✓
	索引	✓	✓	✓
	アフィニティ	✓	✓	✓
	テーブルパーティショニング	✓	✓	✓
クエリ言語	TQL	✓	✓	✓
	SQL	✓	✓	✓
NoSQLインタフェース	Java	✓	✓	✓
	C言語	✓	✓	✓
NewSQL(SQL) インタフェース	JDBC	✓	✓	✓
	ODBC		✓	✓
WebAPI		✓	✓	✓
時系列データ	時系列分析関数	✓	✓	✓
	期限付き解放機能		✓	✓
	長期アーカイブ		✓	✓
クラスタリング	機能クラスタ構成		✓	✓
	分散データ管理		✓	✓
	レプリケーション		✓	✓
運用管理	ローリングアップグレード		✓	
	オンラインバックアップ		✓	✓
	エクスポート / インポート		✓	✓
	運用管理GUI		✓	✓
セキュリティ	運用コマンド		✓	✓
	信暗号化 (TLS/SSL)		✓	✓
オンプレミス環境	認証機能 (LDAP)		✓	
	オンプレミス環境	✓	✓	
クラウドサービス				✓

2016年公開。近年、DB-Engineランキングで急上昇



02

GridDBを使うメリット



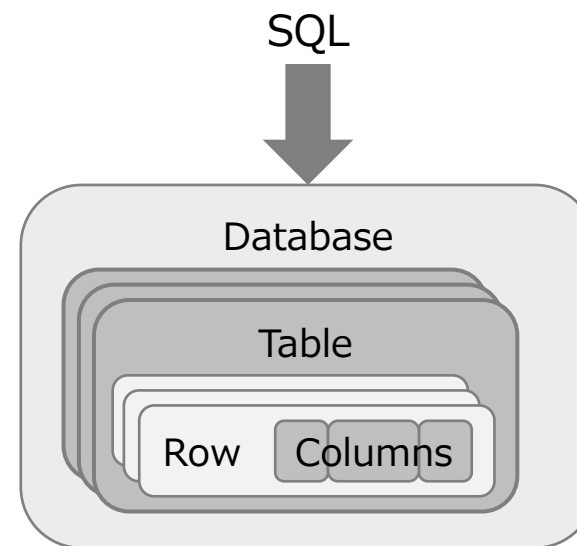
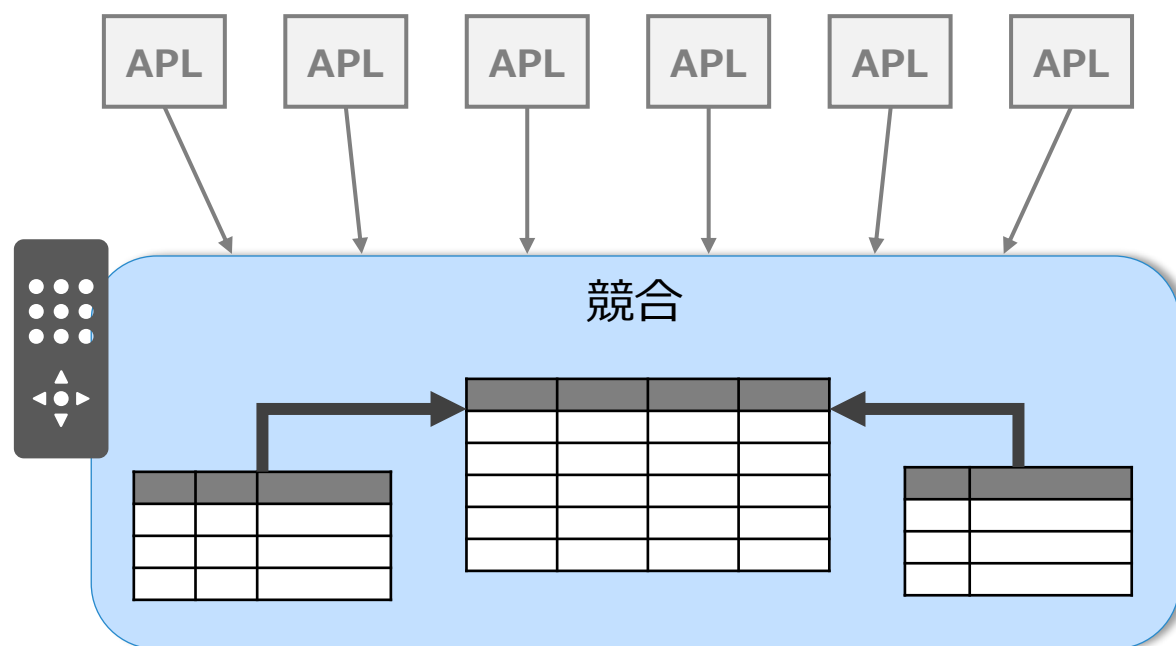
リレーショナルデータベース

Pros.

- ① ACID
- ② リレーショナルスキーマ
- ③ SQL
- ④ 強い一貫性

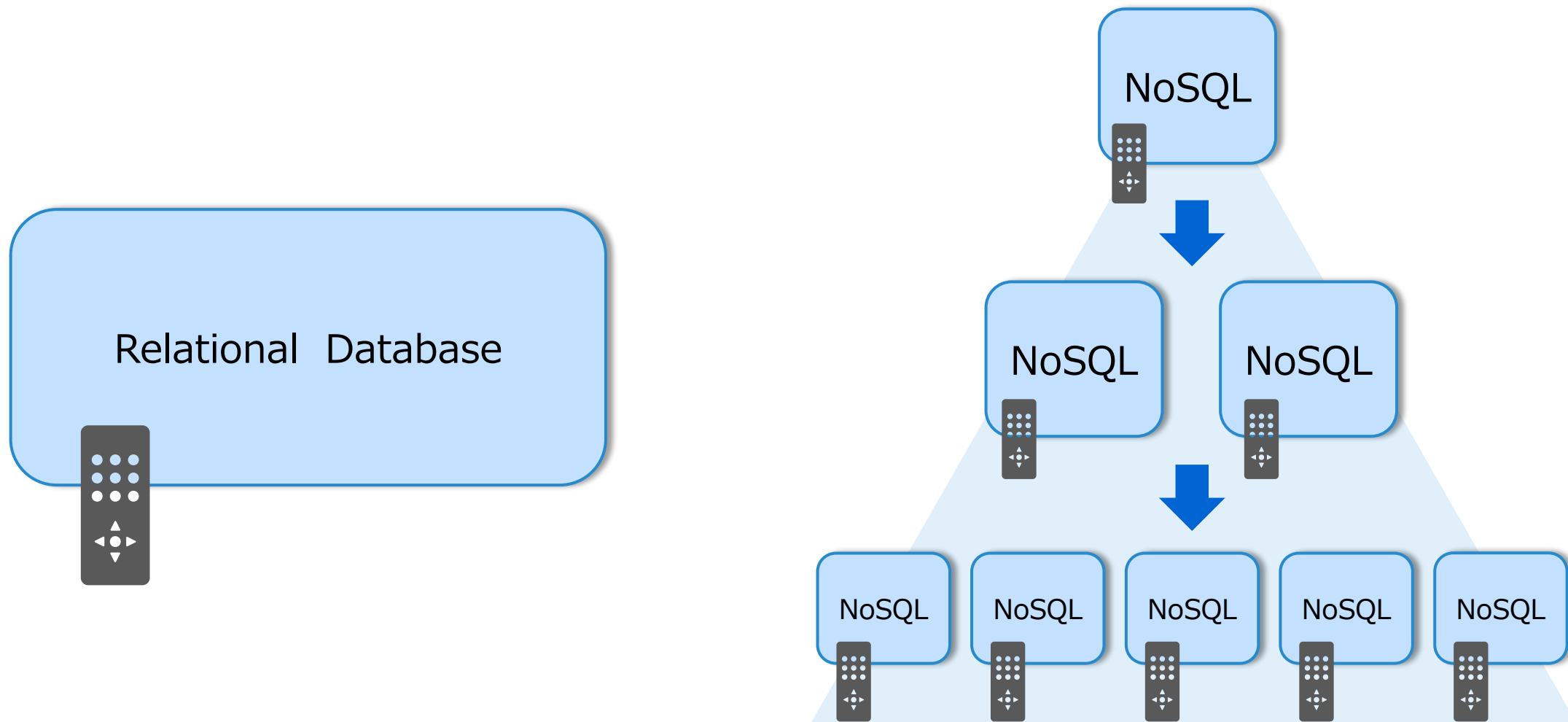
Cons.

- ① スケーラビリティ欠落
数十サーバ以上の環境において
ホリゾンタルなスケール困難
- ② 1サーバ当たりの基本性能



NoSQL (Not Only SQL)

ホリゾンタルなスケールアウトを志向した



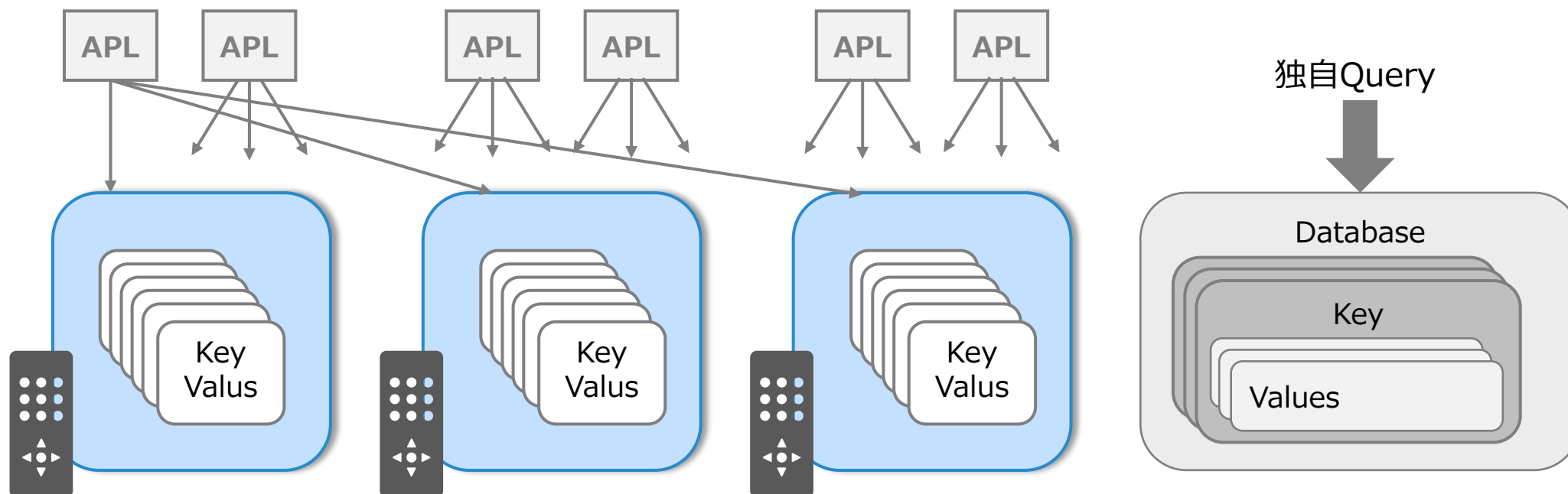
NoSQL (Not Only SQL)

Pros.

- ① 高スケーラビリティ
- ② 高アベイラビリティ

Cons.

- ① スキーマレス
- ② 独自のインターフェイス
- ③ ACID



SQL、NoSQL

	某NoSQL	某NoSQL	SQL(RDB)
問い合わせ言語	独自I/F	独自I/F	SQL
スキーマ言語	無い	無い	SQL(DDL)
データ一貫性	参照一貫性 テーブル内一貫性	緩やかな一貫性 or 参照一貫性	厳格な一貫性
1台あたりの性能	普通	一貫性を上げると低速	実績あるRDBは ある程度
高可用性	ノード分散による 冗長性有り	ノード分散による 冗長性有り	別の仕掛けが必要
拡張性	スケールアウトする手段が不明	スケールアウト困難 特にオンラインスケールアウト	スケールアップ

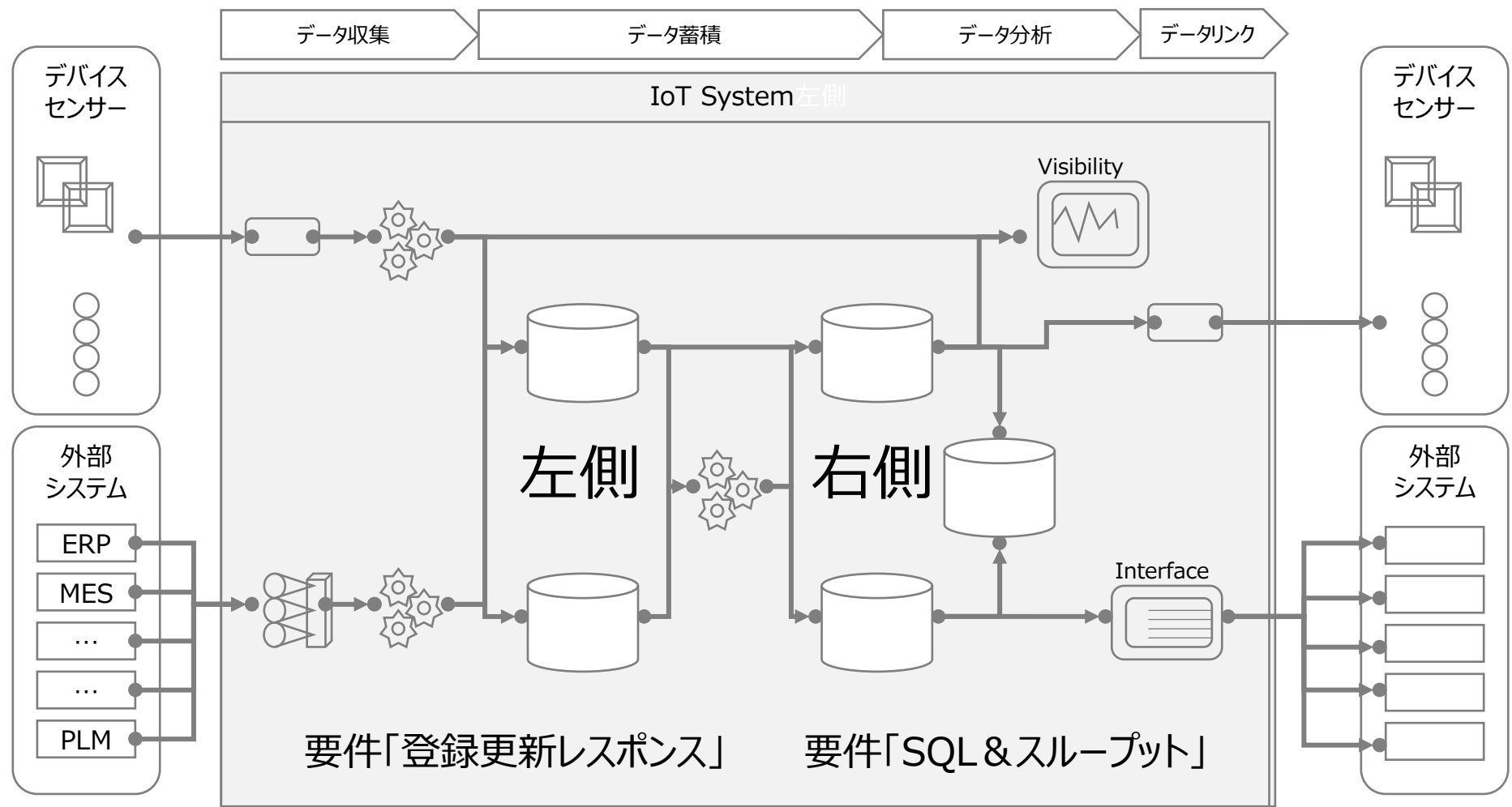
補完関係にある

	某NoSQL	某NoSQL	SQL(RDB)
問い合わせ言語	独自I/F	独自I/F	SQL
スキーマ言語	無い	無い	SQL(DDL)
データ一貫性	参照一貫性 テーブル内一貫性	緩やかな一貫性 or 参照一貫性	厳格な一貫性
1台あたりの性能	普通	一貫性を上げると低速	実績あるRDBは ある程度
高可用性	ノード分散による 冗長性有り	ノード分散による 冗長性有り	別の仕掛けが必要
拡張性	スケールアウトする手段が不明	スケールアウト困難 特にオンラインスケールアウト	スケールアップ

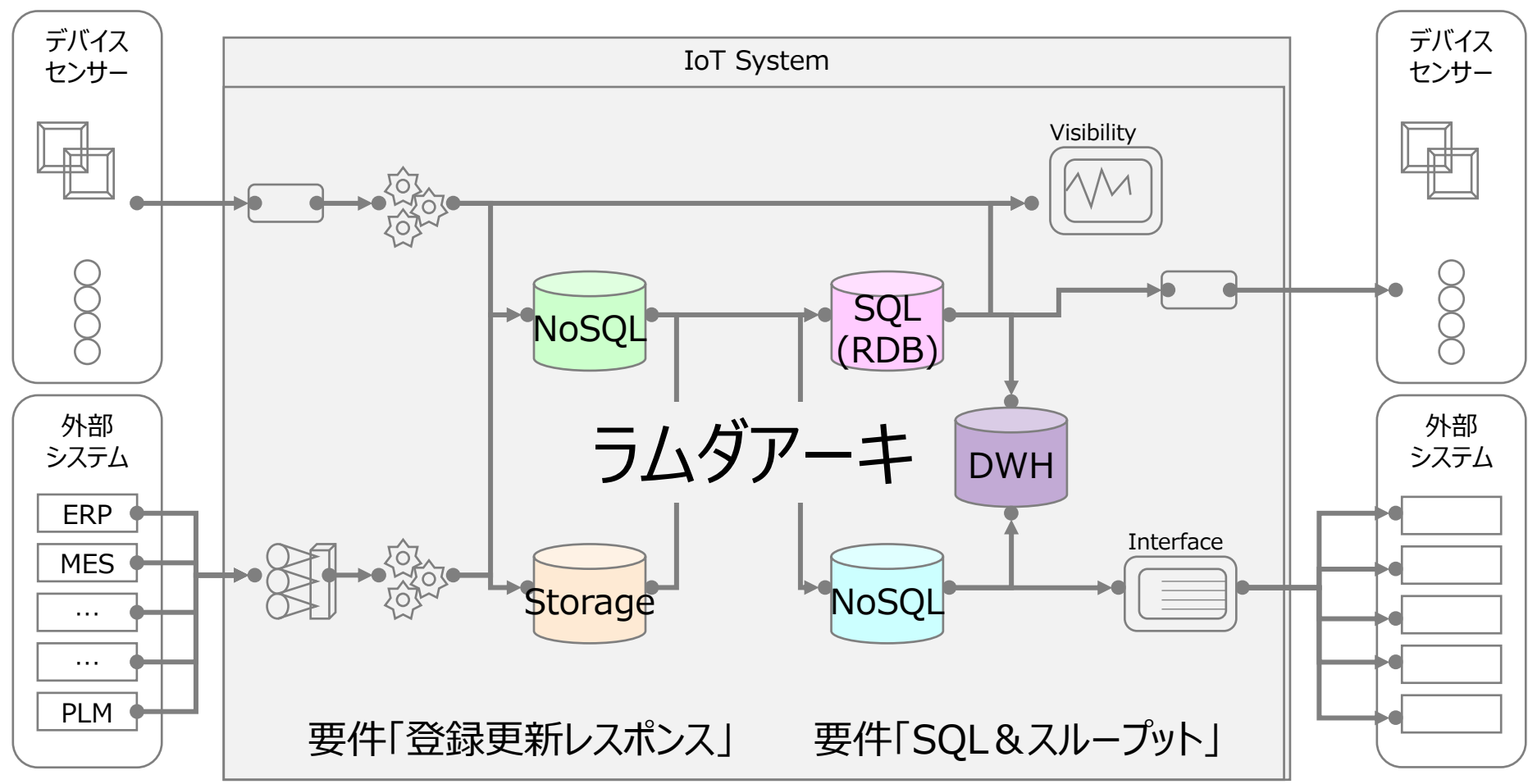
テーブル間一貫性を除けば、全てをカバーする

	 GridDB	NoSQL	SQL(RDB)
問い合わせ言語	SQL + 独自I/F	独自I/F	SQL
スキーマ言語	SQL (DDL)	無い	SQL (DDL)
データ一貫性	参照一貫性 テーブル内一貫性	BASE理論 緩やかな一貫性	厳格な一貫性
1台あたりの性能	イベント駆動による 最大限の性能	SW次第	実績あるRDBは ある程度
高可用性	通常、高速の 2レプリケーション	ノード分散による 冗長性有り	別の仕掛けが必要
拡張性	自律的な スケールアウト	スケールアウト SW次第	スケールアップ

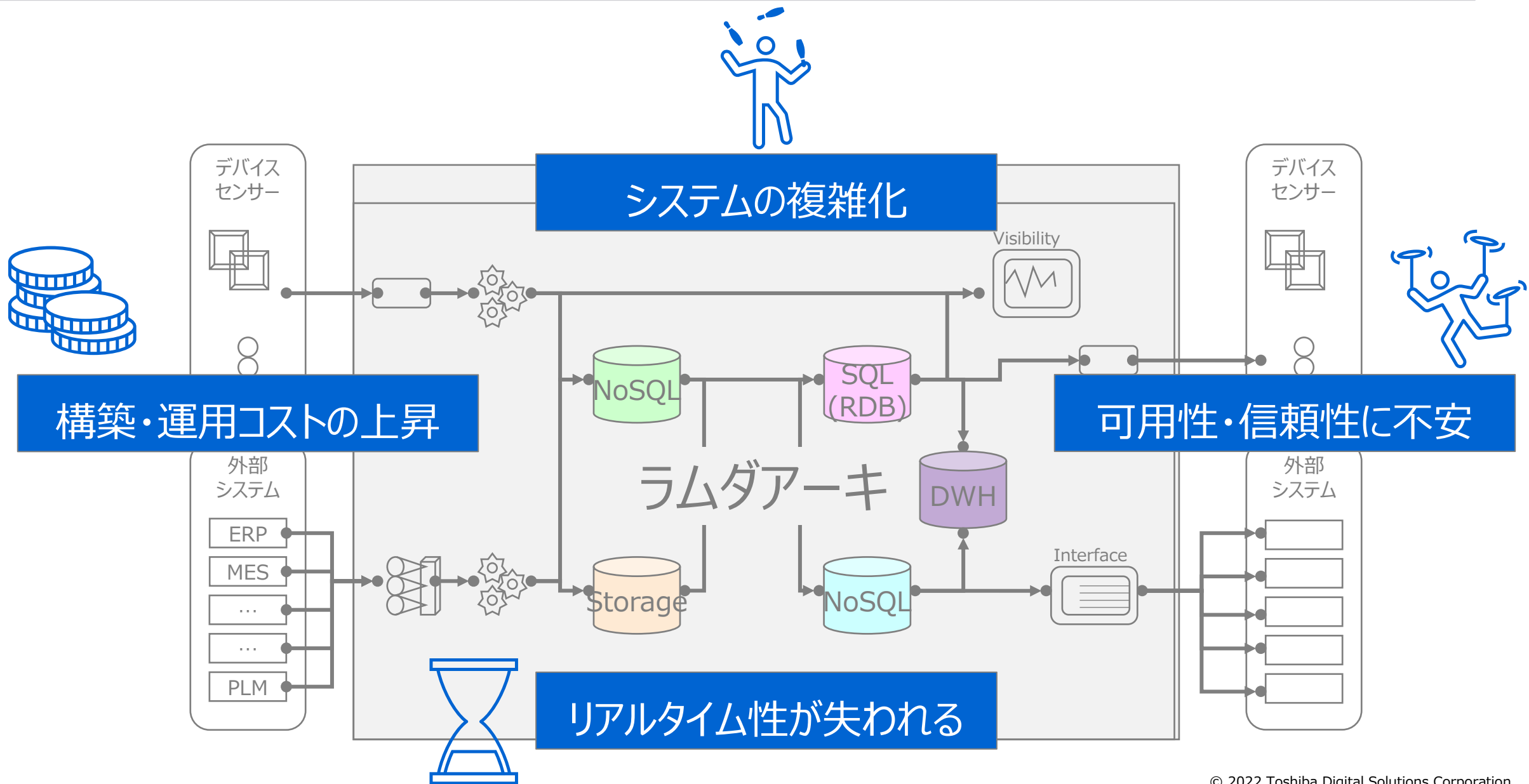
IoTシステムの内部構成



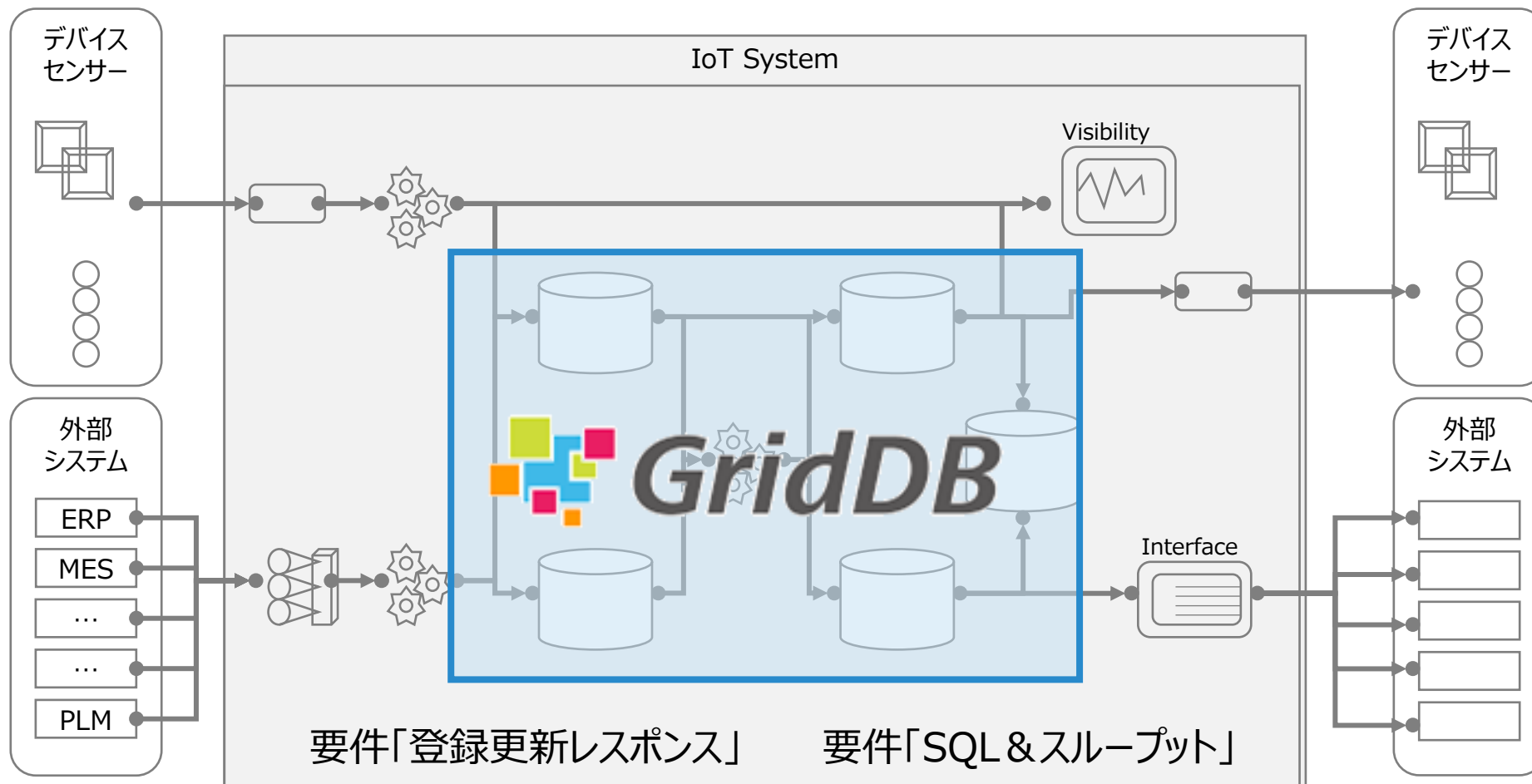
異なる特性を持つ複数のDBの使い分け



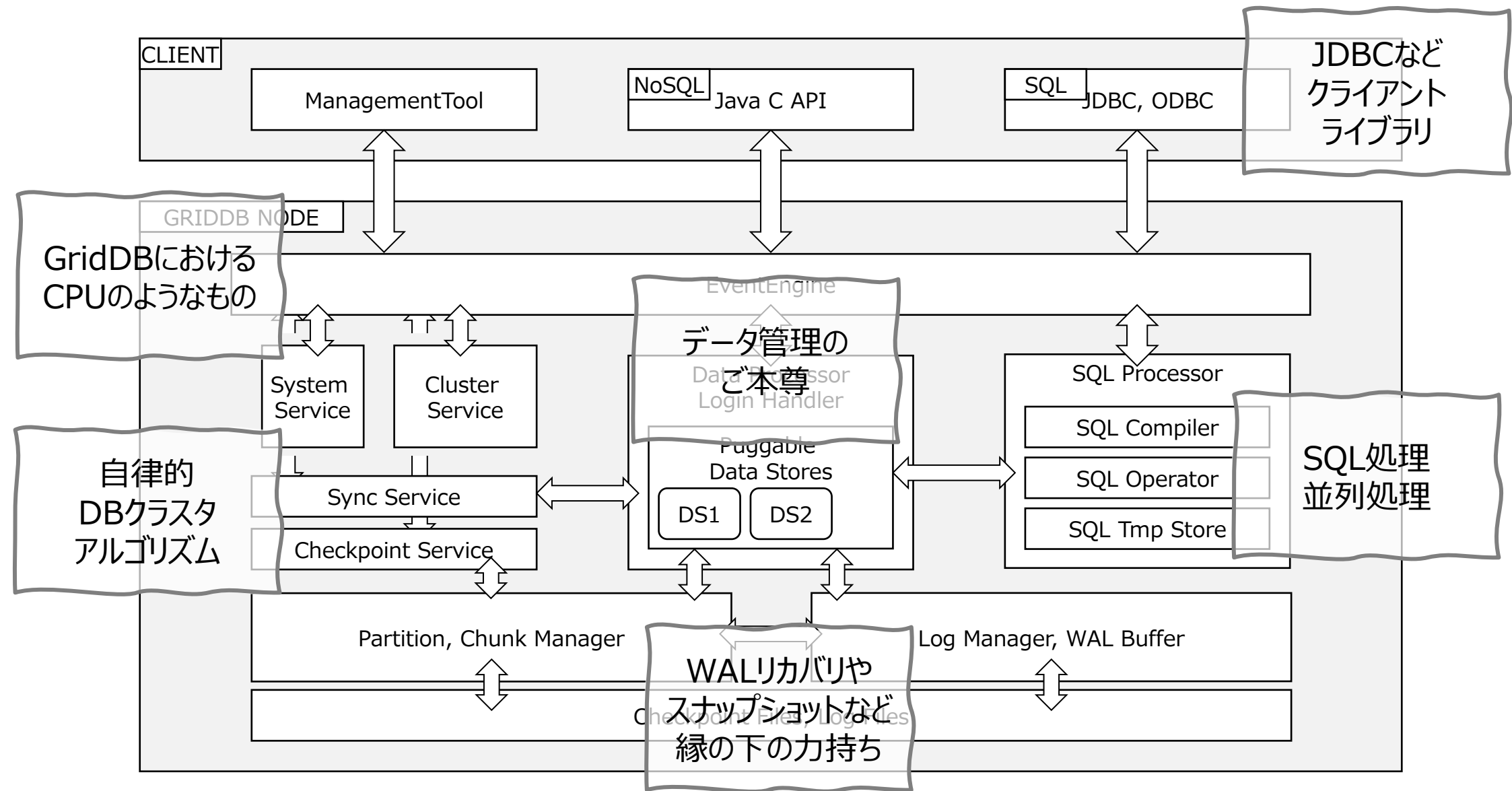
使い分けがなぜ問題なのか？



これらの課題を忙殺されずに、ビジネスの付加価値向上に注力できる



内部構成



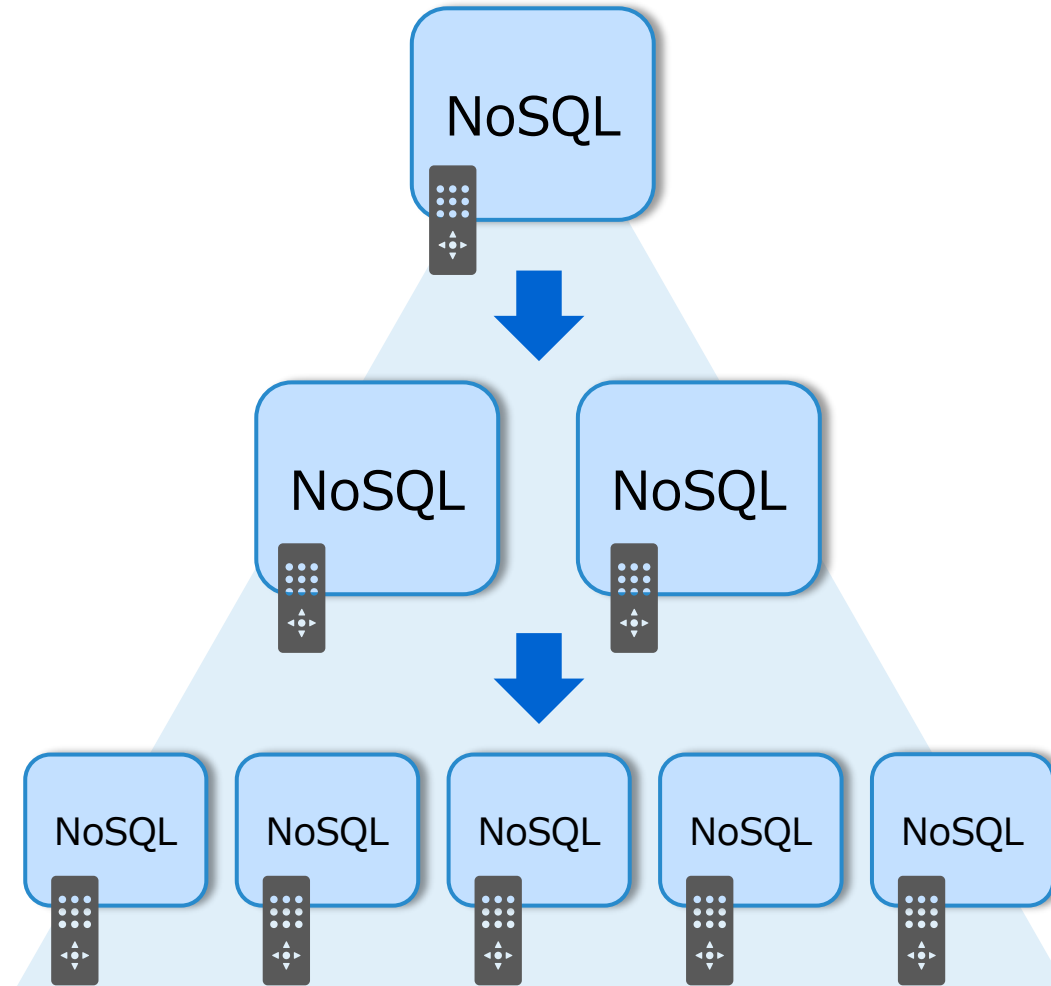
03

自律的なDBクラスタ

効率良く複数サーバを利用してパフォーマンスアップ



効率良く複数サーバを利用してパフォーマンスアップ°



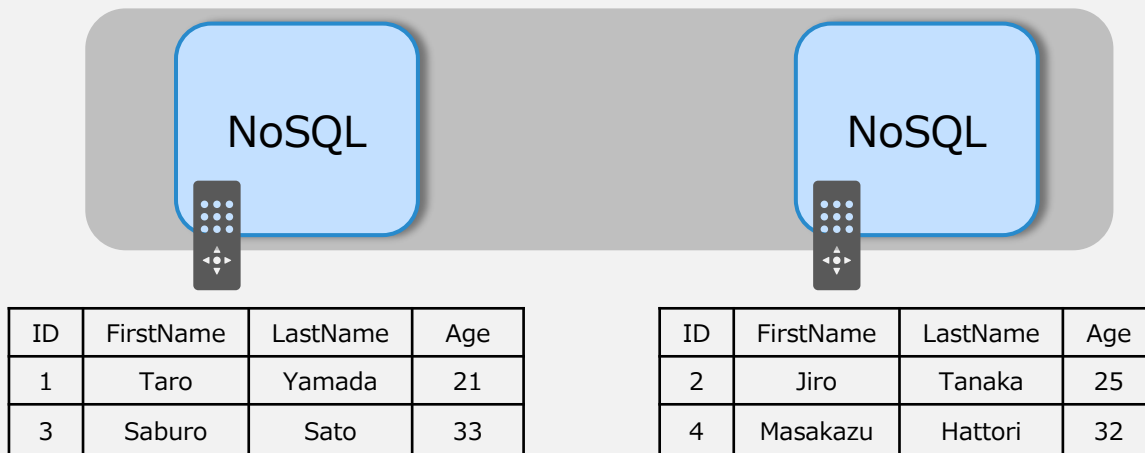
スケールアウトDBの基本概念

(1) シャーディング(Sharding) …複数サーバへのデータ分散

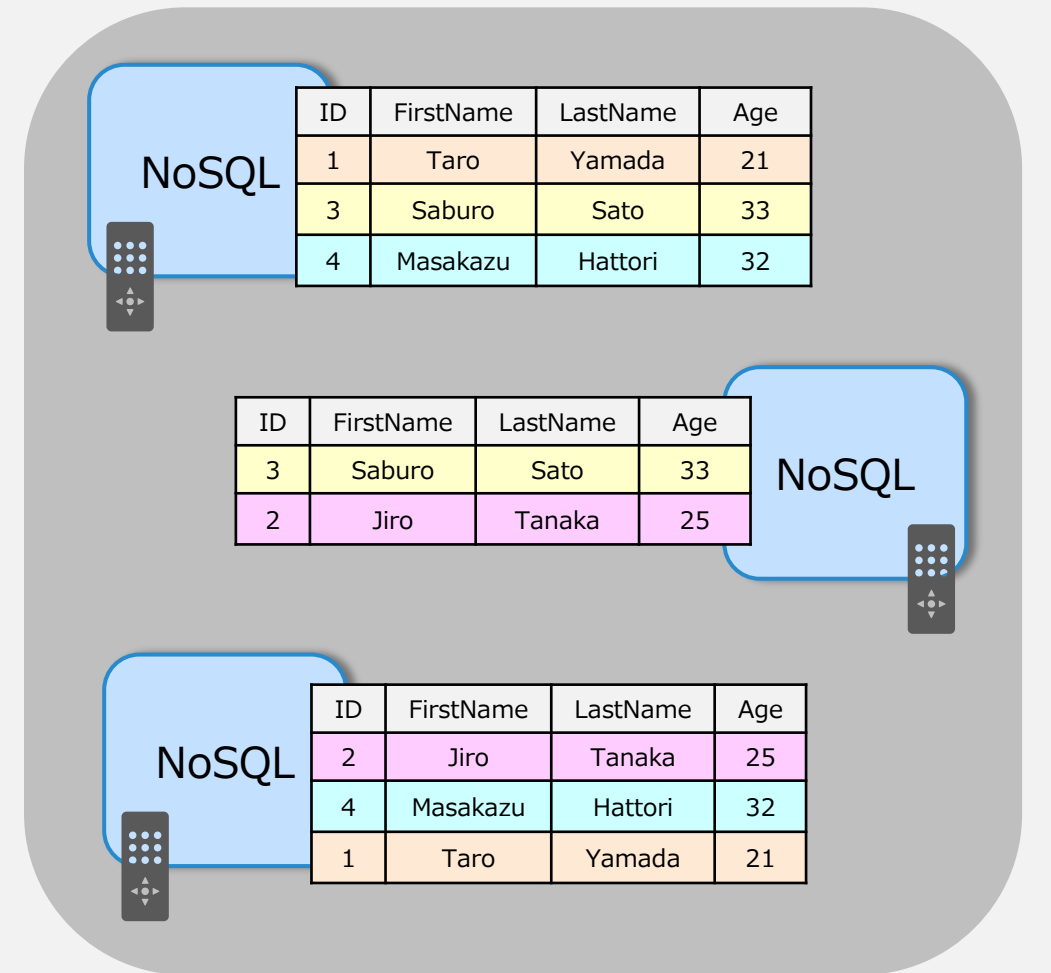
ID	FirstName	LastName	Age
1	Taro	Yamada	21
2	Jiro	Tanaka	25
3	Saburo	Sato	33
4	Masakazu	Hattori	32

オリジナルテーブル

Horizontal Partitions



(2) レプリケーション(Replication) …障害に備えたデータ複製



スケールアウトDB

	従来のスケールアウト型DB	GridDB
初期クラスタ	問題無し	問題無し
サーバ増設	リシャードが遅い オフラインに戻す必要あり	オンラインスケールアウト
クラスタ縮退	さらにNGなケースも	オンラインスケールダウン
障害後の現状復帰	オフライン＆マニュアルで復帰	最適なクラスタ維持

データサービスを支えるDB。DBライフサイクルの観点で考えるべき

自律的DBクラスタを支える技術群

	GridDB
初期クラスタ	問題無し
サーバ増設	オンラインスケールアウト
クラスタ縮退	オンラインスケールダウン
障害後の現状復帰	最適なクラスタ維持

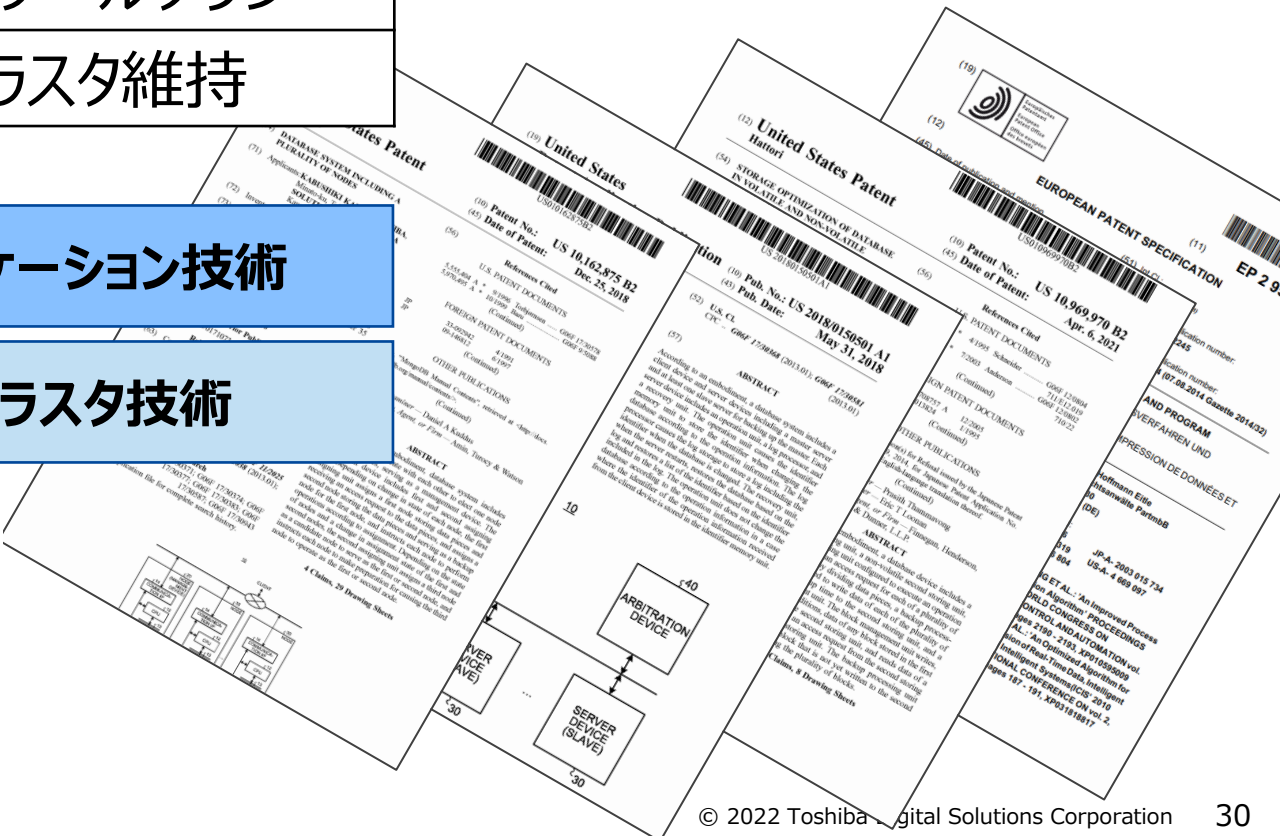
①プランニング技術

②高速レプリケーション技術

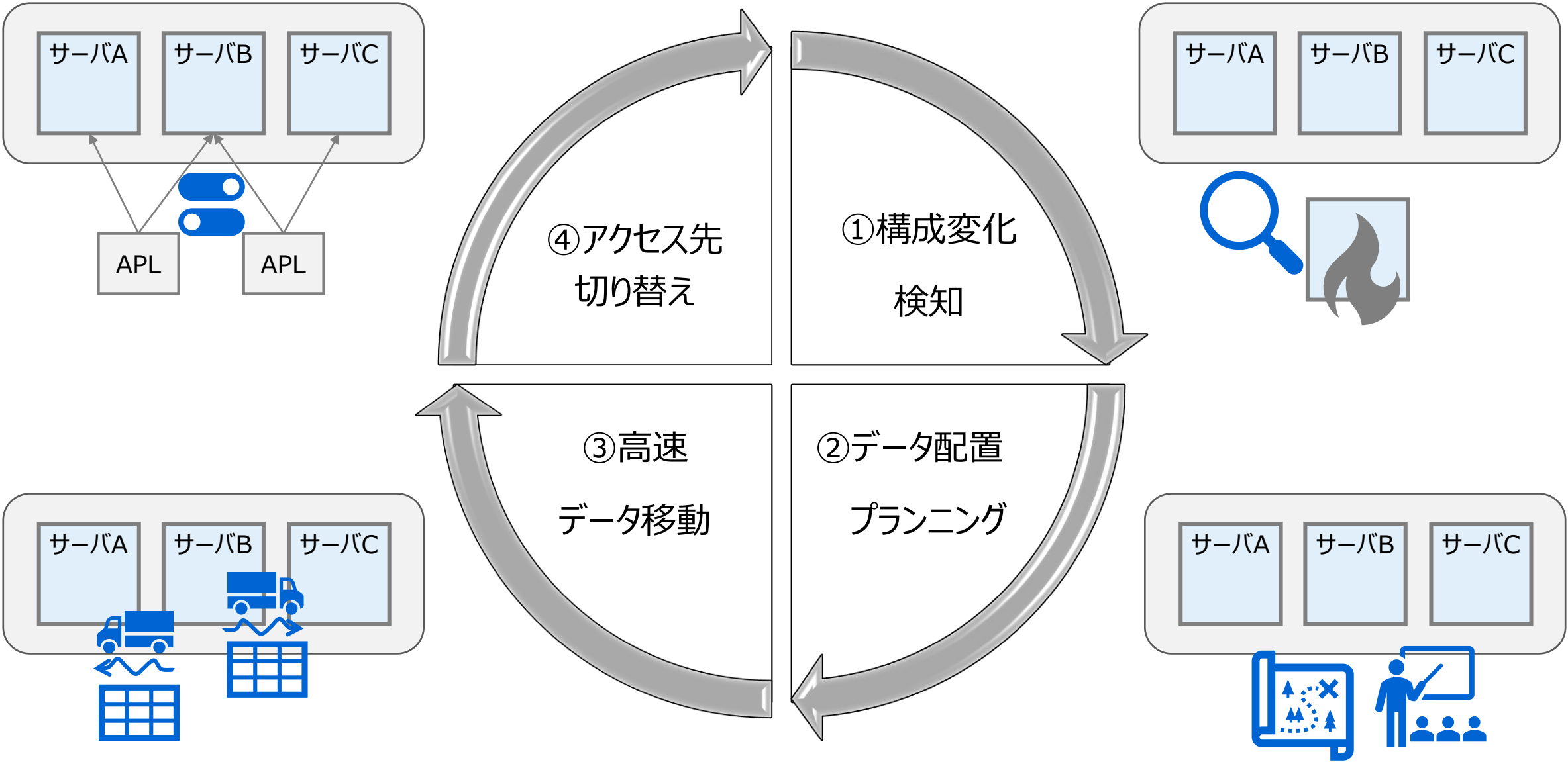
③オートシャーディング技術

④マスタレスクラスタ技術

“Database system including a plurality of nodes”
“Database system, server device, computer program product,
and information processing method”
US10162875B2、US10685041B2、US20180150501A1、etc
日米欧等特許済



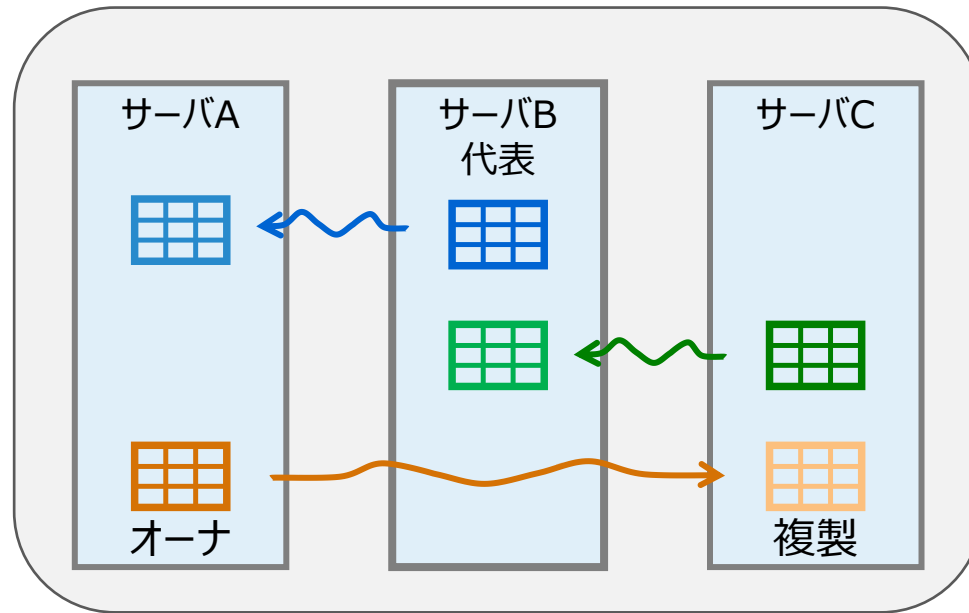
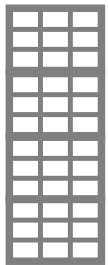
アイデアの骨子



オンラインスケールアウト・イメージ

- 代表サーバ選出フェーズ
サーバ間の選挙により選出
当該サーバ障害後は再選挙

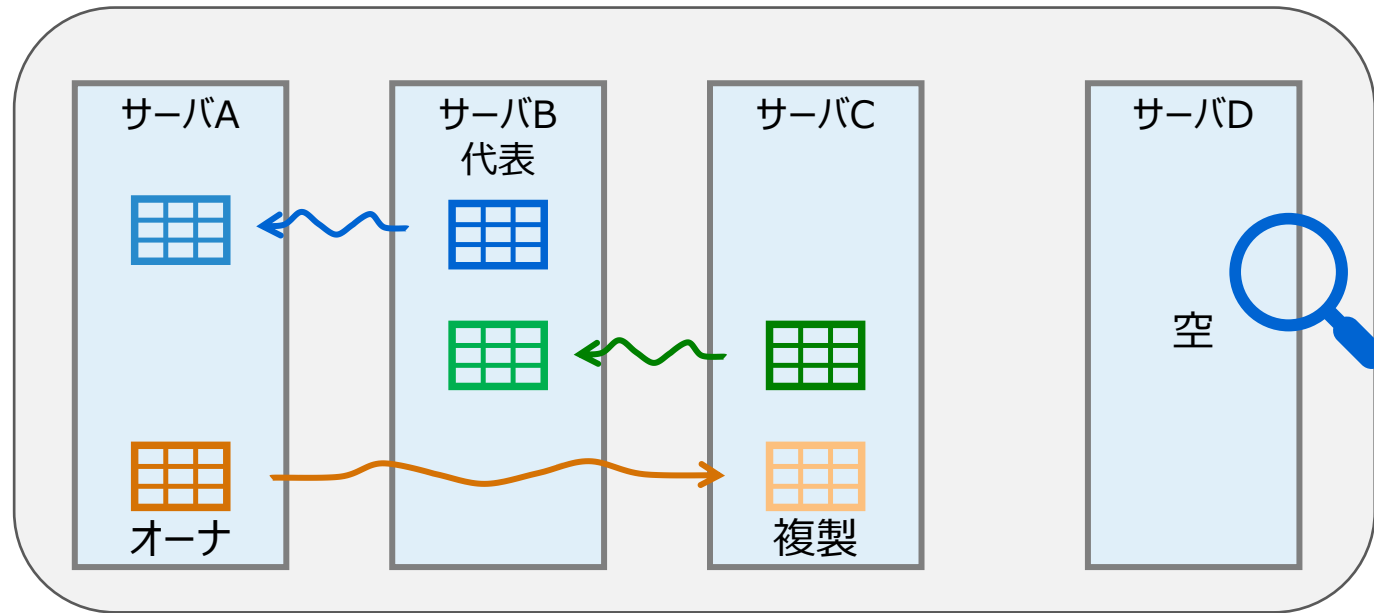
シャーディング



レプリケーション

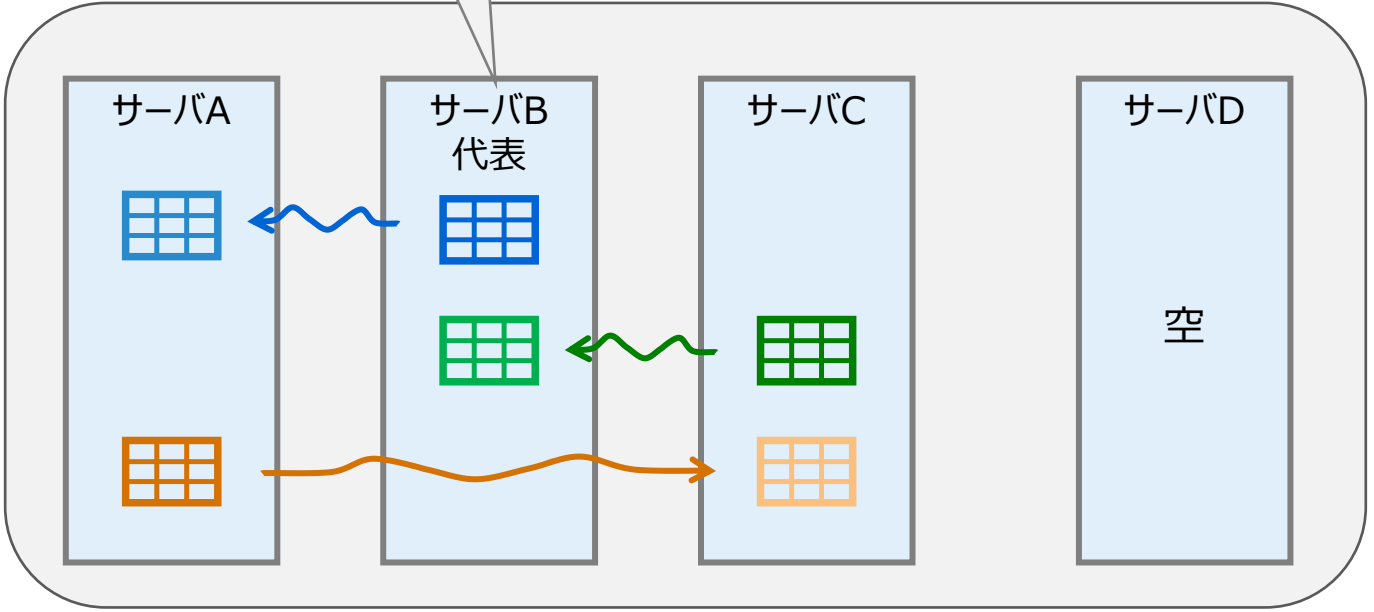
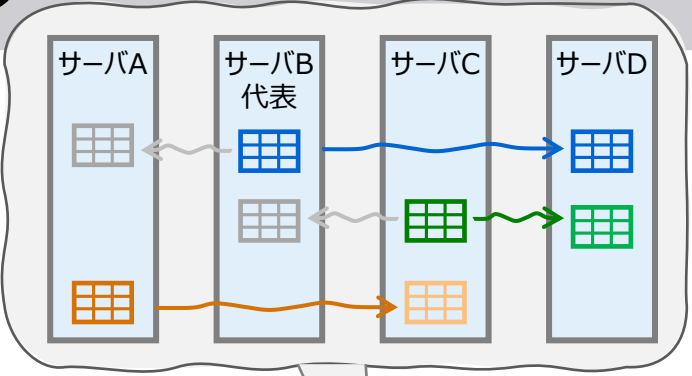
オンラインスケールアウト・イメージ

①構成変化検知
(代表サーバが)
クラス構成変更の検知



オンラインスケールアウト・イメージ

②データ配置プランニング
(代表サーバが)
データ配置変更プランの作成

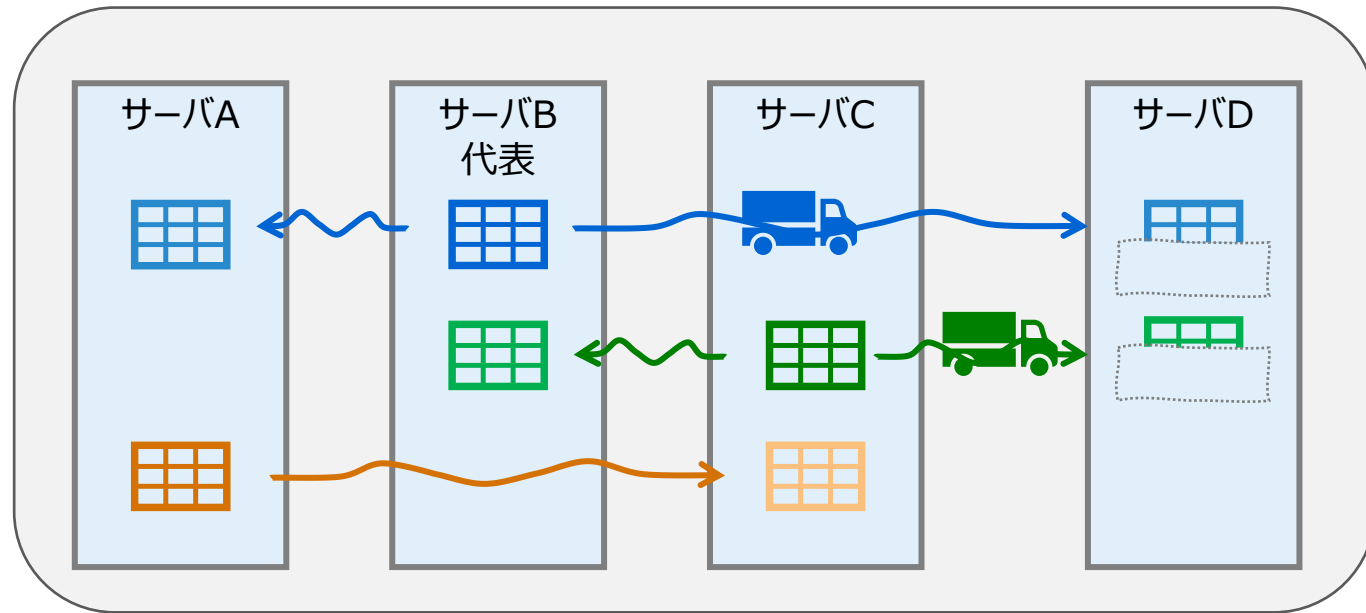


オンラインスケールアウト・イメージ

③高速データ移動

プランに従い（関連サーバ間で）

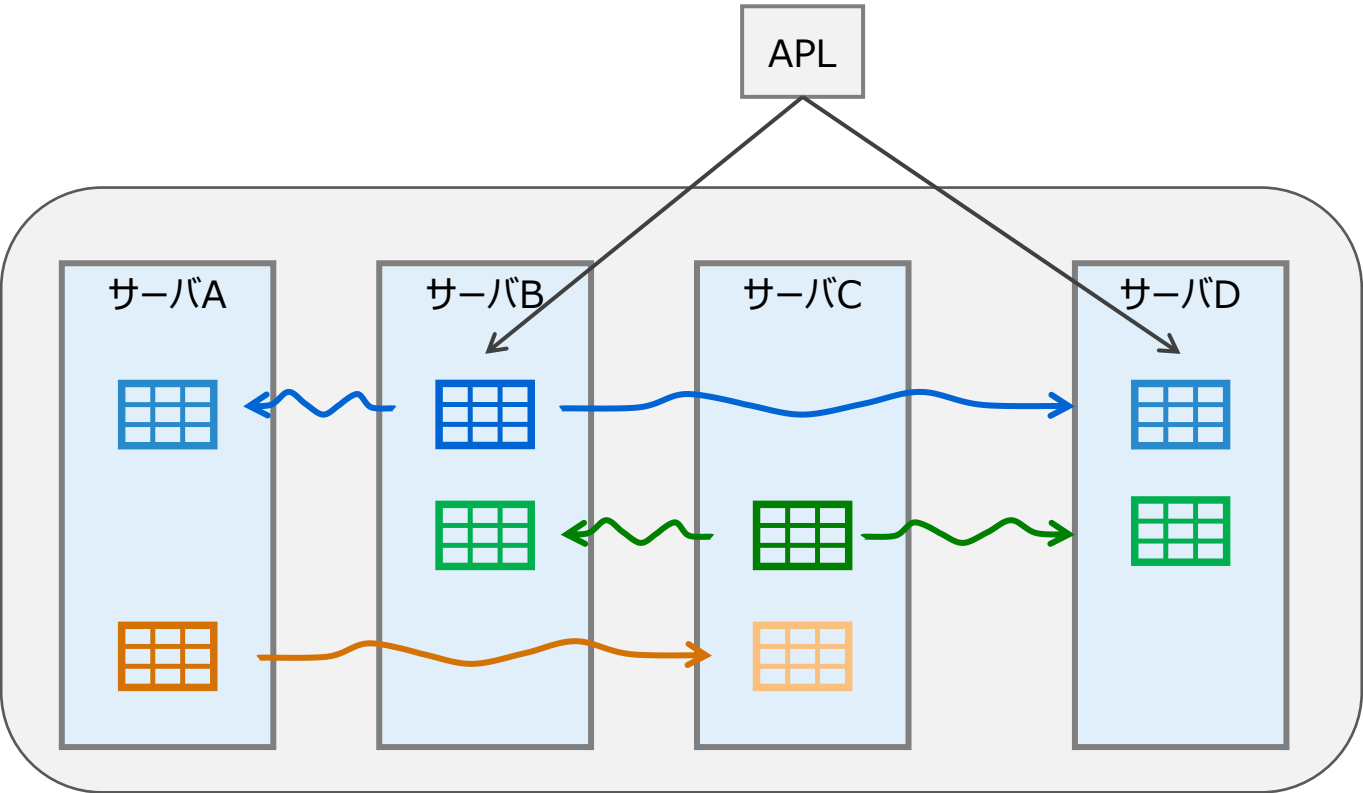
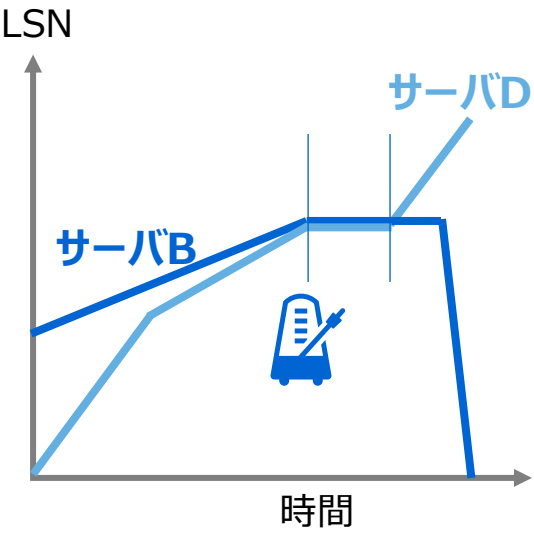
- データ送受信により複製作成
- 高速化のため、以下の3手順
(A)ブロックイメージ送受
(B)WALログ送受
(C)トランザクション情報含む



オンラインスケールアウト・イメージ

④アクセス先切り替え

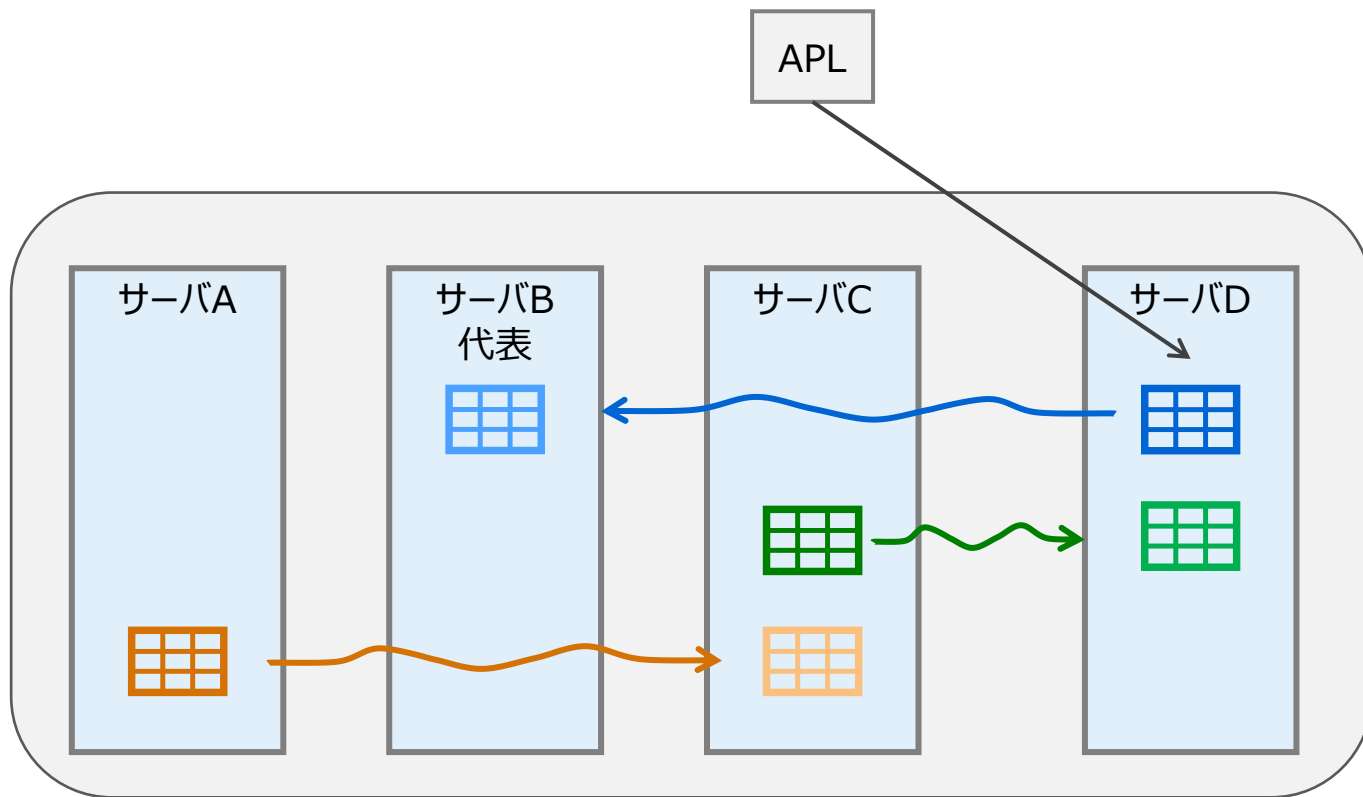
(関連サーバが)
タイミングをとったうえで
アプリ側に接続先切り替えを通知



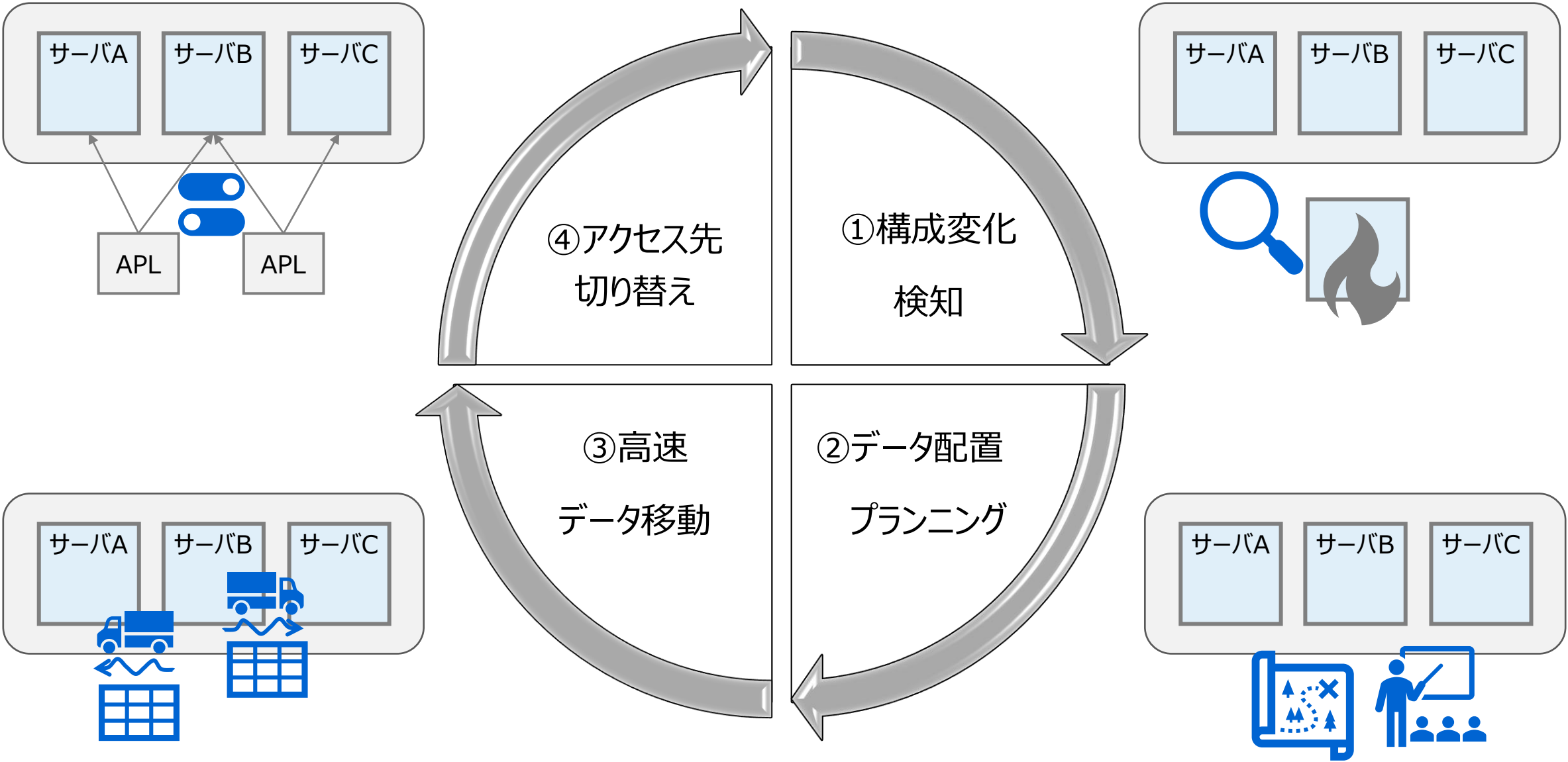
オンラインスケールアウト・イメージ

④アクセス先切り替え

(関連サーバが)
タイミングをとったうえで
アプリ側に接続先切り替えを通知



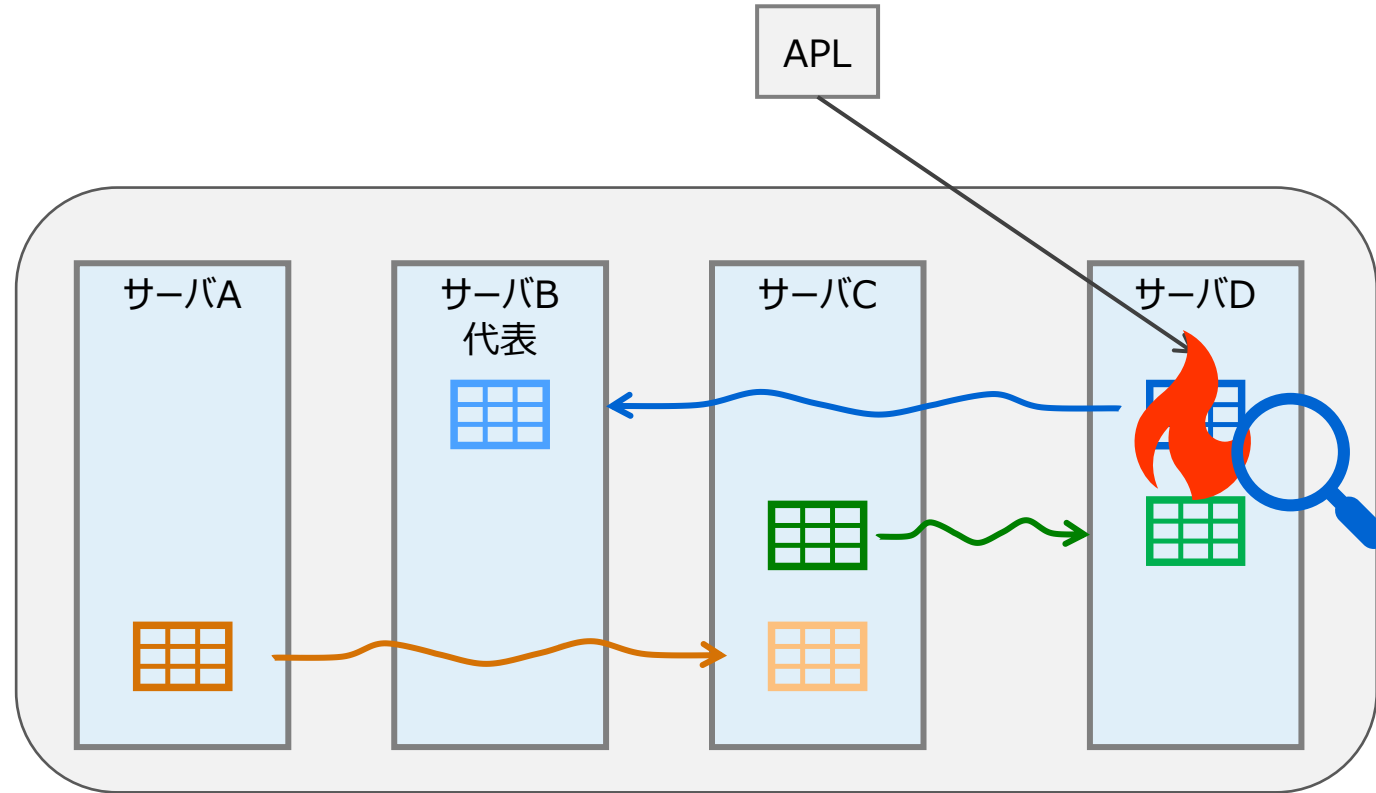
データライフサイクルの観点



サーバ障害フェールオーバー・イメージ

①構成変化検知
(代表サーバが)
クラスタ構成変更の検知

②データ配置プランニング
(代表サーバが)
データ配置変更プランの作成



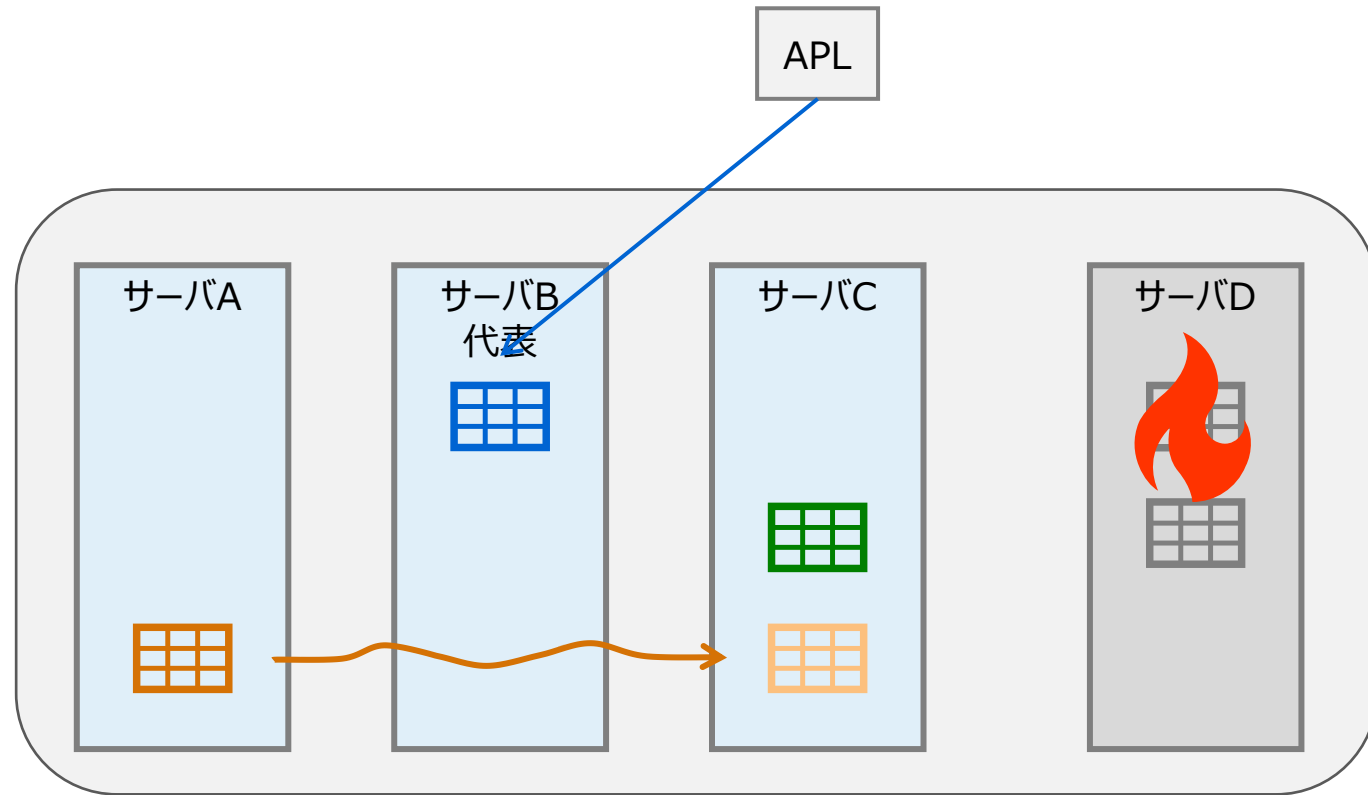
サーバ障害フェールオーバー・イメージ

①構成変化検知
(代表サーバが)
クラスタ構成変更の検知

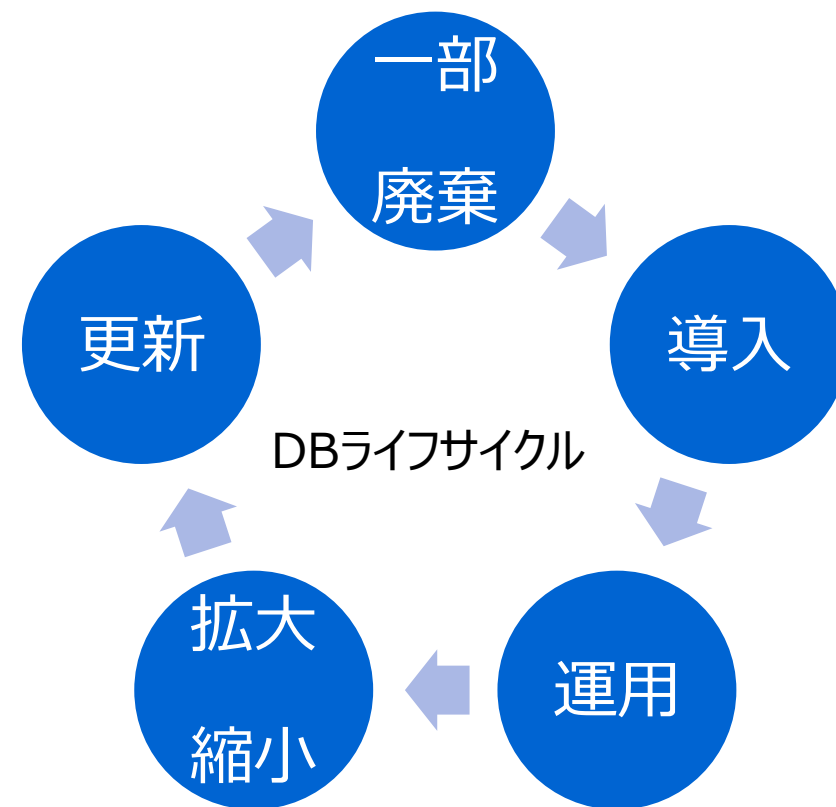
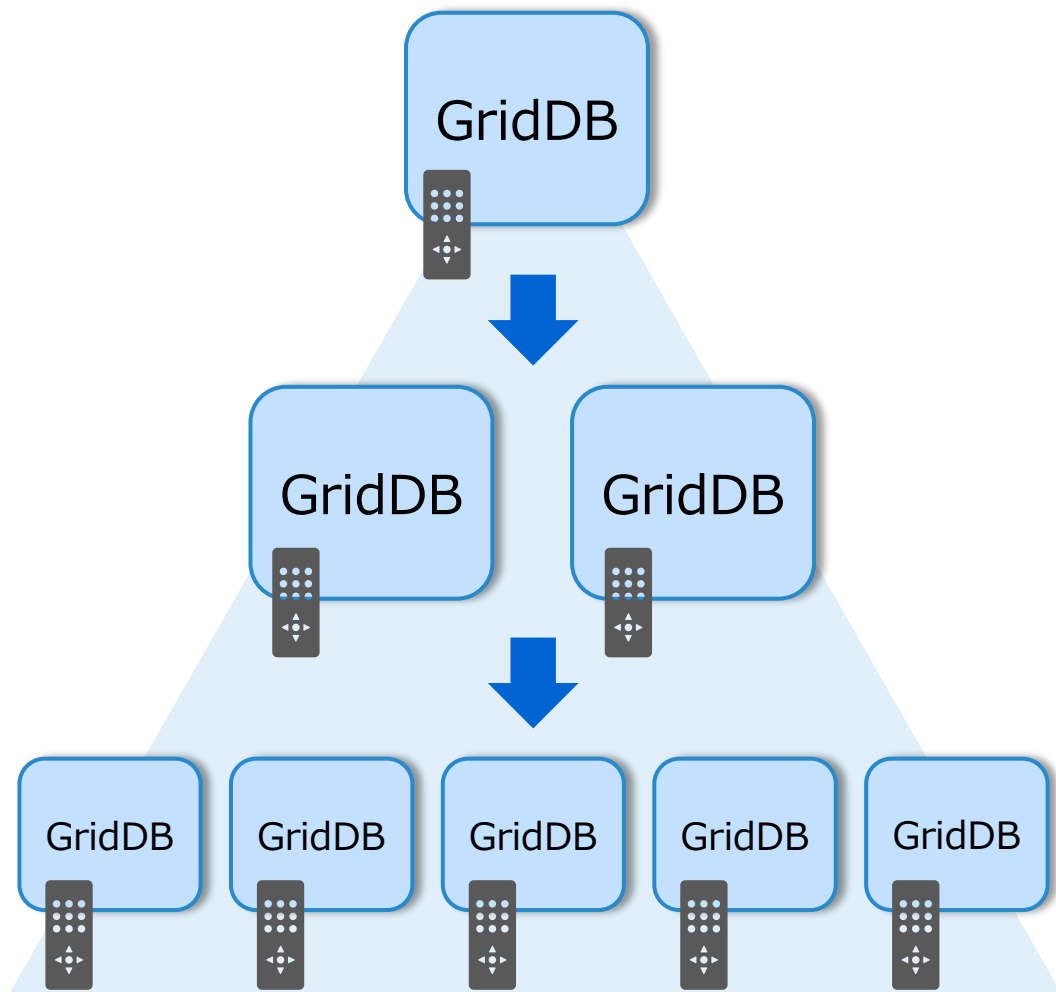
②データ配置プランニング
(代表サーバが)
データ配置変更プランの作成

④アクセス先切り替え

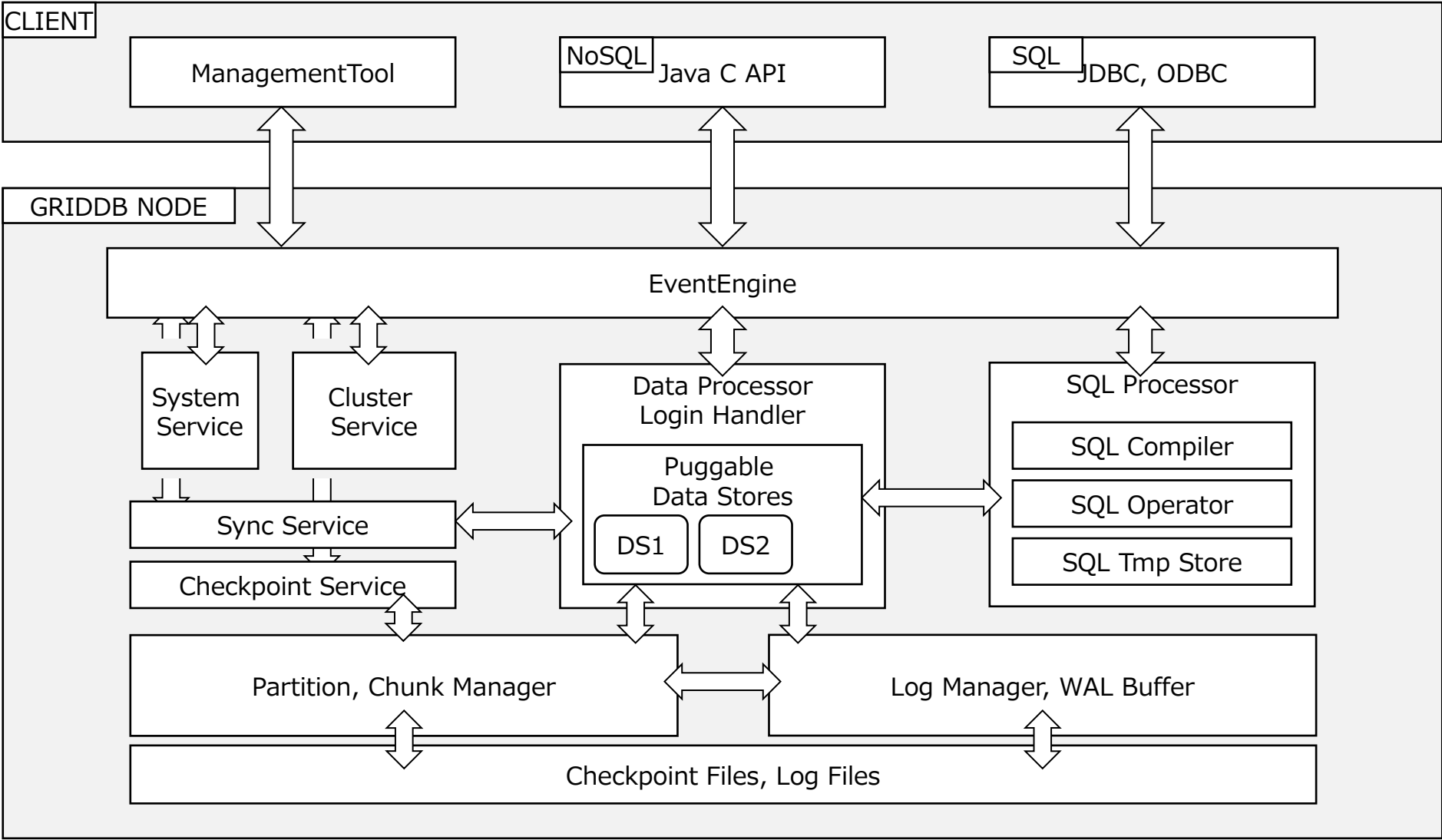
(関連サーバが)
タイミングをとったうえで
アプリ側に接続先切り替えを通知



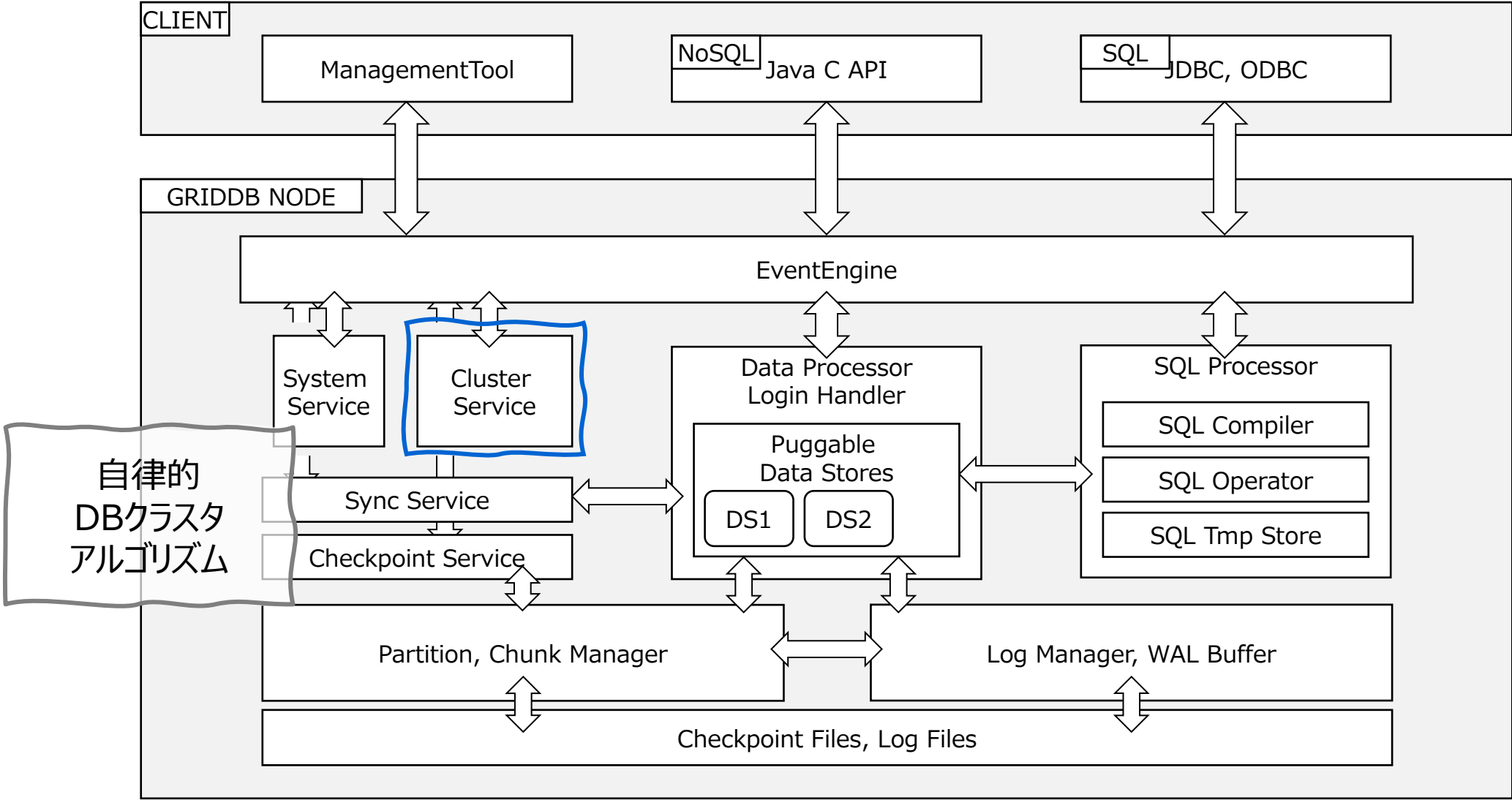
DBライフサイクルを支えるGridDB



内部構成



内部構成



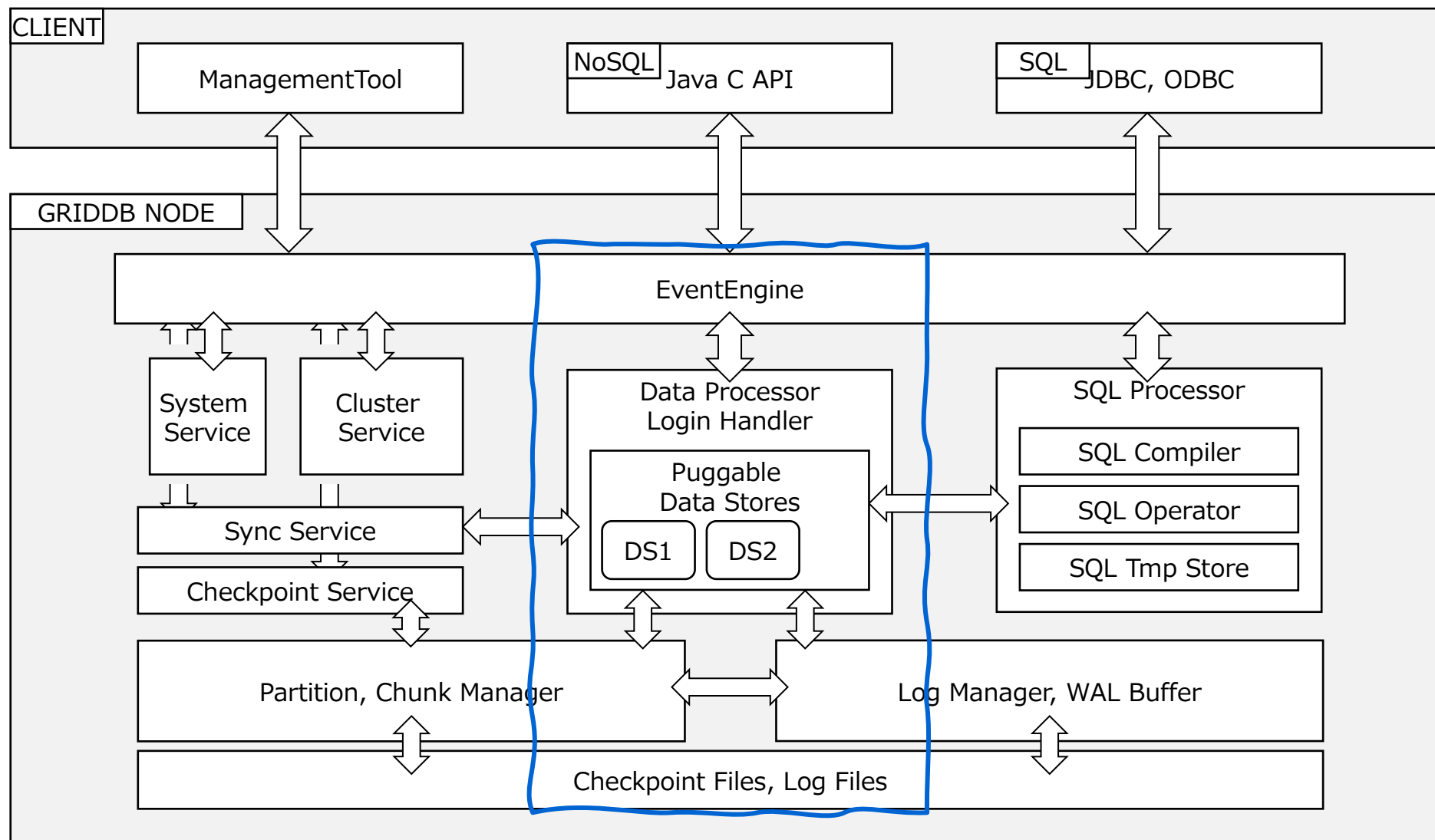
04

イベント駆動アーキテクチャ

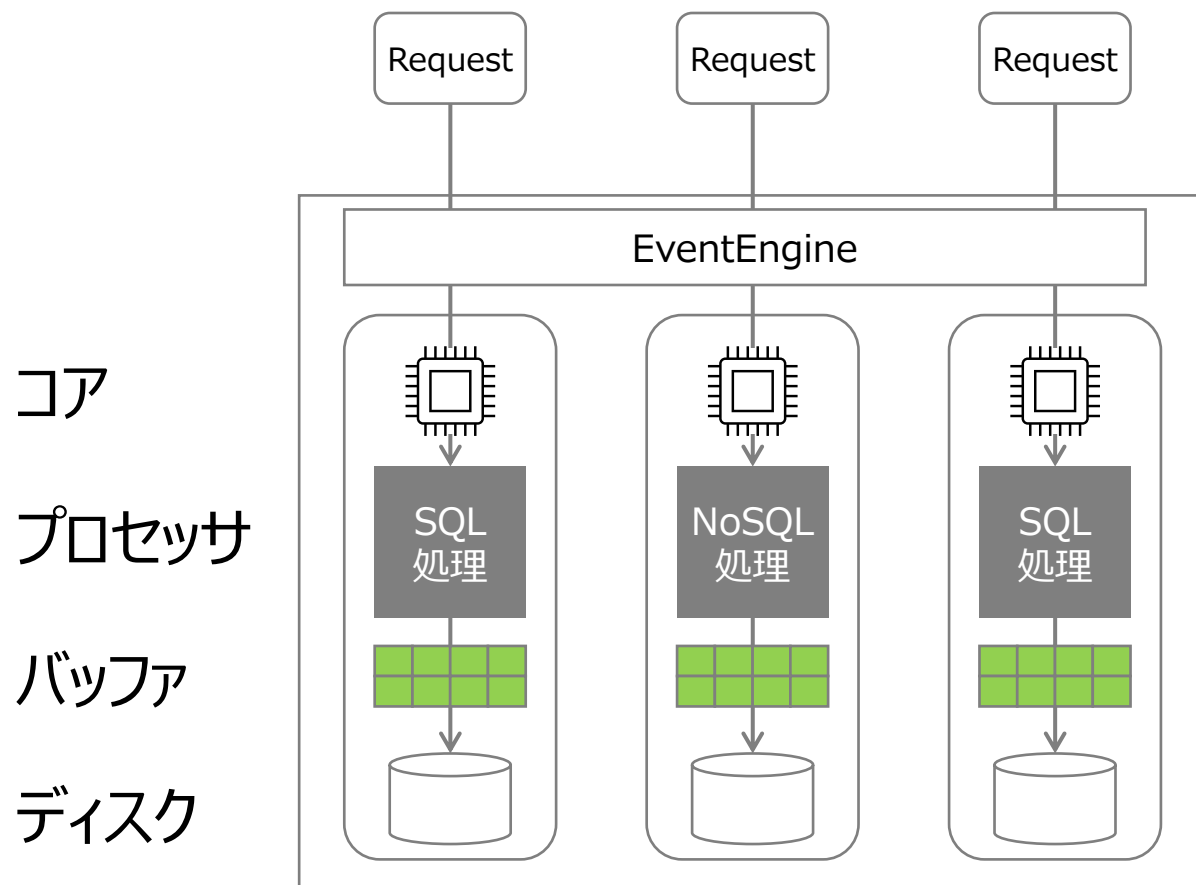
効率良く複数コア、メモリを利用してパフォーマンスアップ



内部構成



効率良く複数コア、メモリを利用してパフォーマンスアップ



CPU関係

- スレッドへプロセッサ割り当て
- 要求をスレッドへ振り分け
- 時分割による同時実行

メモリ関係

- ロックフリー化
- バッファ処理の軽量化

ディスク関係

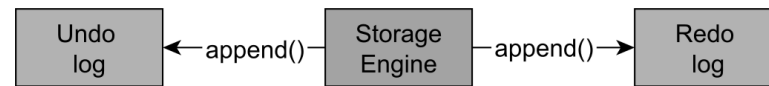
- リカバリ処理の軽量化
- ブロック書き込みの最適化

既存DBの構成

コア

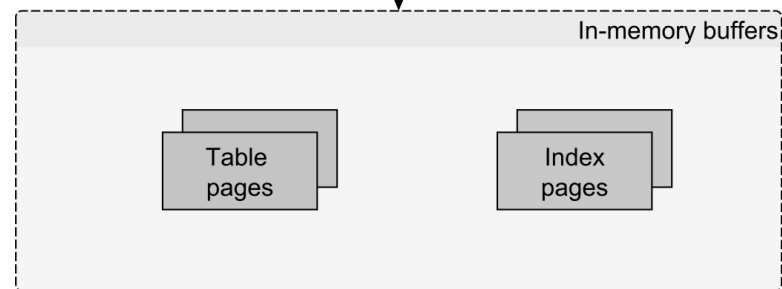


プロセッサ



CRUD

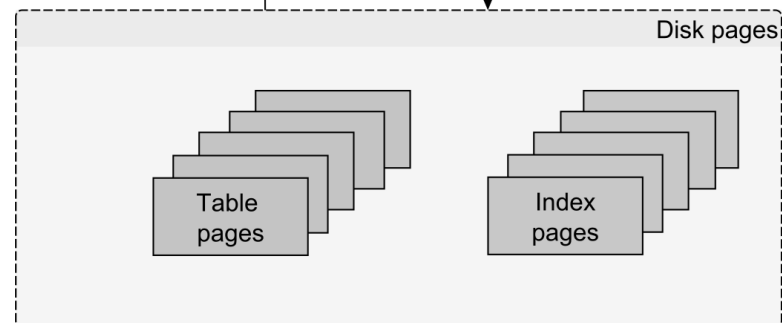
バッファ



map()

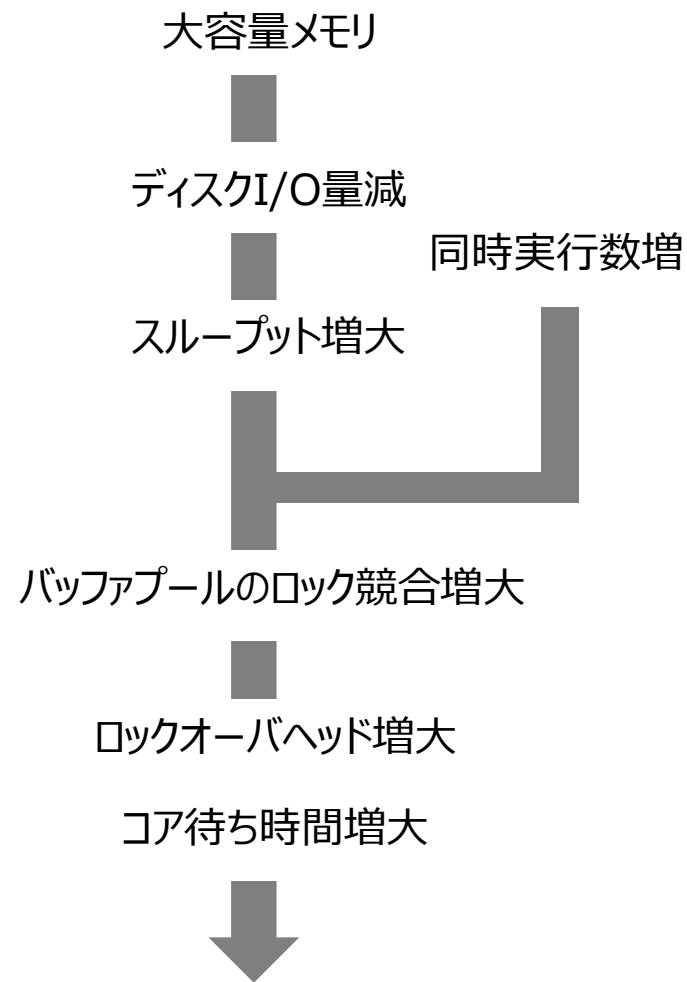
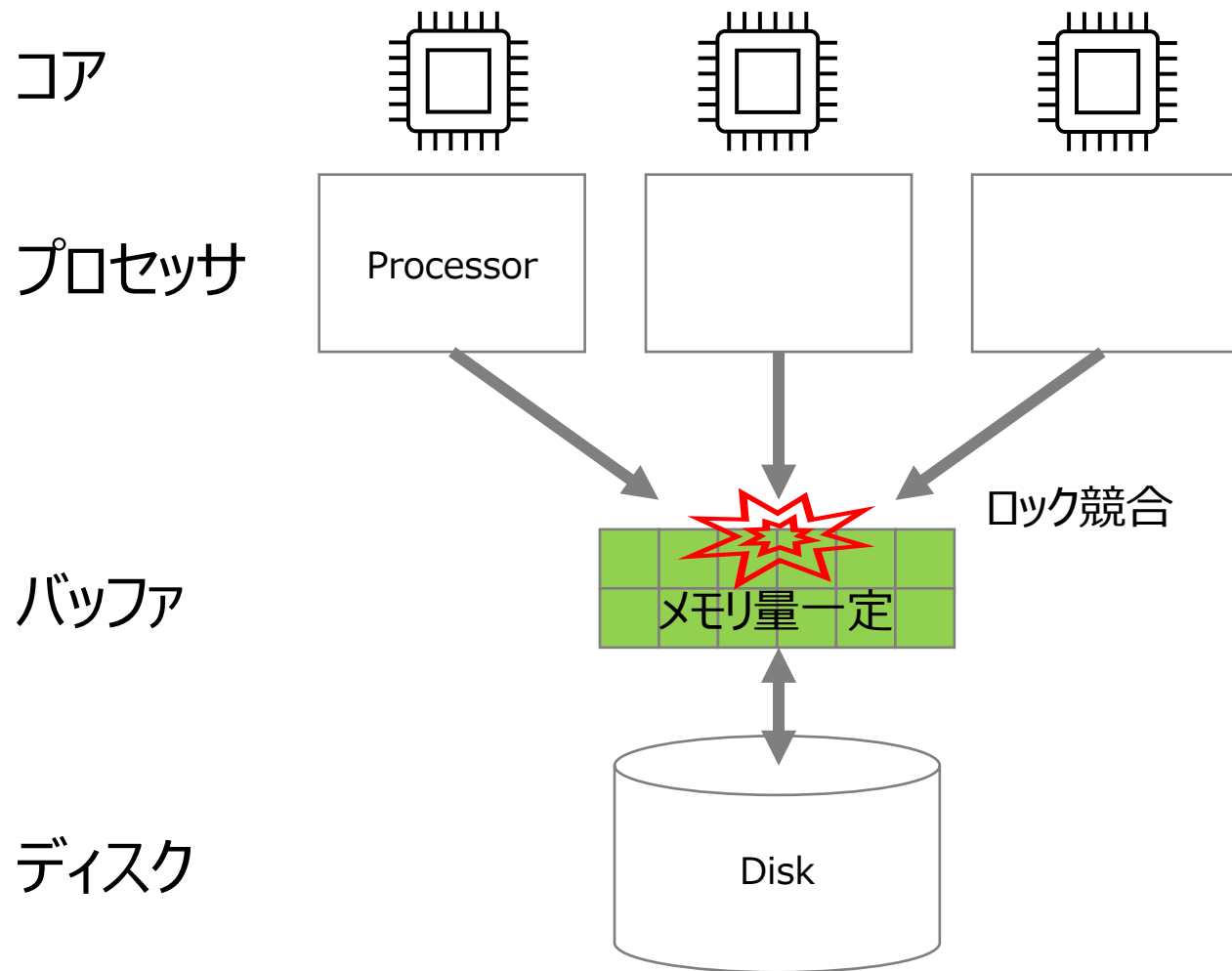
sync()

ディスク



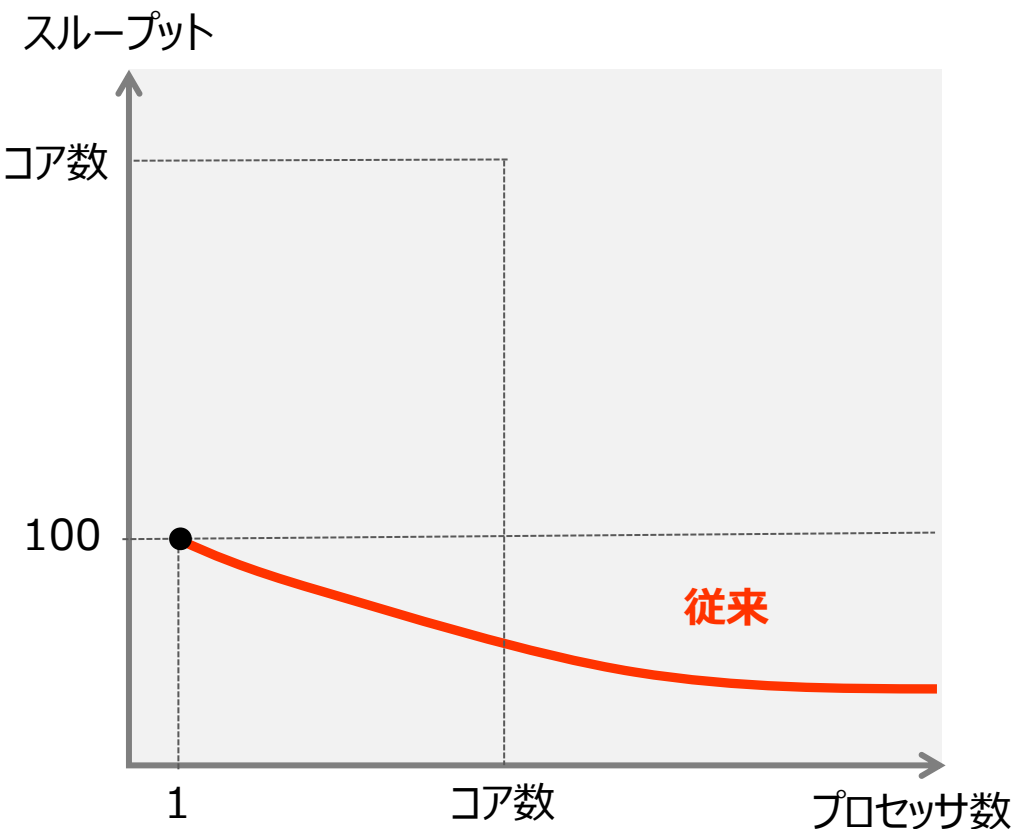
<https://vladmihalcea.com/tips-oracle-jpa-hibernate/>

既存DBの構成

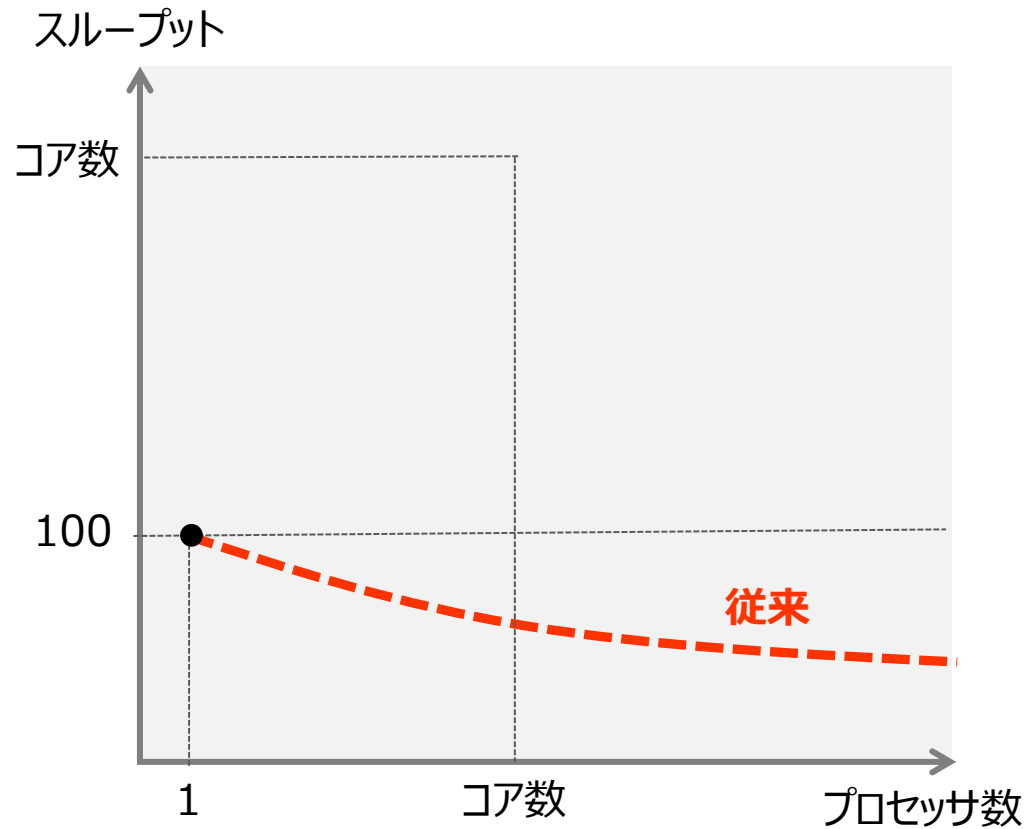


期待したパフォーマンスが出てこない

インメモリ時の実験データ

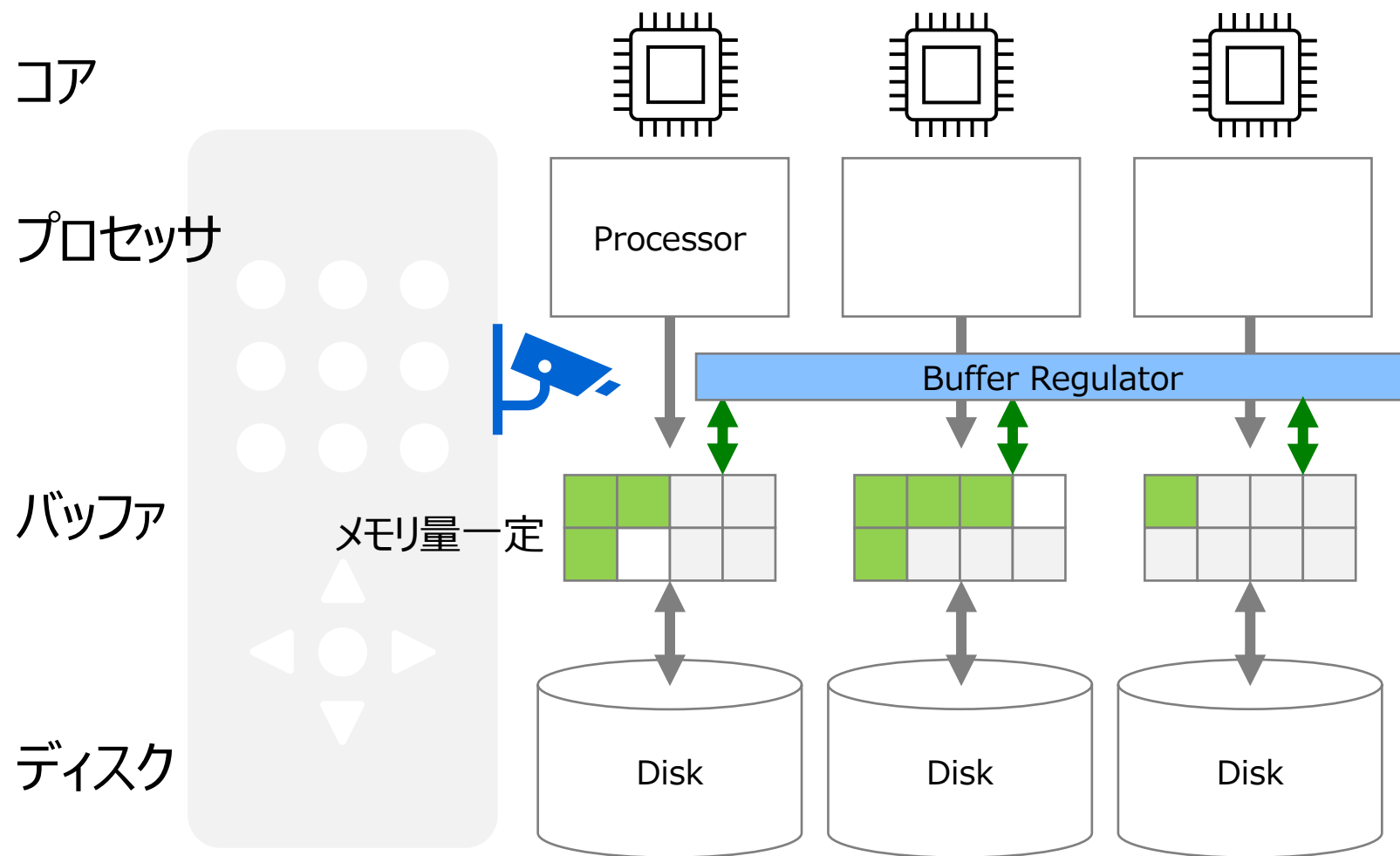


登録性能



検索性能

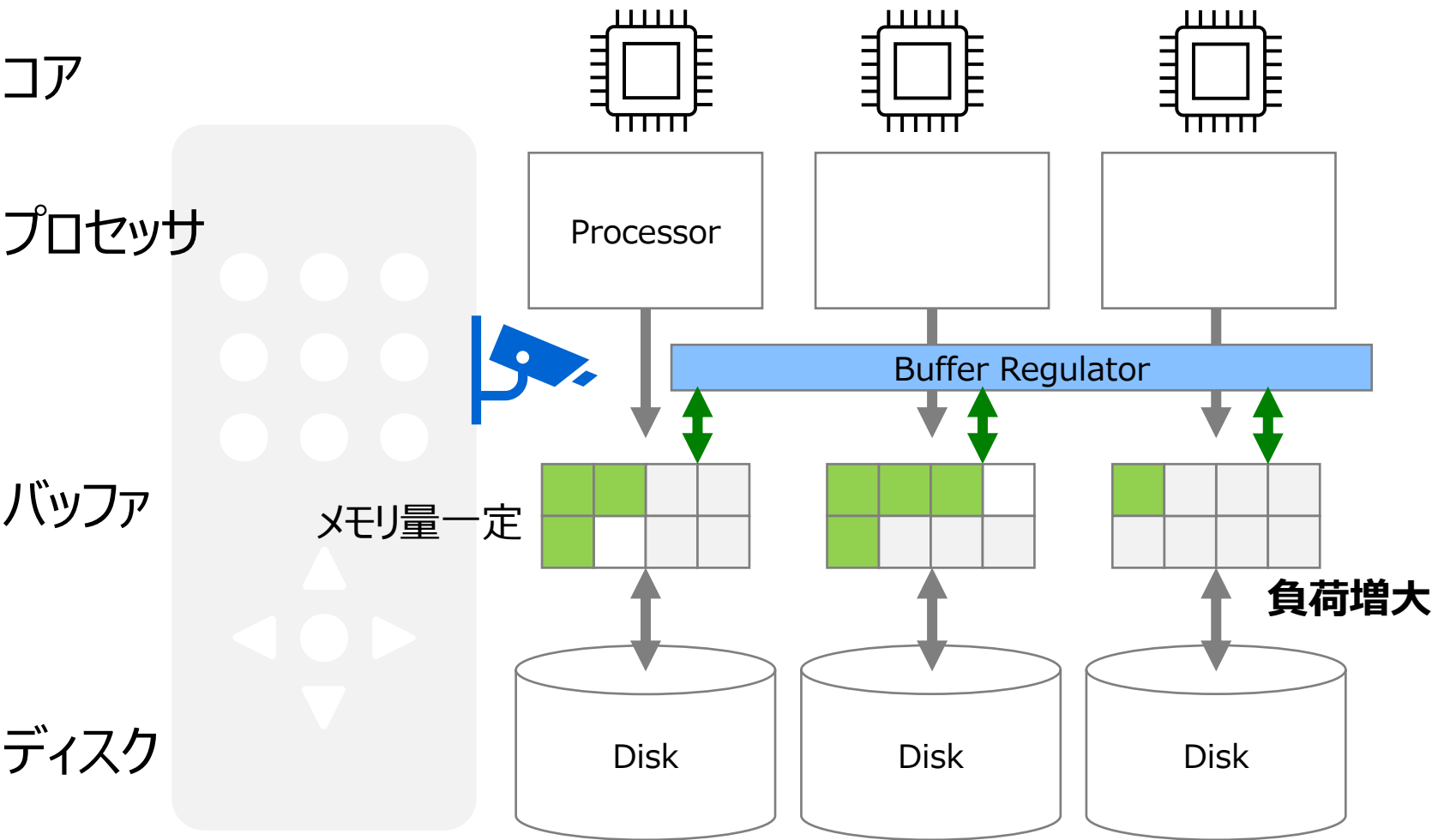
透過的メモリシェアリング、一種のメモリ融通



メモリ融通の一種

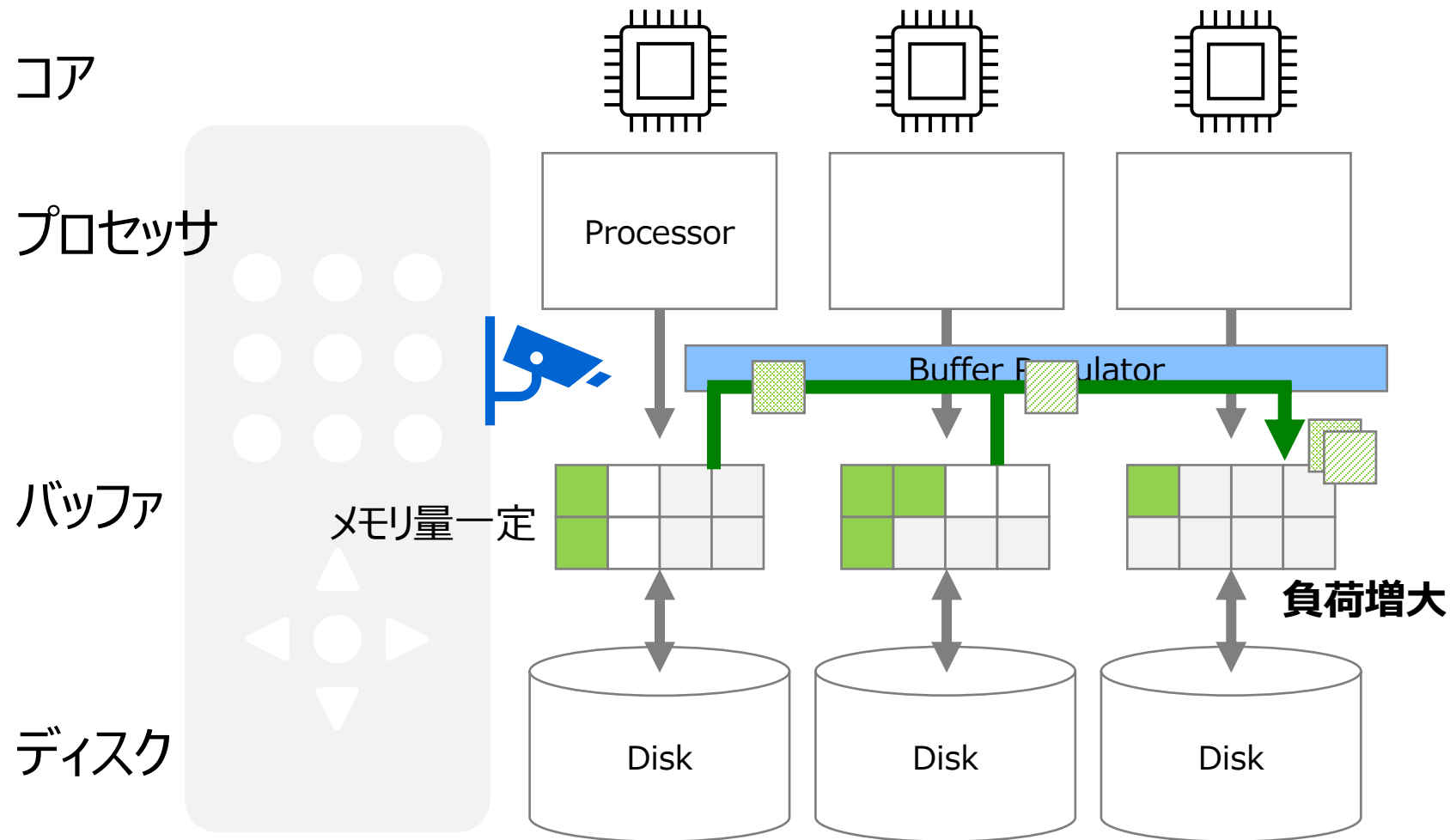
- メモリ利用状況を監視
- メモリー割当不足を予測
- バッファプール上の使われなさそうなメモリー特定
- メモリの貸し借り実施
- 衝突をできる限り削減

GridDBの場合



- メモリ融通の一種
- メモリ利用状況を監視
 - メモリー割当不足を予測
 - バッファプール上の使われなさそうなメモリー特定
 - メモリの貸し借り実施
 - 衝突をできる限り削減

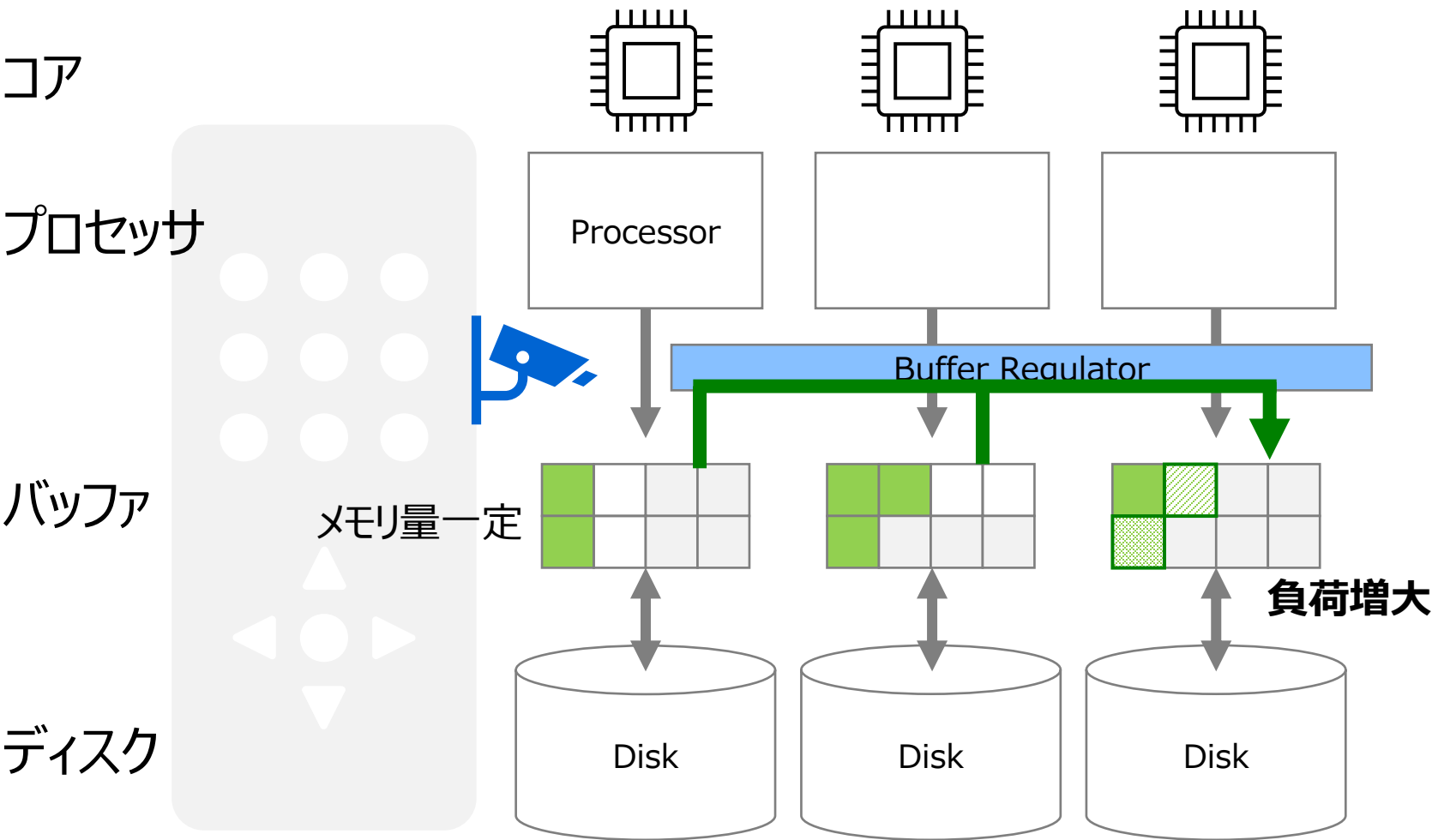
GridDBの場合



メモリ融通の一種

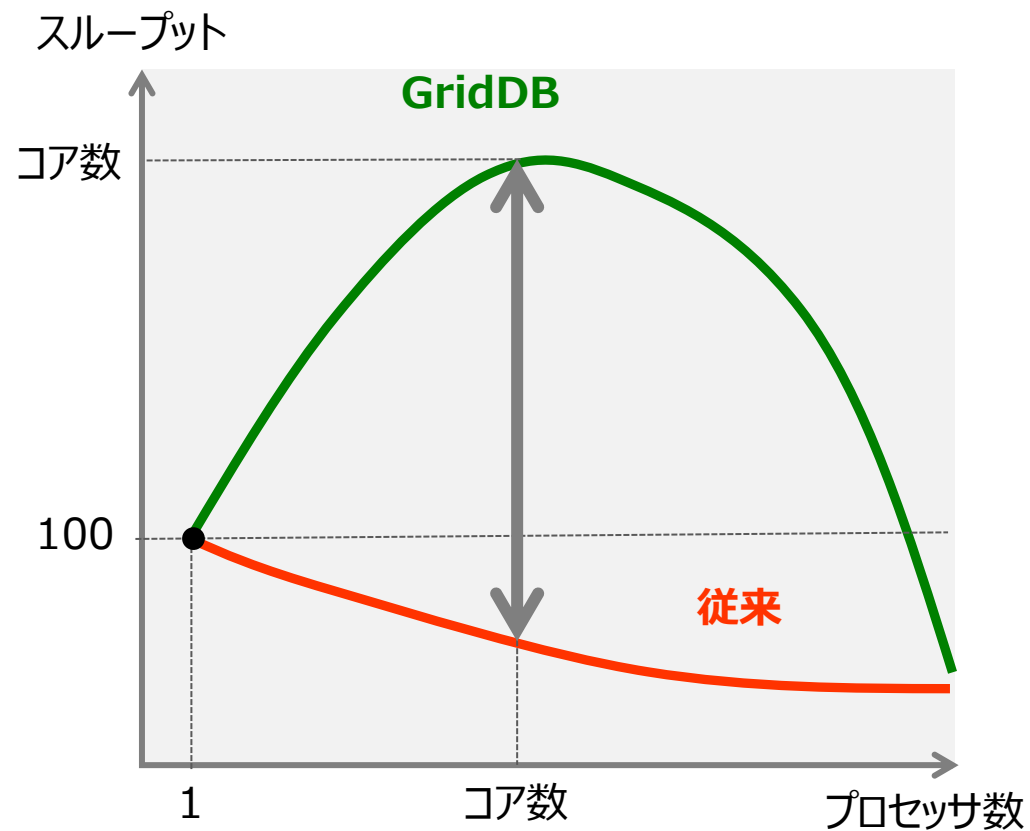
- メモリ利用状況を監視
- メモリー割当不足を予測
- バッファプール上の使われなさそうなメモリー特定
- メモリの貸し借り実施
- 衝突をできる限り削減

GridDBの場合

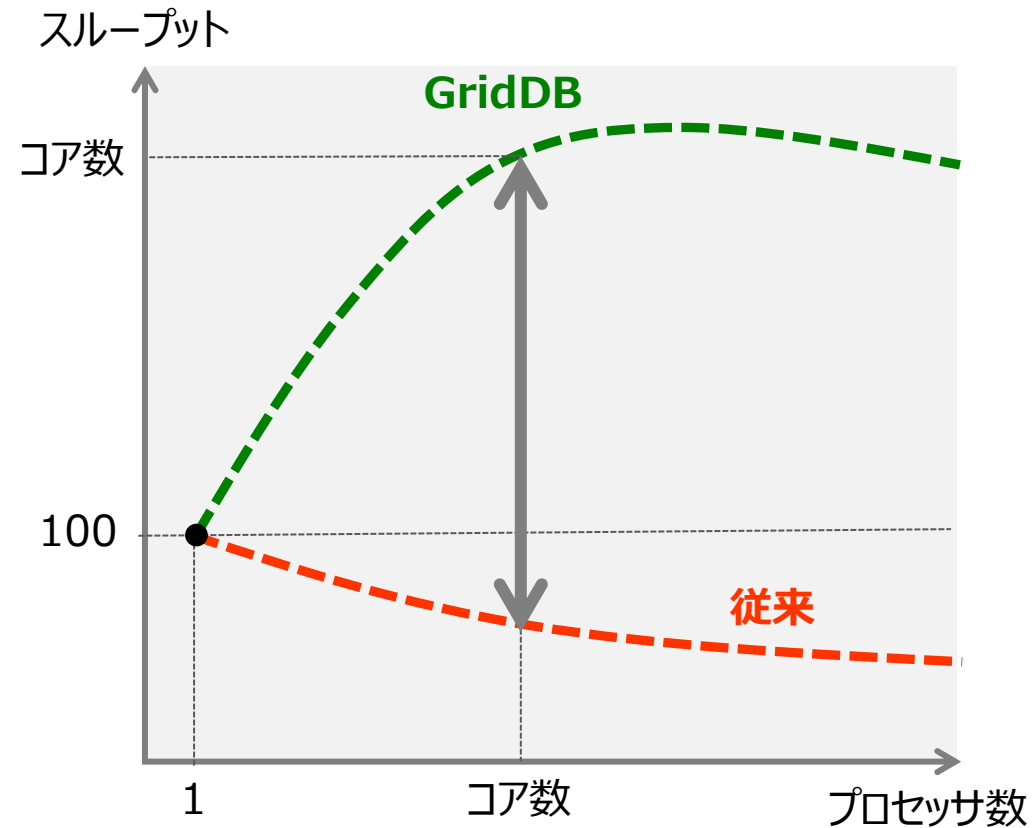


- メモリ融通の一種
- メモリ利用状況を監視
 - メモリー割当不足を予測
 - バッファプール上の使われなさそうなメモリー特定
 - メモリの貸し借り実施
 - 衝突をできる限り削減

数倍から数十倍の差



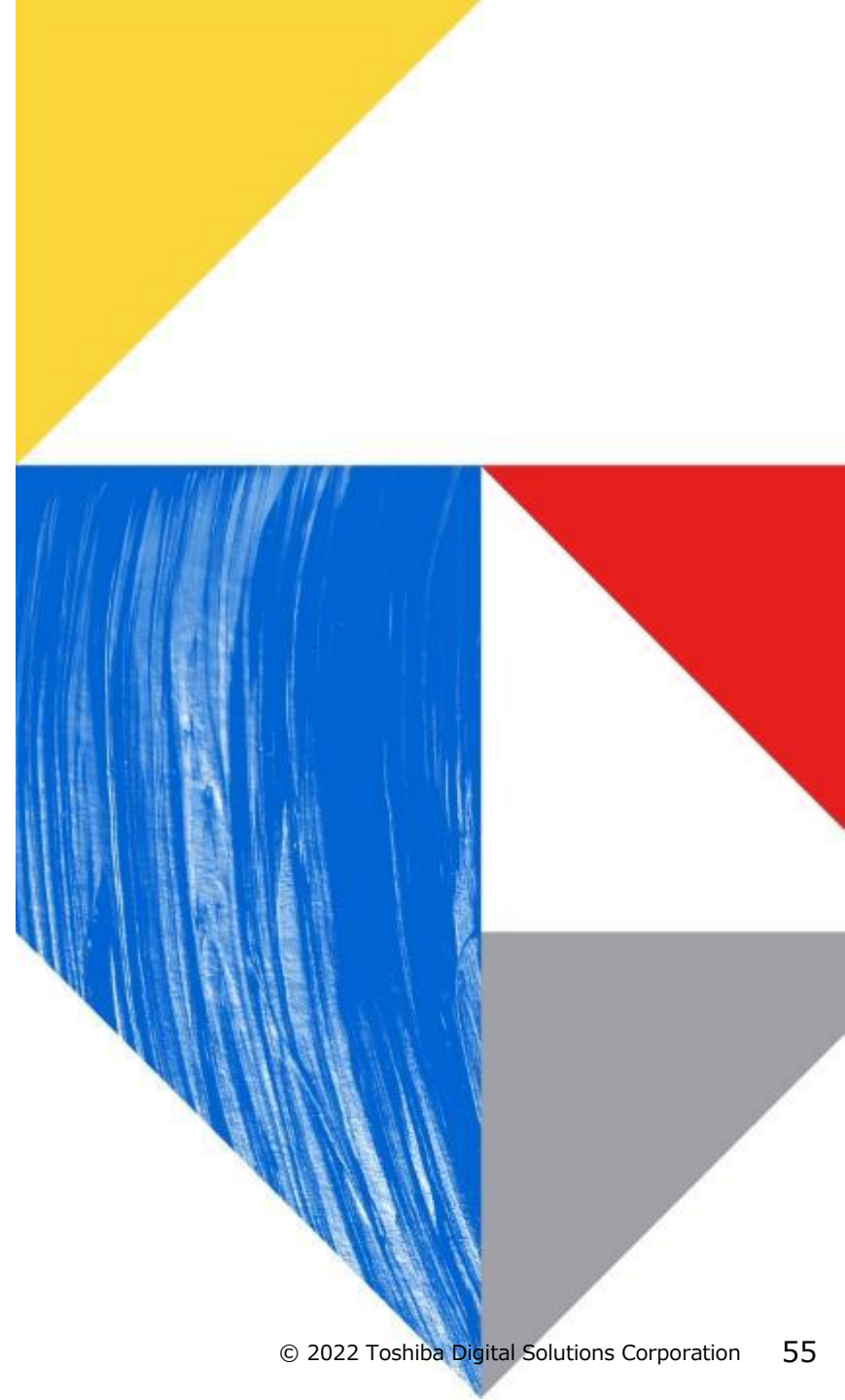
登録性能



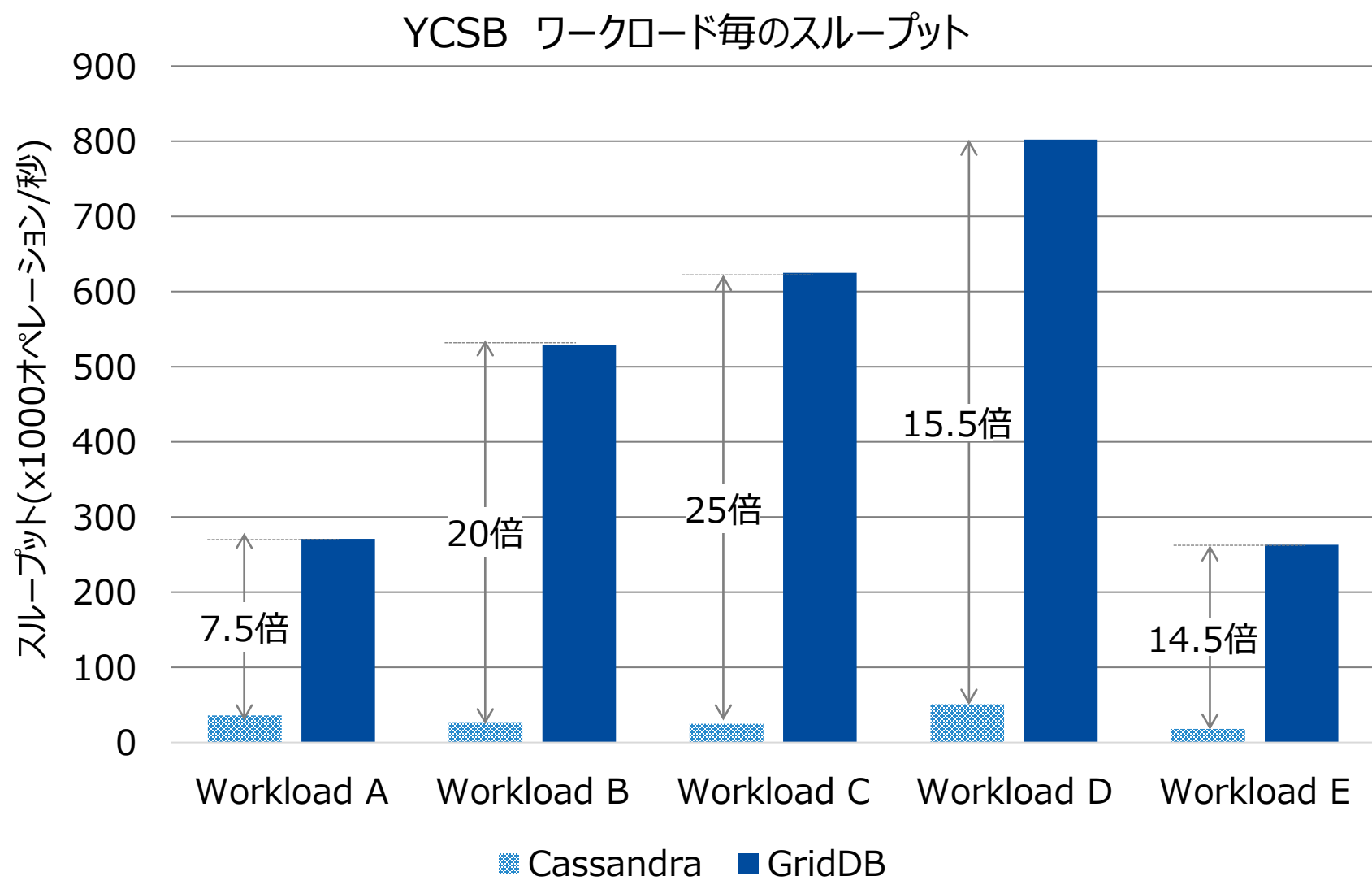
検索性能

05

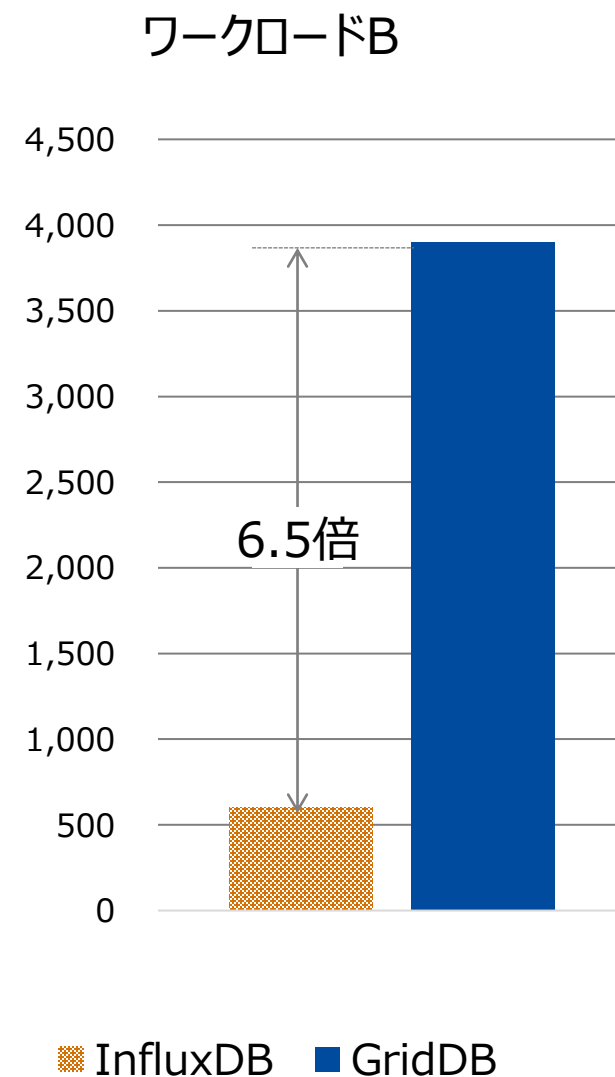
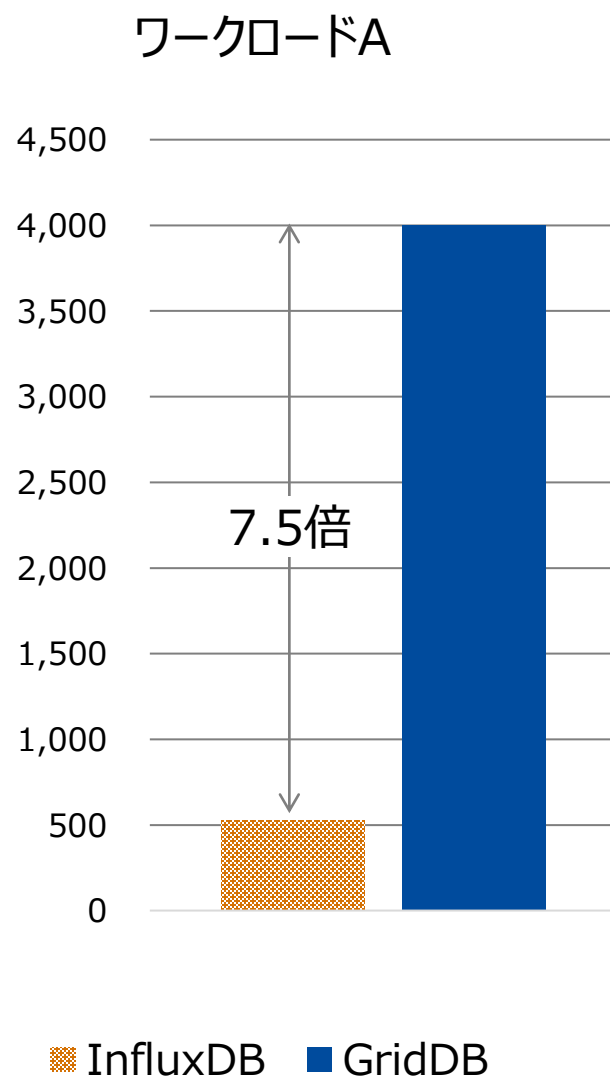
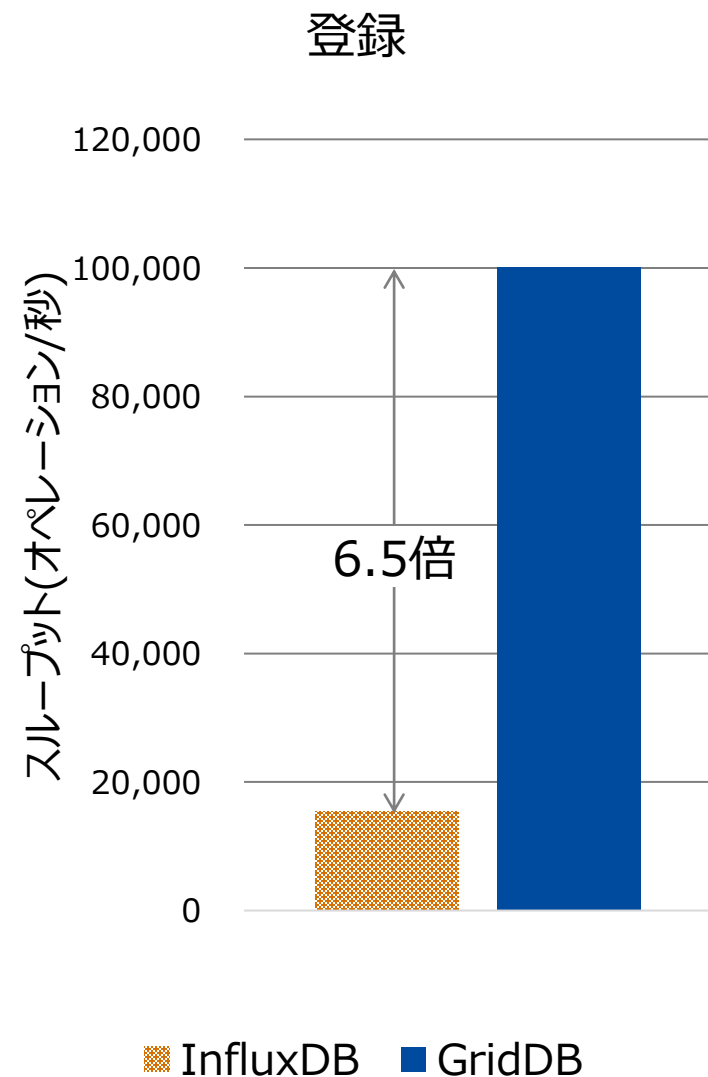
他DBとのベンチマーク



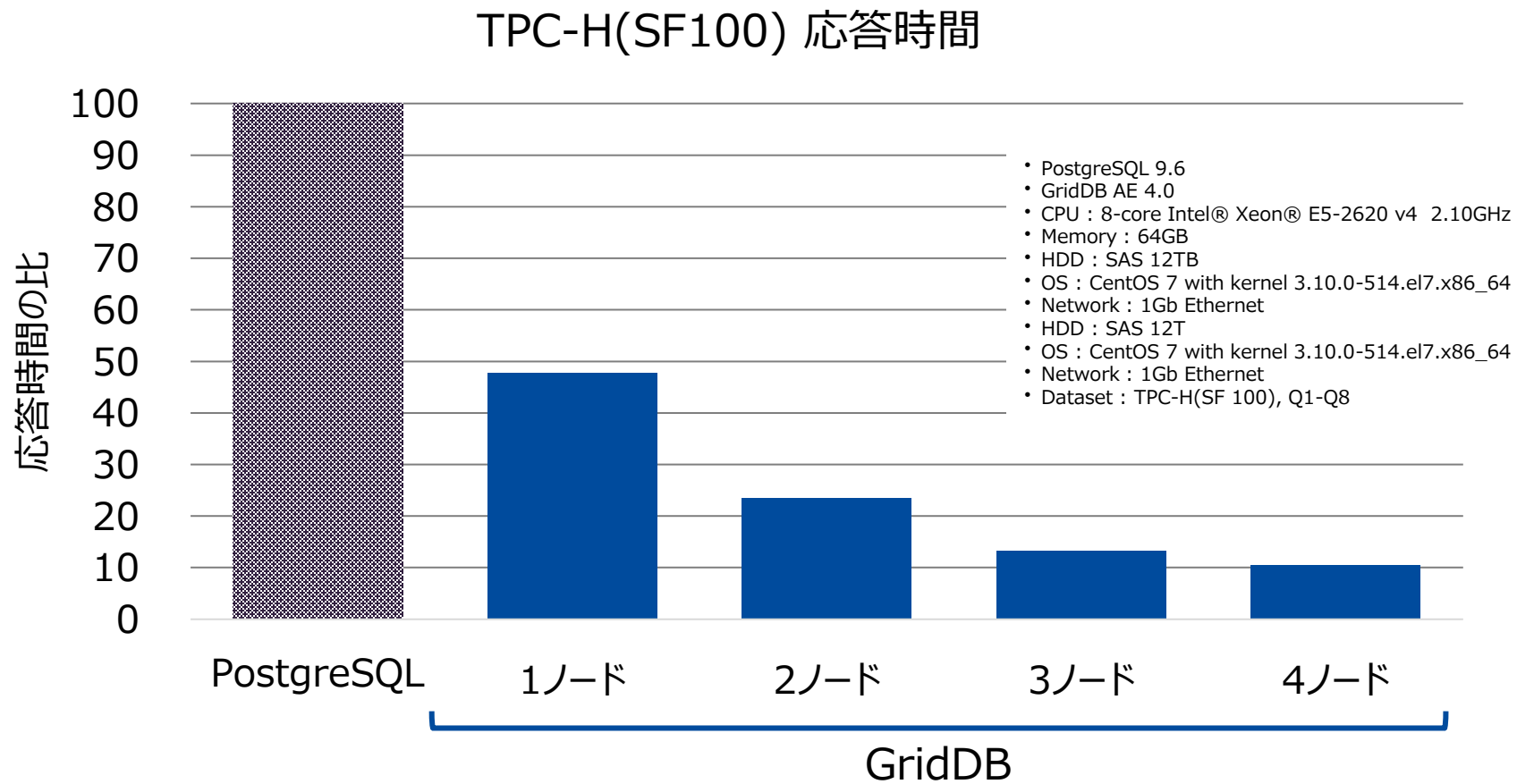
ファーストデータ処理の性能



ファーストデータ処理の性能

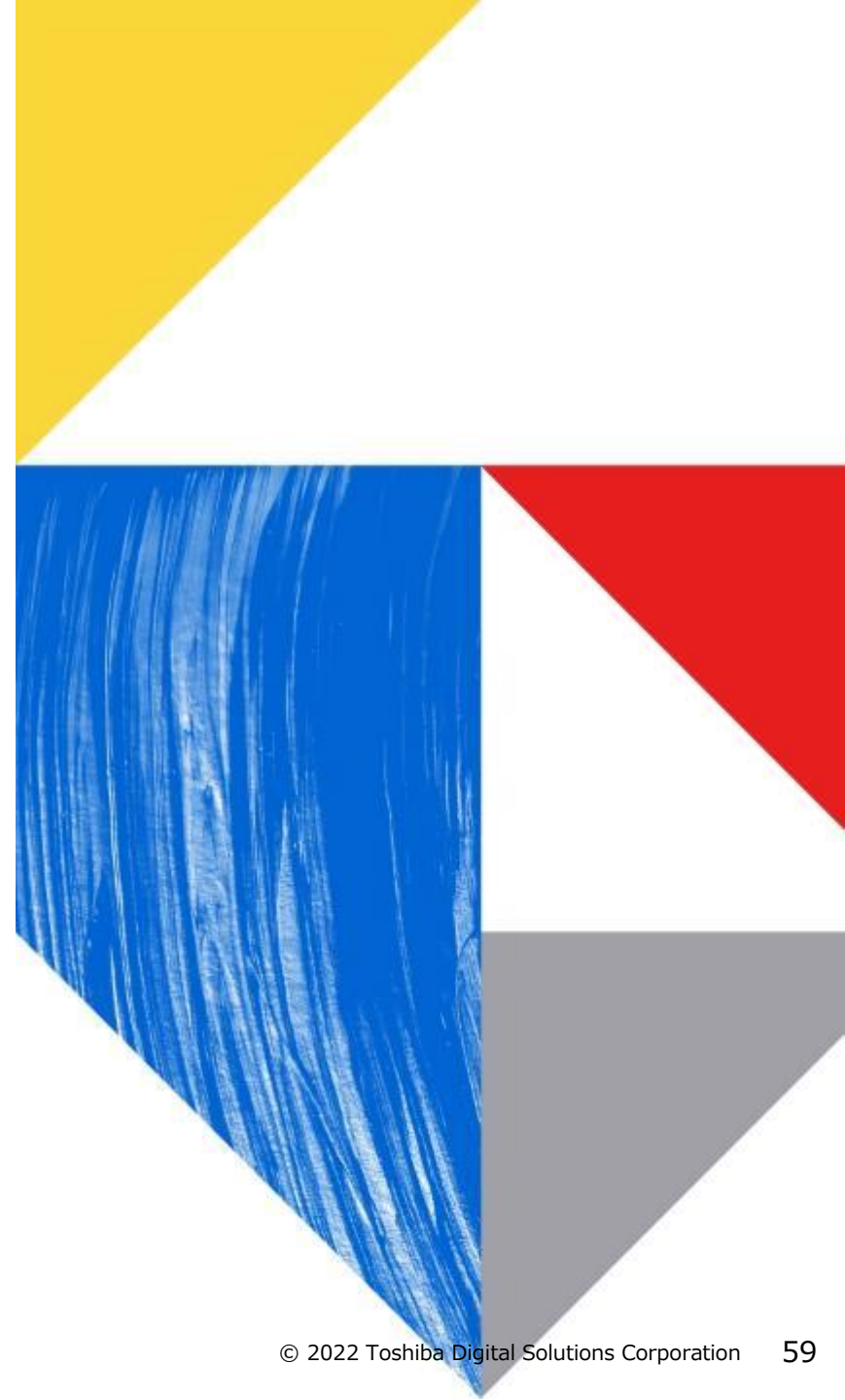


ビッグデータ処理の性能



06

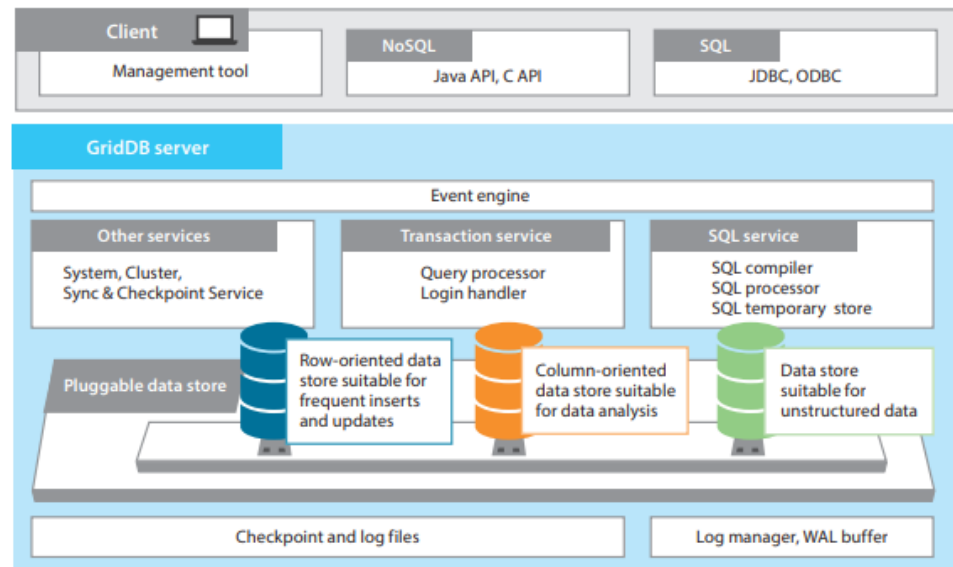
おわりに



2022/04 GridDB 5.0 EE リリース

- アーキテクチャを刷新し、複数のデータモデルを扱えるプラグابلデータストアを実現
 - 独自の高効率チェックポイントアルゴリズム技術 HCAL (Highly efficient Checkpoint Algorithm for Large-scale data) の開発
 - 17%～46%（平均26%） TPC-H性能改善

5.12 GridDB V5.0 Database for IoT and Big Data Featuring Revamped Architecture



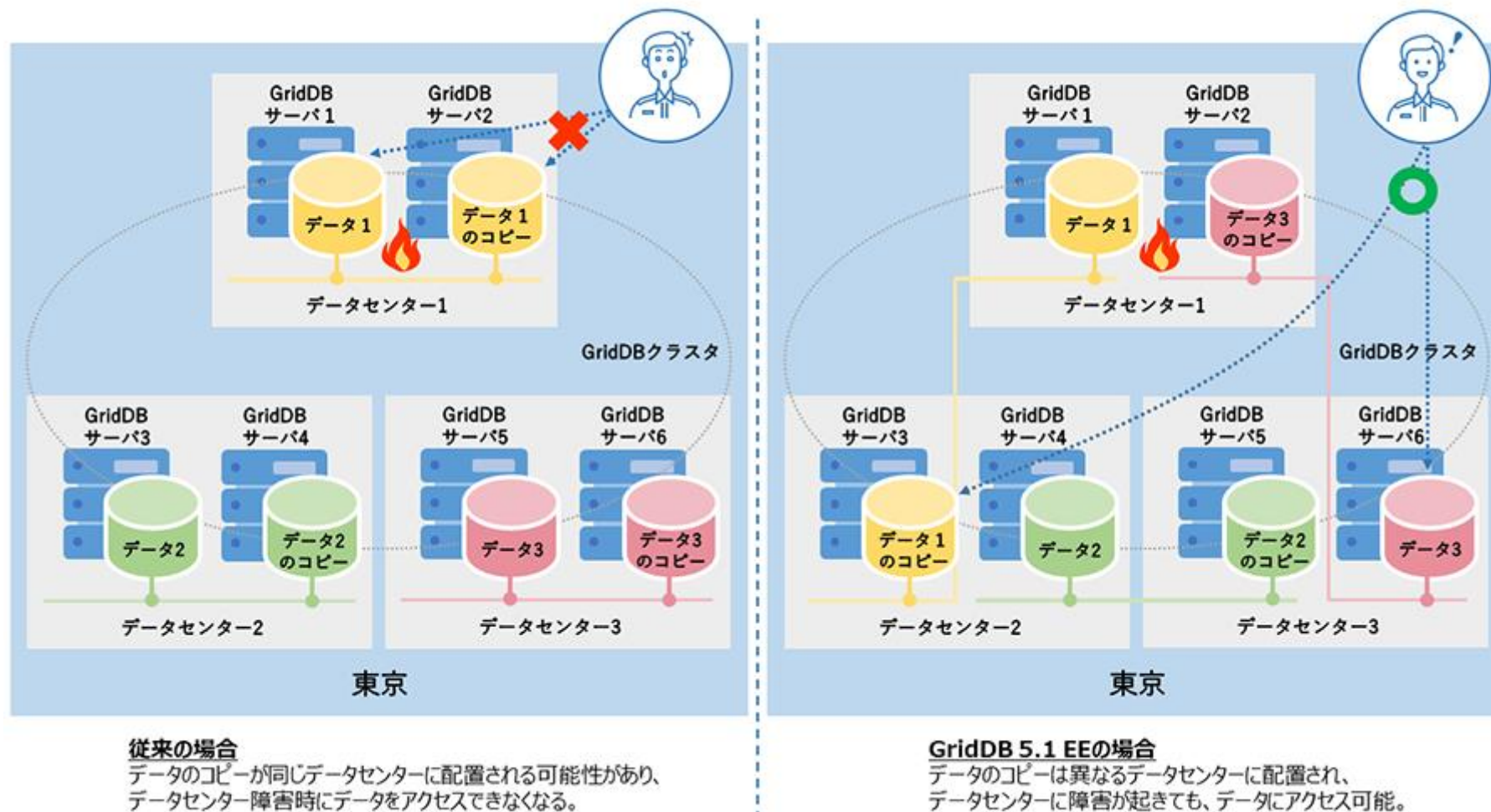
NoSQL: Not Only SQL
JDBC: Java Database Connectivity
ODBC: Open Database Connectivity

WAL: write-ahead logging
API: application programming interface



2022/08 GridDB 5.1 EE リリース

- パブリッククラウドのデータセンターに障害や災害が起きてもDB継続使用を可能



2020年6月、SQLインターフェースのソースコード無償公開

2020年6月17日 【テクノロジー】

東芝DSL、IoT向けDBのSQLインターフェースを無償公開

NEXT MOBILITY編集部

<https://github.com/griddb>

[Tweet](#)

GridDB Developers ドキュメント ブログ コミュニティ リソース 🔍 EN ダウンロード

Go *Faster*. Grow **BIGGER**.

東芝のGridDBは、IoTやビッグデータに適した高スケーラブルな
時系列NoSQLデータベースです。

[詳しくはこちら](#) [ダウンロードはこちら](#)

ご清聴ありがとうございました。