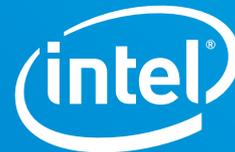


## ソリューション概要

ほぼリアルタイムのデータ・ウェアハウス  
Swarm64で高速化したデータ分析とインテル® FPGA



# Swarm64のS64DA\*と インテル® FPGAが実現する 革新的な高速データ分析ソリューション

## ほぼリアルタイムの大規模なワークロードに対するデータ処理と分析の高速化

このソリューション概要では、革新的なテクノロジーへの投資を通じて、研究やビジネス上の課題を解決する方法について紹介します。組織の中で担当される役割に合わせて、次のような内容が把握できます。

### ・ ビジネス戦略:

データ・ウェアハウスや分析にS64DA\*とインテル® FPGAを使用して、十分な研究やビジネス上の成果を達成する方法について、理解を深めることができます。

### ・ テクノロジーの決定:

S64DA\*とインテル® FPGAの使用が、大規模なデータセットの保管や分析を行うさまざまな組織や業界に対して、いかに価値をもたらすかを理解できます。

## 概要

高可用性の提供やストレージの管理から、ほぼリアルタイムの分析の実施まで、データのITに対する需要は増え続けています。リレーショナル・データベースとSQLは、引き続きエンタープライズ・クラスのデータ分析にとって中枢要素です。Swarm64は、ほとんどの一般的なデータベースやストレージのアプリケーションと連携するPostgreSQL\*に革新的なアドオンを提供して、大量かつ高速のデータ処理を可能にし、新しいITシステムの導入に特有なリスクとコストを解消します。最も重要なのは、S64DA\*が、要求の厳しいワークロードのデータ処理および分析を大幅に高速化するよう設計されていることです。S64DA\*は、最新世代のインテル® FPGAを使用して、分析ユースケースに適した業界トップクラスのパフォーマンスを提供します。

数十億もの行や1秒当たり数百万もの挿入まで拡張できるS64DA\*によって、大量の構造化/非構造化データを実用的なインサイトに変えることができます。S64DA\*では、ネットワーク接続の監視、多数のインテリジェント・デバイスからのセンサー出力の収集、顧客の行動データまたはデータセンターやマシンパークからのテレメトリー・データの取り込みなど、幅広いIoT機能がサポートされています。Swarm64のほぼリアルタイムの大規模データセットに対する高速分析によって、脅威の検出、顧客との対話、予測メンテナンスなどのユースケースにおいて迅速な対応が可能になります。

このソリューションは、コスト効率に優れた使いやすいソリューションで、PostgreSQL\*、MariaDB\*、MySQL\*などほとんどの一般的なデータベースにシームレスにアドオンできます。

Swarm64の革新的な方法は、データベースのデータ構造と処理中のデータフローの最適化を基にして、クエリー中のデータ確認を少なくし、処理を必要とするデータをシステム内で高速に移動します。

インテル® FPGAは、I/Oスループットを向上し、CPUが処理する必要のあるデータ量と演算数を削減します。インテル® FPGAの処理と独自のデータ構造がソフトウェア・レイヤーによってラッピングされたアドオンになっており、このアドオンが標準インターフェイス経由でサポート対象のデータベースとリンクします。

Swarm64ソリューションでは、CPUとインテル® FPGAの間のシームレスな連携を実現し、ネットワークからアクセスするストレージや一般的なクラウド・インフラストラクチャーのストレージにおける、レイテンシーの増加と帯域幅の制限の問題を解消しています。これによって、ストレージをコンピューティングから切り離すことが可能となり、柔軟なリソース運用と優れたコスト・パフォーマンスを実現できます。



## ソリューションの利点: データ分析の高速化

- 高速で発生するデータに対する複雑なクエリーをほぼリアルタイムで実現
- 標準的なデータ・ウェアハウス・ワークロードを2倍以上高速化<sup>1</sup>
- データ・ウェアハウス・ワークロードを運用サーバーと同じサーバー上で実行
- 幅広い分析ユースケースに対して、消費電力対性能比と価格対性能比の最適化を実現
- PostgreSQL\*, MariaDB\*, MySQL\*でほぼリアルタイムの分析を実現

## ビジネス課題: 時間とコストのかかるデータ・ウェアハウス

金融や医療から小売、製造まで、さまざまな組織や業界が、大規模なデータセットに対して複雑な分析をほぼリアルタイムで実行する必要に迫られています。しかし、現在のリレーショナル・データベースは、大量かつ高速なデータをほぼリアルタイムという制限の下で処理するには適していません。リレーショナル・データベースの機能は、適切なデータセットに対しては高いパフォーマンスを発揮しますが、センサーからの入力や顧客トラッキングなどの高速なデータや何十億もの行があるテーブルに対してはすぐに限界に達してしまいます。

さらに、分析専用のサーバーを運用サーバーに追加することで、運用コストも増加します。一部のColumnStoreデータベースでは、すべてのSQLクエリーをサポートしていない場合もあります。

## ユースケース

- 高速のリアルタイム・データで複雑なデータ分析を実行する
- 分析と運用データベースを同一サーバー上で実行する
- データ・ウェアハウス・データベースを2倍以上高速に実行する<sup>1</sup>

## FPGAとは

FPGA (Field Programmable Gate Array) は、デバイス内部の機能の大半を変更できるように設計された半導体集積回路(IC)です。こうした変更は、設計エンジニアが行うことも、プリント基板(PCB)のアセンブリー・プロセス中に行うことも、さらには機器が顧客に出荷された後、フィールド(現場)で行うこともできます。こうした柔軟性や、アルゴリズムの処理を高速化する能力によって、FPGAは、トランザクションなどに必要な複雑かつ大規模な分析やほぼリアルタイムの分析に極めて重要な存在となっています。

### インテル® FPGA搭載ソリューションでメリットを得られる対象

- ほぼリアルタイムの大規模なデータセットを分析する組織や業界
- OEMおよびソリューション・プロバイダー
- クラウド・サービス・プロバイダー

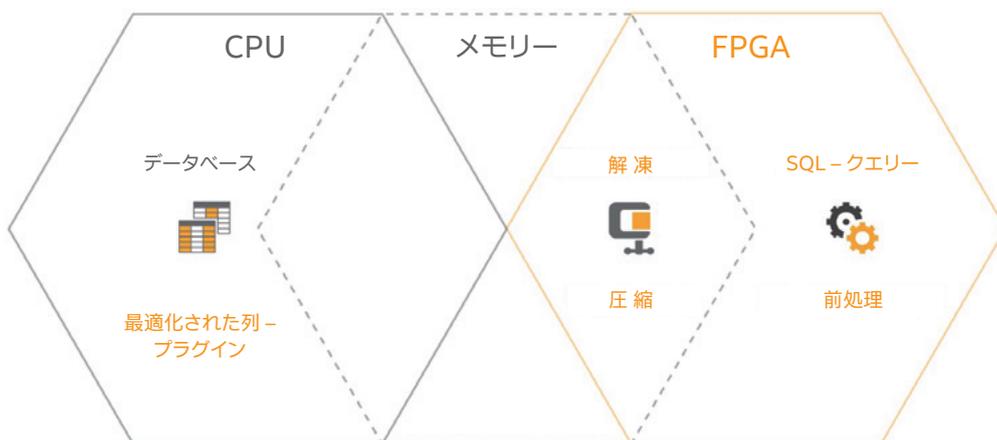
[詳細情報](#)

## ソリューションの価値: インサイトの加速

S64DA\*は、一般的なオープンソースのリレーショナル・データベースへのアドオンであり、透過的にデータ分析処理のパフォーマンスを向上します。処理のパイプラインの中心には、インテル® PCIe FPGAカードを使用します。S64DA\*によって、PostgreSQL\*単独で達成できるデータ・スループットよりも高いスループットを実現し、分析の実行時間を高速化できます。

### プロセスや資産をリアルタイムで制御

- 収集したデータを使用して、製品に付加価値を付けられます。
- ほぼ瞬時に脅威を検出し、阻止できます。
- イノベーションを加速できます。
- 既存のITやスキルを使用してビジネス目標を達成できます。



S64DA\*とインテル® FPGAの統合によって、SQLデータベース内のローデータを変換し、利用しやすい高速の分析を実現

**ほぼリアルタイムで顧客のニーズを検出 / 対応(企業、金融取引からeコマースまで)**

- ニーズの把握を向上し、ビジネスの成長に必要な情報を獲得できます。
- より多くのデータポイントを使用し、社内でも、より深い分析を低コストで実施できます。

**スマートグリッド、データ・ネットワーク、通信においてトラフィック分析から価値を創造**

- 高速に発生するデータストリームについて深い分析を実施できます。
- ほぼリアルタイムに対応することで、低コストで1資産当たりの売上を向上できます。

**ソリューション・アーキテクチャー : S64DA\***

S64DA\*は、データ・ウェアハウス・パイプラインの主要な要素に対処することによって、データの処理、取得、分析を高速化するよう設計されています。

1. インテル® FPGAを使用して、データを大幅に圧縮します。インテル® FPGAの極めて優れた適応性により、数GBのデータに対する複雑なアルゴリズムをほぼリアルタイムで適用することができます。クエリーエンジンは、関連するデータのみを確実にCPUに返します。
2. 最適化された列の定義により、S64DA\*アドオンは特定のクエリーに必要なデータのみを選ぶことができます。その結果、速度が向上するとともに、ストレージの大幅な低減とネットワーク・コストの削減を実現できます。
3. データベース・クエリーを実行する場合、最初の処理ステップをS64DA\*に「プッシュダウン」することができます。S64DA\*では、これらの処理ステップがストレージレイヤーの近くで実行されます。データのうち関連する部分のみが渡されます。
4. このようにして、メインメモリー内の処理およびキャッシュの要件が大幅に削減され、それによってクエリーのパフォーマンスが向上します。また、価格が低く、消費電力が少ないサーバー上でも、同じレベルのクエリー・パフォーマンスを実現できます。

**結果**

S64DA\*は、最新世代のインテル® FPGAを使用して、データ・ウェアハウスや分析のユースケースに最適化したパフォーマンスを提供します。

データの 取り込み速度が 最大 11 倍 <sup>2</sup>	分析 パフォーマンスが 最大 12 倍 <sup>3</sup>	データベースの サイズ削減が 最大 1/5 <sup>4</sup>	行数は 数千億まで 拡張可能
---	--	--	----------------------



インテル® FPGAのアクセラレーションによるS64DA\*アドオンは、既存のデータベース・ソフトウェアとストレージで動作

S64DA\*は、データの取り込みと分析の間、ワークロードをCPUとインテル® FPGAの両方に分割して使用します。FPGAでは、データの圧縮と解凍およびSQLクエリーのフィルタリングとデータの前処理ステップを実行します。CPUでは、クエリーのどの部分をFPGAにオフロードするか計画に集中し、クエリーの残りの部分の実行と、「最適化した列」(S64DA\* OLAPに最適化したインデックスに相当)の維持を行います。

ワークロードをCPUとインテル® FPGAに分割する主な利点は複数あります。インテル® FPGAは、ストレージからデータを取得し、CPUがより効率的に分析を実行できるようにデータを準備して、それをCPUに流します。データをインテル® FPGA上で圧縮することで、スト

**速度と精度の向上**

S64DA\*は、次の条件のうちいずれか1つ(できれば複数)を満たす場合に、最高のパフォーマンスを提供します。

- 高速でのデータ挿入(最大で1秒当たり数百万行)
- 非常に大規模なデータセット内で特定の情報の検索または範囲の集約を行う場合(価格範囲の検索、日付範囲の分析など)
- ほぼリアルタイムという要件(新しい情報をデータベースが受け取ってから、クエリーがそのデータを利用できるまでのタイミンが1秒未満)
- 大規模なデータセットにおける高速インデックス付けとコンパクトなストレージが重要視される場合
- 大規模なデータセットを使用する一般的なデータ・ウェアハウスを行う場合
- すでにPostgreSQL\*, MySQL\*, またはMariaDB\*を使用し、提供されるパフォーマンスに限界を感じている場合
- I/O帯域幅に制限があり、I/Oレイテンシーが高い環境で作業している場合(ローカルストレージのないクラウド・インスタンスなど)

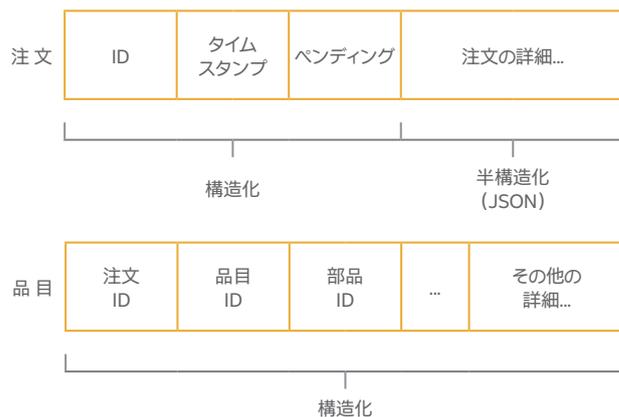
レージへの転送と再ロードが高速化され、ストレージ容量の要件が低減されます。インテル® FPGAによって、データは指定の基準に基づいて確実に圧縮され、提供されるので、関連のあるデータのみがCPUに移動し、それによって全体的な負荷が低減します。CPUは用意されたデータをキャッシュに収まる分量で受け取るので、アルゴリズムの処理が高速化され、場合によっては高速アルゴリズムの選択が可能になります。

CPUとFPGAは連携して、相互のデータ移動を効率的に行います。

つまり、S64DA\*とインテル® FPGAを使用すると、膨大なデータセットをメインメモリーにロードしてその一部を取り出すよりも、処理が大幅に速くなります。また、より適切なキャッシュによりCPUの効率が向上します。圧縮を使用するという事は、一度にメインメモリーに取り込めるデータ量が増えることを意味し、結果として効率が向上します。さらに、CPUに流れるデータの量が削減され、CPUはパフォーマンスの向上により役立つアルゴリズムを選択できるようになります。これらすべての要素が合わさることで、現在のインフラストラクチャーを壊すことなく、卓越した高速化を実現できます。

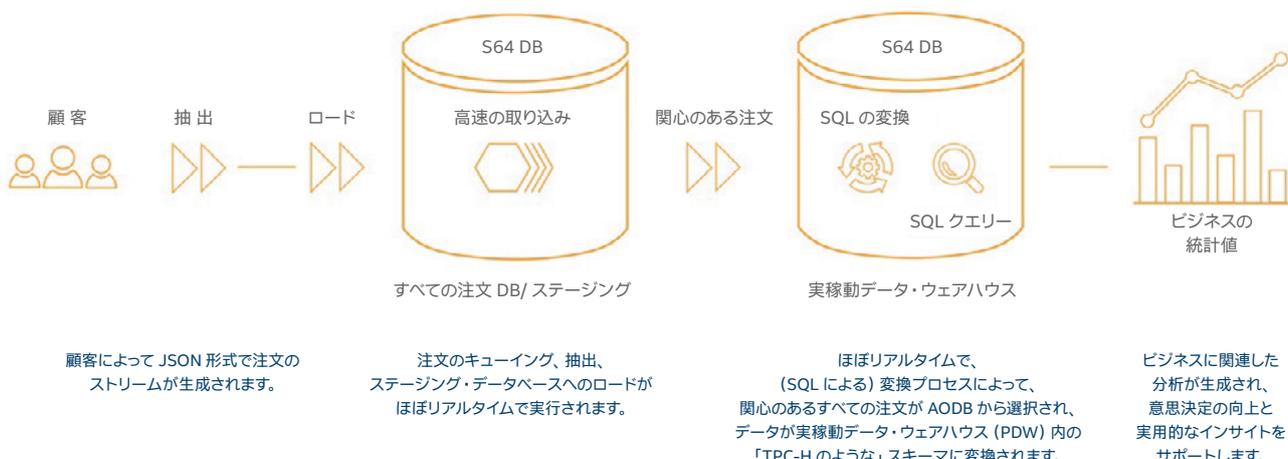
S64DA\*は、PostgreSQL\*とは外部データラッパー経由で、MariaDB\*またはMySQL\*とはストレージ・エンジン・インターフェイス経由で連携します。パフォーマンスの最適化は、インテル® FPGA、インテル® PAC、PAC 仮想化環境 (Open Programmable Acceleration Engine (OPAE)) によって実現されます (インテル® FPGAは、FPGAカードを搭載した対応サーバーによって完全に統合されています)。S64DA\*が、インテル® FPGAを制御して管理を簡素化し、エンドユーザー・アプリケーションに対しては完全にトランスペアレントとなります。

例えば、ほぼリアルタイムのコマース・ユースケースでは、構造化データベースのフィールド内の指定要素 (時刻、注文ID、注文ステータスなど) については、データがPostgreSQL\*経由で一貫した構造に収集され、その他のすべての注文詳細はJSON形式で収集されます。S64DA\*は、このローデータをTPC-Hのようなスキーマに変換して、SQLデータベース処理に備えます。S64DA\*はアドオンなので、SQLデータベースの変更は不要です。



ローデータはSQLデータベース処理向けに収集され準備されます

抽出、変換、ロード (ETL)処理を最新化して、ローデータをSQLデータベース内で変換することによって、Swarm64は圧倒的な優位性を提供します。Swarm64とインテルのソリューションは、画期的なパフォーマンスを提供するだけでなく、維持が容易です。利用しやすい言語を使用しているため、アナリストは必要に応じてスキーマを変更できます。これによって、アナリストはこれまで以上にデータ分析に対する処理能力と制御を獲得し、適切なビジネス指標に素早くアクセスして、意思決定に必要な情報を獲得することが可能になります。



大規模なデータ分析の主なボトルネックに対処して、結果までの時間を短縮するS64DA\*とインテル® FPGA

## ソリューションのコンポーネント

- S64DA\*
- Intel® FPGA
- Intel® Xeon® スケーラブル・プロセッサ

## データ分析向けのソリューションの提供

Intelは、Intel® FPGAが搭載された量産対応 PAC を提供することで、OEM やソリューション・プロバイダーによる統合と検証を簡素化して、市場投入までの期間を短縮できるよう支援しています。

Intel® FPGA は、ニーズの進化に合わせて、アクセラレーションのレートをプログラムおよび再プログラムする柔軟性を提供します。そして、すべてを、固定機能の GPU で達成できる消費電力よりも低い消費電力で実現できます。また、ハード化された浮動小数点 DSP の提供により、複雑なマシンラーニング・アルゴリズムも並列処理を使用して高速化できます。Intel® FPGA と Intel® Xeon® スケーラブル・プロセッサの組み合わせにより、小さなフットプリントで低レイテンシーの実装を提供することが可能になります。

## 結論

データに基づくインサイトは、今日の企業や業界が競争を継続する上で不可欠な要素となっています。Swarm64 と Intel の革新的なソリューションは、データ・ウェアハウスと分析に模範的な効率と柔軟性をもたらし、データ・サイエンティスト、IT 担当者、アナリスト、そして最終的にはエンドユーザーにメリットを提供します。

## 関連情報

Intel® FPGA の詳細については、[こちら](#)を参照してください。

Swarm64 のイノベーションについては、<http://swarm64.com/> (英語) を参照してください。

## OEM とソリューション・プロバイダーに対する FPGA の利点

柔軟性	<ul style="list-style-type: none"> <li>• デバイスの電源を入れるたびに、FPGA の機能を変更することができます。設計エンジニアが変更や試行を行うときは、新しいコンフィグレーション・ファイルをデバイスにダウンロードするだけです。</li> <li>• FPGA に対する変更は頻繁に行うことができ、PC ボードの変更が必要ないため、コストもかかりません。</li> </ul>
市場投入期間の短縮	<ul style="list-style-type: none"> <li>• FPGA デバイスは汎用品として入手できます (一方、ASIC の製造サイクルには数カ月を要します)。</li> <li>• FPGA の柔軟性により、OEM 各社は、その設計が正しく機能し、テストが完了した時点で、直ちにシステムを出荷し、ソリューションを早く市場に投入することができます。</li> <li>• FPGA によって CPU のオフロードと高速化が実現されることで、システム全体の処理速度が向上します。</li> </ul>
集積化	<ul style="list-style-type: none"> <li>• 簡素化されたホワイトボックスの統合。</li> <li>• FPGA 内部の機能が増えるほど、基板上のデバイス数が減ります。そして、デバイス故障の可能性が低減する分、信頼性が向上します。</li> <li>• 現在の FPGA には、オンダイ・プロセッサ、28 Gbps (またはそれ以上) のトランシーバー I/O、RAM ブロック、DSP エンジンなどが含まれています。</li> </ul>
総保有コスト (TCO)	<ul style="list-style-type: none"> <li>• FPGA なら、プロトタイプ・システムを出荷してフィールド試験を行うことができるだけでなく、大量生産に入る前でも、迅速な変更が可能であるため、リスクを低減できます。</li> <li>• 長期のデバイス供給 (15 年以上) をサポートしています。</li> <li>• 基板上の電子デバイスの 1 つが生産終了 (EOL) になっても、OEM 生産機器の再設計や再認定のコストがかかりません。</li> <li>• ASIC は同等の FPGA に比べて単価は低いものの、その構築には、開発費 (NRE)、高価なソフトウェア・ツール、専門の設計チーム、および長い製造サイクルが必要となります。</li> </ul>



<sup>1</sup> ほぼリアルタイムのELTとビジネス分析向けのユースケースのデモ。ワーストケース：PSQLに対して1.8倍の高速化(ベストケースのPSQLとワーストケースのSwarm64のパフォーマンス比較)。システム構成：2基のインテル® Xeon® Gold 6130 プロセッサ(16C/32T、CPU@2.10 GHz)、384 GB(12 x 32 GB)DDR4-2666 ECC RAM、480 GB SSD 4台、120 SSD 2台、インテル® プログラマブル・アクセラレーション・カード(インテル® Arria® 10 GX FPGA搭載版)1枚、デュアルポート10GE NIC 1枚、CentOS® 7.4オペレーティング・システムを搭載した2台のDell® R640サーバー。

<sup>2</sup> テストは次の条件で実施。複数の作業員により25億行を3つのデータベースに取り込む時間を測定。各行は4フィールドから成り、合計サイズが25バイト。取り込んだデータに関する予備情報を収集するテーブル・スキャン・クエリーの時間を測定。4種類のサイズ(極小、小、中、大)の結果セットについてテーブル・スキャン・クエリーの時間を測定。4種類のサイズの結果セットについてインデックス・クエリーの時間を測定。<http://www.altera.co.jp/solutions/industry/computer-and-storage/applications/data-analytics/solutions.html>

<sup>3</sup> 例えば、ドリルダウン・タイプのクエリーの場合、測定に使用したマシンは、次の構成のSupermicro® SuperServer 2028U-TR4+。Super X10DRU-i+メインボード、2基のインテル® Xeon® プロセッサE5-2695 v4、Samsung® 32 GB DDR4-2400 ECC RAM 8個。<http://www.altera.co.jp/solutions/industry/computer-and-storage/applications/data-analytics/solutions.html>

<sup>4</sup> S64DAをPostgreSQL® 9.6.3およびMariaDB® 10.1.21と比較。Swarm64のスタッフメンバーによる比較。外部監査なし。ネットワーク情報データの取り込みとクエリーを同時に行う、Swarm64の「ネットワーク・セキュリティ・ユースケース」の実行。<http://www.altera.co.jp/solutions/industry/computer-and-storage/applications/data-analytics/solutions.html>

性能に関するテストに使用されるソフトウェアとワークロードは、性能がインテル® マイクロプロセッサ用に最適化されていることがあります。SYSmark® や MobileMark® などの性能テストは、特定のコンピューター・システム、コンポーネント、ソフトウェア、操作、機能に基づいて行ったものです。結果はこれらの要因によって異なります。製品の購入を検討される場合は、他の製品と組み合わせた場合の本製品の性能など、ほかの情報や性能テストも参考にして、パフォーマンスを総合的に評価することをお勧めします。詳細については、<http://www.intel.com/benchmarks/> (英語) を参照してください。

インテル® テクノロジーの機能と利点はシステム構成によって異なり、対応するハードウェアやソフトウェア、またはサービスの有効化が必要となる場合があります。実際の性能はシステム構成によって異なります。絶対的なセキュリティを提供できるコンピューター・システムはありません。詳細については、各システムメーカーまたは販売店にお問い合わせいただくか、<http://www.intel.co.jp/> を参照してください。

Intel、インテル、Intel ロゴ、Arria、Xeon は、アメリカ合衆国および/またはその他の国におけるIntel Corporationまたはその子会社の商標です。

\* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

©2018 Intel Corporation. 無断での引用、転載を禁じます。

337425-001JA