

## Amit Kapila を紹介

## 技術者 Blog

Amit Kapila

FUJITSU Limited  
Software Products Business Unit Data Management Division Senior Director  
PostgreSQL Committer and Major Contributor

## はじめに

オープンソース技術の発展が技術革新の推進力となり、近年、データの活用が飛躍的に拡大し、オープンソースの中でもリレーショナルデータベースの重要性が高まっています。リレーショナルデータベースの革新に大きく貢献したものの 1 つが PostgreSQL の成長です。

PostgreSQL は、標準 SQL への準拠と拡張性に重点を置いたオープンソースのオブジェクトリレーショナルデータベース管理システムです。

富士通は PostgreSQL に 18 年以上携わっており、コミュニティでも積極的に活動しています。コミュニティメンバーによる透明性の高い開発とフィードバックの繰り返しにより、PostgreSQL は、金融、政府、教育などで使用される世界で 4 番目に人気のある ODBMS (Source DB Engines) になりました。

コミュニティへの継続的な取り組みをさらに強化するため、PostgreSQL のコミッターである Amit Kapila 氏を富士通に迎えました。Amit 氏は、データベースの中核機能に携わり、20 年間のキャリアを持つデータベースのエキスパートであり、SQL エンジン、ストレージエンジン、レプリケーションについて深い専門知識を持っています。2012 年から PostgreSQL に積極的に従事しており、パラレルクエリ、ハッシュインデックスの持続性、性能測定への貢献が認められています。

## 私と PostgreSQL コミュニティー

最初に、富士通の皆様の暖かい歓迎に感謝申し上げます。入社して数カ月経ちますが、とても楽しく過ごしています。

私のキャリアの中には、いくつかの重要なマイルストーンがあります。NIT Allahabad 大学（インド）でコンピューターサイエンス・エンジニアリングでの功績が認められたことを皮切りに、Oracle 勤務時にデータベースの世界に足を踏み入れたところから、データベースへの興味が広がりました。

おそらく最も重要なマイルストーンは PostgreSQL コミュニティーへ紹介されたことです。ここが私の居る場所だとすぐに思いましたし、私に生産的で魅力的な環境を提供してくれると思いました。これが PostgreSQL におけるメジャーコントリビューターおよびコミッターになるきっかけとなり、個人的な目標を達成するだけでなく、周りの人の PostgreSQL のプロフェッショナルとなる目標を達成する手助けもできました。PostgreSQL コミュニティーでは、Tom Lane 氏や Bruce Momjian 氏のような素晴らしい人々がコミュニティの発展と成長に貢献してくれたことを非常に幸運に思います。また、私は他のコミュニティメンバー、特に Robert Haas 氏、Heikki Linnakangas 氏、Andres Freund 氏、Thomas Munro 氏、Simon Riggs 氏と一緒に働くことで恩恵を受けてきました。PostgreSQL コミュニティーは急速なペースで成長を続けており、この成長に魅力を感じ、多くの組織がオープンソースを採用しようとしています。

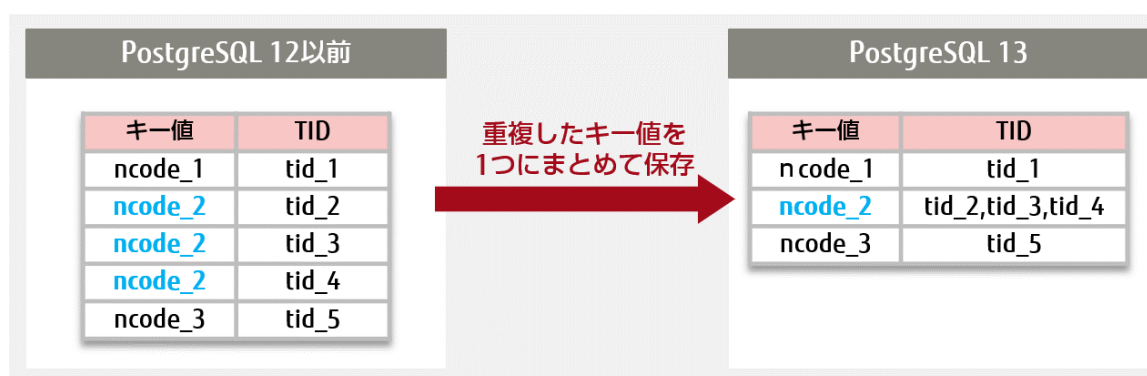
## 富士通と PostgreSQL 開発

富士通に来てまだ日が浅いですが、入社前も入社後も富士通の基本的な認識は変わっていません。富士通は、PostgreSQL コミュニティにおける著名なコントリビューターであり、この貢献が何年も前から続いていることを尊敬します。彼らは小さなバグ修正や中規模の機能だけでなく、最近のプラガブルストレージにおいても大きな貢献をしています。富士通は OSS カンファレンスにも積極的に参加しており、コミュニティの拡大に貢献しています。PostgreSQL の継続的な開発状況から見て、富士通は常に積極的な役割を果たしています。

最近リリースされた PostgreSQL 13 の中で最も有用だと思われる機能をいくつか紹介したいと思います。

### B-Tree インデックスの重複排除

この機能では、重複したキーを 1 つにまとめて保存できます。例えば、多くの人が米国とインドに滞在しているような国別コードをキーとした（同じ国のキー値が複数存在する）表があるとします。この機能を使うと、その複数のキーを 1 つにまとめて保存できます。



これを B-Tree インデックスの重複排除と呼びます。ある特定のケースでは、これによってインデックスサイズが 40% 減少するとの報告もあります。したがって、この機能は、データベース内に多数の一意でないインデックスを持つユーザーまたはワークロードに役立ちます。

### 並列バキューム

私はこの機能の設計、レビューやコードの投稿に直接携わっていました。

この開発によってバキューム速度を最大 3 倍にすることができます。迅速なバキュームが必要な場合、並列バキュームを利用することで非常に速く処理できます。

大規模なトランザクション・ワークロードによってデータベースが肥大化し、データベースが満杯で動かないに近い状態になったら、その肥大化を迅速に取り除く必要があります。

私がこの機能でとても気に入っている点の 1 つは、1 つのバキュームプロセスが使う以上のメモリー量は使わないことと、I/O（基本的にはコストベースのバキュームパラメーターに準拠）が増えないことです。

並列処理に関する過去の経験を考慮して、この機能を慎重に設計しました。

### その他の機能

パーティションまたはハッシュ集約が含まれる特定の領域でのクエリのパフォーマンスが向上しました。

また、統計処理が改善され、ユーザーは WAL データに関してどの程度の I/O が発生しているかを知ることができ、システムのチューニングに役立てることができます。

特に、ロジカルレプリケーション、パラレル処理、シャーディング（スケールアウト）、性能全般、スケーラビリティの改善などの分野では、富士通が PostgreSQL の開発にさらに貢献する可能性は大きいと思います。富士通は PostgreSQL の主要な機能強化に注力することで、より大きな役割を果たすことができると 생각합니다。私はこれを、PostgreSQL が成長を続けるだけでなく、より大規模なアプリケーションで使えるようにするための機会だと考えています。また、これを発展させて、富士通が PostgreSQL

のコントリビューターから成る強力なチームを構築し、コミュニティと富士通の両方がその拠点を拡大し、可能性を最大限に引き出すのを支援していきたいと思います。

## 仕事以外での楽しみ

---

私は富士通の家族の一員であることにワクワクしていますが、仕事以外では、実際の家族や友人と時間を過ごすのが好きなことをお伝えします。暇なときには、リーダーシップに関する伝記や本を読むのが好きです。そしてもちろん、ソフトウェアエンジニアリングに関する資料を読むのも好きです。

引き続き、PostgreSQL についての私の考えを継続的に伝えることを楽しみにしています。

2020 年 10 月 16 日