db tech showcase 2021　【A7】

アーキテクチャを一新した IoT/ビッグデータ向けデータベースGridDB

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

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
- 2020年、日本データベース学会よりGridDB事業化により
  弊社が「業績賞」受賞

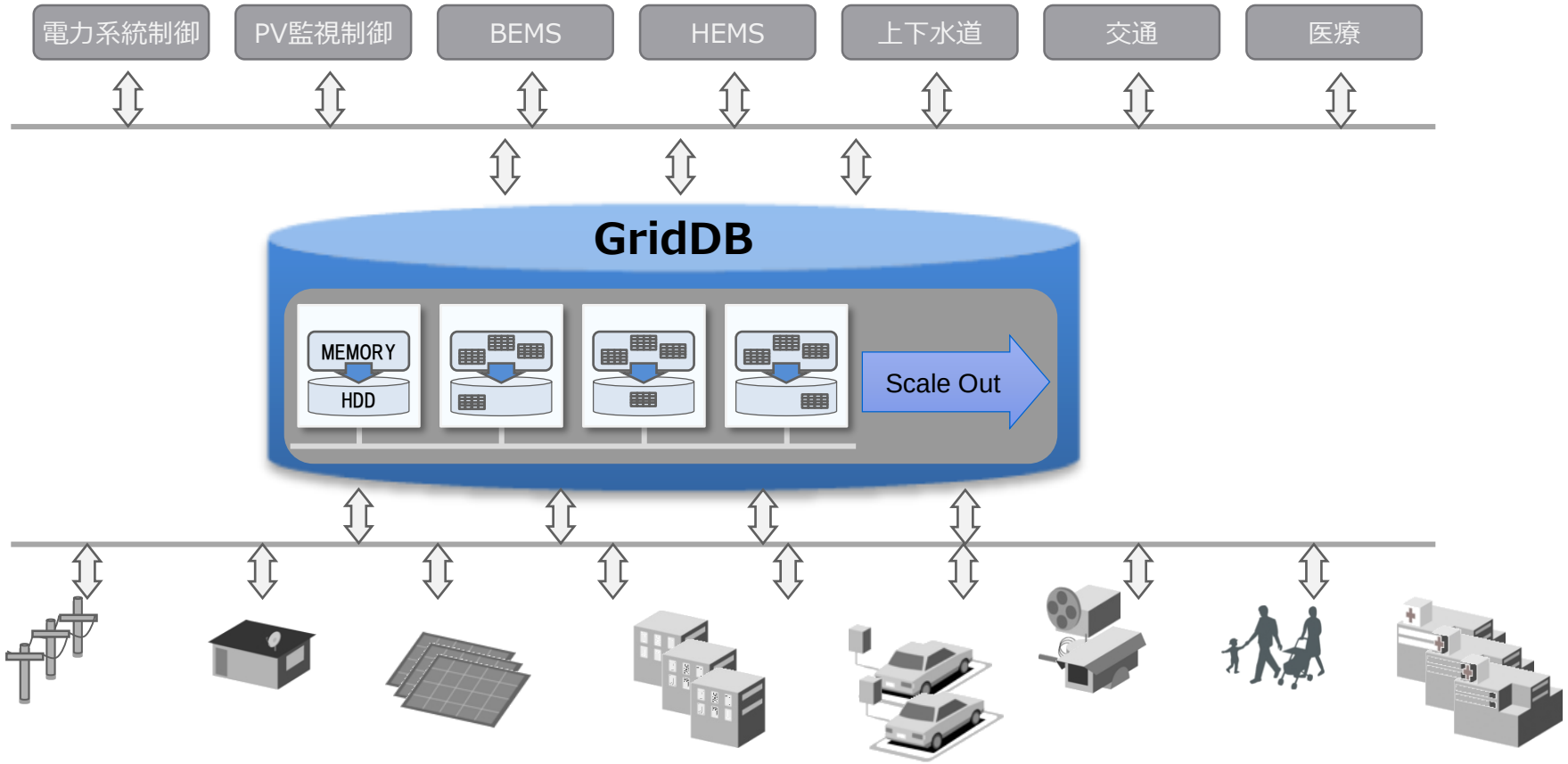
アジェンダ

- スケールアウト型データベース GridDB
- SQLを補完するNoSQL
- NoSQLを超えたGridDB
- アーキテクチャを一新したV5

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

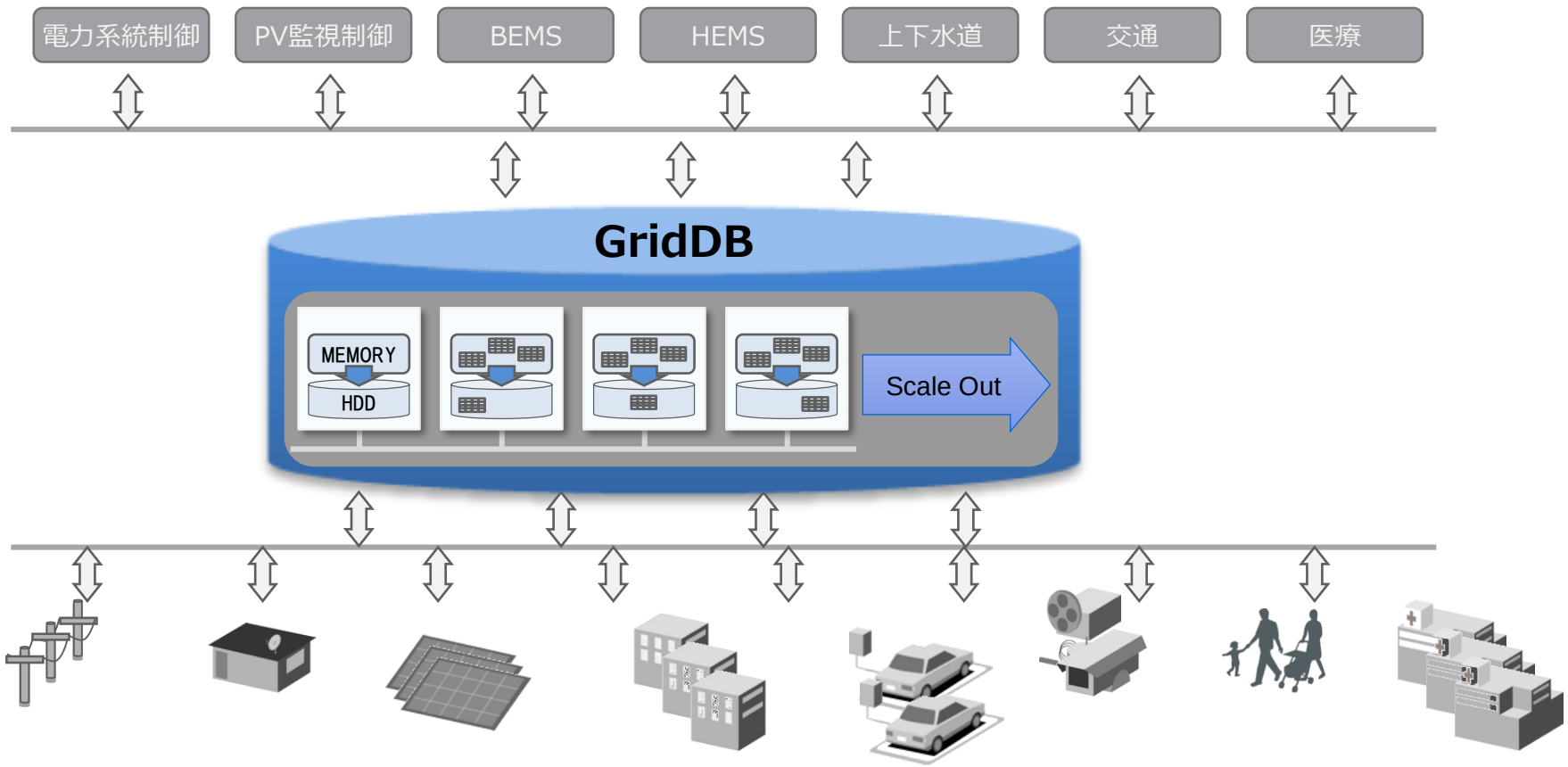
大規模IoT等、高い性能・可用性が求められるシステムに適用（現在バージョン４.6）

2013年、GridDB（NoSQLのみ)上市
2015年、GridDB/NewSQL(SQLインターフェイス機能)
2017年、SQLの並列分散処理化
2019年、1ノードあたりペタバイト級のデータ管理に対応、など
単体製品、DBaaSサービス、Meister DigitalTwinの基盤などを通してビジネス展開



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

- 大規模IoT等、高い性能・可用性が求められるシステムに適用 （現在バージョン４.6）
- 実適用データ規模　数TB～数PB



エディション体系

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

【A23】

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

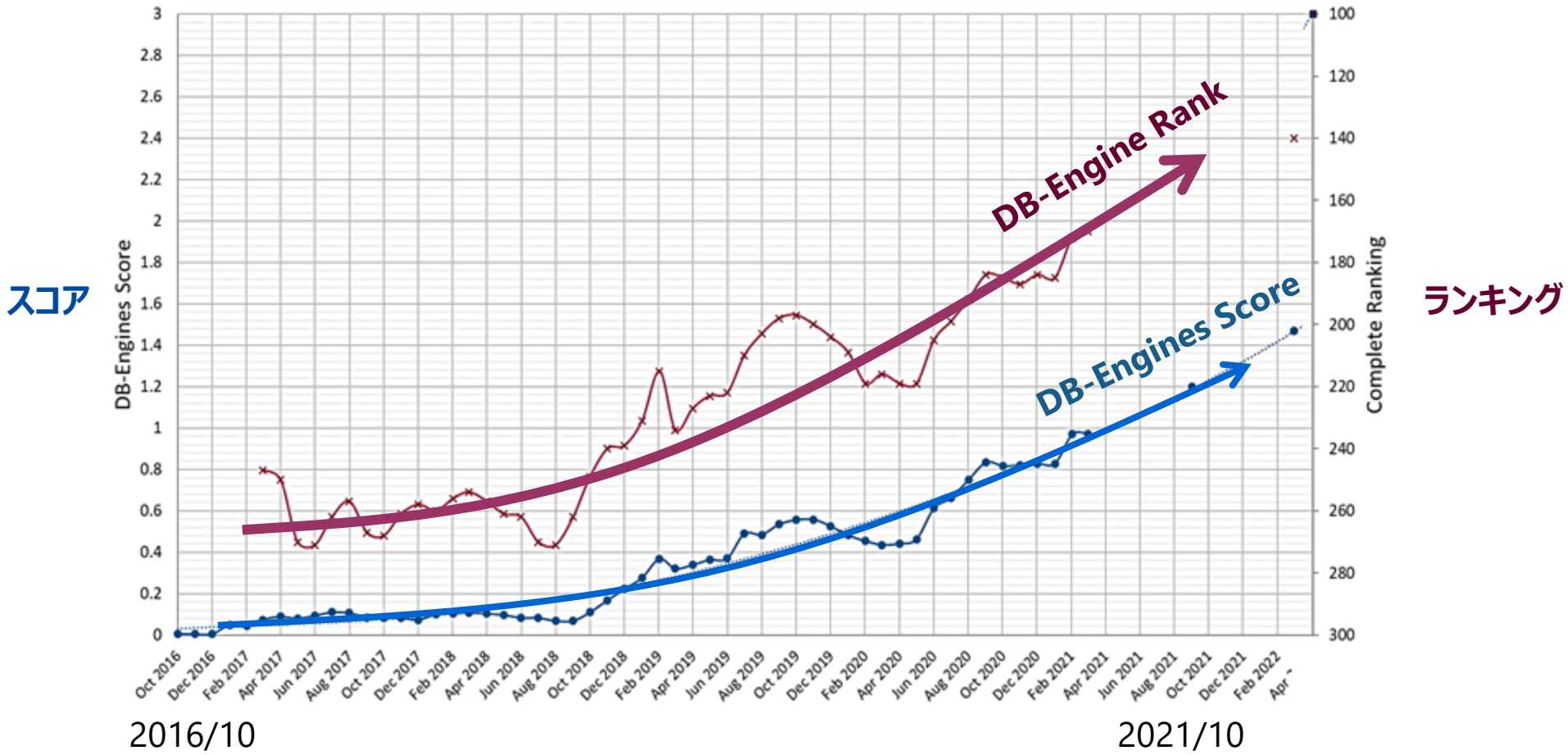
各エディションの違い

- 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		✓	✓
	運用コマンド		✓	✓
セキュリティ	信暗号化 (TLS/SSL)		✓	✓
	認証機能 (LDAP)		✓	
オンプレミス環境	オンプレミス環境	✓	✓	
クラウドサービス	クラウドサービス			✓

日本初OSS　GridDB Community Edition

2016年公開。近年、DB-Engineランキングで急上昇

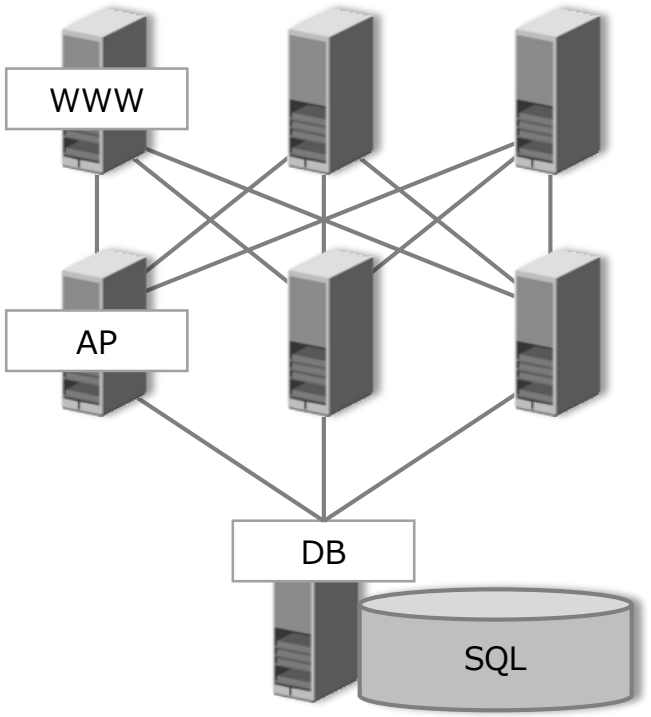


01

SQLを補完するNoSQL

NoSQL（Not Only SQL）の存在理由

スケールアップは物理的な限界やコスト高となる

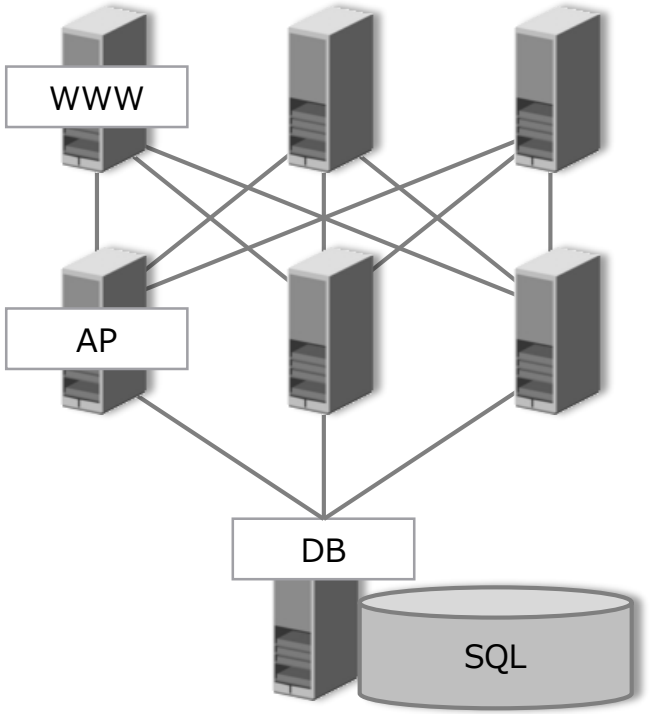


厳格な一貫性

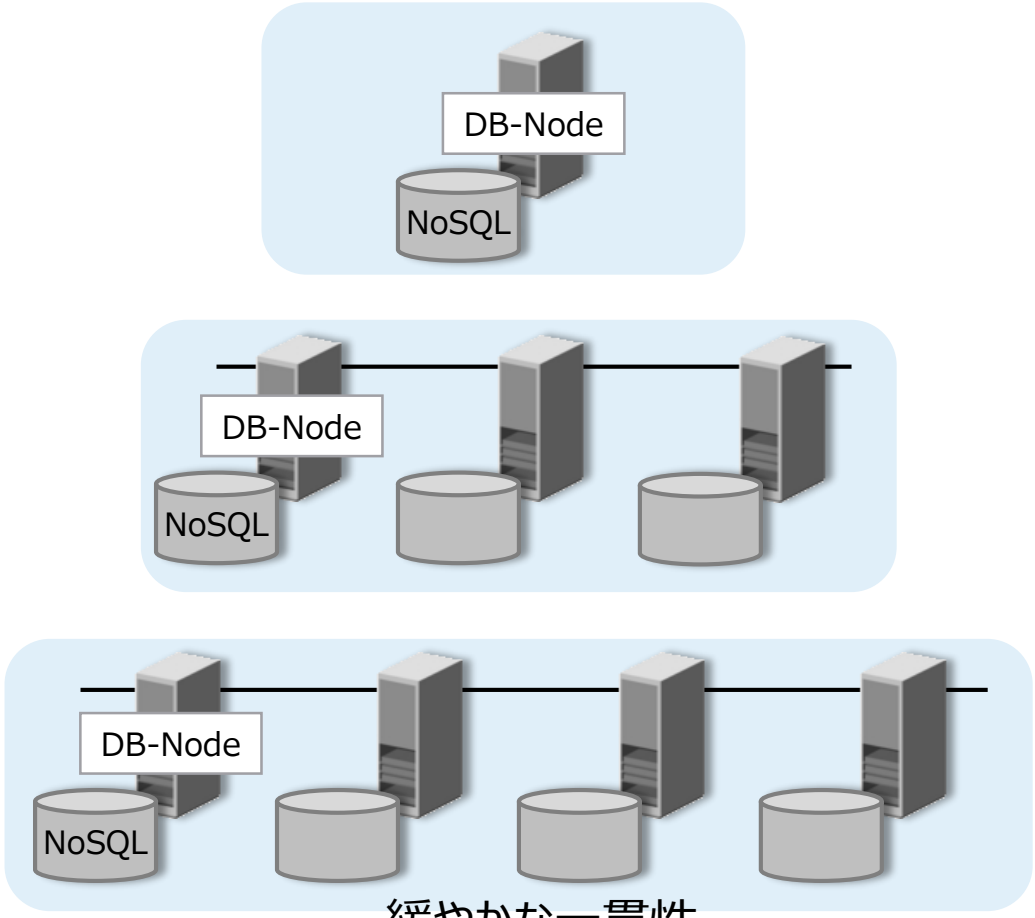
Webノード、APノードと異なり、
DBのスケールアウトは非常に難しい。

NoSQL（Not Only SQL）の存在理由

- ノード台数増で対応できるスケールアウトがお得
- その代わり、データー貫性の要件を緩和する必要



厳格な一貫性



緩やかな一貫性

NoSQLとSQLの関係

SQLとNoSQLは補完関係にある

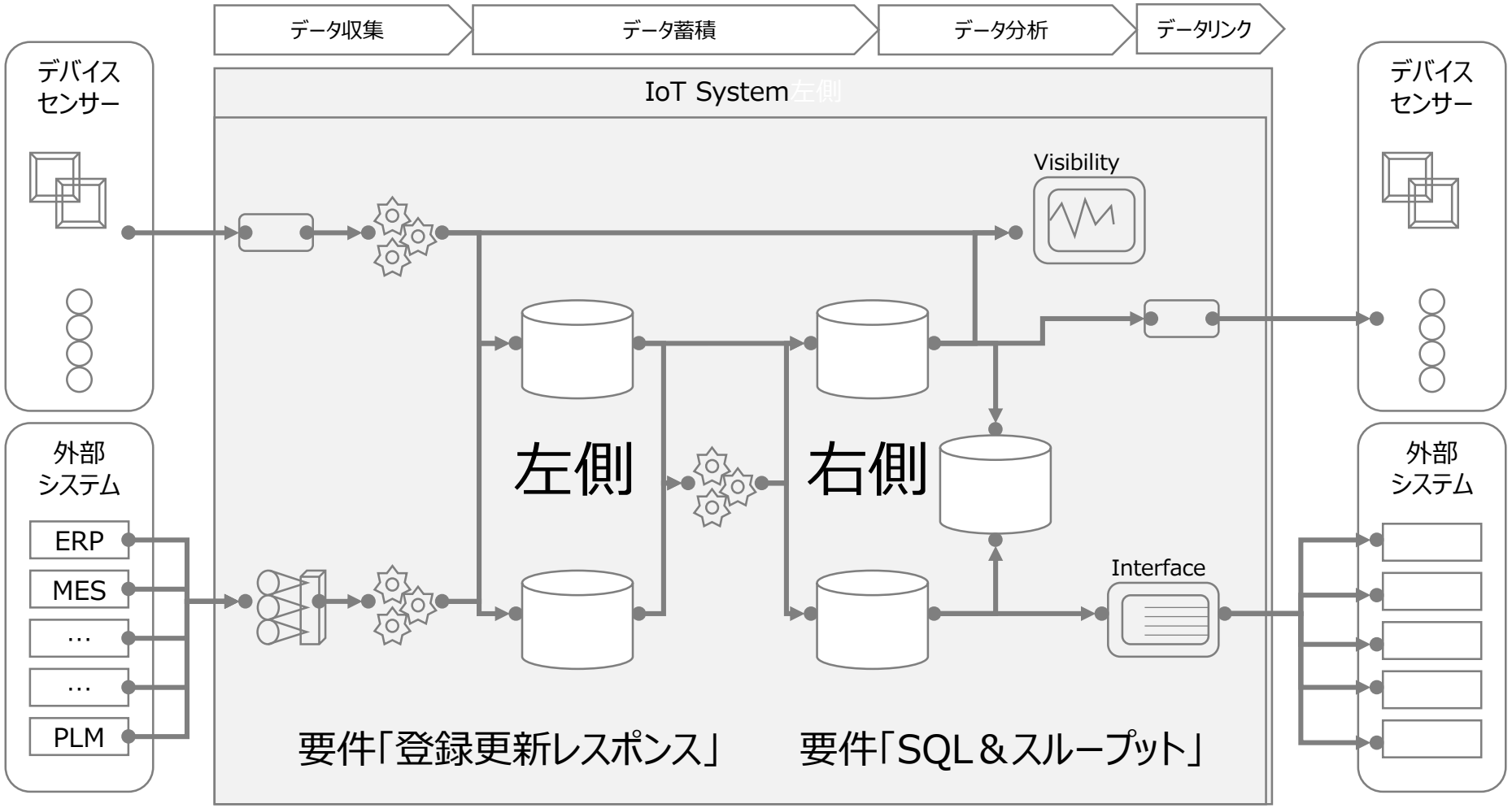
	某NoSQL	某NoSQL	SQL(RDB)
問い合わせ言語	独自I/F	独自I/F	SQL
スキーマ言語	無い	無い	SQL(DDL)
データ一貫性	参照一貫性 テーブル内一貫性	緩やかな一貫性 or 参照一貫性	厳格な一貫性
1台あたりの性能	普通	一貫性を上げると低速	実績あるRDBは ある程度
高可用性	ノード分散による 冗長性有り	ノード分散による 冗長性有り	別の仕掛けが必要
拡張性	スケールアウトする手段が不明	スケールアウト困難 特にオンラインスケールアウト	スケールアップ

NoSQLをベースとしたスケールアウト型DB

SQLとNoSQLは補完関係にある

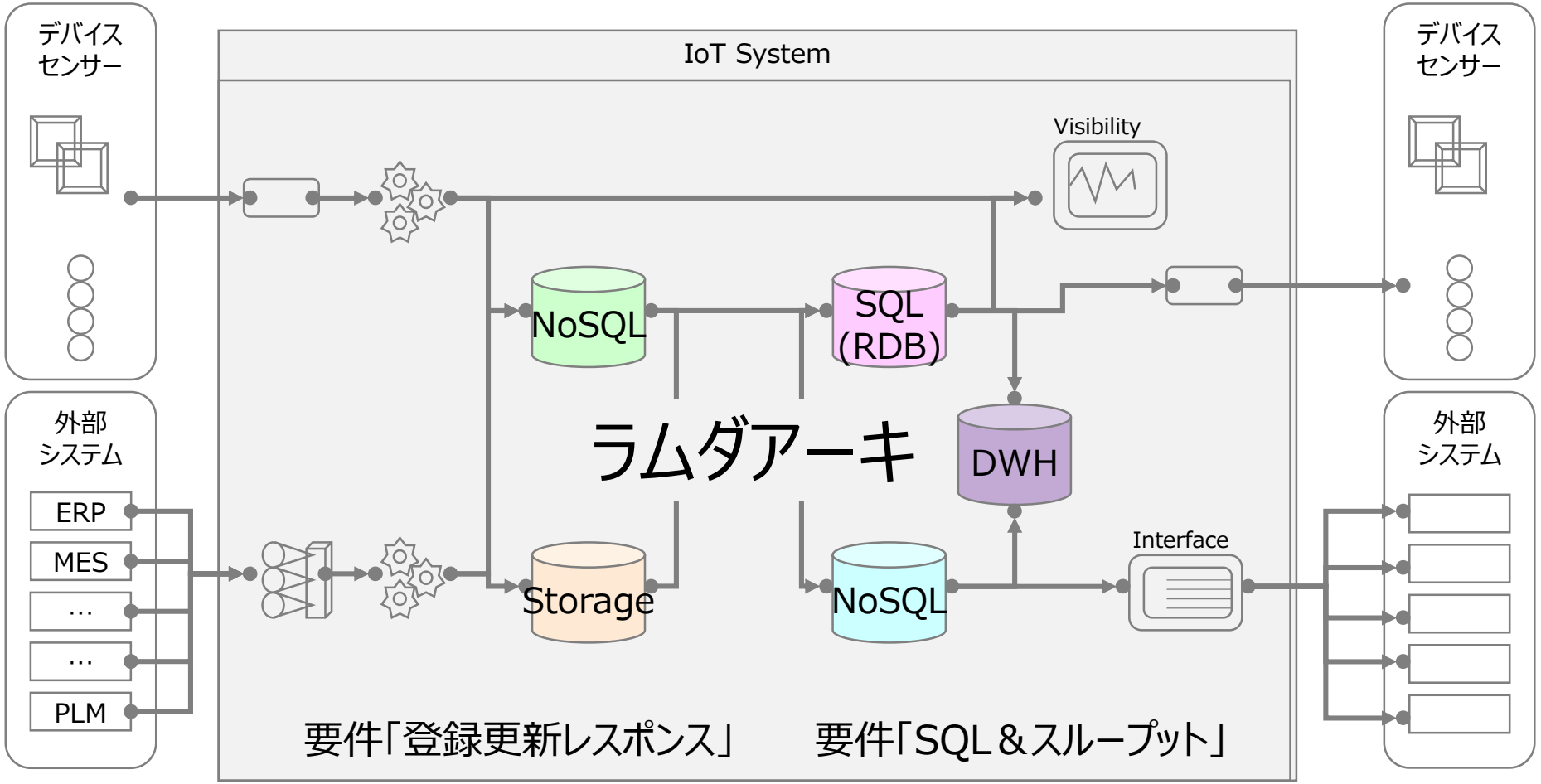
	某NoSQL	某NoSQL	SQL(RDB)
問い合わせ言語	独自I/F	独自I/F	SQL
スキーマ言語	無い	無い	SQL(DDL)
データ一貫性	参照一貫性 テーブル内一貫性	緩やかな一貫性 or 参照一貫性	厳格な一貫性
1台あたりの性能	普通	一貫性を上げると低速	実績あるRDBは ある程度
高可用性	ノード分散による 冗長性有り	ノード分散による 冗長性有り	別の仕掛けが必要
拡張性	スケールアウトする手段が不明	スケールアウト困難 特にオンラインスケールアウト	スケールアップ

IoTシステムの内部構成



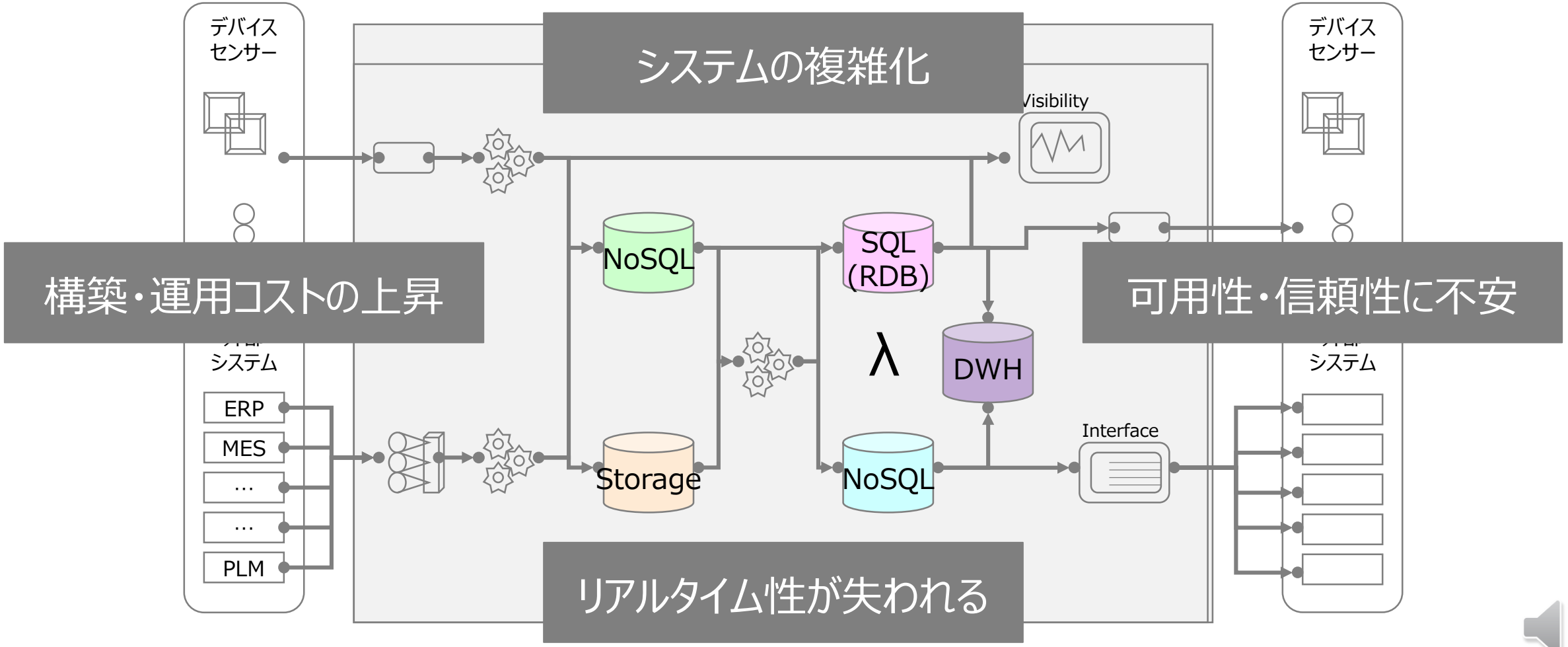
IoTシステムにおけるDB

異なる特性を持つ複数のDBの使い分け

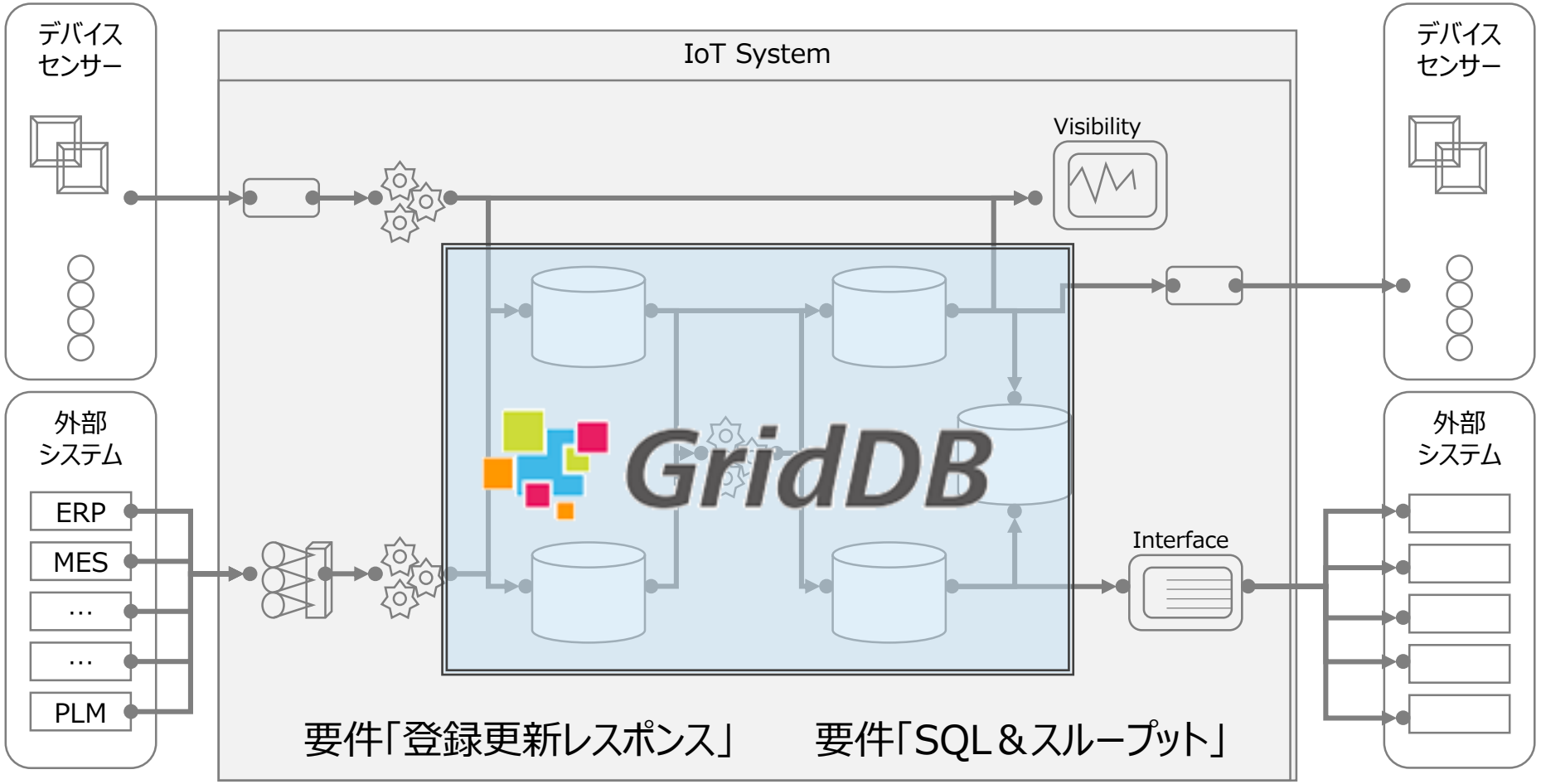


IoTシステムにおけるDB

DBの使い分けがなぜ問題なのか？



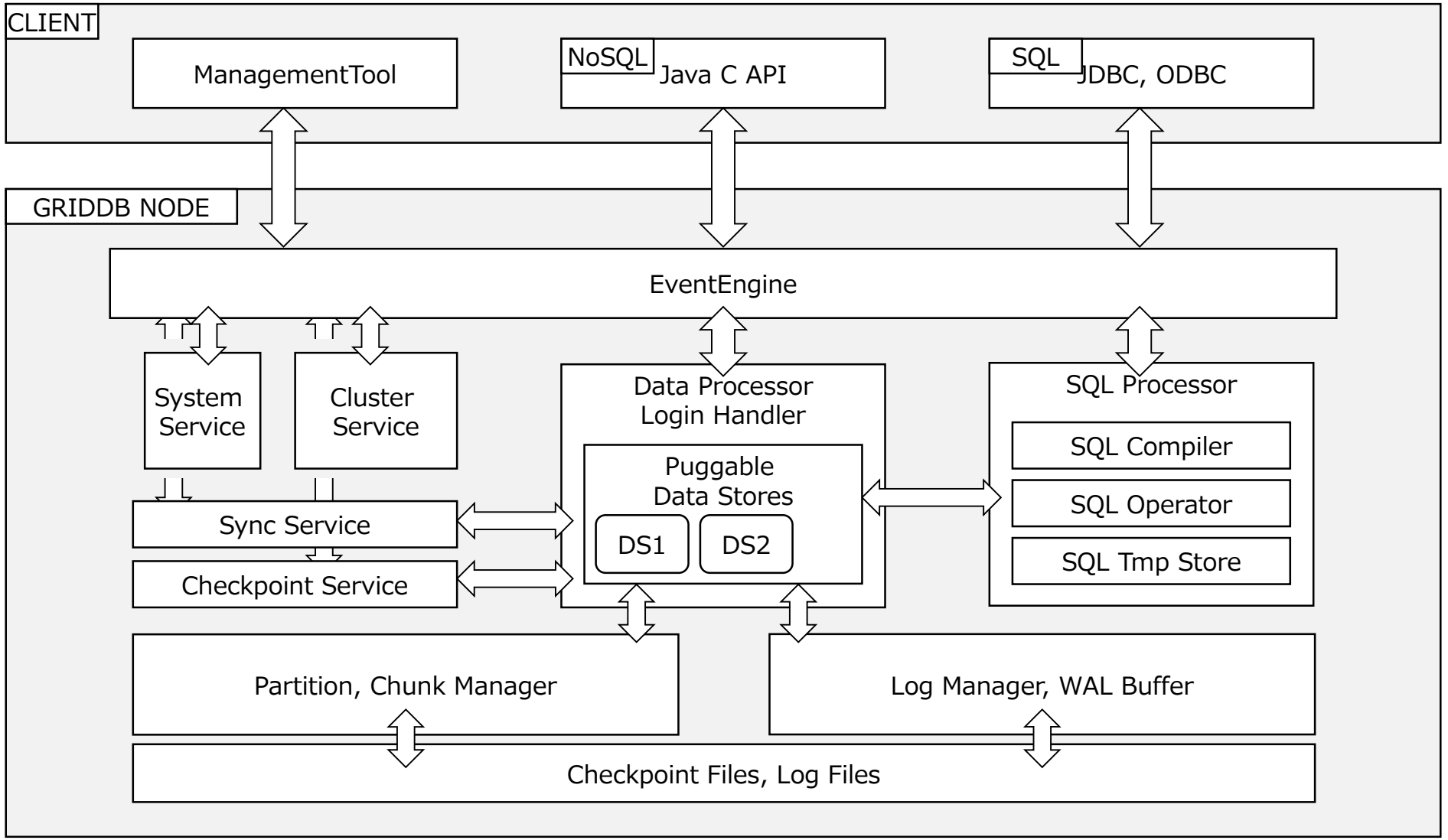
GridDBならば、複数のDBを使わなくても良い。



02

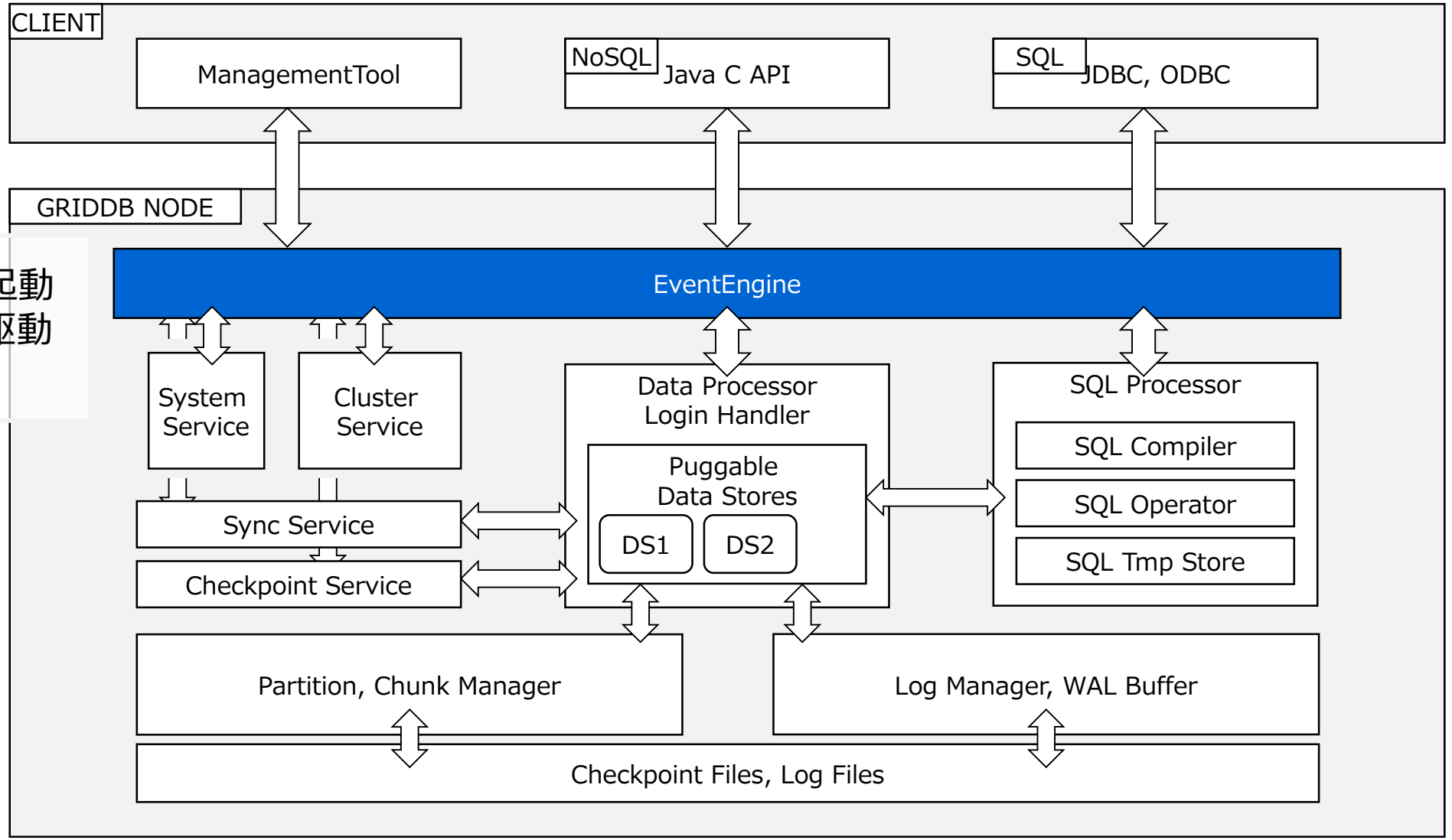
NoSQLを超えたGridDB

V5アーキテクチャ

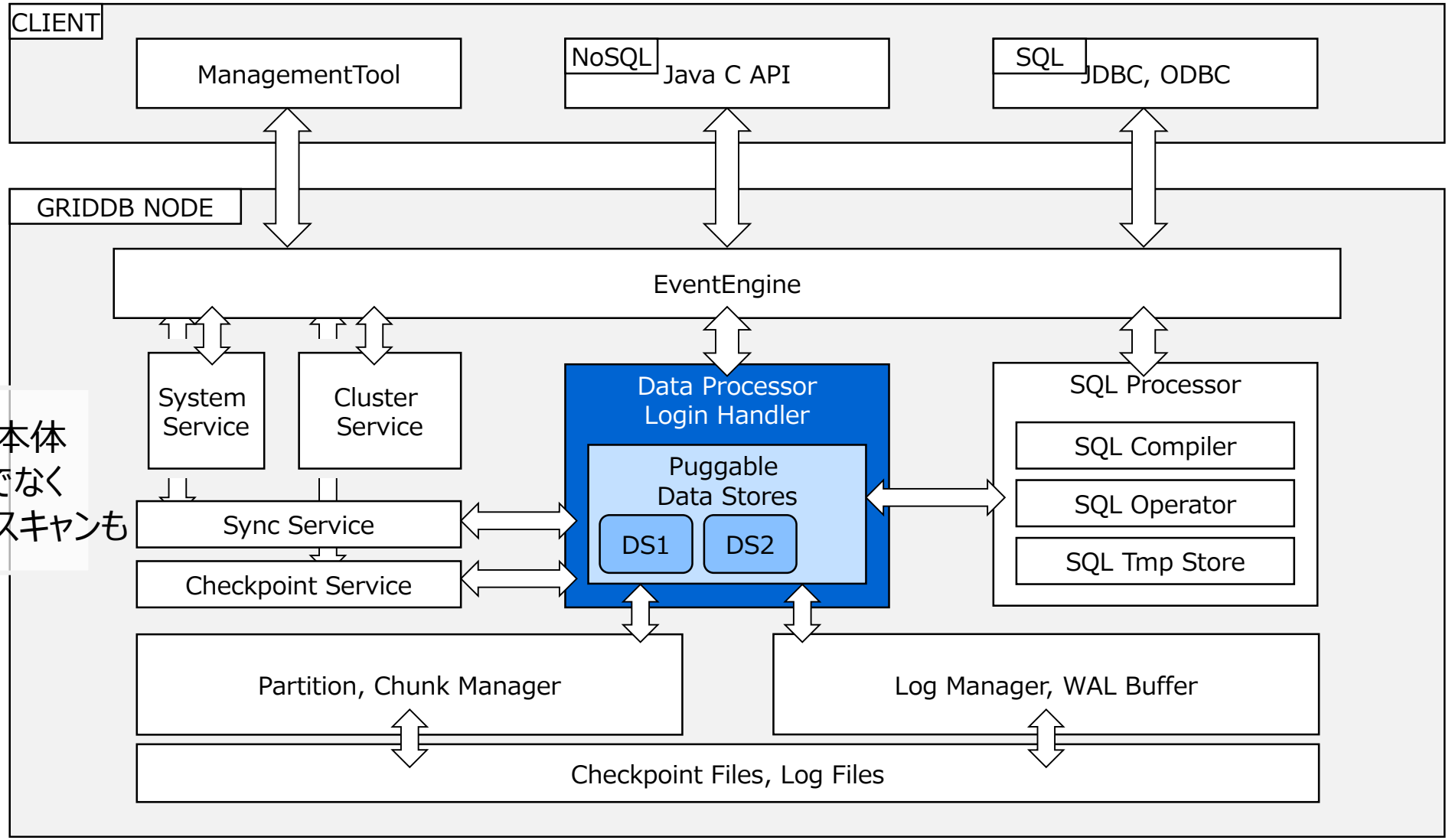


V5アーキテクチャ

各種処理の起動
全てイベント駆動
一種のOS

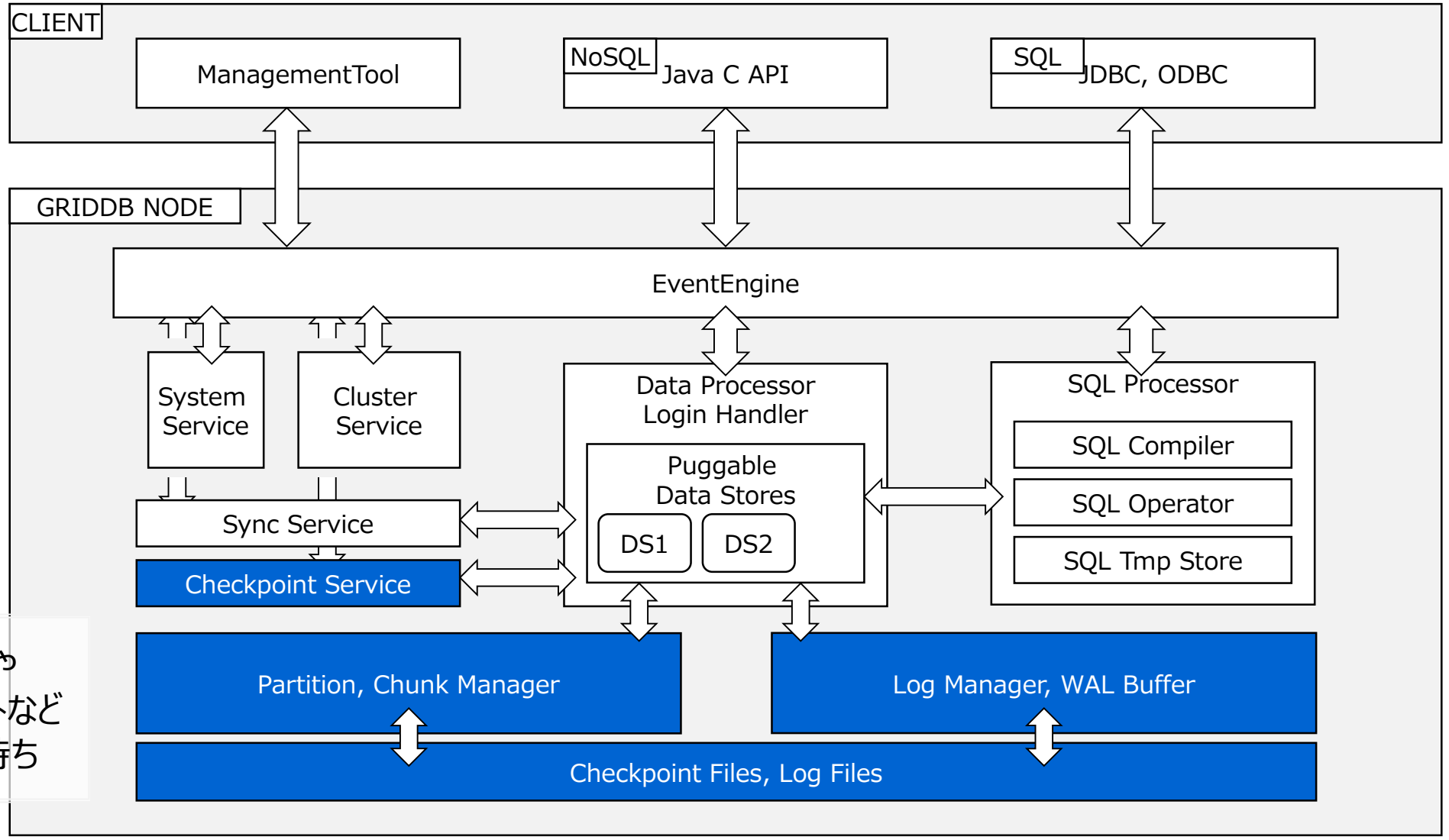


V5アーキテクチャ



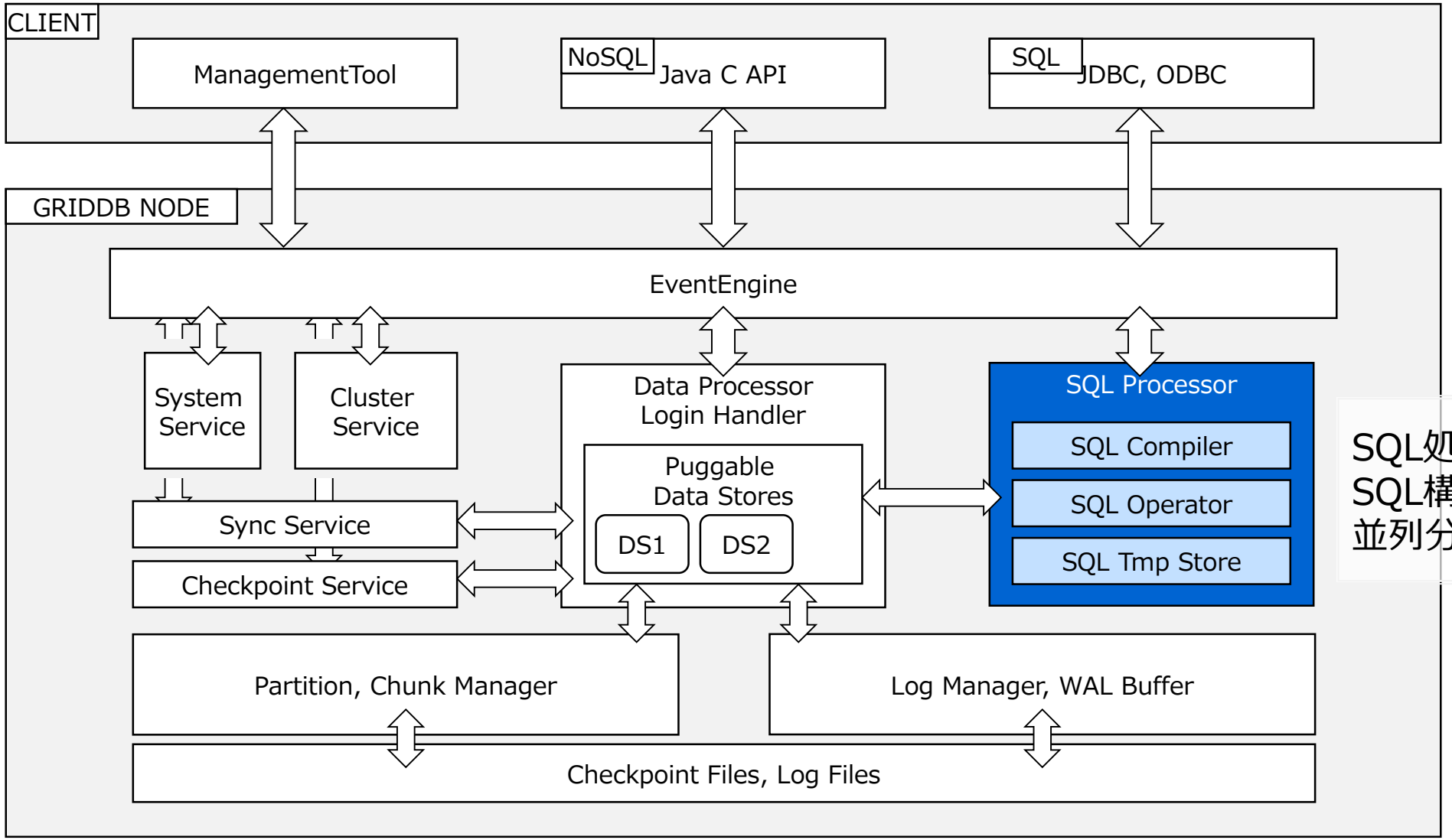
データ処理の本体
NoSQLだけでなく
SQLのデータスキャンも

V5アーキテクチャ



WALリカバリや
スナップショットなど
縁の下の力持ち

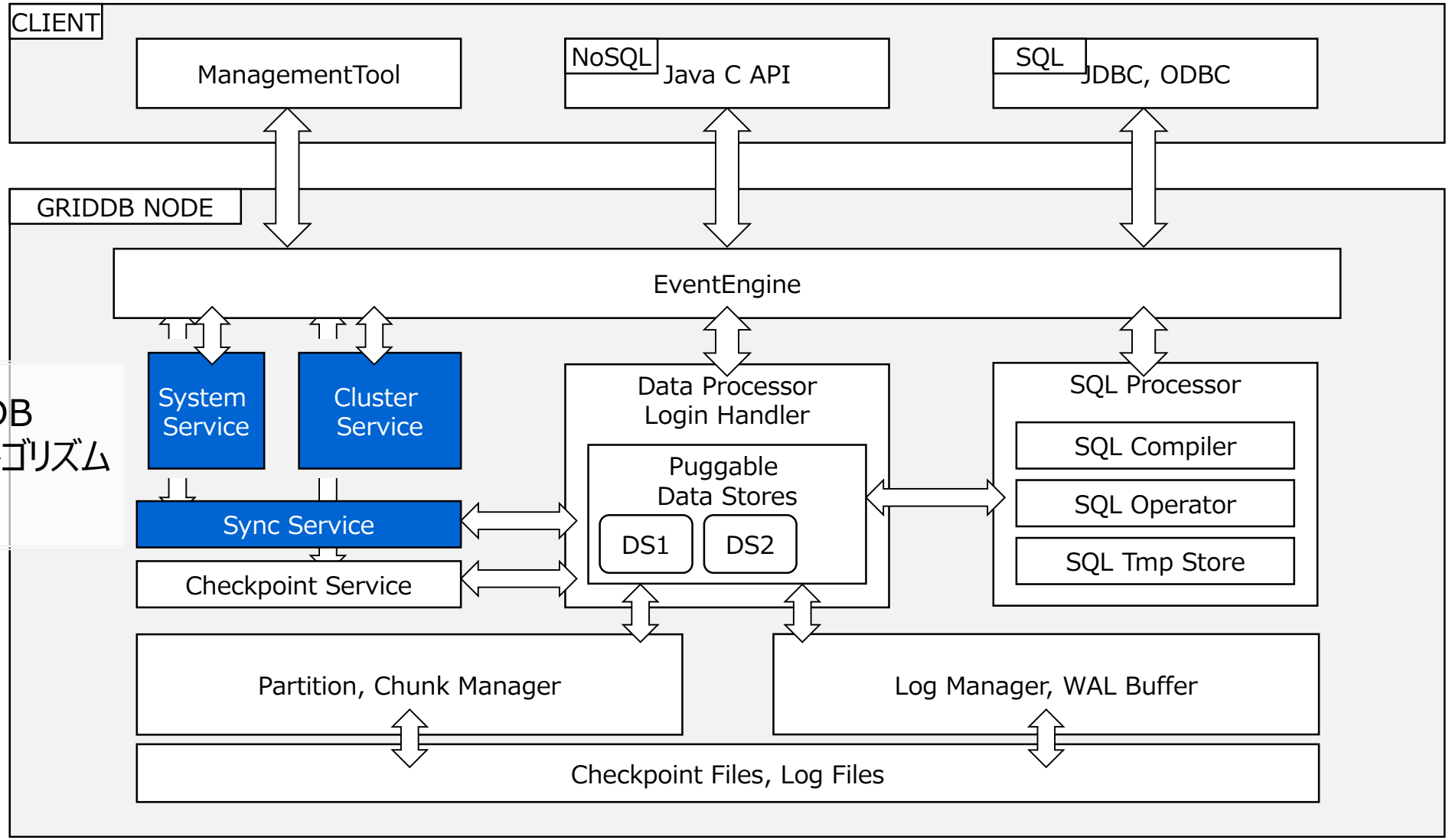
V5アーキテクチャ



SQL処理
SQL構文解析から
並列分散処理まで

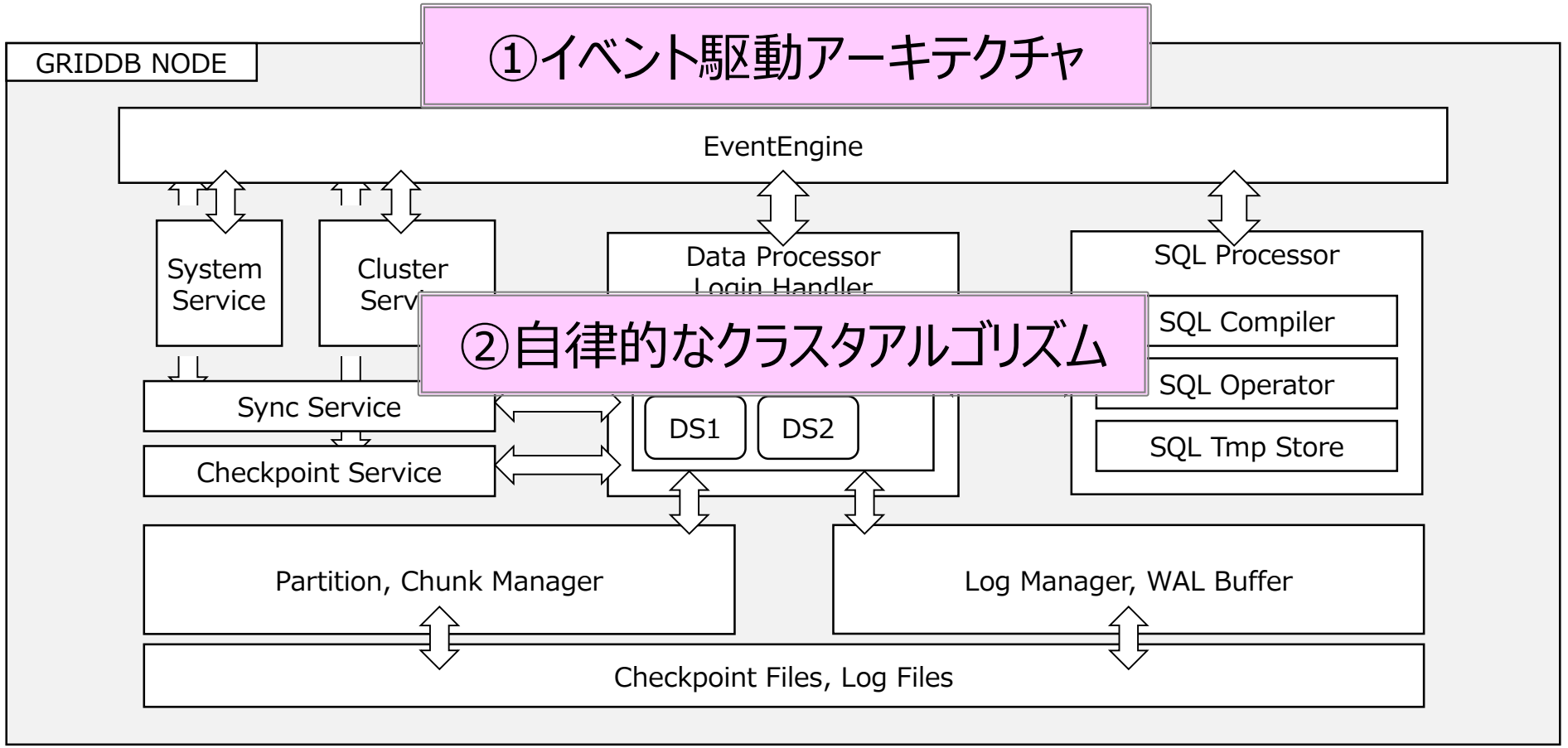
V5アーキテクチャ

自律的なDB
クラスタアルゴリズム
が本尊



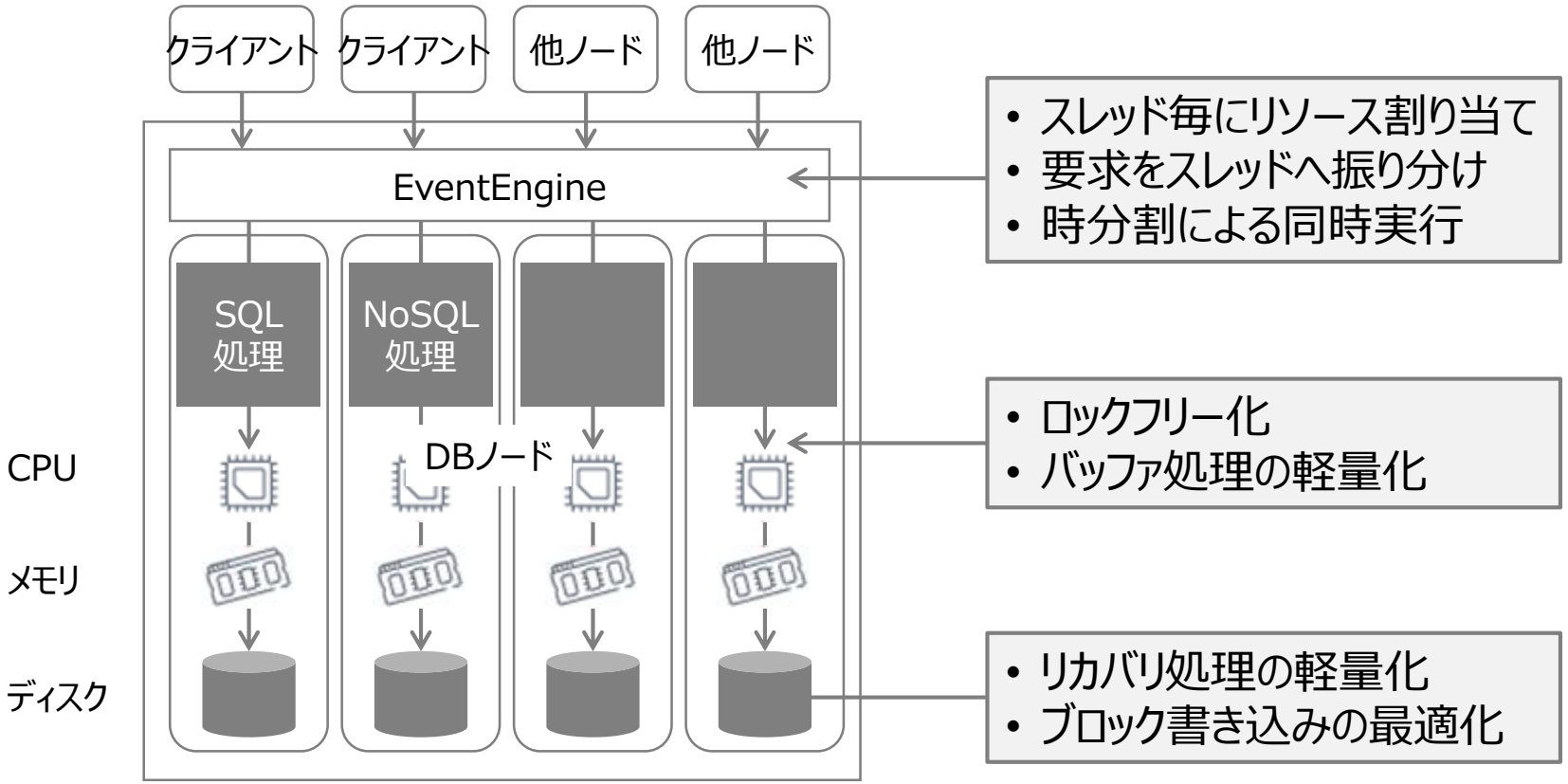
コア技術

これまで培ってきた2つのコア技術



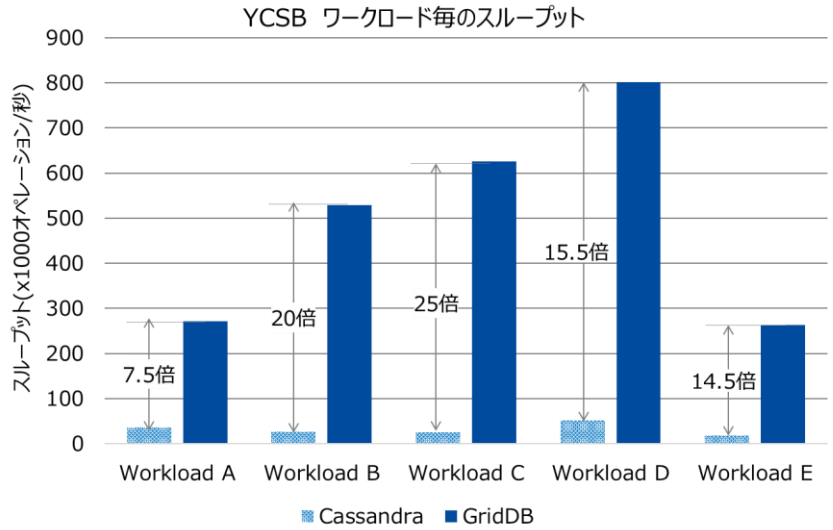
①イベント駆動アーキテクチャ

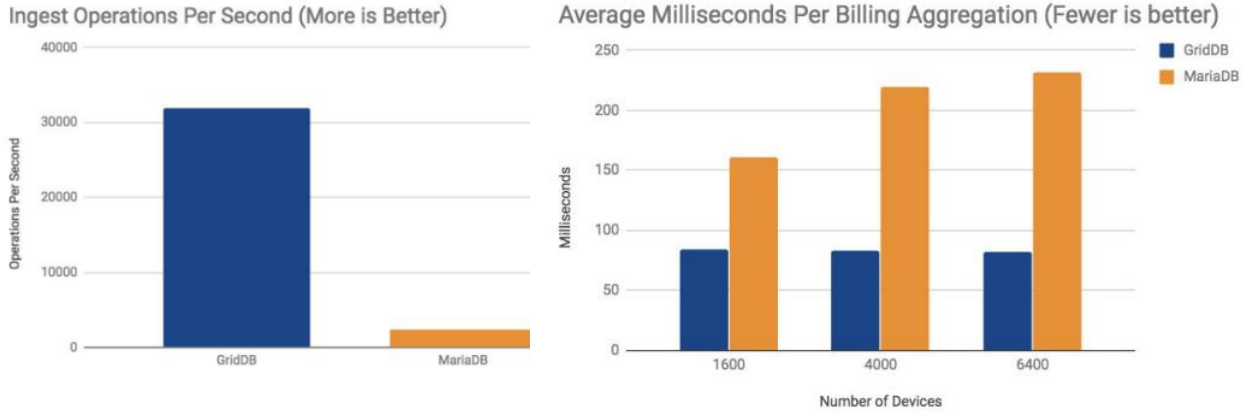
- CPUのマルチコア，メニーコア化を前提
- 非同期的なデータ処理を絶え間なく実行するイベント駆動方式



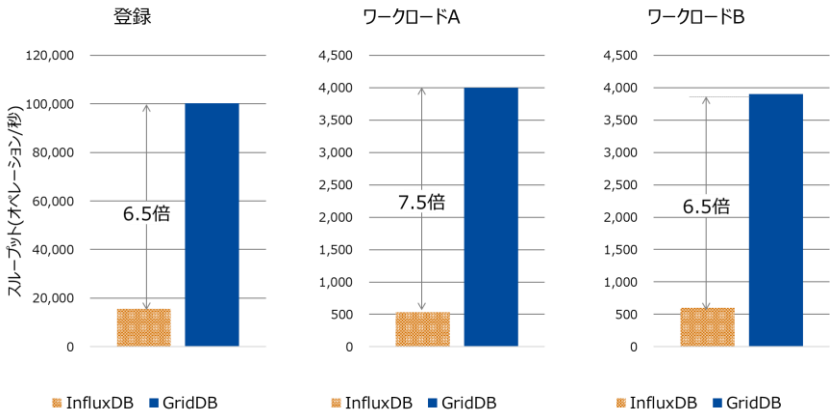
ベンチマーク

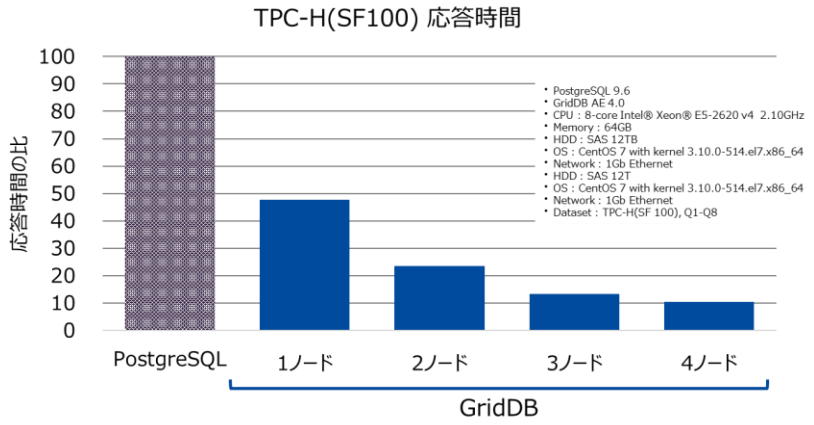
ノードあたりの性能、クラスタの性能。各種ベンチで他DBを圧倒





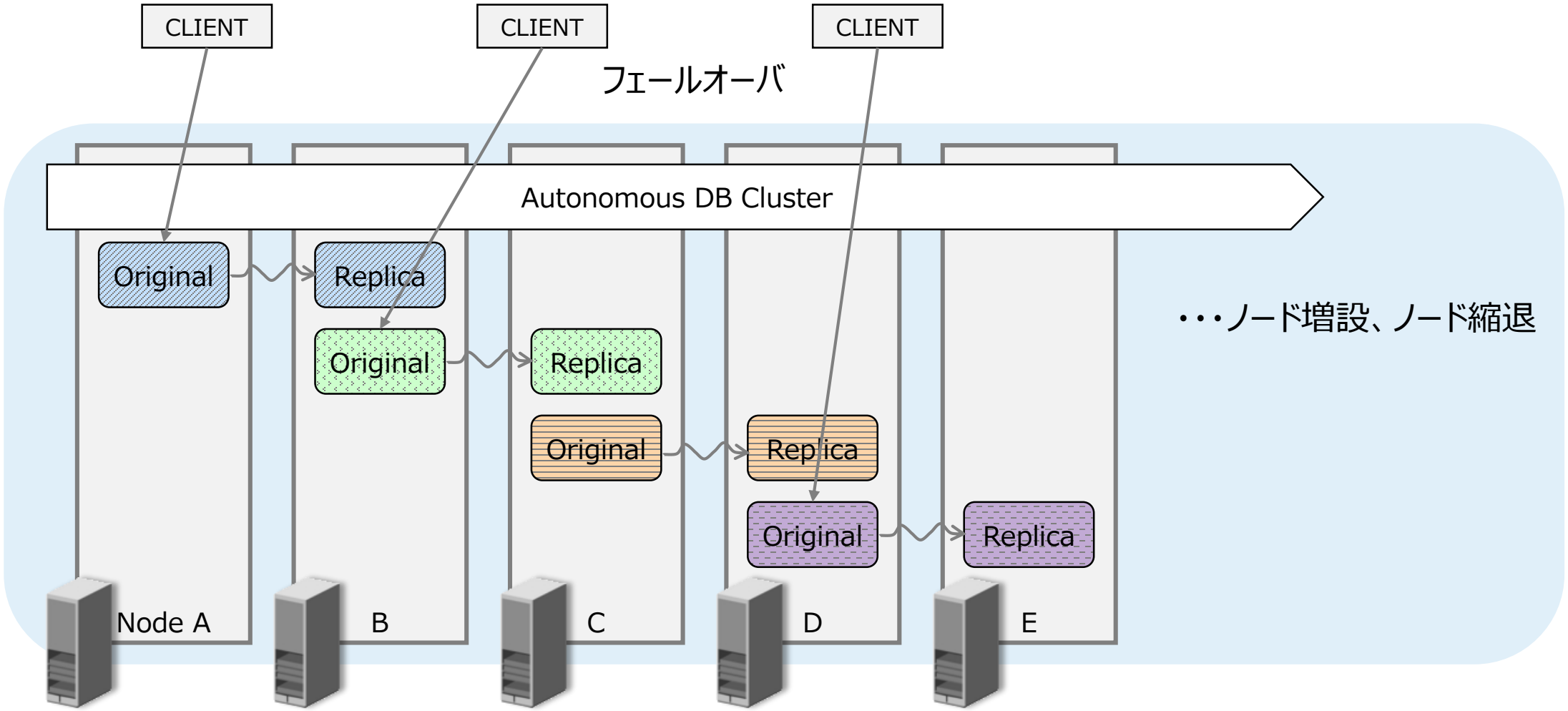
https://griddb.net/en/docs/Benchmarking_Application_GridDB_MariaDB.pdf





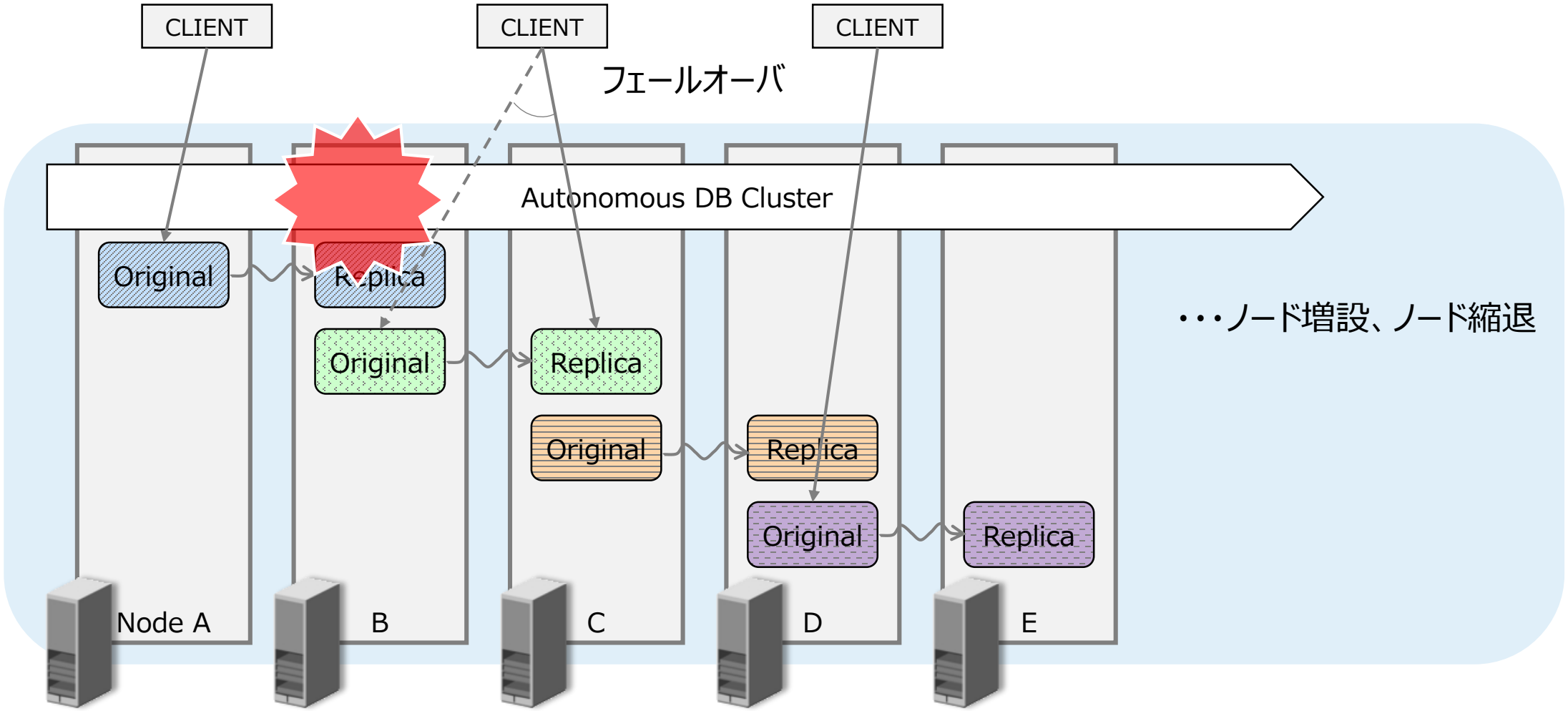
②自律的なクラスタアルゴリズム

シャーディングとレプリケーションを制御する自律的なクラスタ

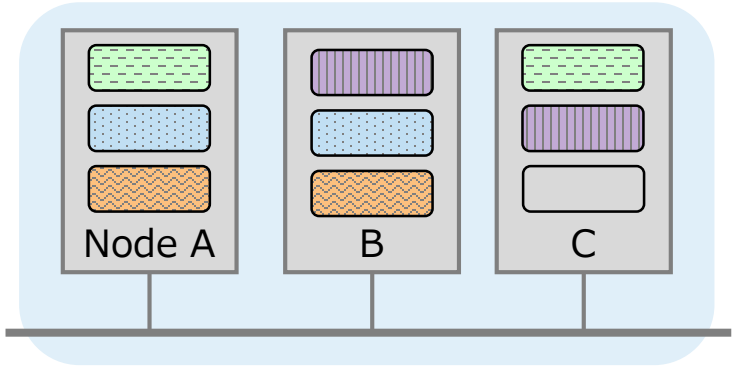


②自律的なクラスタアルゴリズム

シャーディングとレプリケーションを制御する自律的なクラスタ

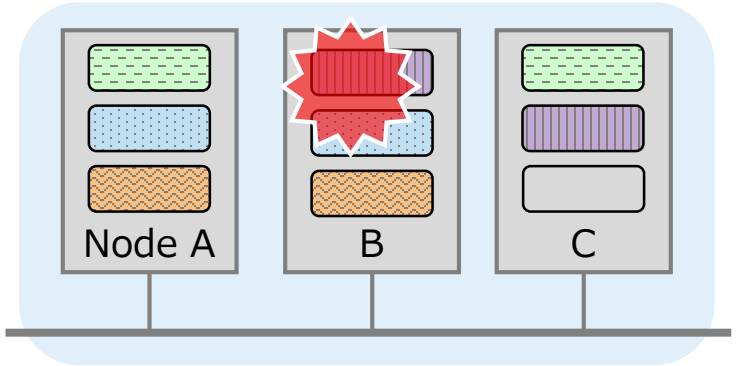


②自律的なクラスタアルゴリズム



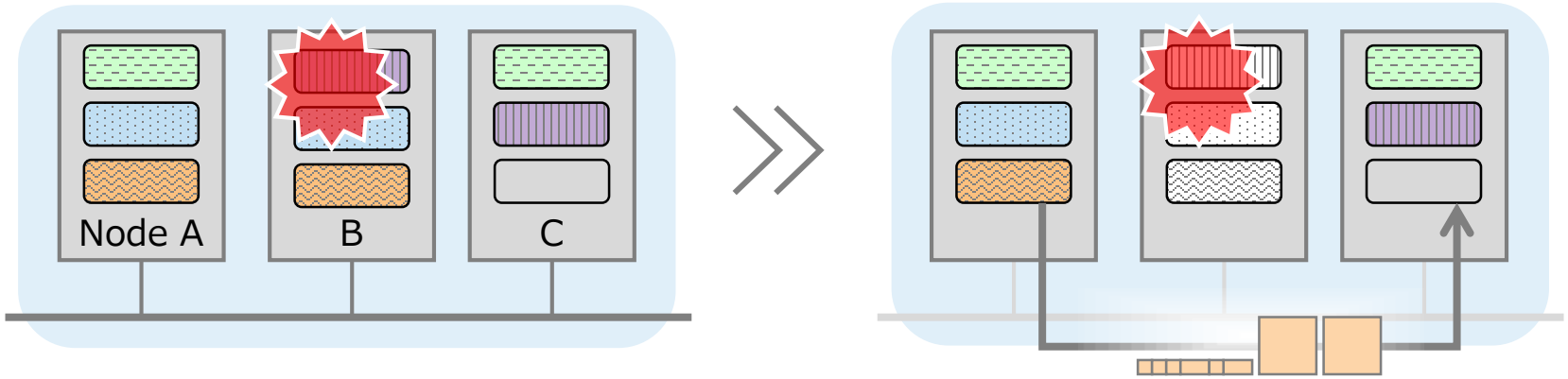
3ノード構成
データ複製度：2

②自律的なクラスタアルゴリズム



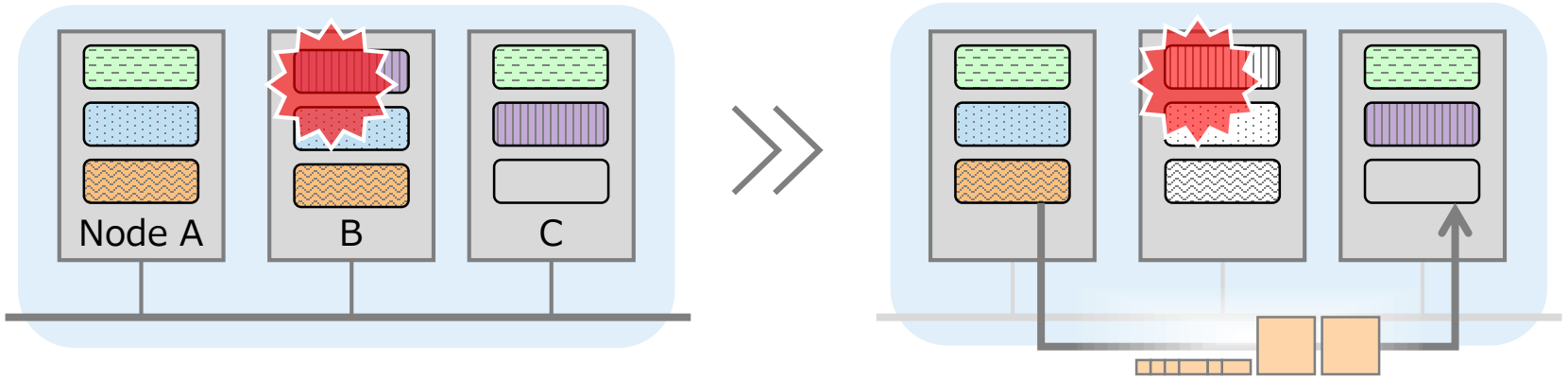
ノード障害発生！
データ複製が減った！可用性が落ちる！

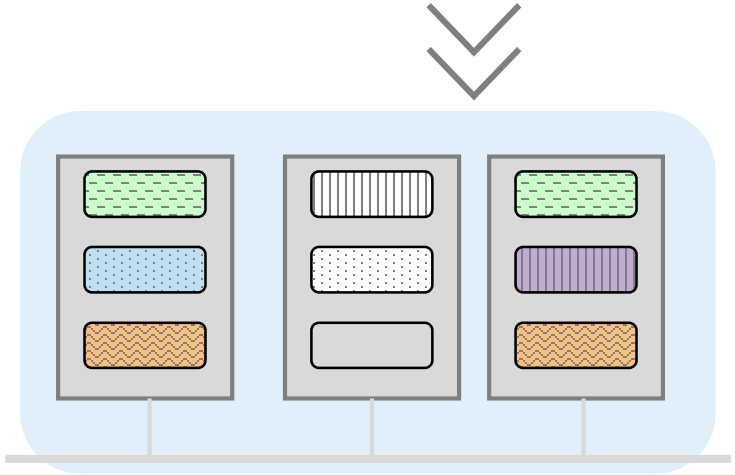
②自律的なクラスタアルゴリズム



データの複製を作り始める。
WALログ転送だけでは時間がかかる。
メモリイメージとWALログの合わせ技で高速化する。

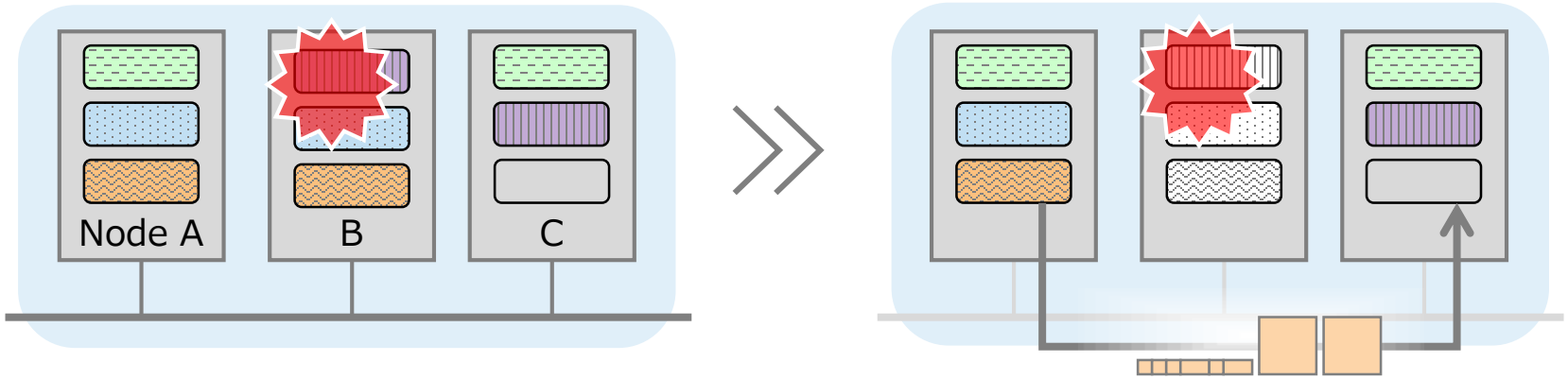
②自律的なクラスタアルゴリズム



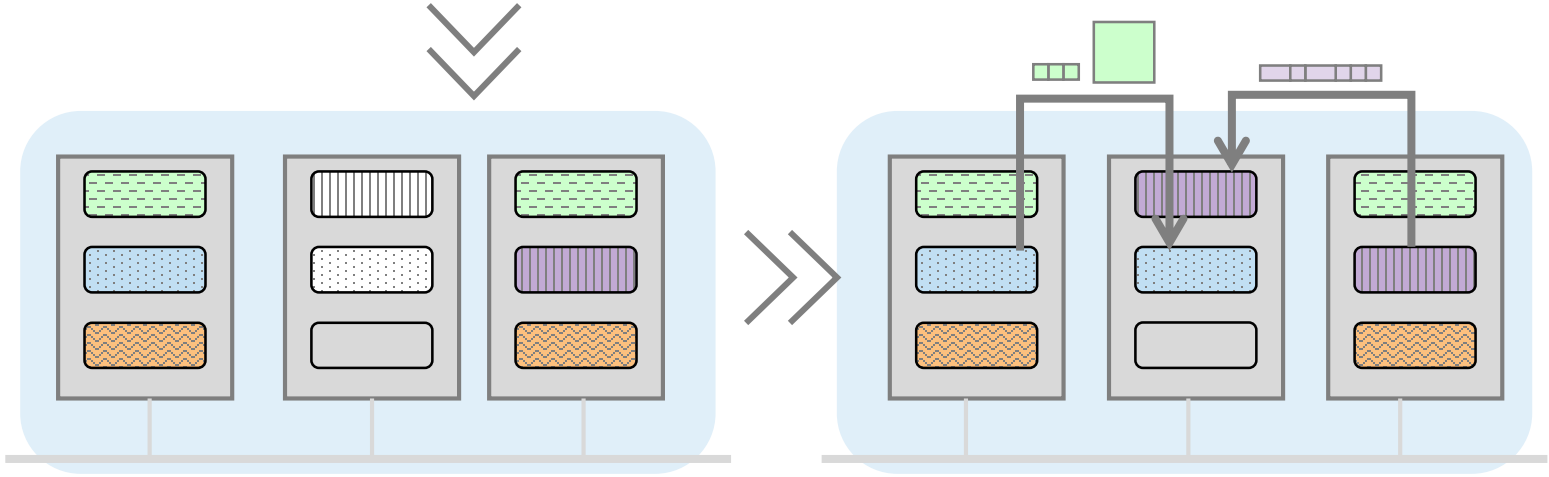


データ複製の成功
クライアントのアクセスを切り替えさせる。

②自律的なクラスタアルゴリズム



オンラインスケールアウト



この動作の繰り返し。
可用性を維持できた！

NoSQLをベースとしたスケールアウト型DB

テーブル間一貫性を除けば、全てをカバーする

	GridDB	NoSQL	SQL
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)
データー貫性	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

NoSQLをベースとしたスケールアウト型DB

テーブル間一貫性を除けば、全てをカバーする

	GridDB	NoSQL	SQL
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

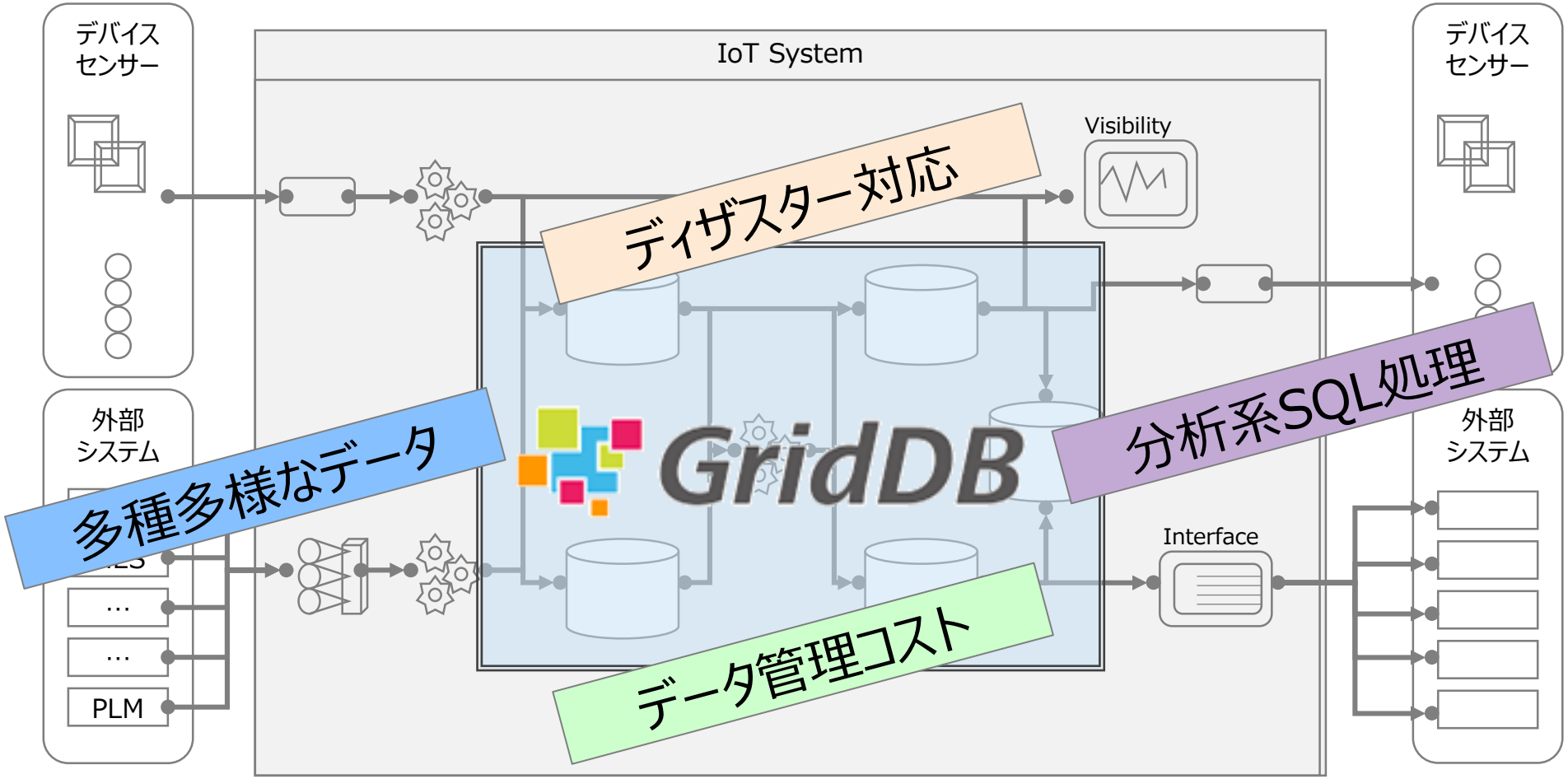
03

新アーキテクチャV5

ー V5シリーズの狙い ー

適用領域拡大に伴う新たな要求

以下の要望に対応するため、アーキテクチャー新



GridDBへの要求の再掲 “多安分耐”

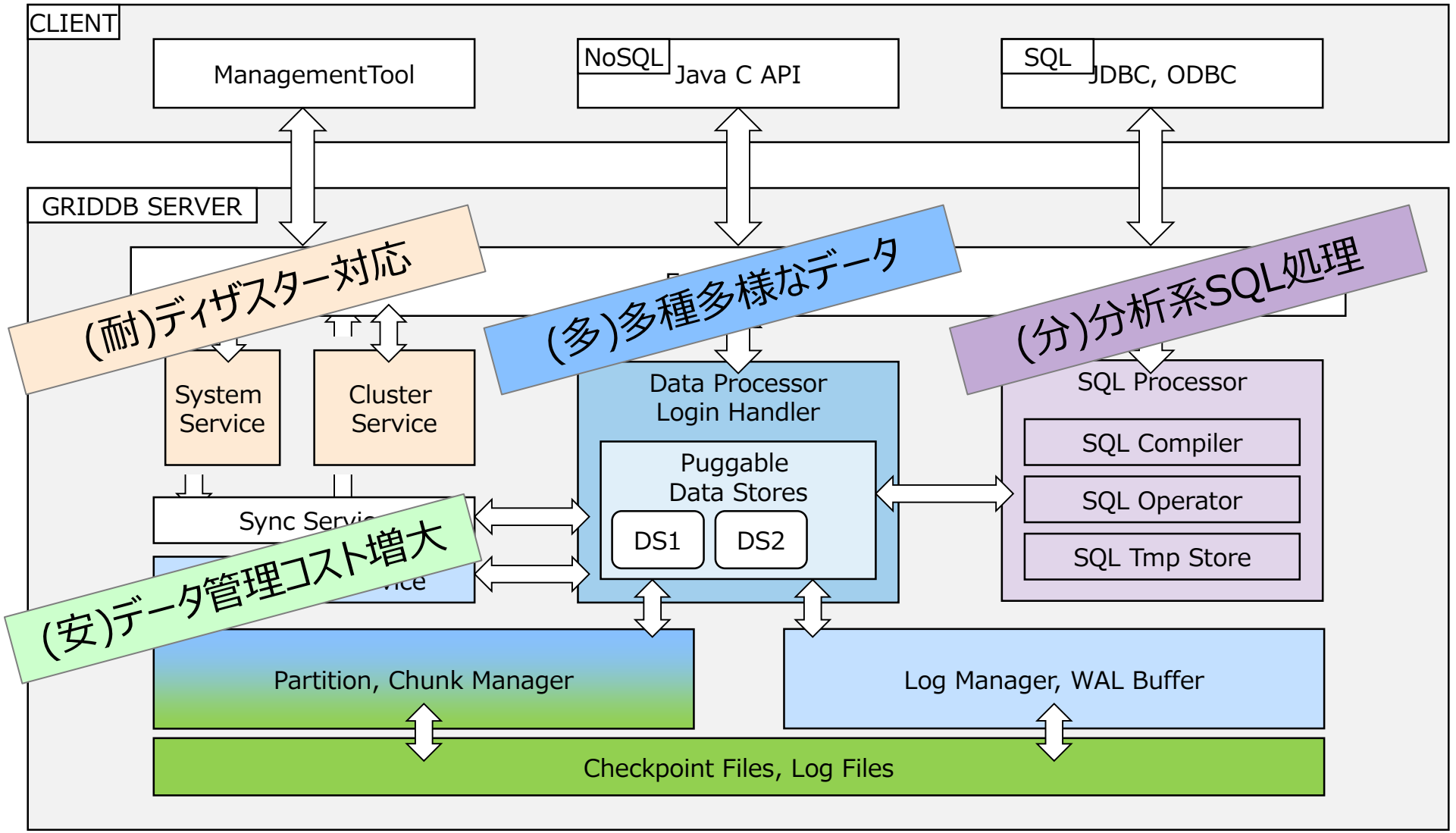
（多）多種多様なデータの最適管理

（安）データ管理コストの削減

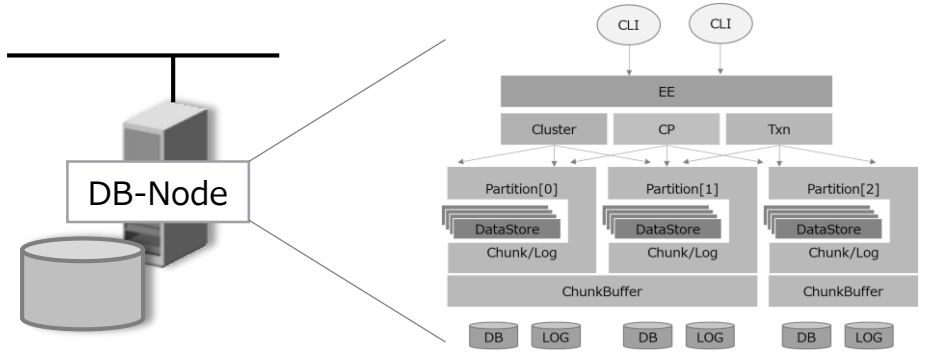
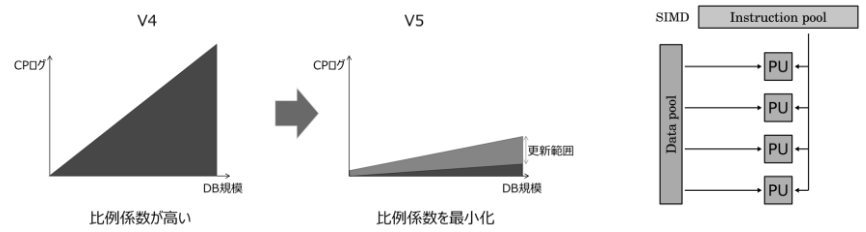
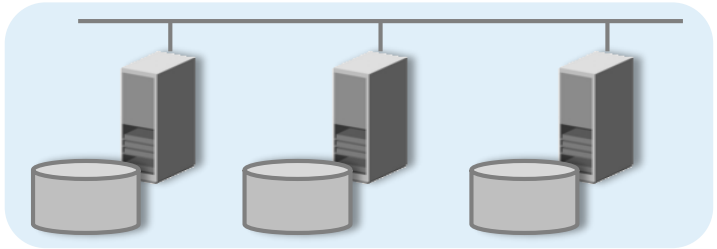
（分）分析系SQLの大幅な性能改善

（耐）ディザスター機能の提供

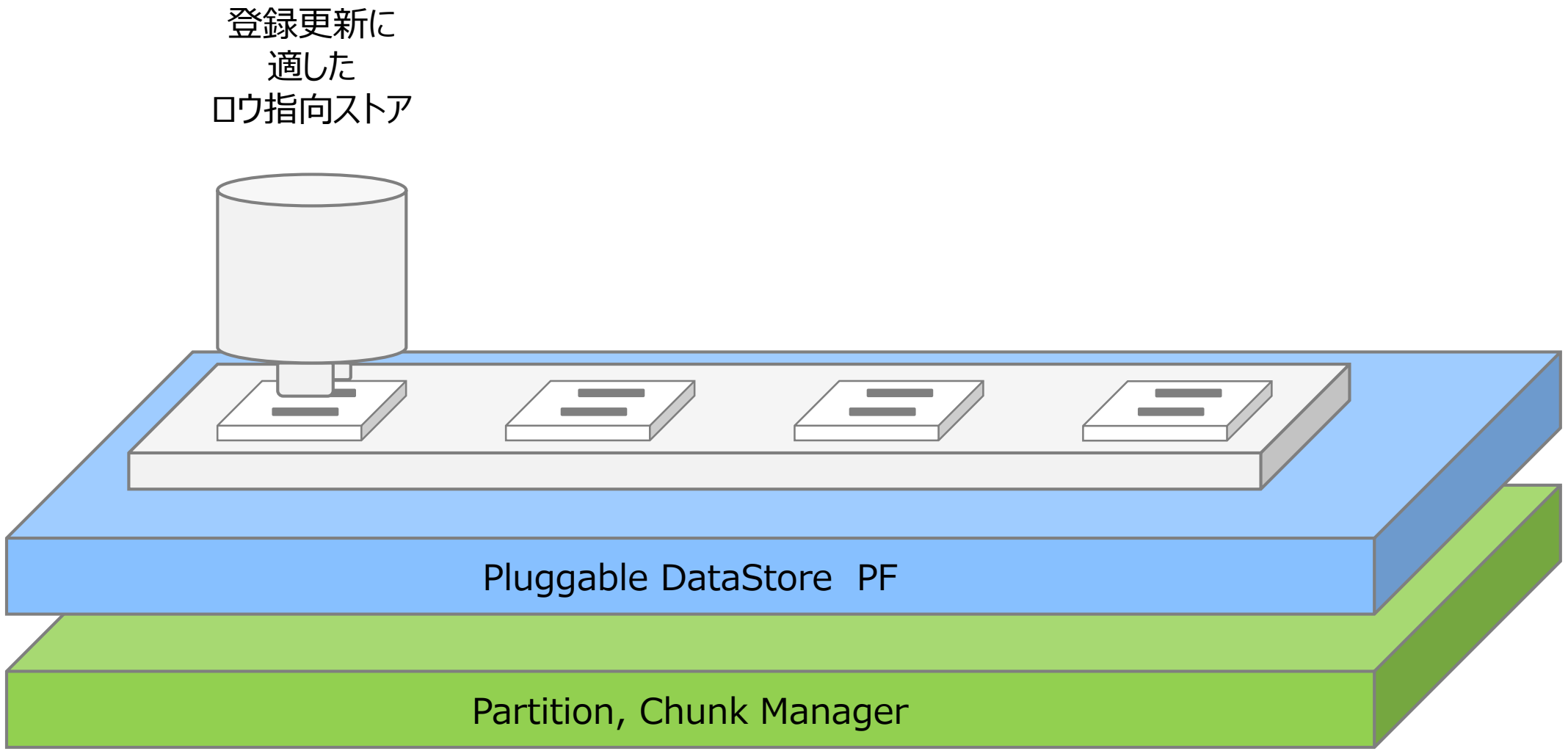
アーキテクチャとの対応関係



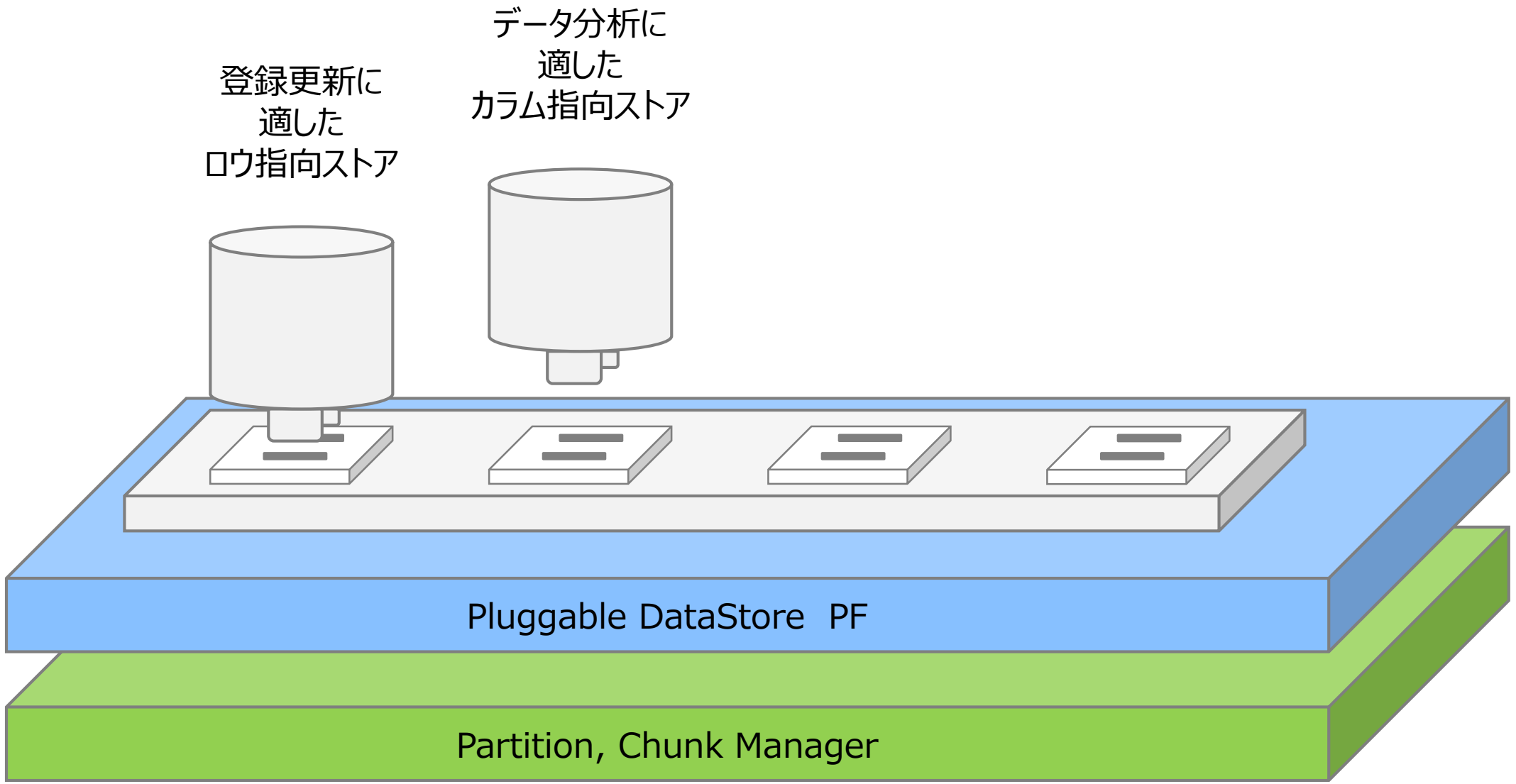
（多）多種多様なデータの最適管理

機能	達成レベル	イメージ図
プラガブルデータストア 5.0	• リレーショナルモデル準拠 • 例えば、カラムナストア • 任意データモデル • 例えば、オブジェクトストア	
ペタバイト対応強化 5.0	• ログサイズ削減(1/N倍) • メモリサイズ削減	
各種高速化 5.0	• チェックポイント高速化 • DB削除、テーブル削除、スキャン(N倍)の高速化 • 一部処理のSIMD化やアルゴリズム変更によるI/O高速化	
クラスタスナップショット	• Copy-On-Writeに基づく • クラスタ全体で瞬時バックアップ	

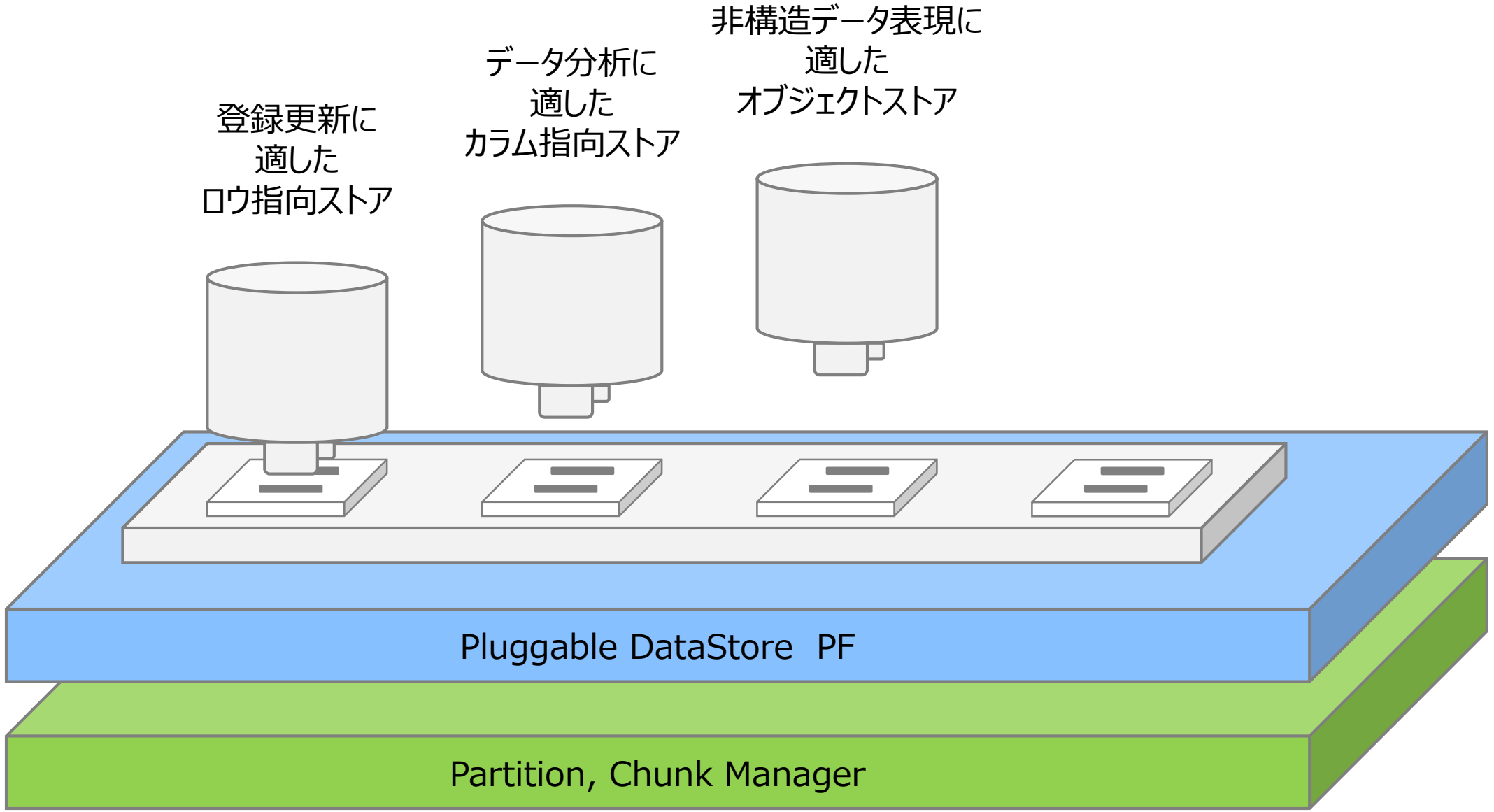
プラガブルデータストア



プラガブルデータストア



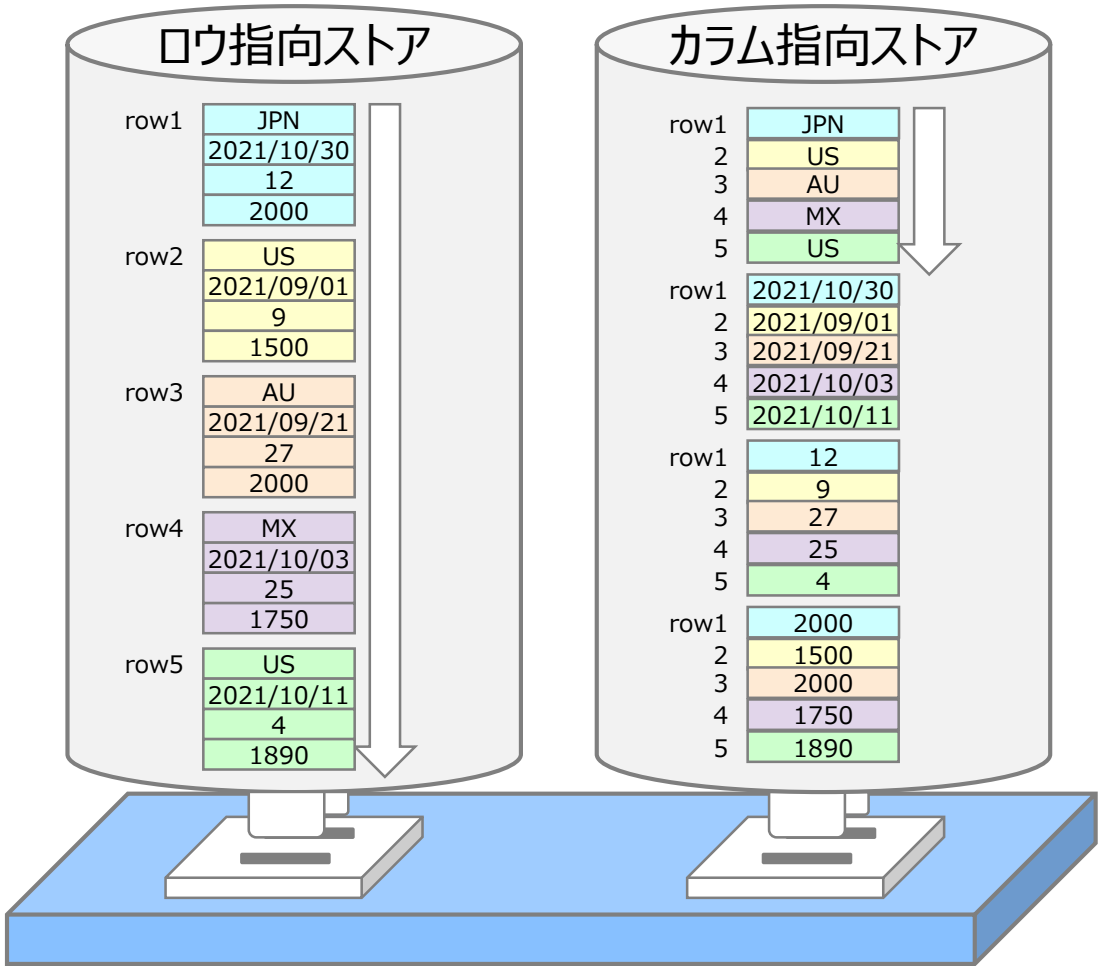
プラガブルデータストア



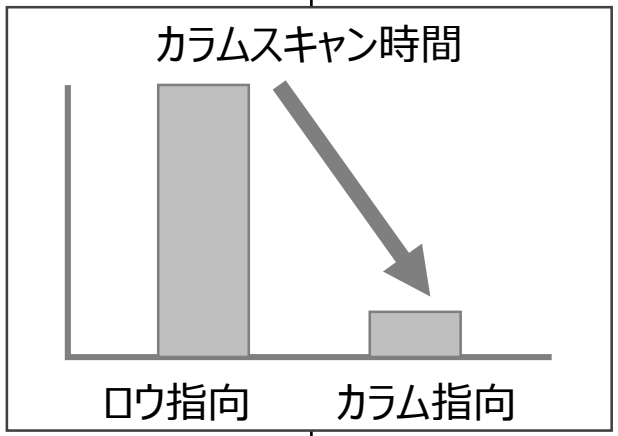
プラガブルデータストア　データストア選択のメリット

	Country	Date	Quantity	Price
row1	JPN	2021/10/30	12	2000
2	US	2		1500
3	AU	2		2000
4	MX	2021/10/03	25	1750
5	US	2021/10/11	4	1890

入力テーブル



	ロウ指向ストア	カラム指向ストア
登録更新	○	×
レコード取得	○	△
カラム参照	△	○
集計	△	○
サイズ	△	○

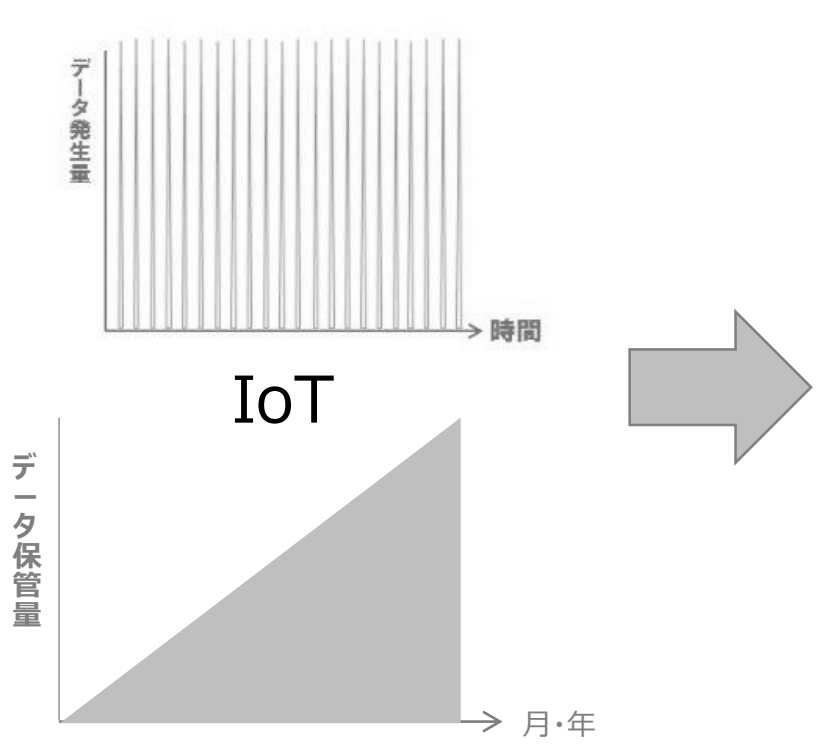


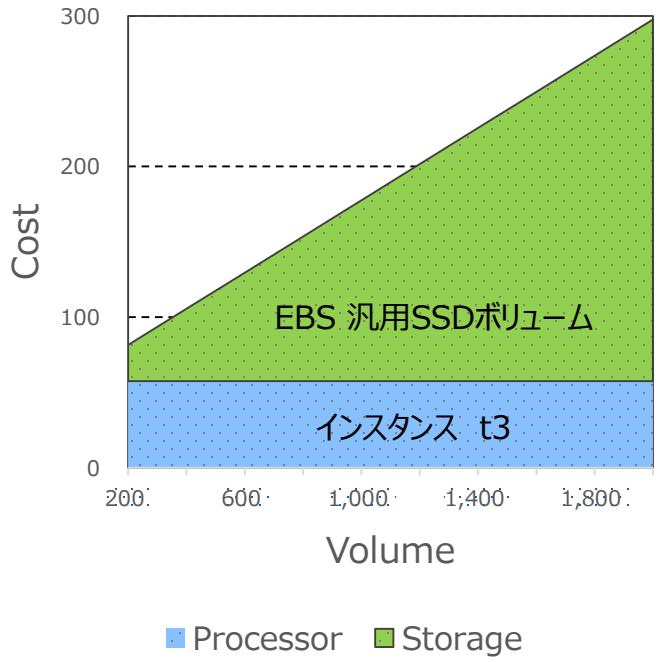
ロウ指向とカラム指向の優劣

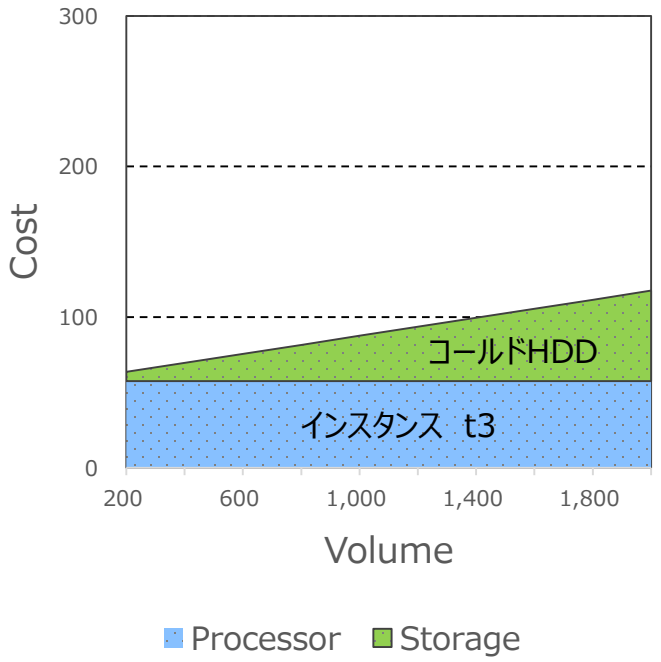
（安）データ管理コストの削減

× IoTにおいて、ストレージコストが全コストを支配

× 一方、ストレージの性能とコストはトレードオフ





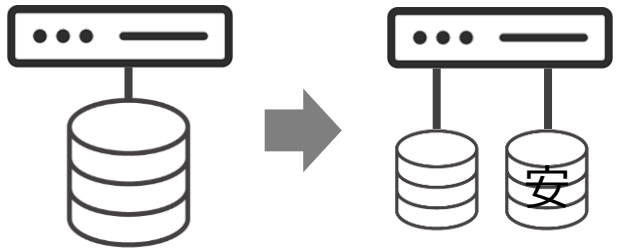


（安）データ管理コストの削減

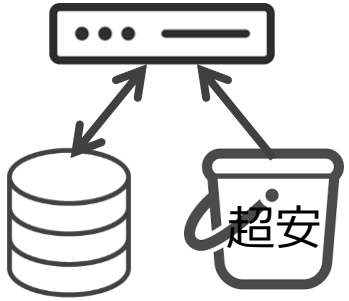
高速ストレージと格安ストレージの使い分けが必要

→　いろいろなタイプの使い分けを提供

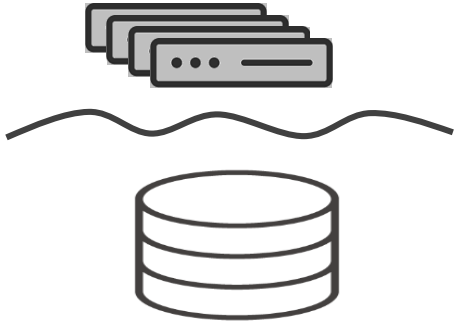
①GridDBデータファイル仮想化によるストレージミックス機能



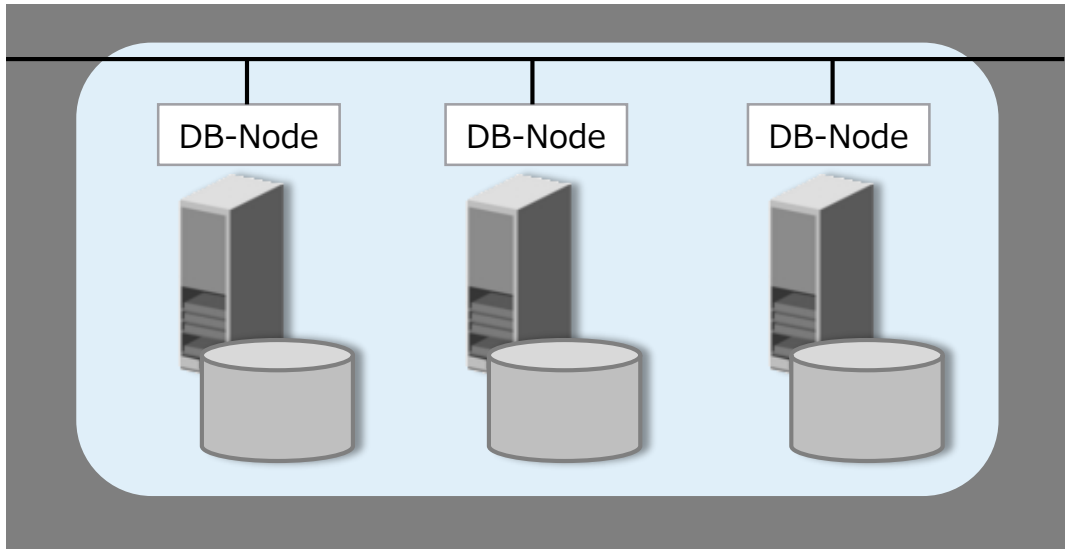
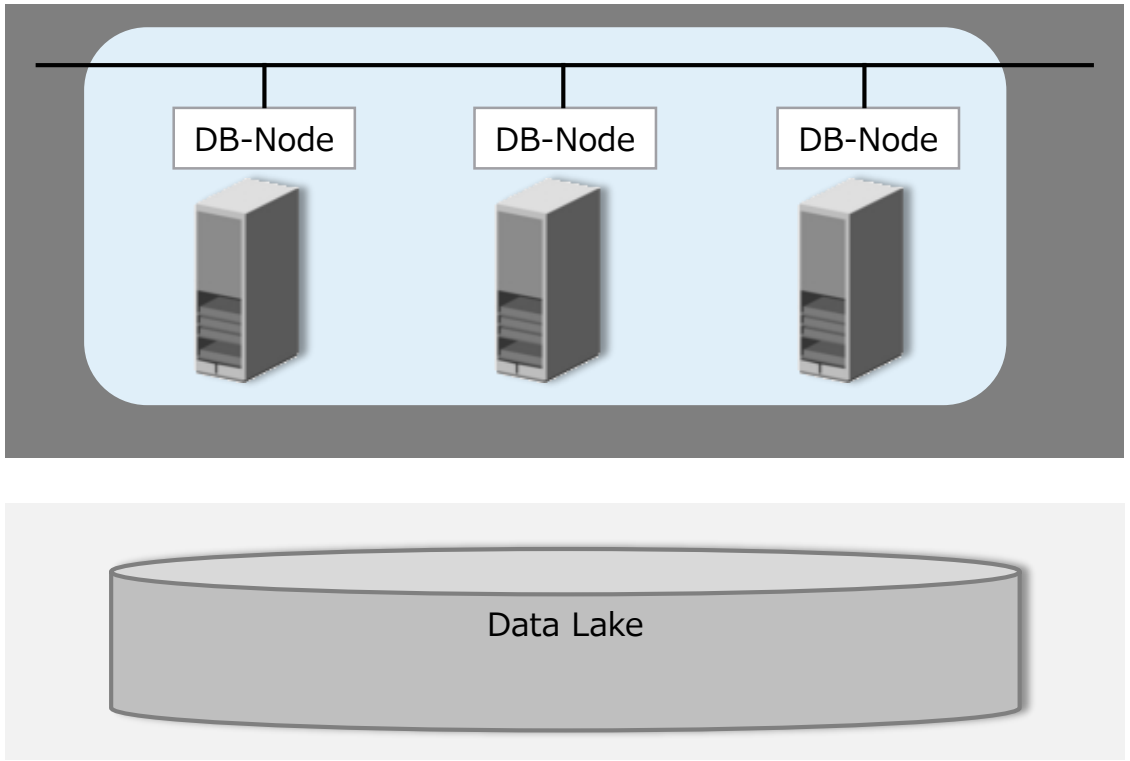
②外部テーブル参照機能



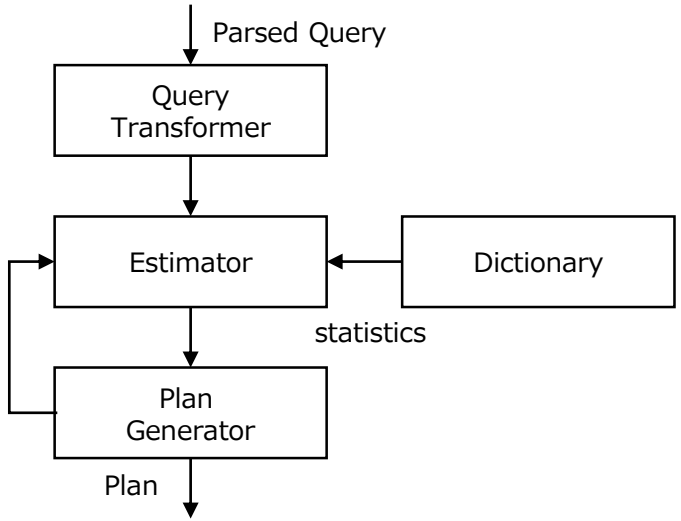
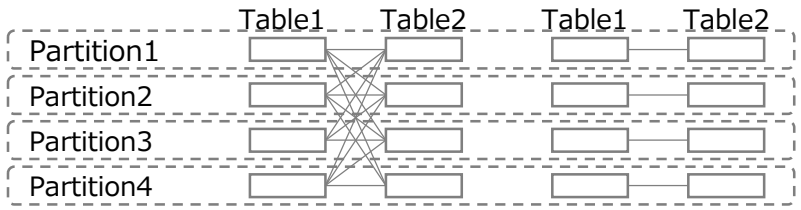
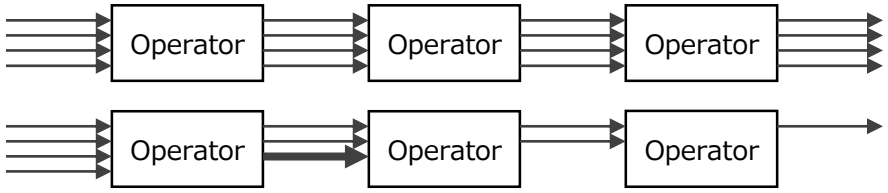
③ストレージ分離機能（サーバレス一種）



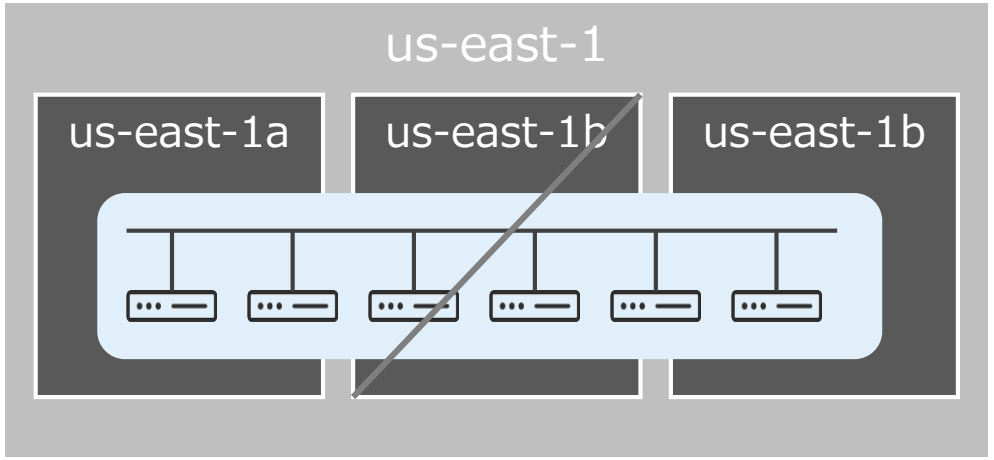
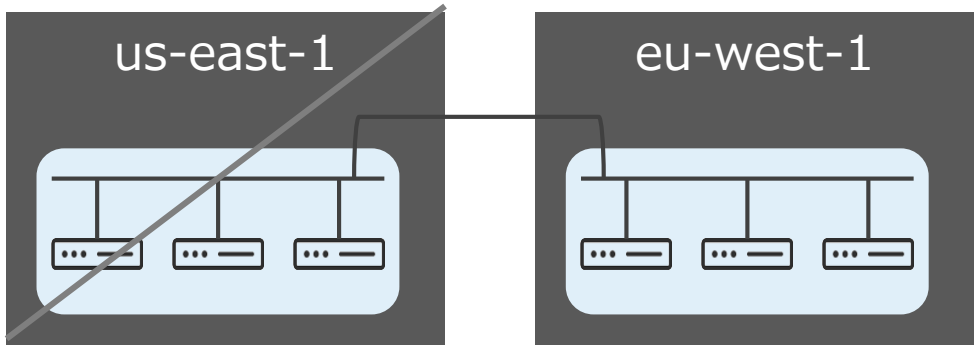
③ストレージ分離機能

非共有アーキテクチャ	ストレージ分離 ディスク共有アーキテクチャ
• 高いパフォーマンス	• 低コスト • 負荷変動に対する弾力性
	

（分）分析系SQLの大幅な性能改善

機能	達成レベル	イメージ図
SQLコンパイラ コストベース最適化 （CBO）	• ジョイン最適化 • オペレータ間最適化 • 統計情報精細化	
並列分散SQL処理 高速化	• 分散ハッシュジョイン（x10～） • 分散ハッシュグループ（x10～）	
パイプライン制御強化	• ノード間データ流量制御	

（耐）ディザスター機能の提供

機能	達成レベル	イメージ図
耐ゾーン障害	• レプリカ増加	
	• レプリカ不変 • ゾーンアウェアなノード配置	
耐サイト障害	• アクティブクラスタ＆コールドクラスタ • 停止期間：～1Day	
	• アクティブクラスタ＆アクティブクラスタ • 停止時間：～1Hour	

おわりに

- V5アーキテクチャー一新により、開発効率アップ
- Community Editionも含めて、リリースサイクルを短縮
- 今後、3エディションでリリースされるV5シリーズにご期待ください。

TOSHIBA

ご清聴ありがとうございました。