Amazon Aurora アーキテクチャ概要

アマゾン ウェブ サービス ジャパン株式会社
データベース スペシャリスト ソリューション アーキテクト
齋藤 航

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

内容についての注意点

本資料では2020年8月1日時点のサービス内容および価格についてご説明しています。
最新の情報はAWS公式ウェブサイト (http://aws.amazon.com/) にてご確認ください。

- 資料作成には十分注意しておりますが、資料内の価格とAWS公式ウェブサイト記載の価格に相違があった場合、AWS公式ウェブサイトの価格を優先とさせていただきます
- 価格は税抜表記となっています。日本居住者のお客様がご利用される場合、別途消費税をご請求させていただきます

AWS does not offer binding price quotes. AWS pricing is publicly available and is subject to change in accordance with the AWS Customer Agreement available at http://aws.amazon.com/agreement/. Any pricing information included in this document is provided only as an estimate of usage charges for AWS services based on certain information that you have provided. Monthly charges will be based on your actual use of AWS services, and may vary from the estimates provided.

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

本セッションについて

対象者

- Amazon Aurora のアーキテクチャにご興味をお持ちの方

ねらい

- どのようにして Amazon Aurora が高い可用性や、パフォーマンスを実現しているかをご理解いただくことで、安心して Amazon Aurora をご利用いただけるようになる

前提知識

- データベースの基礎知識

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

アジェンダ

- Amazon Aurora とは
- アーキテクチャ概要
- パフォーマンス
- 使いやすい機能

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

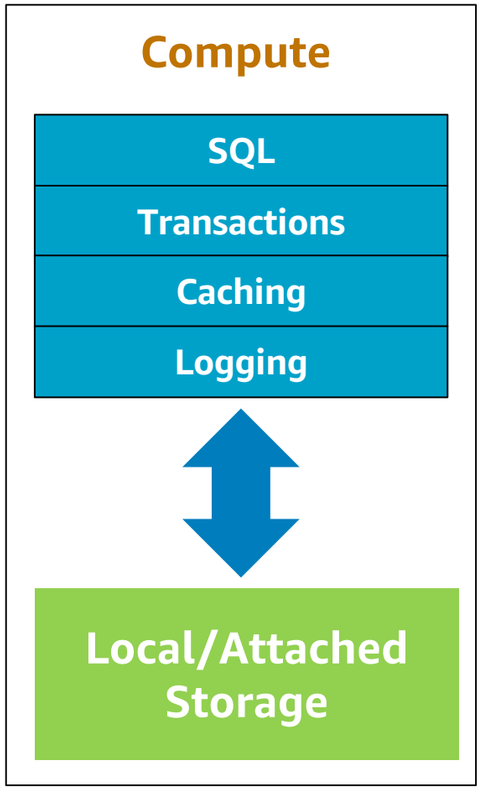
Amazon Aurora とは

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

従来のデータベースアーキテクチャ

全ての構成要素が1つにまとまっている
モノリシック な構成

ローカルストレージでパフォーマンスと可用性、
耐久性を両立させる構成

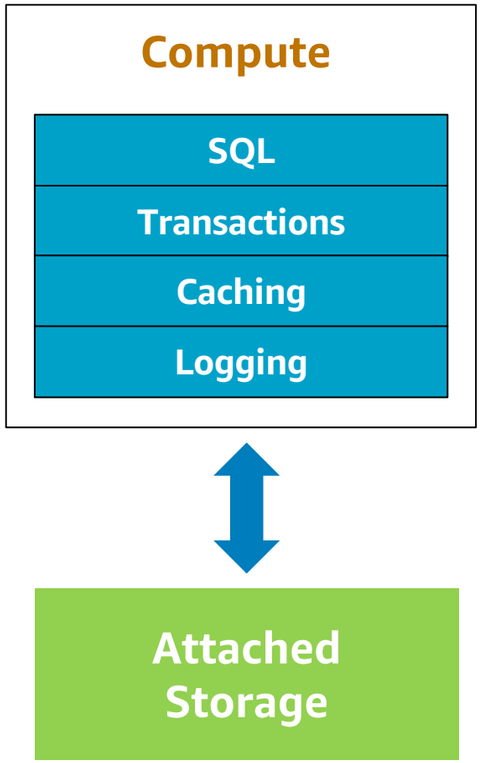


© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

クラウドでのデータベース

コンピュートとストレージは異なるライフサイクルを持ち、
全く異なる要件を満たす必要がある

- インスタンス障害や再配置の可能性
- インスタンス停止の可能性
- インスタンスのスケールアップ/スケールダウン
- インスタンスのスケールアウト/スケールイン



コンピュートとストレージは
スケーラビリティ、可用性、耐久性を考慮して
最適に分離 される必要がある

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

リレーショナルデータベースを再考する

現代にデータベースを設計するならどうするか？

- 少なくとも、1970 年代とは設計は異なるはず

- 以下のような機能を備えたい
  - 簡単にスケールアウト可能
  - 自己修復機能を持つ
  - AWS の既存のサービスを活かすことができる

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

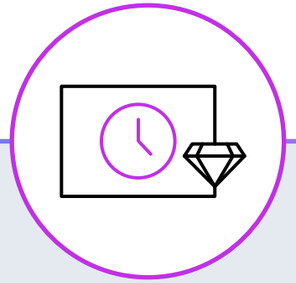
Amazon Aurora

クラウド向けに再設計された MySQL, PostgreSQL と互換性のある RDBMS
コマーシャルデータベースの性能と可用性を 1/10 のコストで

優れた性能と拡張性

標準的なMySQLと比べて5倍、
標準的なPostgreSQLと比べて
3倍のスループットを実現;
リードレプリカを最大 15 個
追加してスケールアウト可能

高可用性と耐久性



耐障害性、自己修復機能を
兼ね備えたストレージ;
3 つのAZにわたり、6 個のコ
ピーを保持; Amazon S3への
継続的なバックアップ

高い安全性

ネットワーク分離、
保管時/通信の暗号化

フルマネージド

ハードウェアのプロビジョニ
ング、ソフトウェアのパッチ
適用、セットアップ、構成、
バックアップといった
管理タスクからの解放

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Amazon Aurora Innovations

Re-imagining database for cloud

1. スケールアウト, 分散, マルチテナントデザイン
2. AWS サービスを活用したサービスオリエンテッドアーキテクチャ
3. フルマネージド・サービス - 自動化されたタスク

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

スケールアウト, 分散, マルチテナントデザイン

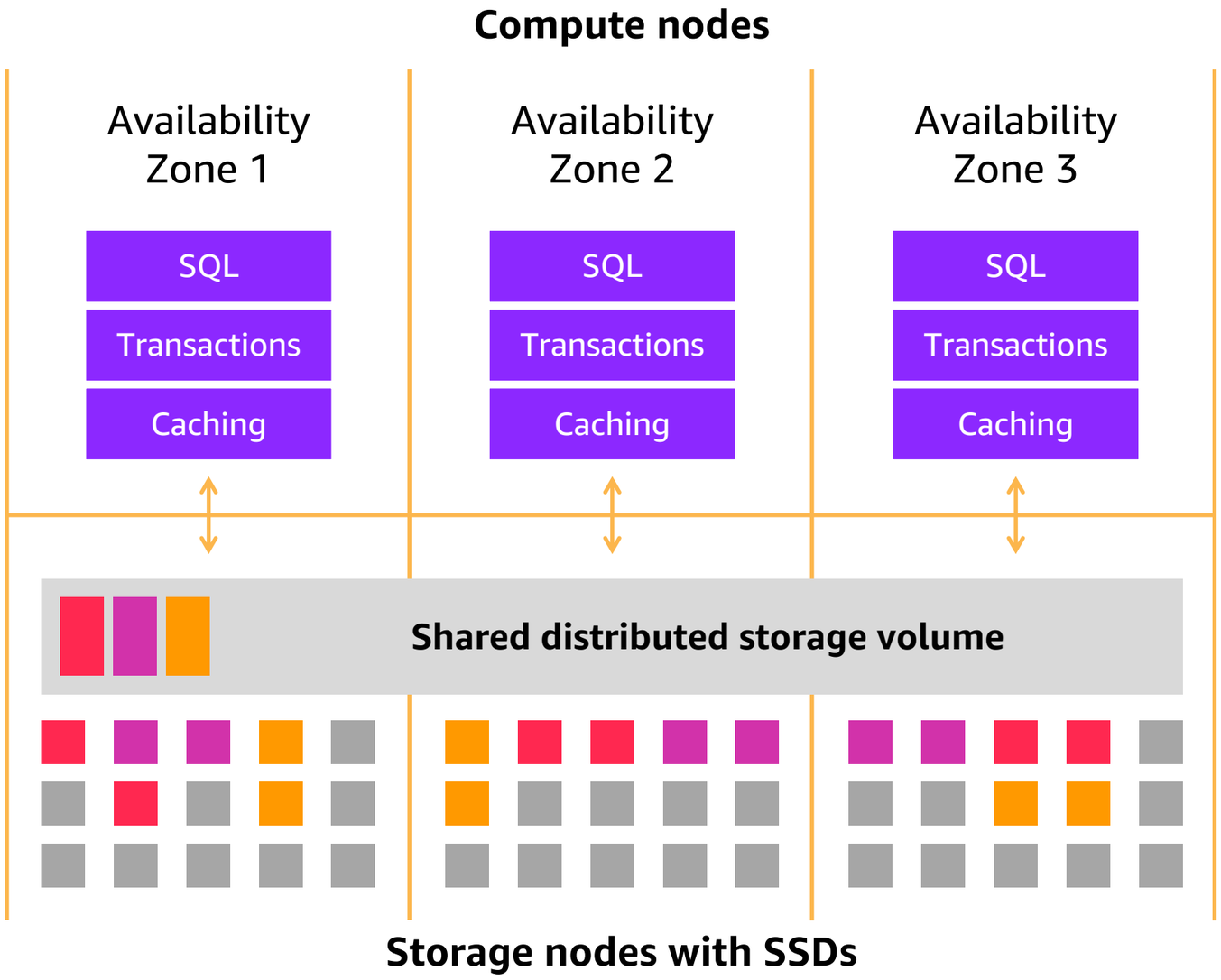
Aurora scale-out, distributed architecture

データベース用に設計された
専用の log-structured 分散ストレージシステム

3つの異なるアベイラビリティゾーンに分散された
数百のストレージノードにストライピングされた
ストレージボリューム

データは10GBのプロテクショングループの単位で
書き込まれる。自動的に最大64TBまで拡張

AZ+1 の障害から保護するためにデータを
各アベイラビリティーゾーンに2つのコピー、
リージョン内で計6つのコピー



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

AWS サービスを活用した サービスオリエンテッドアーキテクチャ

Lambda function

AWS Lambda イベントを stored procedures/triggers から実行

Amazon S3

Amazon Simple Storage Service (Amazon S3) データをロード, S3を利用したバックアップデータの保存、リストア

AWS Identity and Access Management

AWS Identity and Access Management (IAM) roles を利用してデータベースユーザの認証

Amazon CloudWatch

システムメトリクスやログを CloudWatchへアップロード

Amazon SageMaker

Amazon SageMaker, Amazon Comprehend との統合: SQL を利用した連携

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

フルマネージド・サービス – 自動化されたタスク

You

- スキーマのデザイン
- クエリの作成
- データベースの最適化

AWS

- 自動化されたフェイルオーバー
- バックアップとリカバリ
- 分離とセキュリティ
- コンプライアンスへの準拠
- 数クリックでスケール
- 自動化されたパッチ適用
- 拡張モニタリング
- OS, DBのインストール

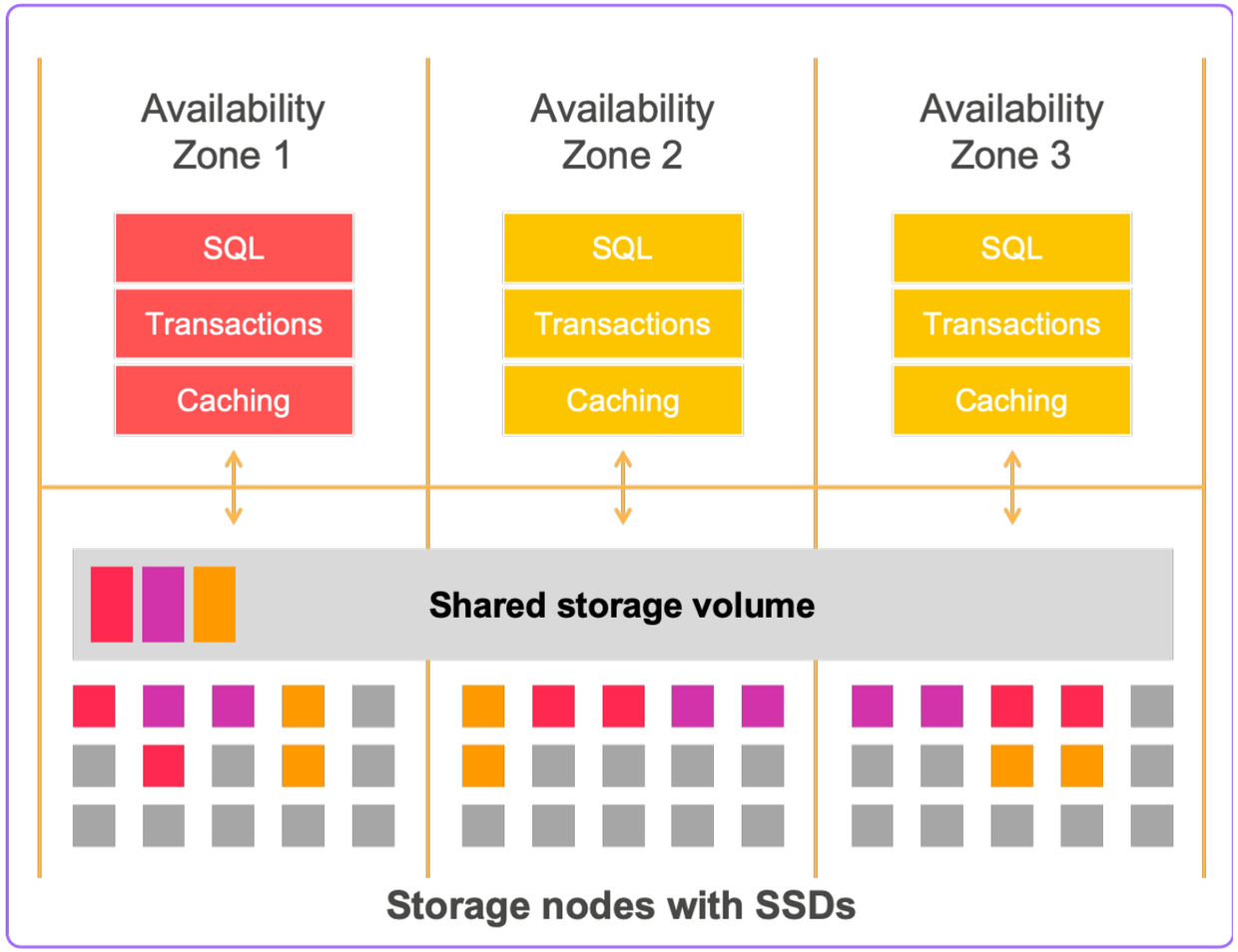
データベース管理者は、時間のかかる
データベース管理タスクから解放され、
アプリケーションやビジネスに集中できる

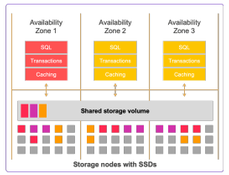
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

アーキテクチャ概要

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Aurora を構成するコンポーネント





- **Amazon Aurora DB クラスター**
  - Amazon Aurora の管理単位
  - Writer インスタンス、Readerインスタンス、クラスターボリュームの総称

- **Writer インスタンス (プライマリ)**
  - 読み込み、書き込みを行うインスタンス

- **Reader インスタンス**
  - 読み込み専用のインスタンス (15台まで作成可能)

- **クラスターボリューム**
  - 3つの AZ 間でレプリケートされる仮想ボリューム
  - Writer インスタンスも Reader インスタンスも同じクラスターボリュームを利用

- **Aurora エンドポイント**
  - Aurora の接続先を示す URL

詳細：http://docs.aws.amazon.com/ja_jp/AmazonRDS/latest/UserGuide/Aurora.Overview.html

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Aurora のストレージ

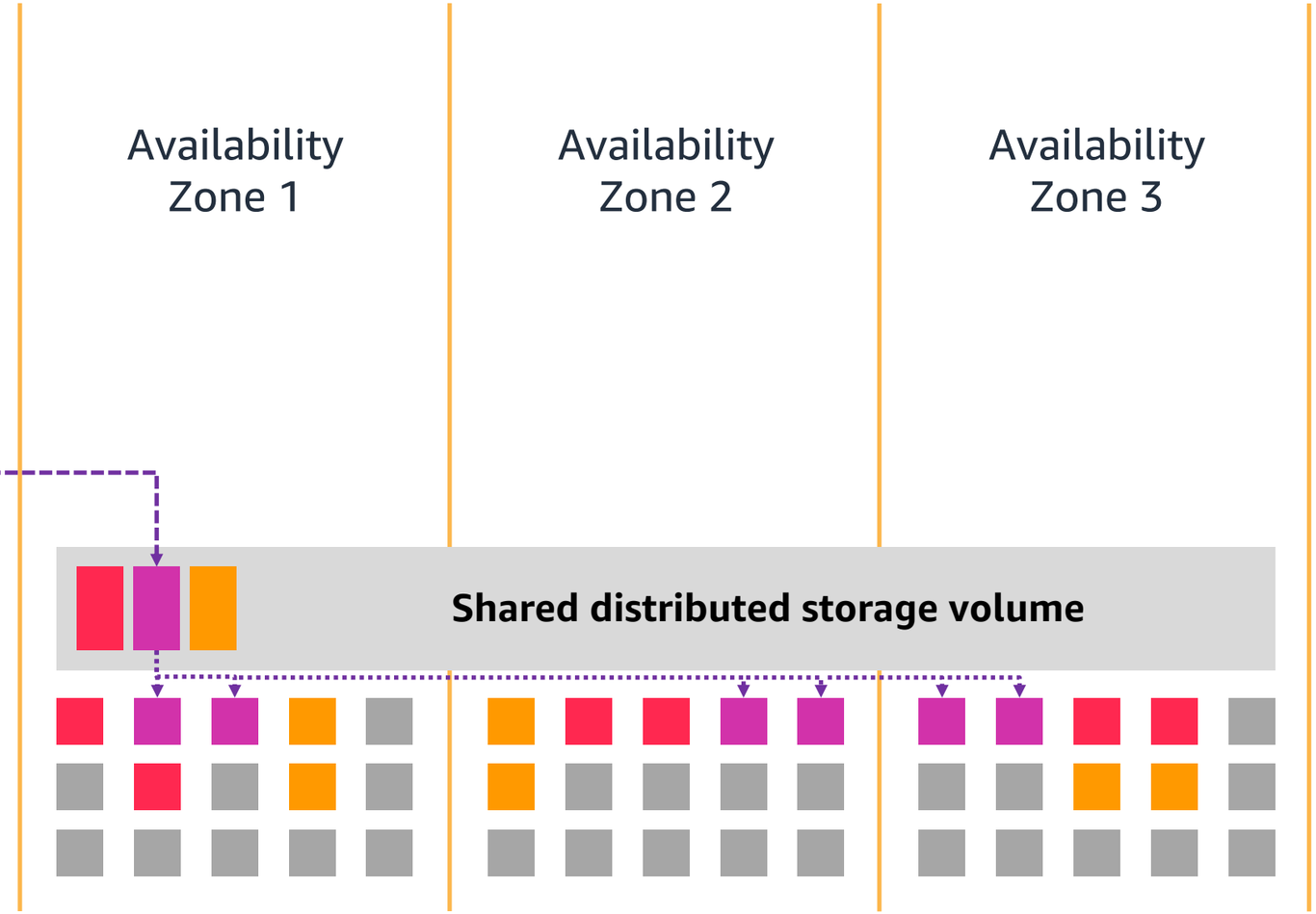
データベース用に設計された
専用の **log-structured 分散ストレージシステム**

3つの異なるアベイラビリティゾーンに分散された
数百のストレージノードにストライピングされた
ストレージボリューム

データは 10GB の プロテクショングループ の
単位で管理される。自動的に最大 64TB まで拡張

AZ+1 の障害から保護するためにデータを
各アベイラビリティーゾーンに2つのコピー、
リージョン内で計6つのコピー

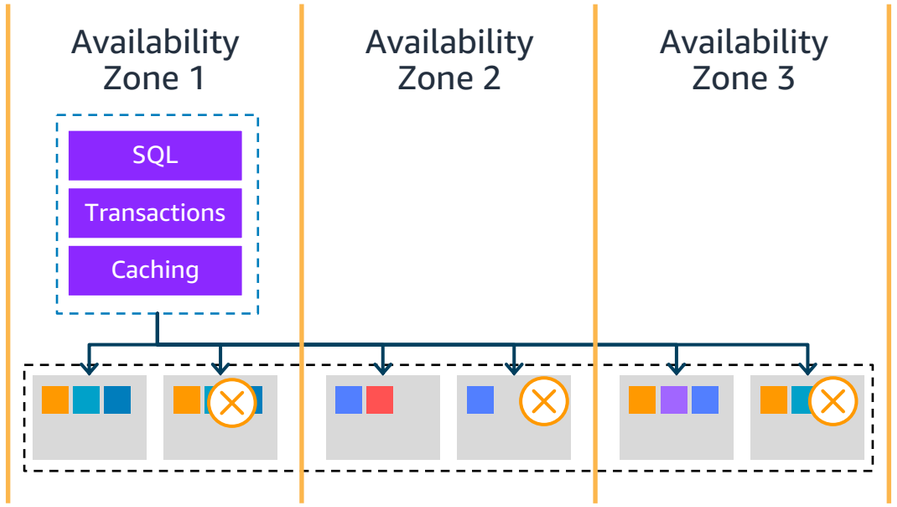
データは継続的に S3 へバックアップ



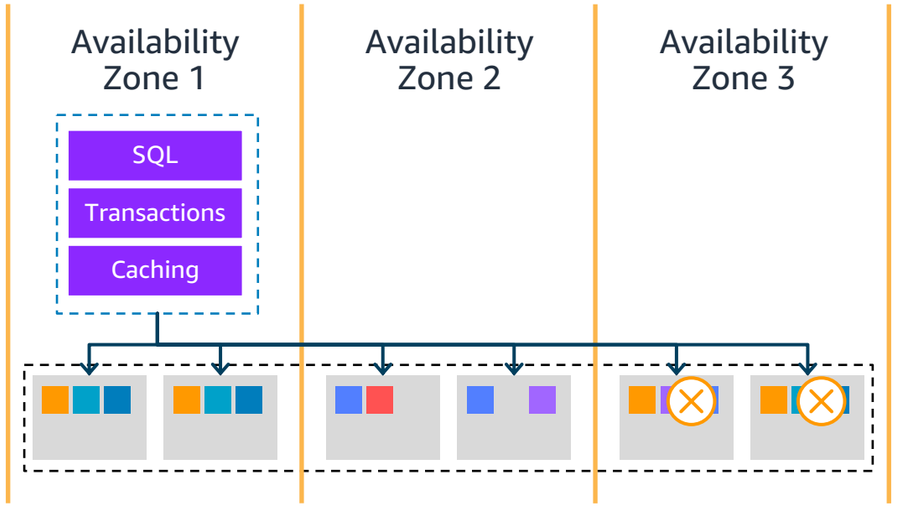
詳細：
https://docs.aws.amazon.com/ja_jp/AmazonRDS/latest/AuroraUserGuide/Aurora.Overview.StorageReliability.html

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

自動修復、耐障害性



Read availability



Write availability

データは6つのノードすべてに非同期・並列で書き込み

書き込みには 4/6 ノードのクォーラム、読み込みには 3/6 クォーラムが必要

各ノード間のデータの欠損、破損は P2P のゴシッププロトコルで確認、修復される

詳細： https://aws.amazon.com/jp/blogs/news/amazon-aurora-under-the-hood-quorum-and-correlated-failure/

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

クォーラムモデル

レプリケーション管理のためのクォーラムのルール

1. **Vr + Vw > V** (読み込みクォーラム Vr 、書き込みクォーラム Vw が、少なくとも1つ共通のコピーを保持)
2. **Vw > V/2** (書き込みクォーラムは、過半数のコピーを保持)

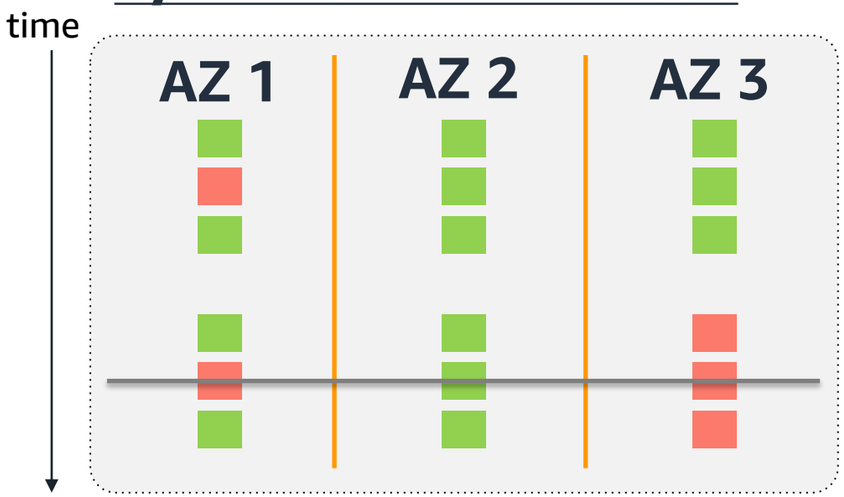
V (コピーの数)	V_w (書き込みクォーラム)	V_r (読み込みクォーラム)
1	1	1
2	2	1
3	2	2
4	3	2
5	3	3
6	4	3
7	4	4

Amazon Aurora におけるクォーラムモデル

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

なぜ、6つのコピーが必要なのか？

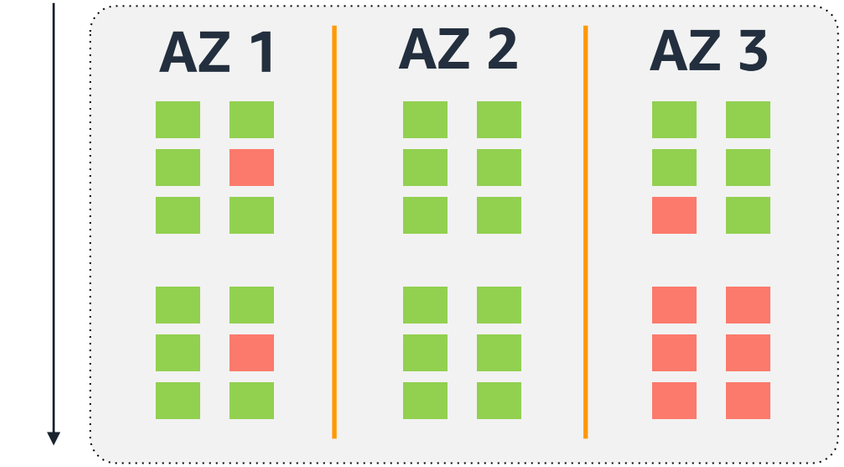
2/3 クォーラムの場合



2/3 read
2/3 write

Quorum break on AZ failure

4/6 クォーラムの場合



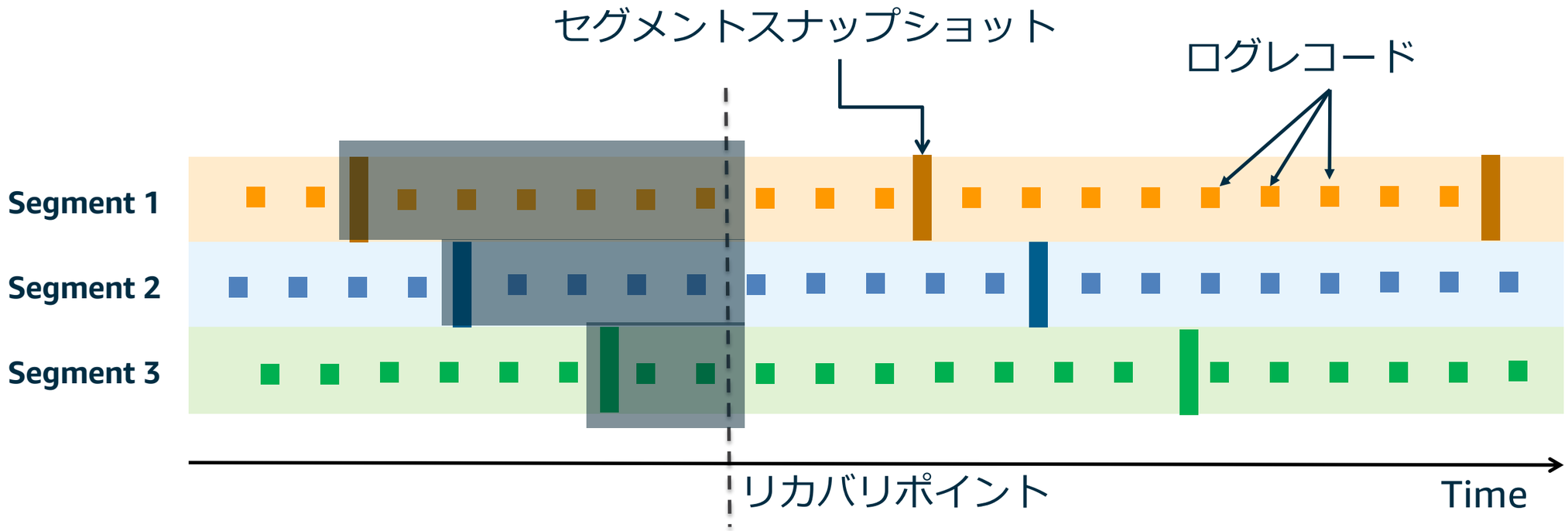
3/6 read
4/6 write

Quorum survives AZ failure

- AWS は大規模環境
- AZ障害は影響範囲が広域
- **AZ+1 の障害**(二重障害)を**許容**し、修復する必要がある
- 3つのAZの場合、6つのコピーが必要

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

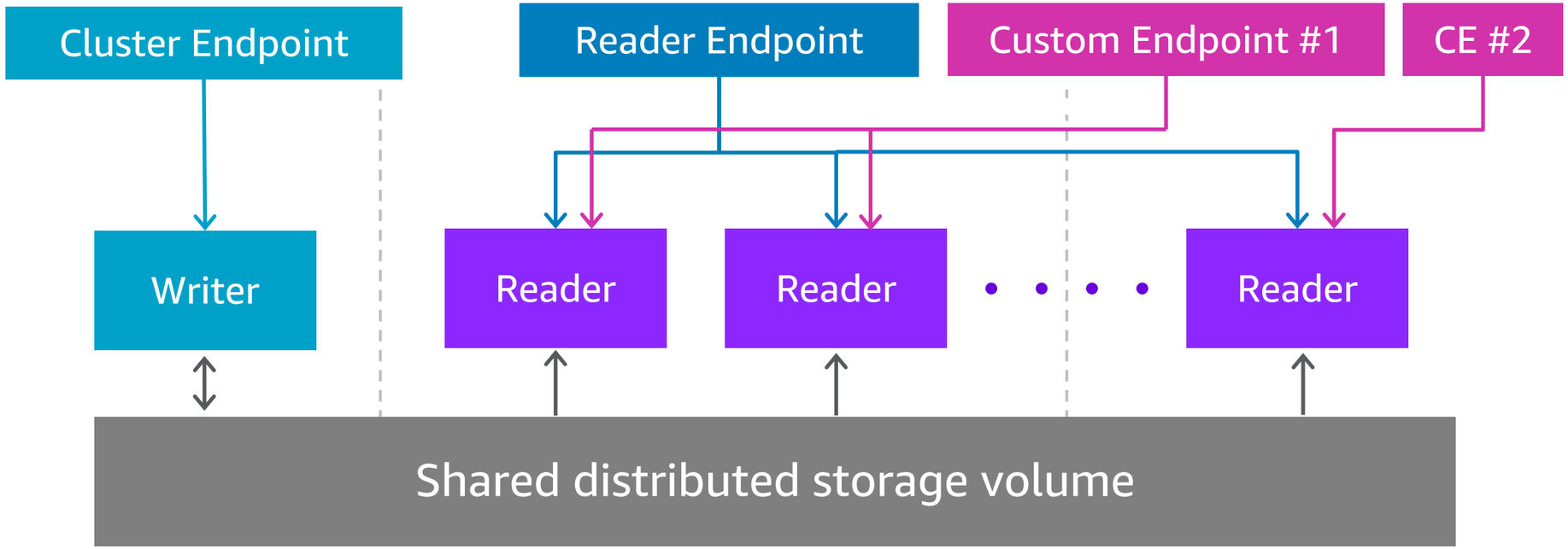
Streaming Backup



- 各セグメントごとに Amazon S3 へ継続的なスナップショットを並列に取得
- Amazon Aurora が使用しているディスクの仕組みによりパフォーマンスへ影響を与えない
- リストア時、並列非同期に適切なセグメントのスナップショットとログを取得し、ストレージへ適用

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Aurora リードレプリカ, エンドポイント

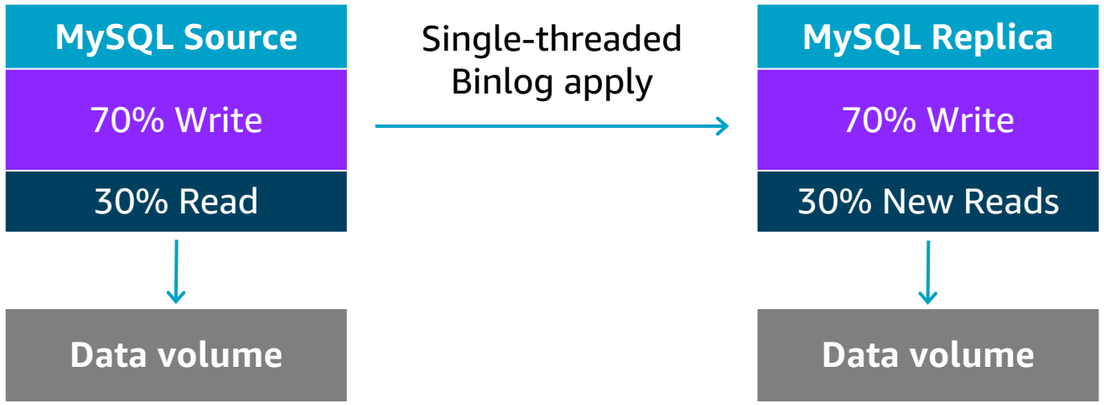


3つのアベイラビリティーゾーンで最大15個の昇格可能なリードレプリカ
リードレプリカは Auto Scaling による自動増減が可能
低遅延な REDO ログベースの物理レプリケーション (通常 20～40 ミリ秒のレプリカラグ)
フェイルオーバーの優先順位を構成可能
ライターとリーダーはカスタムエンドポイントの一部になることが可能

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

従来のレプリカと Aurora リードレプリカの違い

MySQL/PostgreSQL read scaling



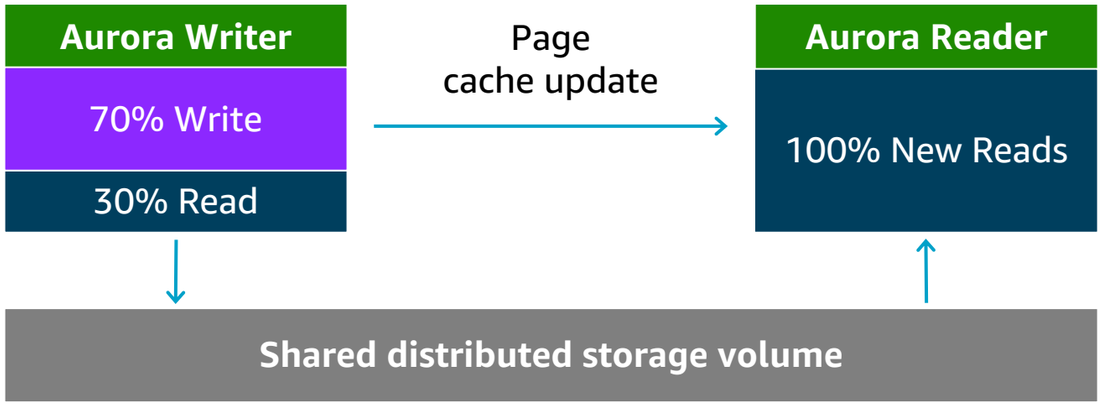
差分 変更を用いた論理レプリケーション

レプリカにソースと 同等の書き込み が発生

独立した ストレージ

ワークロードによっては大きなレプリカラグ

Amazon Aurora read scaling



差分 変更を用いた物理レプリケーション

レプリカに書き込みは発生しない

共有 ストレージ

通常 20~40 ミリ秒のレプリカラグ

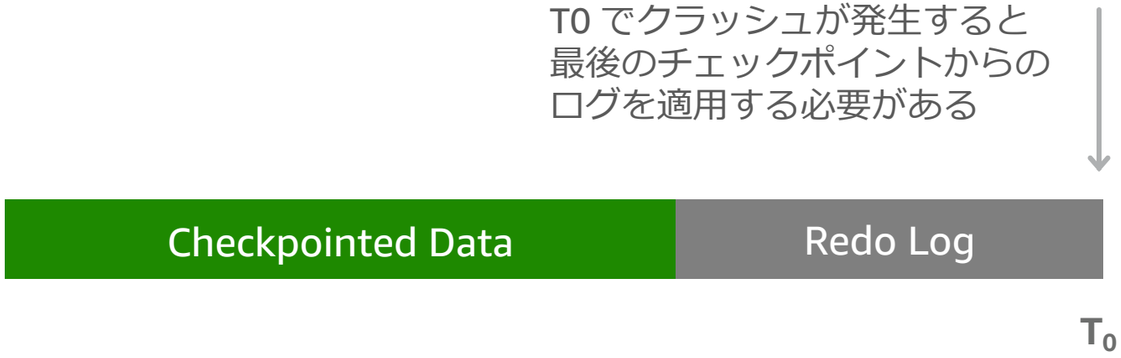
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

高速なクラッシュリカバリ

従来のデータベース実装

最後のチェックポイントから
ログを適用していく

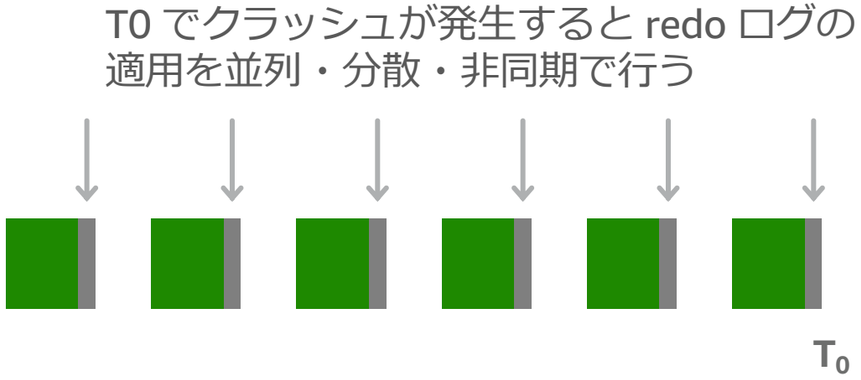
MySQL の場合、処理はシングルスレッドで
多数のディスクアクセスが必要



Amazon Aurora

リカバリ中か否かは関係なく、
ストレージノードで継続的に
ログレコードを再生 (No checkpoint)

並列、分散、非同期で行われる



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

存続可能なキャッシュ

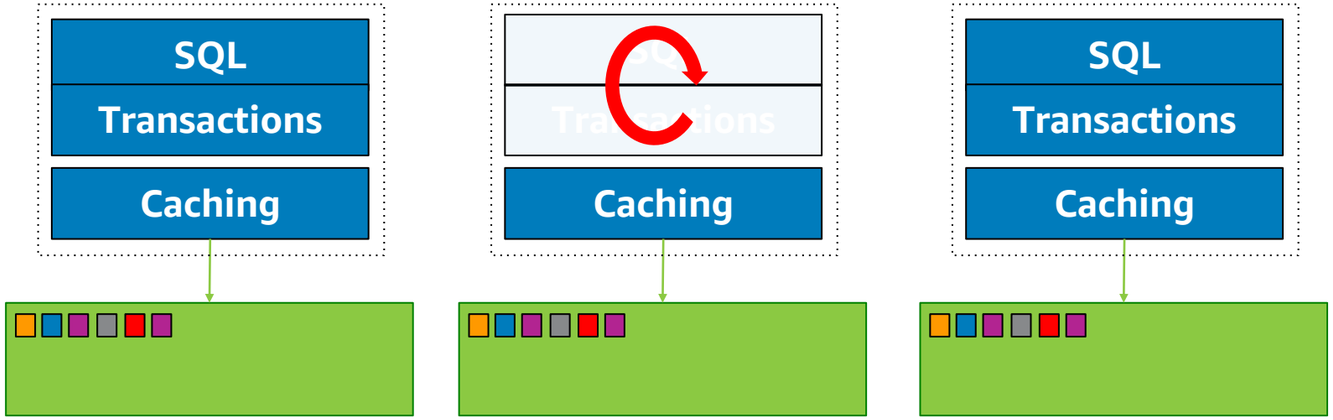
データベースのプロセスから
キャッシュを分離

データベースの再起動が発生しても、
キャッシュは Warm 状態を維持

再起動後でもオンキャッシュの処理を
素早く再開可能

高速なクラッシュリカバリ
＋ 存続可能なキャッシュ
＝ DB障害から素早く簡単なリカバリ

キャッシュプロセスを DB プロセスの外に置くことで、
DB プロセスの再起動でもキャッシュが残る



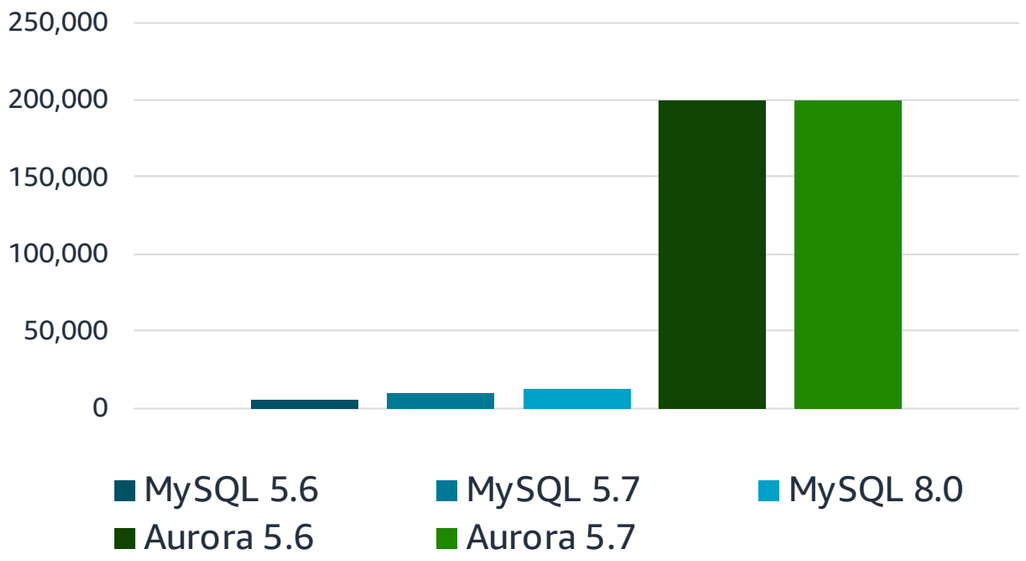
詳細：https://docs.aws.amazon.com/ja_jp/AmazonRDS/latest/AuroraUserGuide/Aurora.Overview.StorageReliability.html

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

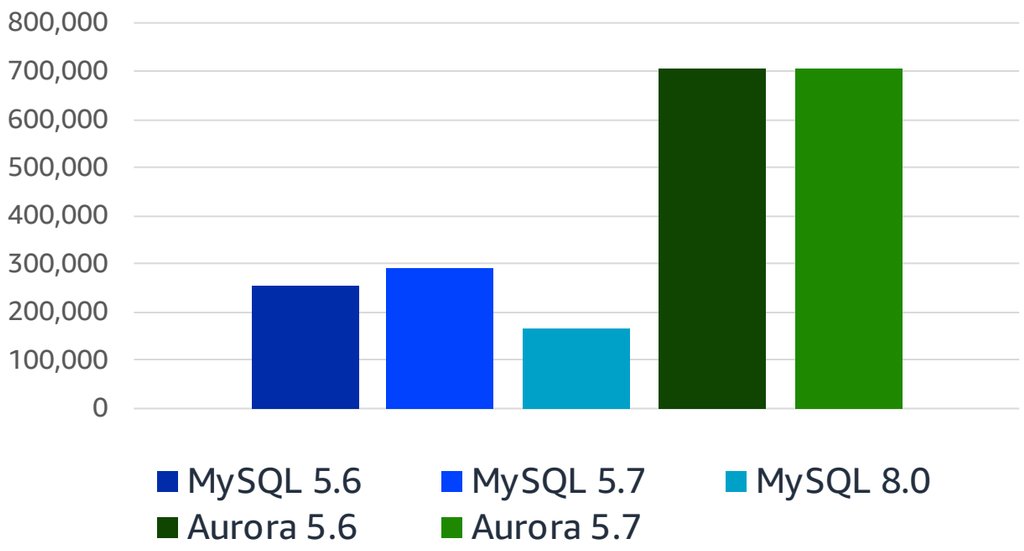
Aurora のパフォーマンス

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Aurora MySQL: 5倍のスループット



Write throughput



Read throughput

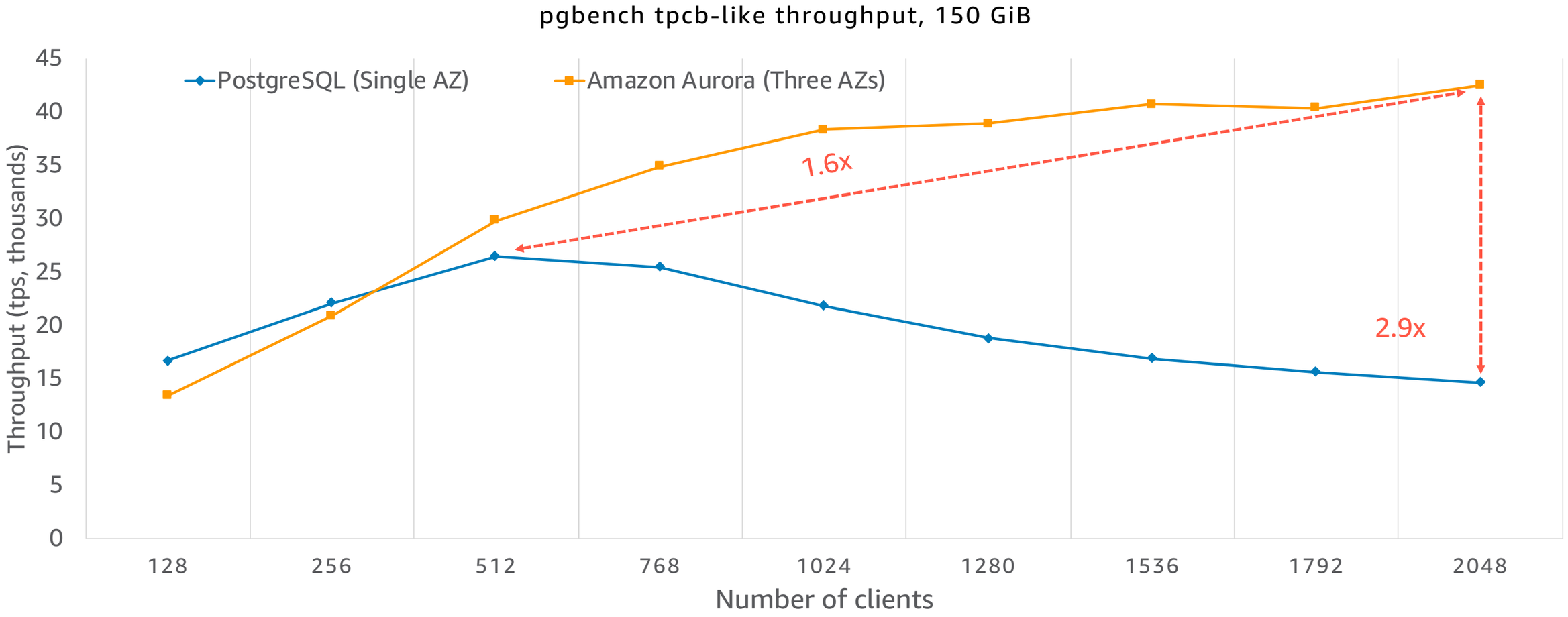
Sysbench: 250テーブル、200,000行/テーブル (R4.16XL)

https://d1.awsstatic.com/product-marketing/Aurora/RDS_Aurora_Performance_Assessment_Benchmarking_v1-2.pdf

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Aurora PostgreSQL: 3倍のスループット

標準の pgbench ベンチマークを実行すると、Amazon Aurora は PostgreSQL の1.6倍のピークスループットと、多数のクライアント接続環境で2.9倍のスループットを実現



https://d1.awsstatic.com/product-marketing/Aurora/RDS_Aurora_PostgreSQL_Performance_Assessment_Benchmarking_V1-0.pdf

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

高負荷な状況でのパフォーマンスの安定性

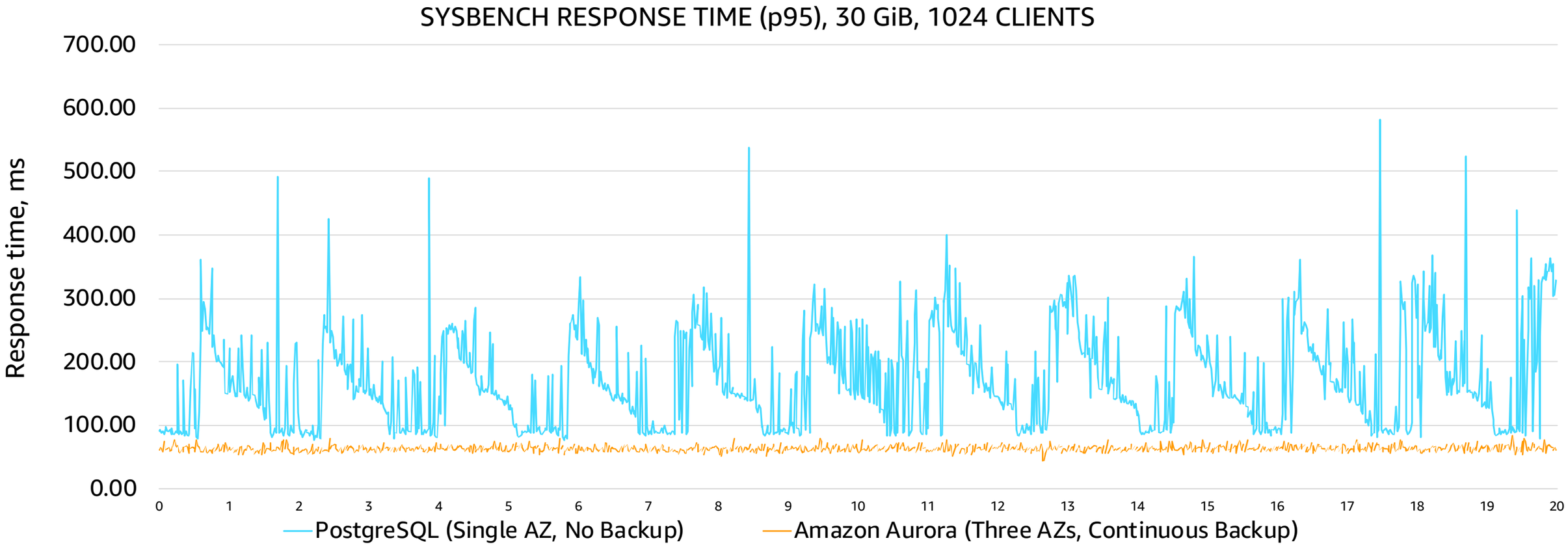
Aurora MySQL は MySQL より最大25倍安定している

Sysbench write-only workload with 250 tables and 200,000 initial rows per table

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

高負荷な状況でのパフォーマンスの安定性

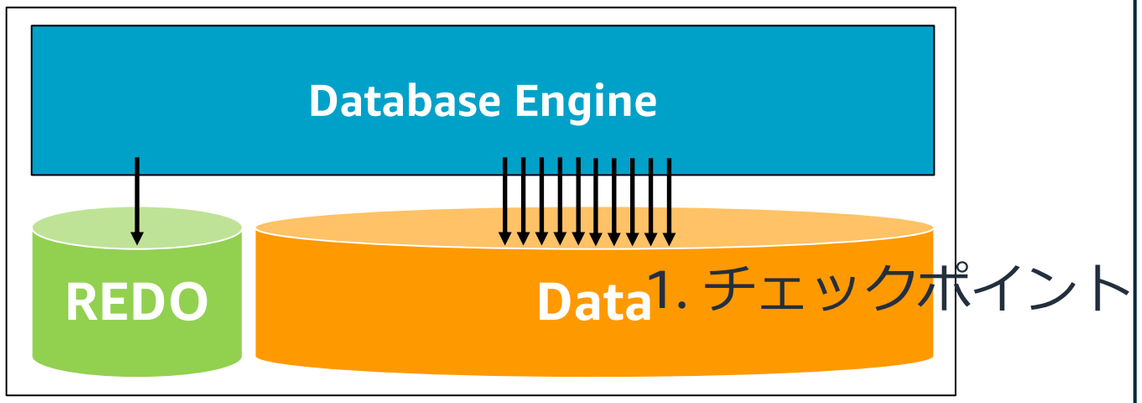
Aurora PostgreSQL は PostgreSQL より最大10倍安定している



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

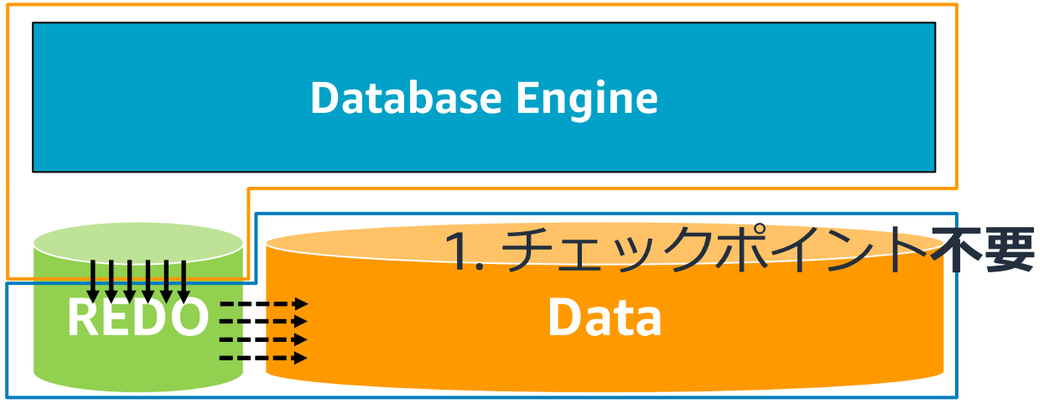
Aurora が高いパフォーマンスを発揮する理由

Traditional Database



2. シーケンシャルな REDO ログ書き込み

Amazon Aurora



2. 非同期、非順序での REDO ログ書き込み

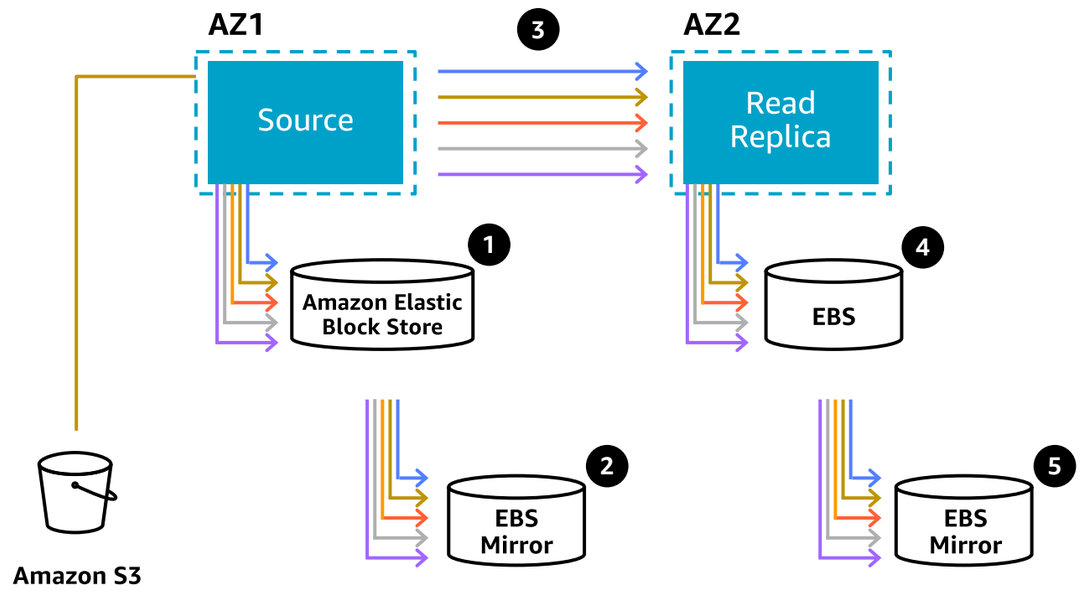
3. クォーラムによる分散

1. チェックポイントが存在しない (I/O 削減によりパフォーマンスの安定性が高い)
2. 非同期、非順序でのログ書き込み (ログ書き込みのスループットの向上)
3. Heavy weight な分散合意アルゴリズムが不要 (クォーラムによる Durability Tracking)

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

MySQL/PostgreSQL vs. Aurora I/O profile

MySQL with replica

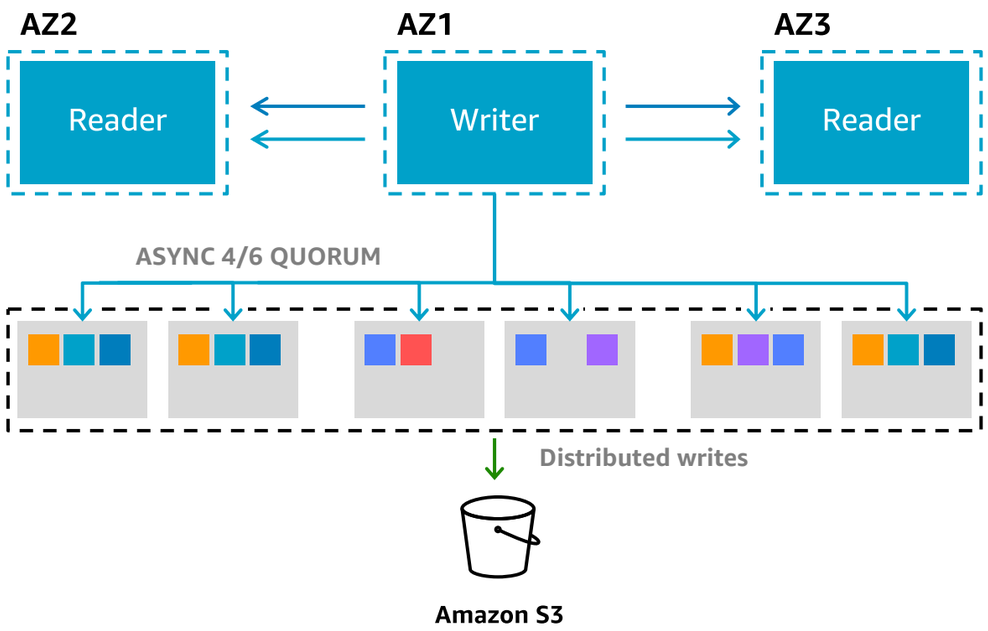


MySQL I/O profile for 30 min Sysbench run

0.78MM transactions

7.4 I/Os per transaction

Amazon Aurora



Aurora IO profile for 30 min Sysbench run

27MM transactions **35X More**

0.95 I/Os per transaction **7.7X Less**

Type of write

Log · Binlog · Data · Double-write · FRM files

(参考) クォーラムにおける読み込みを回避

各データブロックに対して、
クォーラムグループ内の少なくとも
4つのノードが最新のデータを持つ

これら4つのノードのいずれかから
読み取ると、最新のデータが返される

ほとんどのクォーラムベースの
システムでは**読み込みにコスト**がかかる

Aurora は**どのノードが最新で**
レイテンシが少ないかの情報を持っている

読み込みクォーラムは、
修理やクラッシュリカバリに必要

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

使いやすい機能

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

自動化されたストレージ管理

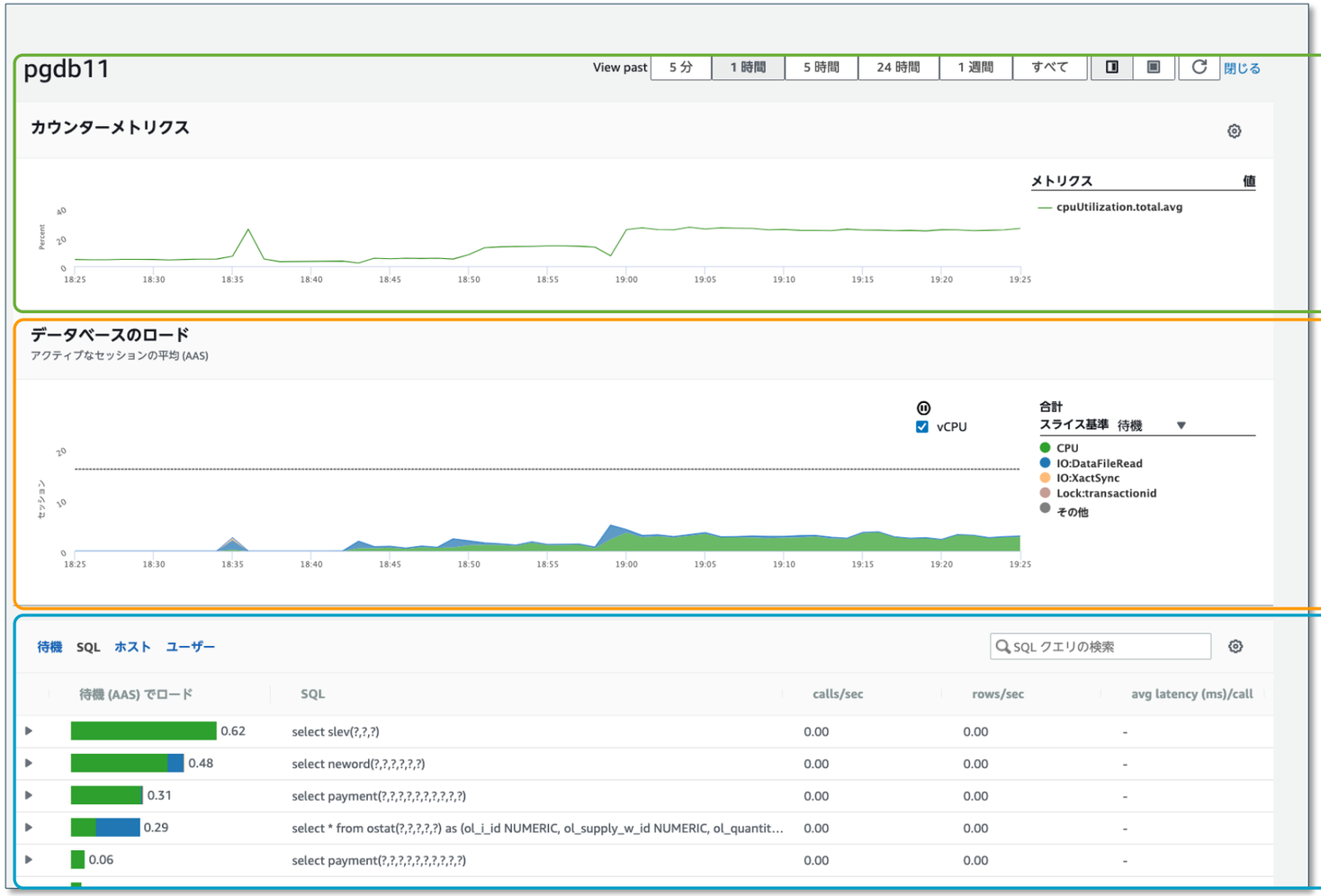
最大 64TB のストレージ – 10GB 単位で自動拡張

- 最大 64 TB の自動ストレージスケーリング - パフォーマンスへの影響なし
- Amazon S3 への継続的な増分バックアップ
- ユーザーのスナップショットを即座に作成 - パフォーマンスへの影響なし
- 自動再ストライピング、ミラー修復、ホットスポット管理、暗号化

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Performance Insights によるワークロード監視

データベース内の性能情報を蓄積してボトルネックを特定(過去データは最大2年保持可能)



カウンターメトリクス

- OSのリソース情報(CPU、Memoryなど)
- DBの統計情報(セッション数、トランザクション数など)

データベースのロード

- 平均アクティブセッション数(AAS)
- CPU時間と待機イベント内訳
- RDS/Aurora全てのエンジンをサポート

ボトルネックの分析軸

- ボトルネックの原因の待機
- ボトルネックとなっているSQL
- 性能影響の高いホスト、ユーザー

詳細：https://docs.aws.amazon.com/ja_jp/AmazonRDS/latest/UserGuide/USER_PerfInsights.html

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Amazon Aurora への移行オプション

ソースデータベース	ソースの場所	推奨オプション
MySQL, PostgreSQL	RDS	• DB スナップショットから直接移行 • リードレプリカを Aurora に作成
MySQL, PostgreSQL, PERCONA, MariaDB	EC2, オンプレミス	• ダンプファイルやフラットファイルを Aurora へ取り込む • MySQL の場合はバックアップから復元、binlog を用いたレプリケーション
ORACLE, Microsoft SQL Server, mongoDB, cassandra, IBM DB2	RDS, EC2, オンプレミス	• フラットファイルを Aurora へ取り込む • SCT でスキーマ変換後、DMS でデータ移行を実施

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

まとめ

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

まとめ

- Amazon Aurora はクラウド時代に Amazon が再設計した RDBMS サービス
  - MySQL, PostgreSQL との互換性
  - コンピュートとストレージの分離
- Amazon Aurora は、実環境でのパフォーマンスにフォーカス
  - 高いクエリ実行並列度で特に効果を発揮
- ハイパフォーマンス・高可用性を実現するための技術的なチャレンジ
  - Log Structured Storage

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

関連資料

Amazon Auroraドキュメント

https://docs.aws.amazon.com/ja_jp/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html
https://aws-ref.s3.amazonaws.com/aurora/Amazon+Aurora.pdf

Blackbelt Seminar

https://aws.amazon.com/jp/blogs/news/webinar-bb-amazon-aurora-mysql-2019/
https://aws.amazon.com/jp/blogs/news/webinar-bb-amazon-aurora-with-postgresql-compatibility-2019/

Amazon Aurora のストレージ実装に関する Blog

https://aws.amazon.com/jp/blogs/news/amazon-aurora-under-the-hood-quorum-and-correlated-failure/
https://aws.amazon.com/jp/blogs/news/amazon-aurora-under-the-hood-quorum-reads-and-mutating-state/
https://aws.amazon.com/jp/blogs/news/amazon-aurora-under-the-hood-reducing-costs-using-quorum-sets/
https://aws.amazon.com/jp/blogs/news/amazon-aurora-under-the-hood-quorum-membership/

Amazon Aurora の論文 - SIGMOD’16, 17

http://www.allthingsdistributed.com/files/p1041-verbitski.pdf
https://dl.acm.org/citation.cfm?id=3196937&dl=ACM&coll=DL

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Thank you!

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.