

LPI-Japan OSS-DB Gold 認定教材

PostgreSQL

高度技術者育成テキスト

NTT ソフトウェア株式会社 [著]

OSS-DB Exam Gold 出題範囲詳細 (2014年9月時点)

| | |
|-----------------------|--|
| <p>運用管理 (30%)</p> | <ul style="list-style-type: none"> ● データベースサーバ構築 重要度:2 <ul style="list-style-type: none"> ➢ サーバ構築における容量見積もり、およびデータベースセキュリティに関する知識を問う <p>主要な知識範囲:</p> <p>テーブル・インデックス容量見積もり</p> <p>セキュリティ</p> <p>通信経路暗号化(SSL)、データ暗号化、クライアント認証、監査ログ</p> ● 運用管理用コマンド全般 重要度:4 <ul style="list-style-type: none"> ➢ データベースの運用管理に関する高度な知識を問う <p>主要な知識範囲:</p> <p>バックアップ、PITR、VACUUM、ANALYZE、REINDEX、自動バキューム、サーバログ管理、ディスク容量監視</p> <p>重要な用語、コマンド、パラメータなど:</p> <p>pg_dump、pg_dumpall、pg_basebackup、PITR、WAL、postgresql.conf、recovery.conf、VACUUM、ANALYZE、vacuumdb、REINDEX、pgstattuple</p> ● データベースの構造 重要度:2 <ul style="list-style-type: none"> ➢ データベースの物理的な構造に関する知識を問う <p>主要な知識範囲:</p> <p>データベースクラスタの構造、プロセス構造、データの格納方法</p> ● ホット・スタンバイ運用 重要度:1 <ul style="list-style-type: none"> ➢ レプリケーションの運用に関する基本的な知識を問う <p>主要な知識範囲:</p> <p>レプリケーション機能の概要、postgresql.conf、recovery.conf の設定</p> |
| <p>性能監視 (30%)</p> | <ul style="list-style-type: none"> ● アクセス統計情報 重要度:3 <ul style="list-style-type: none"> ➢ データベースの利用状況についての統計情報に関する知識を問う <p>主要な知識範囲:</p> <p>以下の関連パラメータ</p> <p>pg_locks、pg_stat_activity、pg_stat_database、pg_stat_all_tables 等の行レベル統計情報、pg_statio_all_tables 等のブロックレベル統計情報</p> ● テーブル / カラム統計情報 重要度:2 <ul style="list-style-type: none"> ➢ プランナが利用するテーブル・カラムの統計情報についての理解を問う <p>主要な知識範囲:</p> <p>pg_class、pg_stats</p> |

| | |
|-------------------------------------|---|
| | <ul style="list-style-type: none"> ● クエリ実行計画 重要度:3 ➢ EXPLAIN が出力する実行計画を読み取り、チューニングを行う 主要な知識範囲: EXPLAIN / EXPLAIN ANALYZE 出力、計画型、EXPLAIN からのチューニング ● その他の性能監視 重要度:1 ➢ 性能監視に関するその他の手法 主要な知識範囲: スロークエリの検出 付属ツールによる解析 |
| <p>パフォーマンス チューニング (20%)</p> | <ul style="list-style-type: none"> ● 性能に関するパラメータ 重要度:4 ➢ データベースの設定パラメータで、特にパフォーマンスに影響を与えるもの、パフォーマンスチューニングの参考になるものに関する理解を問う 主要な知識範囲: 資源の消費 (RESOURCE USAGE)、ログ先行書き込み (WRITE AHEAD LOG)、 問い合わせ計画 (QUERY TUNING)、実行時統計情報 (RUNTIME STATISTICS)、 ロック管理 (LOCK MANAGEMENT) ● チューニングの実施 重要度:2 ➢ データベース、および SQL のチューニングに関する理解を問う 主要な知識範囲: パラメータのチューニング、実行計画のチューニング、 SQL のチューニング、ディスク I/O の分散 |
| <p>障害対応 (20%)</p> | <ul style="list-style-type: none"> ● 起こりうる障害のパターン 重要度:3 ➢ データベースの障害発生時の原因と対策について、適切な判断ができるかを問う 主要な知識範囲: サーバダウン、動作不良、データ消失への対処、OS リソース枯渇 ● 破損クラスタ復旧 重要度:2 ➢ データファイルやログファイルが破損した場合について、適切な対応ができるかを問う 主要な知識範囲: トランザクションログ復旧、システムテーブルのインデックス復旧、開発者向けオプション ● ホット・スタンバイ復旧 重要度:1 主要な知識範囲: レプリケーションログファイル内のエラーメッセージ |

1章 学習の前に 1

学習に便利なツールなど

| | |
|--------------------|---|
| generate_series関数 | 2 |
| psql ¥watchメタコマンド | 3 |
| pg_settingsシステムビュー | 4 |
| iotopコマンド | 5 |
| inotify-tools | 6 |

2章 基本整理 7

データ構造

| | |
|-----------------------|----|
| プロセス構造 | 9 |
| 各プロセスの役割 | 10 |
| PostgreSQLが利用する主なメモリ | 13 |
| データベースクラスタの構造(PGDATA) | 15 |
| テーブルファイル | 19 |
| インデックスファイル | 20 |
| WALファイル | 21 |
| テーブル空間 | 22 |

contribモジュール

| | |
|---------------|----|
| pgbenchとは | 24 |
| oid2nameとは | 30 |
| pgstattupleとは | 35 |

3章 運用管理

41

容量監視

| | |
|------------------|----|
| ディスク容量監視 | 42 |
| テーブル・インデックス容量見積り | 47 |

セキュリティ

| | |
|---------------------|----|
| PostgreSQLのセキュリティ機能 | 53 |
| データ暗号化 | 54 |
| pgcrypto | 55 |
| 通信経路暗号化 | 60 |
| クライアント認証 | 63 |
| 監査ログ | 64 |

バックアップ/リストア

| | |
|----------------|----|
| バックアップ種類/方式の特徴 | 65 |
| pg_basebackup | 69 |
| PITRによるリカバリ | 72 |

メンテナンス

| | |
|-------------------|----|
| VACUUM | 78 |
| トランザクションID周回問題 | 80 |
| vacuumdbコマンド | 83 |
| 遅延VACUUM | 84 |
| 自動バキューム | 87 |
| ANALYZE | 92 |
| 自動バキュームによるANALYZE | 93 |
| REINDEX | 97 |

サーバログ

| | |
|------------------|-----|
| サーバログをどこに出力するか | 102 |
| サーバログをいつ出力するか | 103 |
| サーバログに何を出力するか | 105 |
| サーバログをどのように保持するか | 107 |

レプリケーションとホットスタンバイ

| | |
|--------------------------|-----|
| レプリケーションの目的 | 109 |
| PostgreSQLで利用できるレプリケーション | 110 |
| ストリーミング・レプリケーションの設定 | 111 |
| ホットスタンバイの設定 | 117 |

アクセス統計情報

| | |
|---|-----|
| 関連パラメータ | 124 |
| pg_stat_activity | 125 |
| pg_stat_database | 126 |
| pg_stat_bgwriter | 127 |
| pg_stat_all_tables | 128 |
| pg_statio_all_tables | 130 |
| pg_stat_all_indexes/pg_statio_all_indexes | 131 |

テーブル/カラム統計情報

| | |
|-----------------------|-----|
| pg_class/pg_statistic | 133 |
| pg_stats | 134 |

実行計画

| | |
|--------------------------------|-----|
| EXPLAINおよびEXPLAIN ANALYZEの実行結果 | 139 |
| プランタイプと始動コストの有無 | 141 |
| シーケンシャルスキャン | 142 |
| インデックススキャン | 143 |
| ビットマップスキャン | 144 |
| 並び替え | 145 |
| ネステッドループ結合 | 146 |
| マージジョイン | 147 |
| ハッシュジョイン | 148 |

実行計画のチューニング

| | |
|----------------------|-----|
| INDEXチューニング | 150 |
| ANALYZE | 151 |
| work_mem | 152 |
| effective_cache_size | 153 |
| プランナコスト定数の変更 | 154 |
| プランナメソッド設定の変更 | 155 |
| テーブル結合最適化処理の実行制御 | 156 |

EXPLAINからのチューニング例

| | |
|---------|-----|
| 問題個所の特定 | 157 |
| 原因調査と対応 | 159 |

I/Oチューニング

| | |
|-------------------------------------|-----|
| wal_buffersの値を増加させる | 163 |
| synchronous_commit/wal_writer_delay | 164 |
| commit_delay/commit_siblings | 165 |
| fsync/full_pages_writes | 166 |
| wal_sync_method | 167 |
| チェックポイントチューニング | 168 |
| バックグラウンドライターチューニング | 169 |
| effective_io_concurrency | 170 |
| ディスクI/Oの分散 | 171 |

メモリのチューニング

| | |
|----------------------|-----|
| shared_buffers | 172 |
| maintenance_work_mem | 173 |

ロックのチューニング

| | |
|------------------|-----|
| pg_locks | 174 |
| deadlock_timeout | 175 |

スロークエリの検出

| | |
|---|-----|
| log_min_duration_statement/auto_explain | 176 |
| pg_stat_statements | 177 |

関連システムビューの一覧

178

5章 障害対応 187

障害対処の基本 188

サーバダウン障害への対応

| | |
|----------------|-----|
| サーバダウン時の主な事象 | 189 |
| 動作不良時の主な事象 | 191 |
| OSリソース枯渇時の主な事象 | 193 |

データ消失と復旧

| | |
|-------------------|-----|
| データ消失への対処 | 195 |
| トランザクションログ復旧 | 197 |
| pg_resetxlog | 198 |
| システムテーブルのインデックス復旧 | 200 |
| レプリケーション復旧 | 204 |
| ホットスタンバイ復旧 | 206 |

障害対応向けオプション

| | |
|-----------------------------|-----|
| zero_damaged_pages | 209 |
| zero_damaged_pagesを用いたデータ復旧 | 210 |

OSS-DB Exam Gold 模擬問題集

3章 運用管理

本章では、運用管理について説明をします。

運用管理は出題率が30%と高めであるので試験対策としても重要ですが、実際の運用に際しても知っておくべき項目が多数含まれています。

このため、直近での受験を考えてない方も、しっかりと理解してほしいと思います。

運用管理では、以下の点を解説していきます。

- ・ 容量監視
- ・ セキュリティ
- ・ バックアップ/リストア
- ・ メンテナンス
- ・ サーバログ
- ・ ストリーミングレプリケーションとホットスタンバイ

容量監視

- ディスク容量監視
 - PostgreSQLが正常に動作していることを確認する観点のひとつとしてディスク容量の監視がある
 - 監視すべき主な領域には、以下がある
 - ・ データベース領域
 - ・ WAL領域
 - ・ アーカイブWAL領域
 - それぞれの領域に、どのようなファイルがどの程度作成されるのかを見積もった値をベースとし、運用中の実測値を監視する

まずは容量監視についてです。

データベースにとって、ディスク容量を見積もったり、使用状況を把握することはとても大切です。

見積もりを誤ったり、状況を見誤るとディスク容量がいっぱいになり、データベース停止などサービスに影響を与える状況になることもありますので、正確な知識を身につけてください。

主に監視すべき領域としては、以下があります。

- ・ データベース領域
- ・ WAL領域
- ・ アーカイブWAL領域

それぞれの領域にどのようなファイルが作成されるのか、どの程度の量のファイルが作成されるのかといった情報をもとに見積もりを行い、運用中の実測値との間に大きな乖離がないか監視します。

なお、PostgreSQLのバックアップを取得する領域やサーバログを書き出す領域も監視が必要な領域ですが、ここではその点には触れず後程説明します。

4章 性能監視と パフォーマンスチューニング

本章では、性能監視とパフォーマンスチューニングについて説明します。

安定したサービスを提供するためには、システムが安定した性能を維持できているかを監視するとともに、問題を予測・検出し、その原因を解析できる体制を整えておく必要があります。

PostgreSQLには、問題を検出するための「スロークエリの検出」機能や、原因を解析するために必要な情報となる「アクセス統計情報」等を収集する機能が実装されています。

本章の末尾では、関連するシステムビュー等の実際のクエリ結果と各列の説明の表をまとめて記載しております。

適時ご参照しながら学習を進めてください。

アクセス統計情報

• 概要

- 統計情報コレクタによってデータベースの活動状況が収集される
 - ・ ANALYZEによって収集される統計情報とは別物
- 標準統計情報ビューや統計情報アクセス関数にて収集された情報を参照可能
- デフォルトで有効設定
 - ・ 自動バキュームを有効にするために統計情報が必要
- 公式マニュアルには実行時統計情報と記載

アクセス統計情報は稼働統計情報や実行時統計情報とも呼ばれ、サーバ全体の活動状況の統計情報が収集されます。

アクセス統計情報は統計情報コレクタ (stats collector プロセス) によって自動的に収集されます。また、アクセス統計情報はANALYZE等によって収集される統計情報とは別物であることに注意が必要です。

アクセス統計情報を参照するためには定義済みのビューや関数を利用します。

アクセス統計情報は、常に最新の情報を返すわけではないことに注意する必要があります。個別のサーバプロセスは、待機状態になる直前に統計情報を統計情報コレクタに送信するため、実行中の問い合わせやトランザクションはその時点でのアクセス統計情報には反映されません。

また、統計情報コレクタ自体もおおよそPGSTAT_STAT_INTERVAL (サーバ構築時に変更しない限り500) ミリ秒に一度新しい報告を出力するため、最新の情報よりやや遅れた情報が出力されることとなります。

アクセス統計情報は、同一トランザクションからの参照時において、常に同じ結果を返します。その時点で最新のアクセス統計情報を再取得するためには、新しいトランザクションを生成するか、pg_stat_clear_snapshot関数を呼び出す必要があります。

OSS-DB Exam Gold

模擬問題集

学習にあたって

OSS-DB Exam Gold 対策向けの学習にあたっては、なるべく苦手分野が出ないように、出題範囲を万遍なく学習することを推奨いたします。

LPI-Japan 様の公式サイトでは、OSS-DB Exam Silver は、出題数は 50 問で合格点は 64 点であると公開されています。この点数がパーセンテージを意味しているのであれば、50 問の 64% は 32 問ですので、18 問までは間違えても良いことになります。

対して OSS-DB Exam Gold の場合は、出題数は 30 問で合格点は 70 点であると公開されています。30 問の 70% は 21 問ですので、9 問までしか間違えられないことになります。Silver の半分です。

また、同公式サイトでは、OSS-DB Exam Gold 試験の出題範囲と合わせて各分野の重要度が公開されていますが、「重要度の低い分野の項目は、出題率が必ずしも低いわけではない」と筆者は考えます。

例えば、重要度が最大の 4 である『運用管理用コマンド全般』では、主要な知識範囲や重要な用語として 20 前後もの項目が挙げられています。対して、重要度が最低の 1 である『ホット・スタンバイ復旧』では『レプリケーションログファイル内のエラーメッセージ』のみが主要な知識範囲として挙げられています。ちなみに、重要度を全て合計すると、Silver は 50、Gold は 30 となり、それぞれ出題数と一致します。

重要度の明確な定義は公開されておられませんので断言や保証は致しかねますが、仮に、重要度がその分野からの出題数と一致するのであれば、『レプリケーションログファイル内のエラーメッセージ』について出題される確率は相当高い、という考え方もできるのではないかと思います。

本書の冒頭でも述べましたが、本試験は少し学習すれば誰でも受かるようなレベルの試験ではありません。9 問までは間違えて良いのだから重要度の低い分野は捨てる……などということなく、全問正解する心持ちで試験に臨んで頂ければと思います。

パフォーマンスチューニング

01 シーケンシャルスキャンのみ行われていた選択クエリ高速化のためにインデックスを付与したが、その後のクエリが逆に遅くなってしまった。推測できる原因として**適切なものを2つ**選びなさい

1. random_page_cost の値が適切な値より小さく設定されているため
2. maintenance_work_mem の値が適切な値より小さく設定されているため
3. effective_cache_size の値が適切な値より小さく設定されているため
4. seq_page_cost の値が適切な値より小さく設定されているため
5. 統計情報の精度が低く、適切な実行計画が作成されていないため

02 CLUSTER コマンドに関する説明として**適切ではないものを2つ**選びなさい

1. CLUSTER コマンドの実行後は、該当のインデックスを利用した検索クエリの性能が向上する可能性がある
2. CLUSTER コマンドの実行後は、該当のテーブル更新時にインデックス情報に基づいた処理が行われるため、更新性能が低下する可能性がある
3. 単一のテーブルに対して複数のインデックスが存在する場合、それぞれのインデックスに対して同時に最適化させるための CLUSTER コマンドを発行することはできない
4. CLUSTER コマンドは ACCESS EXCLUSIVE ロックを獲得するため、クラスタ化が終わるまで、そのテーブルに対する読み込みが待機される
5. correlation の値が負数の場合は相関率が低いいため、CLUSTER コマンドの実行を検討すべきである

03 ロックに関する以下の記述で**適切ではないものを2つ**選びなさい

1. pg_locks ビューの current_query 列には、ロックを保持しているクエリが表示される
2. pg_locks ビューには行ロックを保持しているトランザクションは表示されない
3. deadlock_timeout の値を大きくすると、デッドロックが検出されるまでの時間が長くなる
4. deadlock_timeout の値を大きくすると、デッドロックの誤検知の発生確率が高くなる
5. トランザクションが並列実行される状態において、パラメータチューニングによりデッドロックの発生を防止する方法はない

パフォーマンスチューニング

01 正解 1, 5

インデックスを付与したことによりクエリが遅くなったということは、「オプティマイザはインデックスを利用した方がクエリが高速化すると判断したが、その判断に誤りがあった」ということになります。

選択肢 1 は、「random_page_cost の値が適切な値より小さい」場合は、「オプティマイザは実際よりもインデックスアクセスが高速である」と判断することになるため、適切な内容です。

(実際はインデックスアクセスは予想より遅かったのだから、random_page_cost の適切な値は seq_page_cost に対してもっと大きな値であるべきと推測されます)

選択肢 2 は、maintenance_work_mem は本クエリのコスト計算とは無関係であるため、適切な内容ではありません。

選択肢 3 は、「effective_cache_size の値が適切な値より小さい」場合は、「オプティマイザは実際よりもインデックスアクセスが低速である」と判断されるため、適切な内容ではありません。

(実際はインデックスアクセスは予想より遅かったのだから、effective_cache_size の適切な値はもっと小さな値であるべきと推測されます)

選択肢 4 は、「seq_page_cost の値が適切な値より小さい」場合は、「オプティマイザは実際よりもシーケンシャルアクセスが高速である」と判断することになるため、適切な内容ではありません。

(実際はインデックスアクセスは予想より遅かったのだから、seq_page_cost の適切な値は random_page_cost に対してもっと小さな値であるべきと推測されます)

選択肢 5 は、統計情報の精度が低い場合は実際のデータに適していない実行計画が作成される可能性があるため、適切な内容です。

02 正解 2, 5

選択肢 2 は、事前にどのインデックスに対して CLUSTER コマンドが実行されたかはテーブル更新時には判断されないため、適切な内容ではありません。

選択肢 5 は、correlation の値は絶対値が 1 に近いほど相関率が高くなるため、負数が整数より相関率が低いわけではないので、適切な内容ではありません。

他の選択肢は適切な内容です。