

限界を 超える!

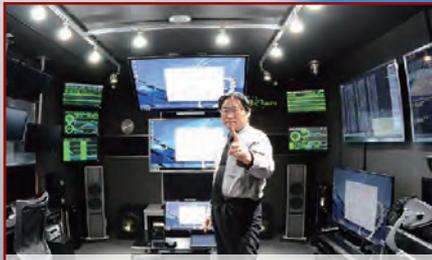
速さを追求した
インメモリデータベース
エンジン開発への挑戦



人物相関図

～Dr.Sum Ver.5.0～

東京本社



執行役員 CTO 島澤 甲

ウイングアークの最高技術責任者。
"Dr.Sum Ver.5.0" を成功させ、ウイングアークを BI分野で世界のトップメーカーにすることを目標に掲げている。常に目標は高く妥協は許さない。

札幌オフィス



開発責任者 笹原 徹

Dr.Sum 開発部の部長で札幌オフィスの責任者。純国産 BI 製品 "Dr.Sum" の初期開発メンバーの一人。もともとは開発エンジニアで、実装が技術思考に偏る傾向がある。それにより顧客志向の島澤と衝突することもしばしば。Dr.Sumのいい面も悪い面も熟知している Dr.Sum の生き字引。

エンジン開発チーム



開発リーダー 橋田 哲尚

Dr.Sum エンジン開発のリーダー。
Dr.Sumのことを思うあまり、ときにキツイ言い方をするが、周りからの信頼は厚い。インメモリエンジンが、大規模データで通用するかどうかの鍵を握る人物である。



開発エンジニア 安達 俊貴

Dr.Sumのインメモリエンジン開発を担当した。三度の飯よりデータベース研究が好きという本物のマニア。普段は口数は少ないが、インメモリエンジンに関しては周りが引くくらい熱弁することも。



開発エンジニア 川代 雄太

開発も兼ねるが、大規模なインメモリエンジンの品質を向上させるため、作業の自動化やインフラ整備に注力する。堅実な仕事ぶりと的確な意見により、若手社員のリーダー的存在である。



開発エンジニア 清水 亨

調査やプロトタイプ開発など根気が必要な作業が多い苦勞人。コンピューターに関する深い知識を持つ、まさに「グル(導師)」と呼ぶにふさわしい人物。



開発エンジニア 木村 聡

安達と並びインメモリエンジン開発の2トップを担う。インメモリエンジンのハイブリッド化や分散処理サーバーの実装を担当した。難易度の高い開発も淡々とこなす職人である。



開発エンジニア 北村 友藝

開発のエンジニア。幅広い開発を担当するユーティリティプレイヤーである。大きな機能から小さな機能まで柔軟に対応し仕事も早い。



開発エンジニア 本田 純子

ユーザーインターフェイスを担当。エンジンのパフォーマンス計測も担当し鋭い切り口の分析は、エンジン開発の窮地を何度も救ってきた。



開発エンジニア 野村 明讓満

若手エンジニア。Ver.5.0では品質エンジニアとして活躍し、テストケース作成量は一番多い。細かい部分に気が回り、数多くの障害を検出してエンジンの品質向上に努めた。



開発エンジニア 三浦 公寿

当初はエンジン開発だったが、担当分を速攻で終わらせて品質エンジニアとして活躍した。SQLに精通しており、その観点のテストはピカイチ。

最速のエンジンを求めて



新生 Dr.Sum への 挑戦は こうして始まった

ウイングアークは2017年10月5日
6年ぶりのメジャーバージョンアップとなる
BI ツールの主力製品 Dr.Sum Ver.5.0 を発売した。
実はこの BI ツール、これまでの Dr.Sum の
延長線上にあるわけではない。
完全なゼロベースで開発された
インメモリデータベースエンジンを実装しているのだ。
「100億件のデータを1秒で実行する」という目標を掲げて始まった
札幌オフィス Dr.Sum 開発グループのチャレンジに迫った。

継ぎ足しを重ねて拡張を図っていくにも限界

JR札幌駅北口から徒歩3分のオフィスビル内に、ウイングアークの札幌オフィスはある。この札幌オフィスこそ、過去10年以上にわたってBIツールの主力製品 Dr.Sumの開発を脈々と続けてきた中心拠点だ。

もっとも、中心拠点と呼ぶには少々気が引ける。世に広く名の知れた外資系ソフトウェアベンダーの開発体制を見ると、シリコンバレーあたりに広々とした専用ラボを構え、ワールドワイドで採用した数千人を超えるエンジニアを擁しているのが当たり前である。これに対して札幌オフィスの開発体制は、貧弱と言わざるを得ないからだ。

Dr.Sumはデータベースエンジン部分とクライアントツールであるDatalizerの大きく2つのコアモジュールから成り立っているが、そのデータベースエンジン部分の設計やコーディング、テスト、アップグレードにあたっている開発チームの体制はわずか10名にすぎない。

外資系の大手ソフトウェアベンダーと比べれば、そもそも手がけているプロダクトのラインナップが根本的に違うという点を考慮しても、札幌オフィスの開発体制はせいぜい町工場のレベルだ。

そんな小さなソフトウェア開発のチームに最初の奇跡を起こしたのが、まさに Dr.Sum の初期バージョンだった。ひとりの伝説的な“天才開発者”が設計したアルゴリズムを基に、世界でも例のないカラムナ型データベース技術を確立し、データ集計の圧倒的な高速化を成し遂げた。その後、日本発のBIツールとしてエンタープライズIT市場で確固としたポジションを獲得し、Ver.4.2まで拡張や改善を重ねてきた。Dr.Sumの歩みについては、いまさら説明するまでもないだろう。

とはいえ、継ぎ足しに継ぎ足しを重ねて拡張を図っていくにも限界がある。10年前には圧倒的だった Dr.Sum のデータ集計スピードだが、気がつけば他ベンダーのBI/データベース製品も遜色のないレベルまで迫ってきている。また、Ver.4.2はマルチビュー構成をとることで数十億件のデータを処理することができるが、昨今の爆発的なデータ量の急増に求められるパフォーマンスとは言えない。

日本のソフトウェア開発者の底力を見せつけたい

そんな停滞感が漂い始めた2014年のある日、技術本部 Dr.Sum 開発部の部長を務める笹原 徹のもとに、本社から新たな開発方針が提示された。

「そろそろ Dr.Sum のデータベースエンジンを全面的に作り直さないか。100億件のデータを1秒で実行できるくらい高速なエンジンを作って、もう一度世間を驚かせたい」

こんなことを思いつくのは、執行役員 CTO 技術本部部長の島澤 甲のほかにはいない。笹原も最初のうちこそ突拍子もない話と聞き流すつもりだったが、次第に前のめりになっていく自分自身を感じずにはいられなかった。

「島澤さん、いつもやり方が上手い……」

実際、そうなのだ。完全に荒唐無稽な要求であるならば、いくら直属の上司からの進言であっても「無理なものは無理」と突っぱねることができる。だが、本気で頑張れば何とかするのはないかという絶妙なところを島澤は突いてくるのだ。いきなり100億件のデータを1秒で実行するのは無理でも、目標を3段階くらいに分けて地道なステップアップを狙うのなら成果を出せるかもしれない。あるいはデータのスケラビリティを拡大するだけなら案外簡単に実現できるかもしれないといった具合だ。

しかも今回は島澤自らがプロトタイプを作って実験し、「ムダを徹底的に排除してプログラムを最適化すれば確実にスピードが上がる」ことに確信を持ったとまで断言する。こうなると、もはや笹原にとっても「やらない」という選択肢はない。

ただでさえBIツールやデータベースの世界は外資系ベンダーが大きなシェアを占めており、さらにその勢いを増している。そうした状況を目のあたりにし、ただ指をくわえて眺めていたのでは、この先 Dr.Sum といえども存在感を失いかねない。

「いつまでも外資系に牛耳られてたまるか。日本のソフトウェア開発者の底力を見せつけてやる！」

ふつふつと湧き上がってくる思いに、笹原の腹は固まった。

【最初の段階でいい加減なものを作ると、10年以上もその苦勞を引きずってしまう】

ただ、笹原には乗り越えなければならないもうひとつの関門があった。Dr.Sum 開発部でデータベースエンジンを担当しているメンバーを動かすためには、まずひとりの人物を“納得”させなければならないのだ。その人物とは、Dr.Sum 開発グループのグループマネージャーを務める橋田 哲尚である。

「本気で Dr.Sum のデータベースエンジンをゼロから作り直すのですか？ 笹原さん、そんなに簡単に言わないでくださいよ。そもそも 100 億件のデータを 1 秒で実行するという目標にどんな価値があるのですか？ それは具体的にはどんなワークロードにおける、どんなパターンのデータ集計を想定しているのですか？ それは多くのユーザーが抱えている業務課題に応える普遍的なアプローチと言えますか？」

思ったとおりの反応だった。一つひとつの物事を腑に落ちるまで理詰めで突き詰め、絶対に安請け合いはしない。橋田とはそういう男なのである。

一方で笹原は、Dr.Sum のデータベースエンジンをゼロベースで作成するという構想そのものを、橋田が否定しているわけではないこともよくわかってきた。

Dr.Sum のデータベースエンジンが限界を迎えていることは、他人に言われるまでもなく橋田自身が肌身で感じていたことだ。しかしゼロベースで作成し直しても、いまがそのタイミングとして本当に正しいのか。新しいデータベースエンジンはどのようなアーキテクチャーが望ましいのか。それを実現するためにどんなアプローチをとるのが効率的なのか——。橋田は必死にその答えを導き出すとしていたのだ。

方向性がいまいなまま雰囲気流されて開発に着手した結果、ふたを開けてみれば無理難題が山積みで、開発現場が大混乱するというのはよくある話だ。だからこそ慎重にならざるを得ないのである。橋田はこう語る、「いちばん難しいのは、いちユーザーの業務にあわせて最適化するのではなく、一つの機能を追加するにしても、それが多くのユーザーの満足度を高められるか、もっとエッジの効いたアプローチはないかを考え抜いて仕様を決めていかなければならないことです。しかも今回データベースエンジンを刷新するとなれば、その基盤は今後少なくとも 10 年以上にわたって使われ、改善や拡張を続けていくことになります。裏を返せば最初の段階でいい加減なものを作ってしまうと、10 年以上もその苦勞を引きずることになります」

【多次元データの「検索」「集計」「集約」を究極的に高速化することが、Dr.Sum の価値】

丁々発止の議論の末、Dr.Sum 開発部の総意として導き出した答えが、データベースエンジンをインメモリで実行するというアプローチだ。

「Dr.Sum のデータベースエンジンにとっての基本は、多次元データの『検索』『集計』『集約』であり、これらの操作をどこまで高速

化できるかが、後にも先にも変わることはない Dr.Sum の価値そのものです。そこに最も早く、効率的に近づくことができる方法として着目したのがインメモリへの実装なのです。既存の Dr.Sum のデータベースエンジンのようなファイル層へのアクセスはとりあえず横に置き、すべてのデータがメモリに載っていることを前提にスピードを最大化することに全力を挙げました」と橋田は語る。

この基本方針のもと、開発グループのメンバーが一致団結した。インメモリデータベースに最適化したアルゴリズムの設計とプログラミングにあたったエンジニアの安達 俊貴は、「これまでの Dr.Sum のデータベースエンジンもユニークな発想を積み重ねた高い価値を世の中に提供してきましたが、自分の手でそれを上回る価値を提供すべくスピードを追求したいと考えました。これもゼロベースで作成し直すからこそやれることで、責任は重大ですが、これ以上面白い仕事はありません」と語る。

一方で、開発グループの“引き締め役”を担ったエンジニアが川代 雄太だ。10 年以上にわたって使われてきた旧バージョンの Dr.Sum のデータベースエンジンは、そのぶん十分に品質も実証されている。しかしその基盤をゼロベースで作成し直すとすれば、当然のことながらそうした実績もすべてリセットされる。ユーザーの現場で想定されるあらゆるパターンの操作に対して新たにテストを実施し、プログラムの不具合（バグ）を洗い出して解消し、品質とパフォーマンスを確保しなければならないのだ。川代はそうした高度な品質を担保するため、効率的な開発作業を支援するインフラやテストツールを整備するとともに、自らが先頭に立って検証を行っている。

「プロジェクトの責任を個人に押し付けるつもりは決していないのですが、発見した不具合についてはあえて無慈悲に開発者本人にチケットを通知し、修正してもらう必要があります。開発メンバーからは嫌がられ煙たがられる存在になるかもしれませんが、この責任だけはしっかり果たさなければなりません」と川代は語る。

こうして完成したインメモリデータベースエンジンを実装し、2017 年 10 月 5 日に正式リリースされたのが Dr.Sum Ver.5.0 というわけだ。

もともと当初からの目標である 100 億件のデータを 1 秒で実行するというスピードを実現するのは容易ではなく、現状ではかなりのハイエンドサーバーを複数台配置したスケラブルな構成を組む必要がある。

その意味では Dr.Sum のインメモリデータベースエンジンはまだまだ道半ばなのだが、一方で現時点においてもスタンダードクラスのシングルサーバー環境で、10～20 億件規模のデータを秒単位で実行できる高速性能を達成しているのも事実である。

「データ活用・分析のコモディティ化を求める Dr.Sum の主要ユーザーの業務にとって、むしろそうしたシングルサーバー環境で発揮されるコストパフォーマンスに優れたスピードこそが、“現実解”になると言えるのではないのでしょうか。もちろんこの先もハードウェア性能は確実に向上していき、そこにインメモリデータベースエンジンの絶え間ない改善や拡張を続けていくことで、さらなるブレークスルーによる BI ツールのイノベーションを世の中に巻き起こしていけると確信しています」と笹原は胸を張る。

いずれにしても Dr.Sum は生まれ変わったばかり。新たな歴史が、いま幕をあけた。

リーダーたちが語る！ Dr.Sum Ver.5.0 開発の核心

Dr.Sum Ver.5.0 のインメモリデータベースエンジンは、旧バージョンからどのような進化を遂げてきたのか。様々な課題を乗り越えてきた苦労話も織り交ぜつつ、リーダーたちに今回の開発のポイントを語ってもらった。

開発グループとデータベースエンジンは 運命共同体

— あらためて伺います。Dr.Sum のデータベースエンジンを、なぜゼロベースで作り直そうと考えたのでしょうか。

笹原 Dr.Sum は歴史の長い製品です。もっと高速化したい、機能強化を図りたいという思いは持っていましたが、10年以上前のコードに手を入れると、どこに影響がでるか分かりません。思い切った手は打てなくなっていたのです。その意味でもデータベースエンジンの全面的な刷新が必要なタイミングでした。

— 今回は特にインメモリへの移行を大きなテーマとしていますね。

橋田 はい。Dr.Sum を通じて提供する価値は何かというと、端的にはデータ集計のスピードです。そして Dr.Sum が今後に向けて訴求していく価値も、やはりスピードでありたいと思っています。これが Dr.Sum に対する一貫した開発コンセプトです。そうした中から出てきたのが、「100 億件のデータを 1 秒で実行する」という目標で、これを達成するためにはディスクをベースとした既存のデータベースエンジンでは不可能。インメモリに軸足を移すのは必然の選択でした。もっともお客様のすべてのワークロードをインメモリに展開できるとは限らないので、既存のデータベースエンジンとハイブリッドで運用できる仕組みを提供しています。

— ちなみに「100 億件のデータを 1 秒で実行する」という目標は、執行役員 CTO の島澤 甲さんが設定したのですか。

笹原 そこは島澤の感覚だったと思います。ただ、一足先に MotionBoard でインメモリ方式を採用している実績があり、簡単なことではないが決して達成不可能な数字ではないという確信はあったようです。

— それにしても、なぜそこまでデータベースエンジンにこだわったのですか。スピードを追求しているのはよくわかりましたが、それならばむしろ、すでに市場に登場している汎用的なインメモリデータベースエンジンを採用するという手もあったかと思えます。そうすれば 3 年間にも及ぶ開発期間をかけることなくスピードアップを実現することができ、ウイングアークとしては BI ツールのユーザーインターフェイスやエクスペリエンスの部分にもっと注力できたのではないのでしょうか。

笹原 そもそもなぜ私たちが Dr.Sum を開発したかということ、日本企業のお客様が求めているような集計ツールがなかったことに端を発します。その状況は現在でも大きく変わっているわけではなく、他社の汎用的なデータベースエンジンでは、お客様の業務の“かゆいところに手が届かない”のです。だからこそ私たちが提供する情報活用ソリューションについては、主要モジュールをすべて内製化すべきという基本姿勢で臨んできました。これからは逆に Dr.Sum のインメモリデータベースエンジンにより汎用的な機能を取り込んでいくことで、超大規模なデータ活用の市場を切り崩していきたいとも考えています。

安達 私もウイングアークがデータベースエンジンを開発することに疑問は持ちませんでした。というより私たちはこれまでずっと Dr.Sum のデータベースエンジンに関わってきたエンジニアのグループで、そこに存在意義を感じています。仮に『もうデータベースエンジンの開発から撤退するので君たちはこれからエクスペリエンスの開発をやれ』と言われたら、会社を辞めると思います。データベースエンジンを高速化することは私にとって最大の仕事のやりがいであり、楽しいことなのです。

— なるほど。皆さんとデータベースエンジンは“運命共同体”なのですね。

橋田 そういっても過言ではないと思います。



開発リーダー
橋田 哲尚

クエリーごとに異なるパターンで いかに高速性を発揮するか

— いくら楽しくてもインメモリデータベースエンジンの開発には様々な苦勞があったかと思います。具体的にどんな困難を乗り越えてきたのでしょうか。

安達 Dr.Sum にとって集計スピードこそが命であるわけですが、あらゆるパターンで性能を発揮できるわけではありません。あるデータセットに対して特定のクエリーを適用した場合に高速化できたとしても、別のデータセットに切り替えた途端に速度がガクンと落ちてしまうのです。そこで別の対処を考える、そういった作業を繰り返しながら全体としてのスピードを上げてきました。

— なぜデータセットやクエリーのパターンによって、集計スピードが変わってしまうのですか。

安達 最も大きく集計スピードに影響するのはカージナリティ（カラムに格納されているデータの種類のどれくらいあるか）です。例えば「所在地」を例にとると、都道府県単位であればデータの種類の種類は47種類しかありませんが、市区町村単位になると1,700以上にも広がります。

橋田 どんなデータセットやクエリーのパターンに対しても汎用的に適用できるアルゴリズムを用意できれば理想的なのですが、安達が言ったように現実にはそうはならず、カージナリティの高低に応じて特殊化された複数のアルゴリズムを使い分けることとなります。

— そうすると当然、全体のコントロールが難しくなりますね。

橋田 そのとおりです。アルゴリズムを特殊化したことでオーバーヘッドが肥大化し、データベース全体のパフォーマンスが低下してしまうケースさえあります。実は外資系の大手ベンダーが提供しているデータベースエンジンも同様に特殊化されたアルゴリズムのかたまりなのですが、大規模な人海戦術でその開発をこなしています。私たちのような少人数の開発体制にとって一番厳しいところですよ。

— どうやってその課題を乗り越えたのですか。

安達 アルゴリズムの設計を大人数で分担するのではなく、私ひとりに任せてもらったことで逆にやりやすい面もありました。集計処理の大まかな流れは一本ですが、集計データを保持するコンテナを動的に切り替えるというアーキテクチャーを考えたのです。詳しいことは話せませんが、処理を始める前にどのコンテナを使用するかなどの分岐の“シナリオ”を決めておき、あとは一気に処理を流していくという仕組みです。

— 淡々と話していますが、かなり高度なテクニックですよ。

安達 頭全体がコンパイラにならないとできません（笑）。

川代 開発言語には C++ を使っているのですが、今回はそのテンプレートを使ったメタプログラミングを多用しています。これは旧バージョンの Dr.Sum のデータベースエンジンではほとんど使っていなかった技法で、そういった面でも安達のチャレンジに開発グループ一同注目しています。



開発エンジニア
安達 俊貴

開発側での手戻りを最小化する テスト自動化の取り組み

— データベースエンジンをゼロベースで作直す、しかもこれまでと違った技法まで駆使したとなると、テストもますます大変ですね。

川代 そのとおりで、新しく作ったものが問題なく動き、最初から正しい結果を返すはずがありません。改善したはずの部分がかってデグレード（以前の状態よりも品質が悪化）をもたらすことも多々あります。社内には品質保証部という専門チームがあり最終的な品質を担保しています。しかし、機能数も膨大なため、いきなり彼らに渡しても“お手上げ”になってしまいます。そこで開発側であらかじめ入念なテストを行い、一定レベル以上の品質を満たしておく必要があります。

— 開発側でどういったテストを強化したのですか。

川代 各担当者がそれぞれ作ったコードに対して、開発グループ内でレビューを行い、そこで OK が出たものを製品モジュールにマージして統合テストを行っています。率直なところ、このプロセスにおいては、これまで十分なテストがなされていませんでした。いざ実行してみたところ動かなかったり、結果に矛盾が生じたりすることがあり、そのたびに個々のコードに立ち返って原因を調べ直すという手戻りが発生していたのです。これは大変な無駄な作業であり、開発スケジュールにも大きな遅延をもたらすため、コードレビューを行う前に自動でテストを行う仕組みを作りました。簡単に言うと、そうした全体的な作業の手戻りをなくすためのインフラ整備に注力してきました。

— 様々なクエリーのパターンに対する網羅的なテストを自動的に実施し、品質維持のための作業を効率化するのはですね。

川代 基本的にはそういう考え方をとっているのですが、クエリーのパターンの組み合わせを単純に展開すると1億以上になってしまうため、ある程度の“足切り”も必要です。その見極めが難題でした。

橋田 まさにそこが QE（品質工学）の核心であり、開発グループ全体をリードしてくれた川代の貢献は絶大です。川代がいなければスケジュールどおりに Dr.Sum Ver.5.0 をリリースできなかったと思います。

汎用データベースとしても活用できる機能の拡張へ

— 少し気が早いかもしれませんが、今後に向けた Dr.Sum の拡張、特にインメモリデータベースエンジンの機能強化の構想をお聞かせください。

橋田 これまでの Dr.Sum のデータベースエンジンは一般的な RDBMS とは違い、集計処理に特化したアーキテクチャーを採用していました。これに対して今回の Dr.Sum Ver.5.0 のインメモリデータベースエンジンは、よりジェネリックなワークロードを意識したアーキテクチャーを設計しており、汎用的なデータベースとしてもそれなりに活用できる機能を付加していく計画です。例えばトランザクション処理を担う更新系の機能、ビッグデータを想定した複数の異なるデータソース（テーブル）の JOIN といった機能についても、できるだけ早期に実装したいと考えています。

安達 個人的な思いとしては、これまでの Dr.Sum のデータベースエンジンでトライしたことがなかった Window 関数を組み込みたいです。クライアントツール側ではなく SQL レベルで機能強化を図ることでこそ、お客様のかゆいところに手が届くようになります。せっかくジェネリックなインメモリデータベースエンジンに生まれ変わったのですから、そのメリットを最大限に活かし、より多様な分析をより容易に実行できる BI プラットフォームに Dr.Sum を発展させていきたいです。



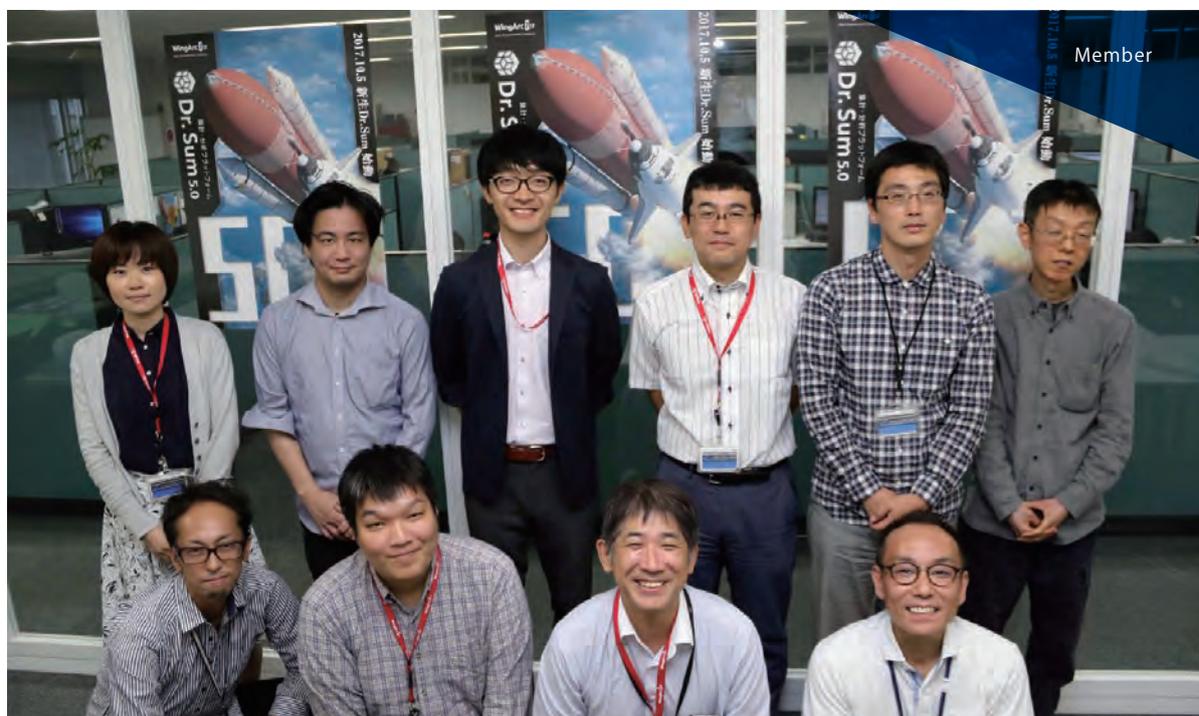
開発エンジニア
川代 雄太

川代 私は開発作業やテストを支援するためのインフラ部分の強化を引き続きやっていきたいと考えています。開発やテストのプロセスを効率化することでコスト削減を促せば、よりコストパフォーマンスの優れた製品を提供するという形で、お客様にも還元することができます。

笹原 ハイブリッド型のエンジンはまだ始まったばかりです。パフォーマンスはもちろんですが、機能的にも絶え間ない進化を続けていきますので、どうかご期待ください。



開発責任者
笹原 徹



Member

