

政策研究大学院大学 様

Google BigQuery+Lookerなら、 永遠に終わらなかった処理が10秒で完了

国立大学法人 政策研究大学院大学(以下、GRIPS)は、Google CloudのBigQueryとBIツールのLookerを組み合わせ、1億7000万件、1.5テラバイトの論文データ等を全件分析できるシステムを作り上げた。ほかに試みたさまざまな手法ではデータ取得すらままならなかったが、ビッグデータ分析へのニーズを唯一満たすことのできるシステムとして、研究活動のスピードアップに大きく貢献している。このプロジェクトでは、システムのPoCフェーズから、NI+CとLooker日本人が全面的にサポートした。



政策研究大学院大学は、政策及び政策の革新に関わる研究と教育を通して、我が国及び世界の民主的統治の発展と高度化に貢献することを目的としています。

政策研究を専門とする大学院で、1997年に国立大学として開学し、世界各国から未来の政策リーダーや研究者が集まる国際的な政策研究・教育の拠点となっています。

民主的統治を担う指導者、政策プロフェッショナルの養成を目的とし、「政策研究科」の1研究科、「政策研究」の1専攻のみを置いています。多様な教育プログラム・コースは、政策研究の進展や社会的変動に伴う行政課題の変化に対応し柔軟に見直され、2020年4月には行政官向けデータサイエンス科目も開講します。

キャンパス : 〒106-8677 東京都港区六本木7-22-1

創 設 : 1997年10月

学 生 数 : 343名(修士課程67%、博士課程33%/2019年5月1日現在)

U R L : <http://www.grips.ac.jp>

導入・取り扱い製品
・Looker
・Google BigQuery

【データが巨大すぎ、BIツールが結果を返せない

GRIPSは、政策研究に特化した大学院大学だ。諸外国の日本担当窓口として働く海外機関の行政官を留学生として受け入れており、新国立美術館の目の前にあるキャンパスは国際色豊かだ。官公庁に勤める行政官のリカレント・エデュケーションの場として機能するとともに、国家の政策に関連するビジネスを展開する企業の社会人など、幅広い層の学生が学究活動に取り組んでいる。

そのGRIPSにおいて、専門職(Cloud Engineering)伊谷 陽祐氏が率いるチームが手掛けている、日本の研究活動が世界の中でどのような位置づけにあり、その強みと弱みを客観的に導き出そうとしているプロジェクトがある。

具体的には、世界中で日々発表される論文、特許、助成金などをすべてデータ化し、データベースへ格納。これらを横断的に分析することで、客観的な証拠に基づくリアルな情報として見られるようにするというものだ。たとえば、過去の政策とそれにひも付く助成金が、論文や特許としてどのように結実したかを探り、その結果として該当する分野において日本の研究活動が進展したのかどうかを把握することも期待できる。

論文データ、特許データ、研究助成データ、国際統計データ、財務諸表データ、法律文書データ、行政文書データなど様々な種類のデータを集積して、超巨大なデータベースが出来上がった。

政策研究大学院大学
専門職(Cloud Engineering)
伊谷 陽祐 氏



伊谷氏は、「当初は様々な制約のなかローカルのMySQLで処理していたのですが、どれだけマシンのスペックを上げててもクエリーがなかなか返ってきません。国内データだけならなんとか我慢できるのですが、海外のデータも加えると実用に耐えられませんでした」と話す。

こうして伊谷氏は、Google CloudのBigQueryを利用することになった。著名なパブリッククラウドのデータハウスを実際にいくつか使ってみたが、どれも満足できる性能ではなかった。そんな中、ビッグデータを高速に扱える性能面に加え、サーバーレスでインスタンス管理などITインフラの整備に時間を使うことなく分析業務に注力したいという思いがBigQueryの採用理由だった。ところが、さらに大きな壁が待っていた。

「エンタープライズでよく使われているBIツールをいくつか試してみたのですが、データが大きすぎて、BIツールが機能しないのです。帰宅前に実行させて、翌日研究室で見るとメモリがオーバーフローしていたこともありまして、データを4時間かけてロードした挙げ句にフェイルしたこともありまして」(伊谷氏)

BIツールはデータを保持すべきではないことに気づいた

伊谷氏はITに詳しく、問題の原因がBIツール側にあると理解した。BigQueryをネイティブに触れば、結果は高速に返ってくるためだ。しかし、そのためには自らコーディングする作業が発生する。BIの良さは、頭に浮かんだ分析を瞬時に実行し、結果を見ながら適切にブラッシュアップし、最終的に正確な結果を導き出すプロセスを高速に回せることだ。コードを書いて分析しても、対話型で仕事を進められない。

とはいえ、要求にこたえてくれるBIツールはない。結論として、当時のIT環境の中では、2つの選択肢しかなかった。1つは、分析対象をBIツールが扱える範囲内に絞り込み、その枠内で分析すること。もう1つは、時間をかけてコーディングし、BigQueryの性能をフルに使って分析することだ。前者は分析者としては満足できるプロセスで仕事を進められる一方、使いたいすべてのデータを使えておらず、完全に客観的かつ正確と言える分析結果にならないリスクがある。後者はコーディングに時間がかかりすぎるため、期限内に研究するテーマを極端に絞り込まなければならない。

苦闘しながら研究を続ける中、NI+CとGoogle Cloud、Lookerが共同開催するセミナーの案内が届いた。「Lookerと呼ばれるBIツールをBigQueryと組み合わせて使えば、ビッグデータを高速に処理できる」という触れ込みだった。専門職(Business Intelligence) 大沼 詩音氏は、「データを切り取って一部だけを分析するのではなく、すべてのデータを使って横断的に分析したいという私たちのニーズにこたえてくれそうなので期待しました」と話す。



政策研究大学院大学
専門職(Business Intelligence)
大沼 詩音 氏

一方の伊谷氏は、すでにLookerについて聞いたことがあった。Lookerは、BIツールだ。ただし、「ふつうのBIツール」とは異なる特色がある。データの入っているデータベースの構造を可視化し、データをビジュアルに表現するツールではあるのだが、処理はすべてデータベース側に行わせるのだ。そのため、性能はデータベースの能力に依存する。「当時は、BIツールはデータを取り込むものだという“昔からの常識”に縛られていました。データを保持しないのであれば、Lookerは何をしてくれるツールなのだろう、という興味から参加を決めました」。

セミナーにはチームメンバーが全員参加し、そのアーキテクチャについて質問もした。「データを持たないことが、本来正しいのだと気づかせてくれて、期待にこたえてくれるような予感がした」(伊谷氏)ため、PoCを実施した。

Lookerは本採用の前にPoCを実施することを推奨している。それには、いくつかの理由がある。1つは、Lookerで扱うデータを蓄積するデータベース側の構造を可視化するために、LookMLという言葉を経験してもらいたいからだ。この部分に対してとっつきにくい印象を受けるかもしれないが、LookMLは直感的な言語であり、それほど難しいくない。むしろ、ここにレイヤーを噛ませることで、新たなデータ項目を得ることができる。集計や計算などの処理をデータベース側にやらせた上で、データソースにない情報を作り出し、Looker上で見られるようにできるのだ。

そして、重要なのもう1つの理由。データソース側のデータ項目とデータの意味を正確に理解するためだ。「データソースに入っているが、何を表しているのかわからないデータ」は意外に多く、それらのデータの意味を紐解けば将来の分析を優位に進められる。ただし、今回のケースでは、データソースの意味をだれもが完璧にわかっていた。そのためPoCフェーズは初歩からスタートしたが、最後には極めて実戦的な内容になった。



政策研究大学院大学
技術職員(Cloud Engineering)
梶 裕仁 氏

技術職員(Cloud Engineering) 梶 裕仁氏は、「チーム作業のためにGitHubを使ったのは初めてでした。NI+CさんとLookerさんには初期にその部分から丁寧にレクチャーしていただいたのは、とてもありがたかったです」と話す。その後、全員が知識を習得し、分担して複数データソースのダッシュボードを作成できるレベルにまで達することができた。

専門職 (Data Engineering) 佐田 真穂氏は、「PoCもそうですが、ハンズオンに参加してさらに理解が深まりました。アフターフォローも含めて感謝しています」と当時を振り返る。



政策研究大学院大学
専門職 (Data Engineering)
佐田 真穂 氏

わずか10秒程度でクエリーが返ってくる

このPoCフェーズで、Lookerが十分にニーズにこたえられることがわかった。これまで途中でメモリがオーバーフローしていたようなクエリーも、10秒以内に返ってくる。膨大なデータをそのまま分析できるため、BIツールの性能という制約条件を充たすためにデータを準備する必要もない。

ちょうどその期間に、緊急案件が飛び込んできた。その際も、Lookerを使って瞬時にアウトプットを出すことができた。Lookerの操作はわずか十数分。分析処理はわずか10秒程度で返ってくるため、ずっとやり続けた本来の分析プロセスをわずかな時間で完結することができた。かつては絞り込んだデータをエクスポートし、それを取り込んだExcelで最終化していたような案件だったが、必要な全データを参照しながら極めて短期間に終わらせることができたのだ。非常に強力な分析パートナーを得たと感じた。

「契約にあたっては、大学法人特有の手続きなどあり、NI+Cさんがスムーズに動いてくれて本当に助かりました。Lookerをより良く使うという観点だけでなく、NI+CさんとLookerさんが協力してサポートしてくれたおかげで、かつては実現不可能だとすら考えていたシステムを手に入れることができました」(大沼氏)

こうして、GRIPSは新たな分析システムを手に入れた。今後は研究活動の飛躍的なスピードアップを期待できそうだ。そして、だれもが使えるダッシュボードへ。権限設定などの問題をクリアできれば、情報検索型のダッシュボードを公開したい考えもある。

伊谷氏は、「われわれの研究成果を、多くの方に閲覧して使っていただけるようなダッシュボードを目指します。BIツールからデータをエクスポートして、Excelで最終化するという作業は、本来おかしいのです。BIなのでから(笑)。LookerとBigQueryなら、それができます」と話してくれた。

<NI+Cより>

今回のプロジェクトで構築をサポートさせていただいたのは、BigQuery+Lookerの実力を存分に発揮できるシステムです。そして、システムを使ってお客様が実施される研究のテーマはユニークで、広く社会の役に立つものです。私たちにとっても、本当にやりがいのある仕事になりました。

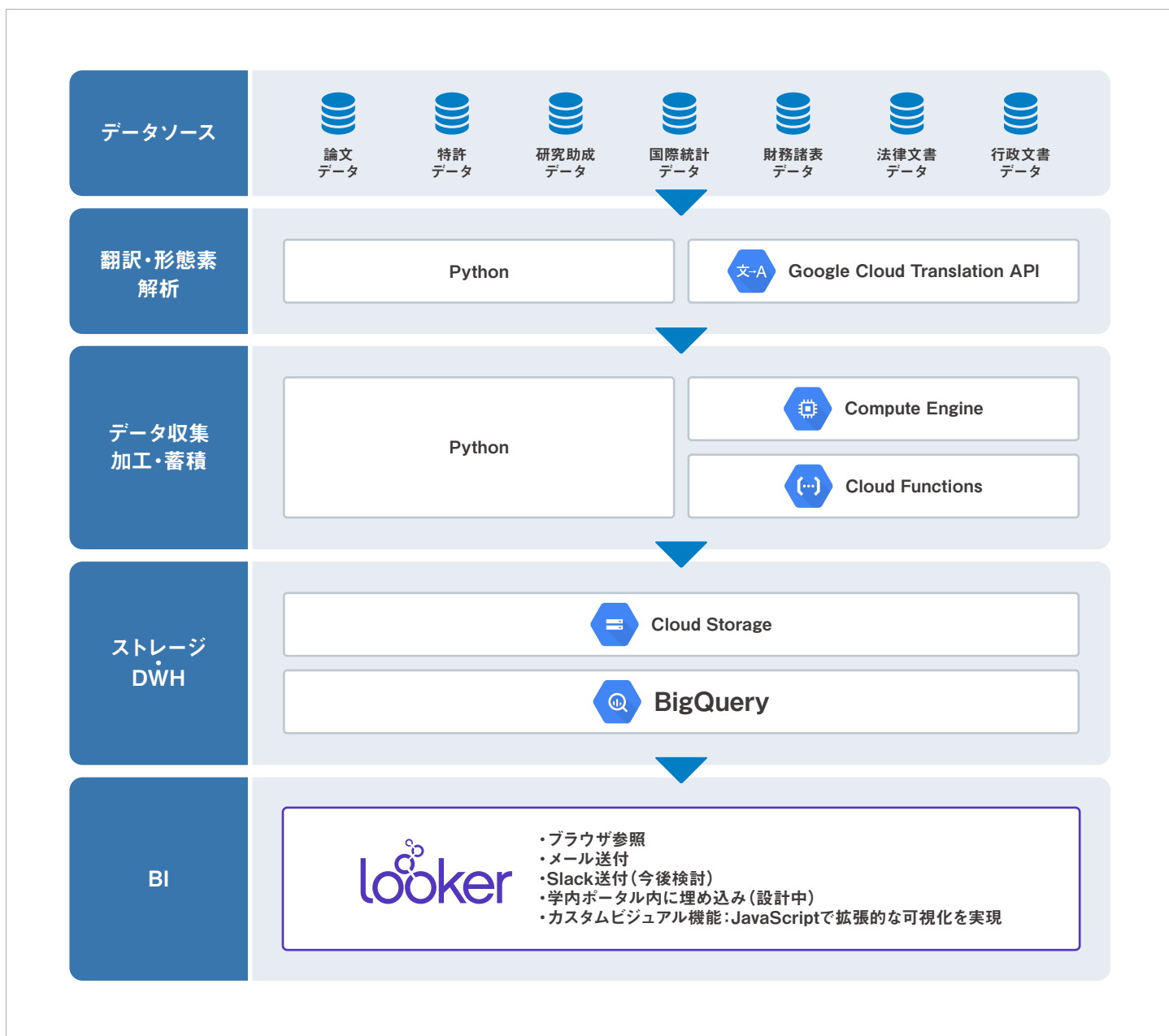
お忙しい中、PoCにも積極的に取り組んでいただきました。GitHubを使ったプロジェクトの進め方やLookMLの習得から、実データを使って画面を作り込んだり、複雑な分析を回したりするところまで。Lookerのコンサルタントの方々に丁寧なサポートをいただきながら、すべてをスムーズに進めることができました。

今後、研究成果を学内外に広めるダッシュボードを公開する予定ということで、Slerとして引き続きサポートしていきたいと考えています。



バリューインテグレーション本部
ソフトウェア・テクニカルセールス部
第1グループ 主査
伊藤 明彦

政策研究大学院大学様システムイメージ図



未来の政策リーダー・研究者の国際的研究拠点

政策研究大学院大学は現在、外国人留学生が学生総数の60%以上を占めており、世界的に見ても国際色豊かな研究教育機関です。現在までに110ヵ国5,000名以上の修了生を輩出しています。教員の多くは海外の名門大学を修了し、政策研究の先進的研究者、または官界・国際開発等の分野における豊富な実務経験を持つ者が国内外から結集しています。

また、修了生の多くがパブリックセクターに勤務しており、自国の政策形成・立案に大きな役割を担う存在として活躍しています。

