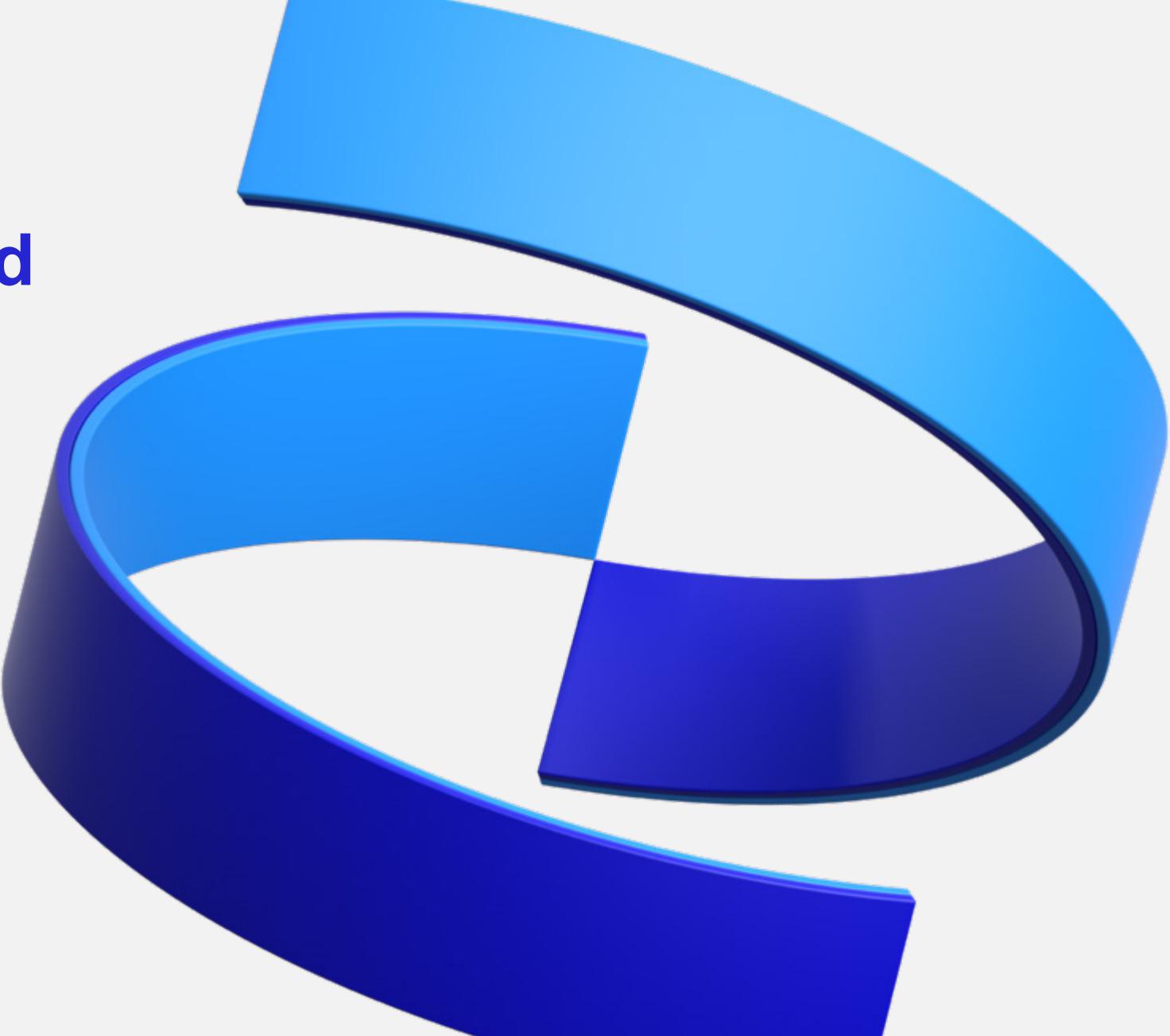


# GA Groton Resource Optimization Workload Management and Analysis Tool (GROWMAT) External Presentation

Samuel Zhang  
Herbie Bingham  
GMP Analytics  
Pfizer Sandwich x Pfizer Groton  
Sandwich, UK  
23/Jan/2024 – 08/Aug/2024



# Business Challenges

## Workload Modelling and Resource Management

1. Uncertainties in workload distribution and schedule after business restructuring.
2. Legacy solution to workload management and scheduling does not scale well, isn't particularly robust, and takes a significant time to compute.
3. There lacks granularity in the legacy report, team leaders don't have the ability to check individual jobs that affect the total workload.



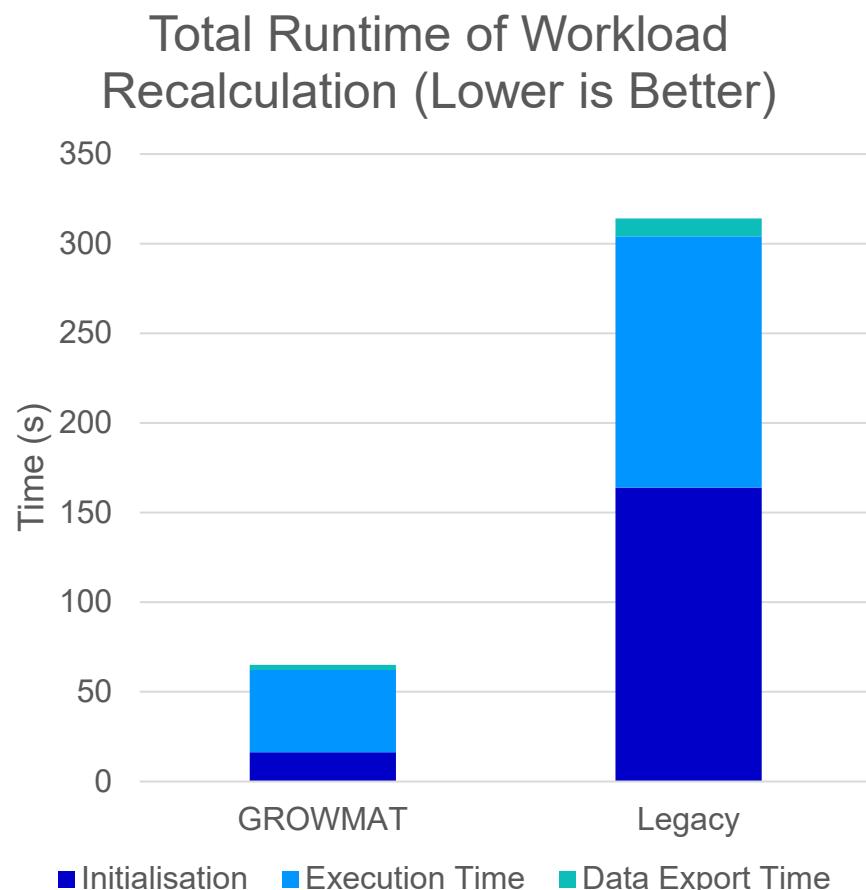
# Business Solution

## Workload Modelling and Resource Management

1. GROWMAT is designed to streamline Good Manufacturing Practice Analytics (GA)'s complex resource and workload management. It is set to replace Pfizer's legacy Excel-based GA Capacity Tool with improved features like advanced visualization through Spotfire, a dedicated website for scheduling, a more reliable codebase, and a support team for ongoing development.
2. Workload data are gathered from a variety of databases ranging from stability testing data from [LIMS](#); release, post packaging, excipients, and manufacturing data from [Clementine](#); to manually added jobs and rolling jobs. A unit of predicted workload **FTEs/day** (Full Time Employee per Day) is calculated for each job from historical data and automatically assigned an expected due date. Through different tabs, managers and team leaders can view not only a general workload distribution among teams, but they also have the granularity of viewing specific jobs affecting workload across time.
3. GROWMAT has two user facing components, a custom website for workload management (input) and a Spotfire view for data visualisation (output).

# GROWMAT Features

Overview of benefits that GROWMAT has over the legacy tool (Groton Capacity Tool)



GROWMAT	Legacy
<input type="checkbox"/> Responsive, simple, and user-friendly interface (Spotfire + Webpage)	<input type="checkbox"/> Clunky, error prone, complicated multi-workbook-dependent Excel interface
<input type="checkbox"/> Much higher granularity in job specific data visualisation	<input type="checkbox"/> Lacks graduality in job specific data visualisation
<input type="checkbox"/> Bihourly automatic Spotfire visualisation update	<input type="checkbox"/> Manual recalculation from analysts required
<input type="checkbox"/> Bihourly automatic workload recalculation	<input type="checkbox"/> Weekly manual workload recalculation from analysts needed
<input type="checkbox"/> Concurrency proof workload scheduling	<input type="checkbox"/> Constant conflicts with concurrency issues, resulting in downtime
<input type="checkbox"/> 180 Days of database rollback, 3 tiers of disaster recovery plan	<input type="checkbox"/> Data rollback dependent on SharePoint only
<input type="checkbox"/> Database aggregation via permanent service account	<input type="checkbox"/> Database aggregation depends on user account
<input type="checkbox"/> Long-term support via Adrian's Team and/or DataSci	<input type="checkbox"/> No long-term support

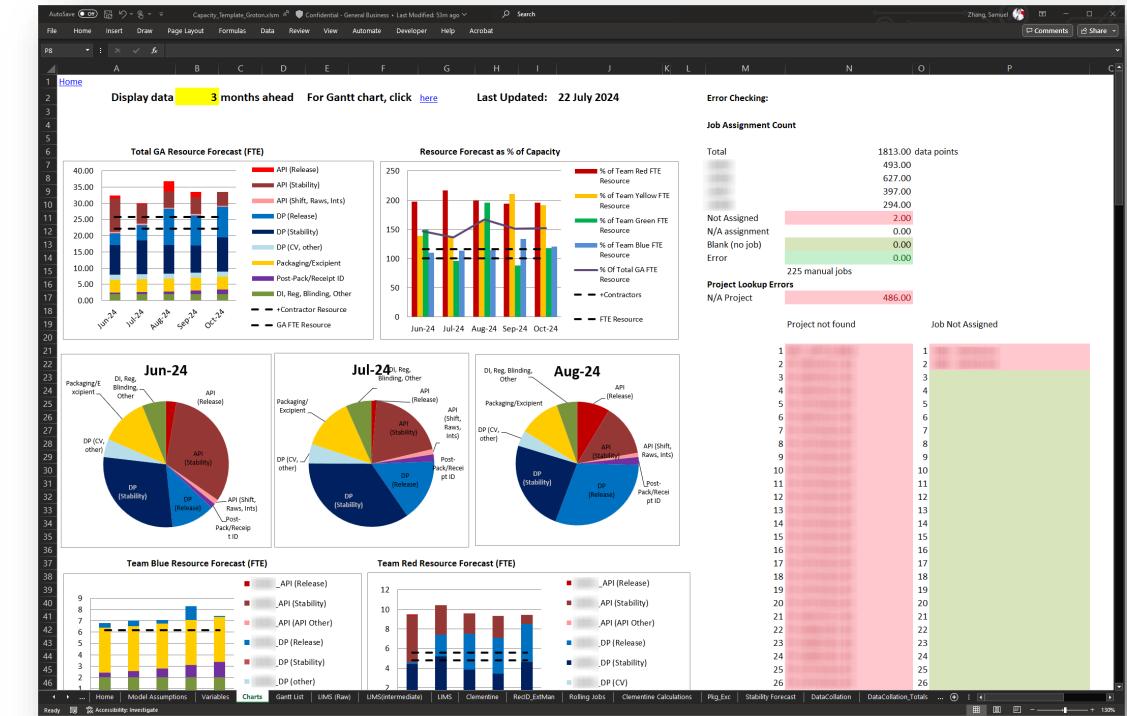
# User Interface (Legacy)

The Legacy Excel based Workload Management Capacity Tool's Interface

## Legacy Excel Workload Manager

This screenshot shows the main interface of the Legacy Excel Workload Manager. At the top, there is a ribbon with standard Excel tabs like File, Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, Automate, Help, and Acrobat. Below the ribbon is a toolbar with various icons. The main area is a large grid table with columns labeled from A to S. The first few rows contain headers such as 'Job Description', 'Team Red', 'Team Yellow', 'Team Green', 'Team Blue', 'Team Orange', 'Add a Manual Job', 'Output for specific', 'Create new site', and 'Lifetime update'. The grid itself contains numerous entries, each consisting of a short description followed by a color-coded cell indicating team assignment (red, yellow, green, blue, orange). The rows are numbered from 1 to over 100, and the columns are labeled A through S.

## Legacy Excel Workload Visualiser



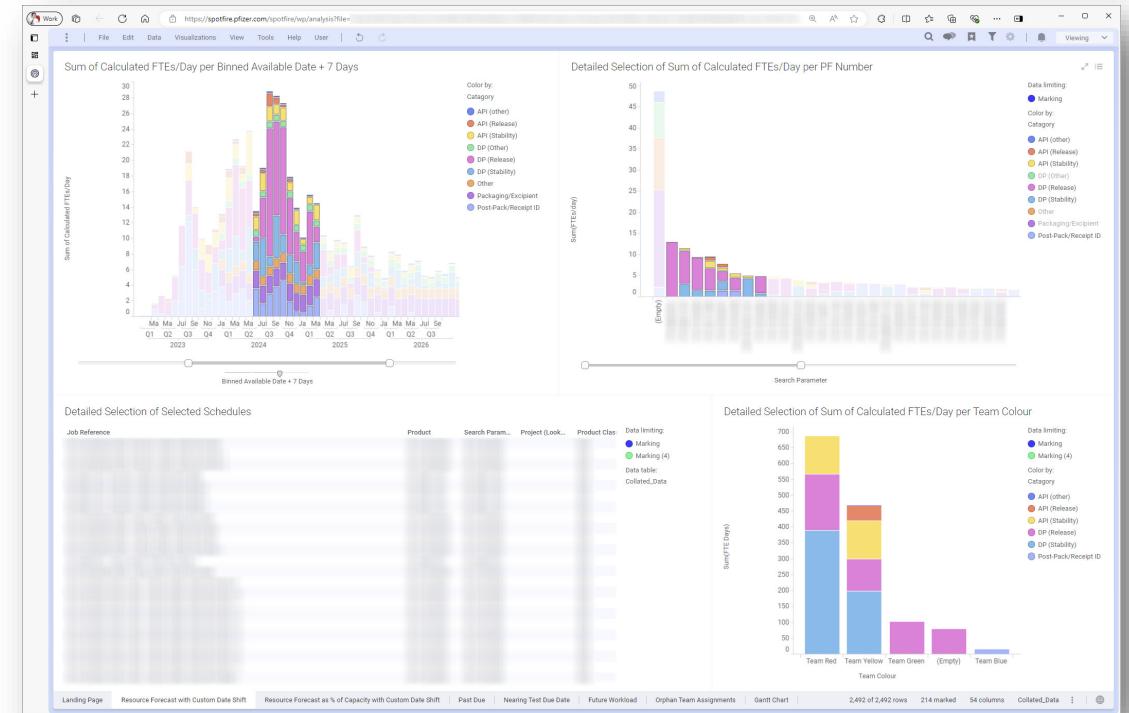
# User Interface (GROWMAT)

GROWMAT's Web Based Interface

## GROWMAT Web Workload Manager

The screenshot shows the GROWMAT Schedule Dashboard. At the top, there are three tabs: Schedule Dashboard, Admin Variables, and Spotfire Dashboard. Below the tabs, there is a section titled "Schedule Dashboard" with the sub-instruction: "View & edit your workload here. Note that this is a shared data source - changes you make here will be permanent & shared across GAI". There are two buttons: "Add a Manual Job" and "Download Excel File". A "Filters & Toggles" section includes buttons for "Team Red", "Team Yellow" (which is highlighted in orange), "Team Green", "Team Blue", "Team Orange", "Groton SDM", "Show Only Edits", and "Clear Filters". A search bar with placeholder text "to search by date, use format 'yyyy-mm-dd'" is present. The main area is a grid of workload items with columns: Job Description, Now, Job Reference, Sample Available, Test due date, Sample Due Date, Manual Due Date, Lifecycle, Analyst, Reviewer, Notes, eQMS record, Number of Samples, Team, and Product. Each item has a small preview icon and a delete button.

## GROWMAT Spotfire Workload Visualiser



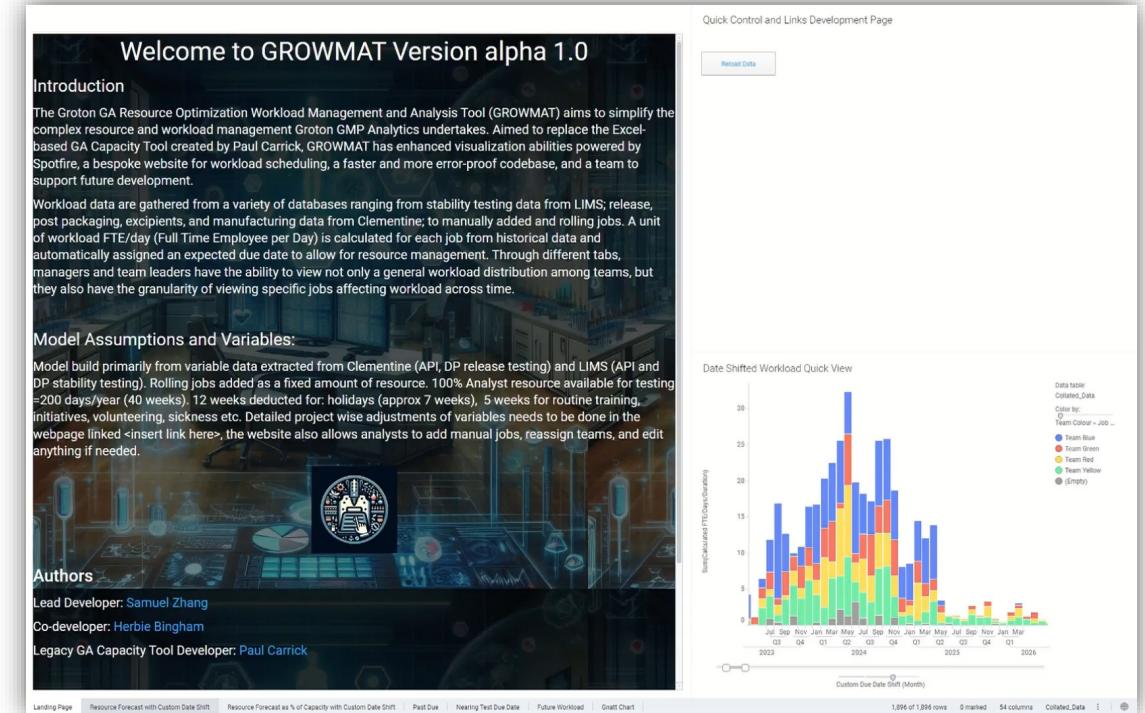
# GROWMAT Spotfire Visualisation 1

## Identifying Scheduled Jobs via FTE/Days

### GROWMAT

- Spotfire's interactive approach allows the user to visualize any specific views independent from affecting source data and other users
- Much improved data aggregation allows a much more granular database where users can find any specific jobs within seconds through selecting visualisations
- Web and cloud-based approach simplifies user experience, allowing analysts to focus more on meaning behind the data rather than bug fixing and running separate visualisations locally.

### Demonstration



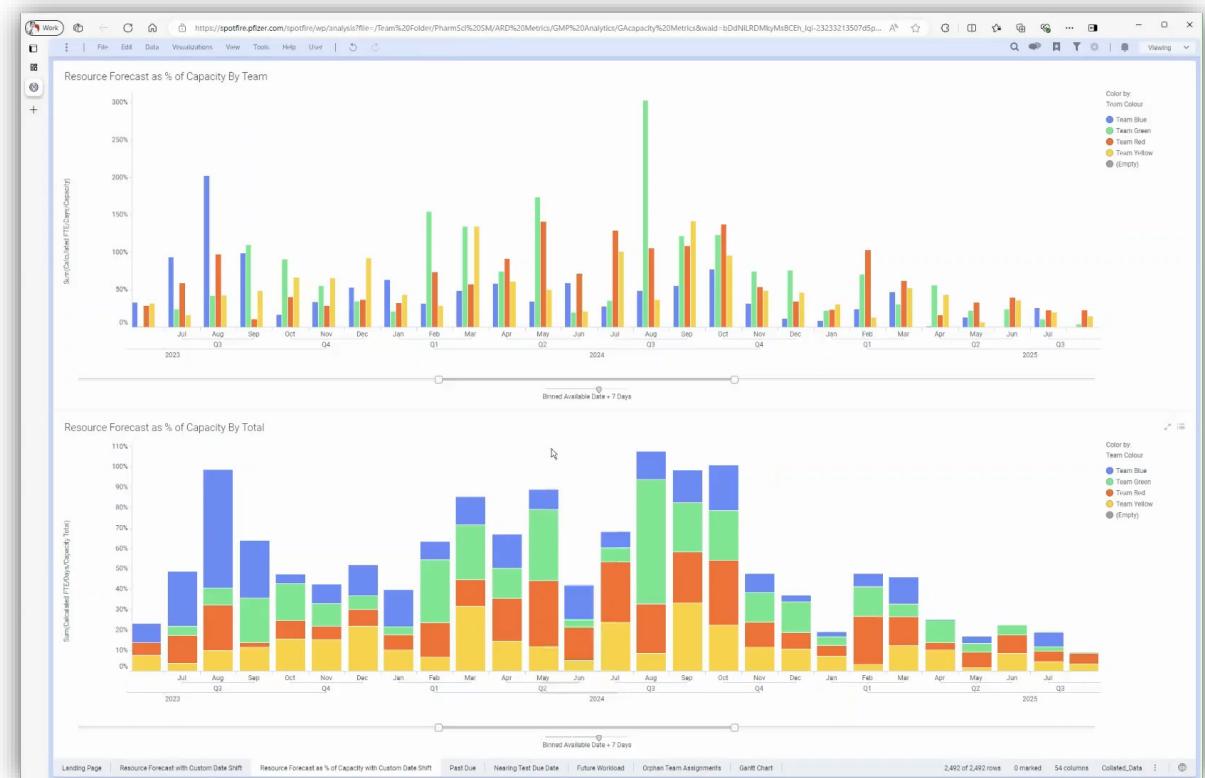
# GROWMAT Spotfire Visualisation 2

## Variety of Workload related Visualisations

### GROWMAT

- Resource Forecast as % of Capacity
  - Allowing team leads and managers to easily identify if the current workload exceeds or is behind the target workload calculated from the current FTE working hours.
- Past Due
  - Quickly identifying any jobs that are at least 30 days past due date and isn't marked as completed
- Nearing Testing Due Date
  - Flags any jobs due within 7 days that is yet to be marked as completed
- Future Workload
  - Tracks any jobs due in the next month

### Demonstration



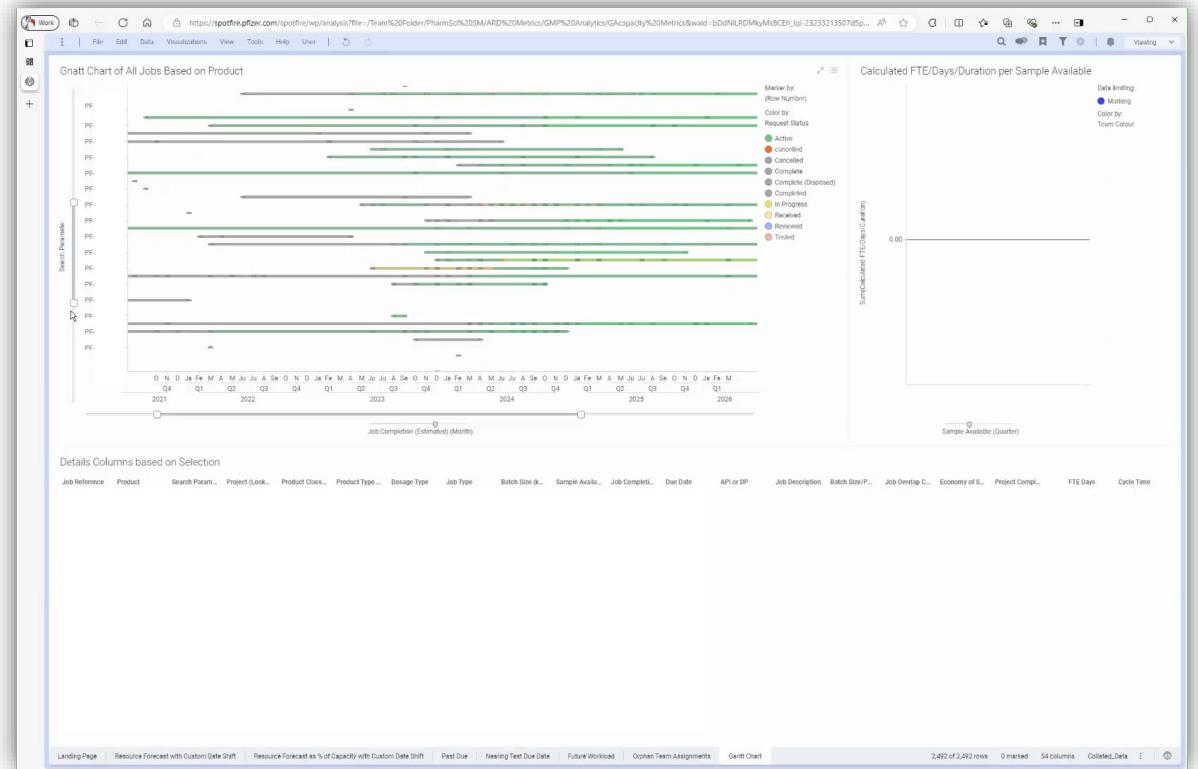
# GROWMAT Spotfire Visualisation 3

## Gant Chart of all Groton GA Products

### GROWMAT

- An overview of all projects (completed or cancelled) currently in GA Groton's portfolio. With time being the horizontal axis, each dots on the project indicates one assignment/task triggered by a due date.

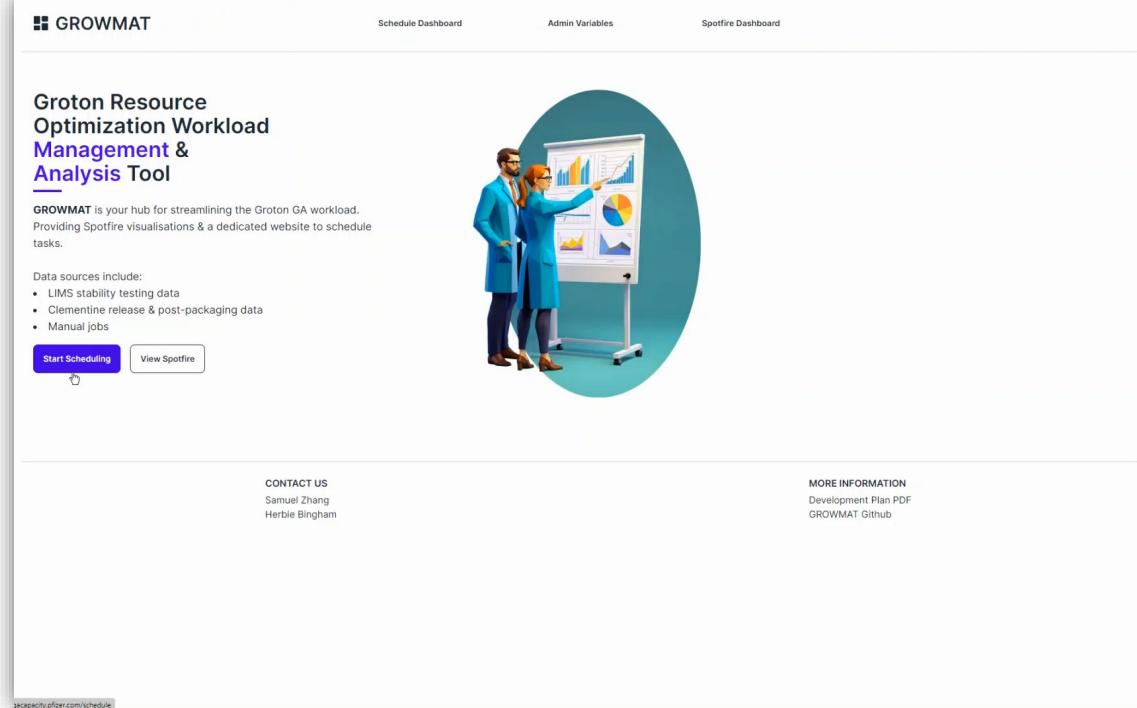
### Demonstration



# GROWMAT Workload Scheduler 1

Search, Filter, and Edit Job

## GROWMAT



The screenshot shows the GROWMAT web application interface. At the top, there is a navigation bar with links for "GROWMAT", "Schedule Dashboard", "Admin Variables", and "Spotfire Dashboard". Below the navigation bar, there is a section titled "Groton Resource Optimization Workload Management & Analysis Tool" which includes a brief description of the tool's purpose and data sources. A large circular graphic features two people looking at a whiteboard with charts and graphs. Below this graphic, there are buttons for "Start Scheduling" and "View Spotfire". At the bottom of the page, there are sections for "CONTACT US" (listing Samuel Zhang and Herbie Bingham) and "MORE INFORMATION" (linking to a Development Plan PDF and GROWMAT Github). The URL "jcapacity.pfizer.com/schedule" is visible at the very bottom.

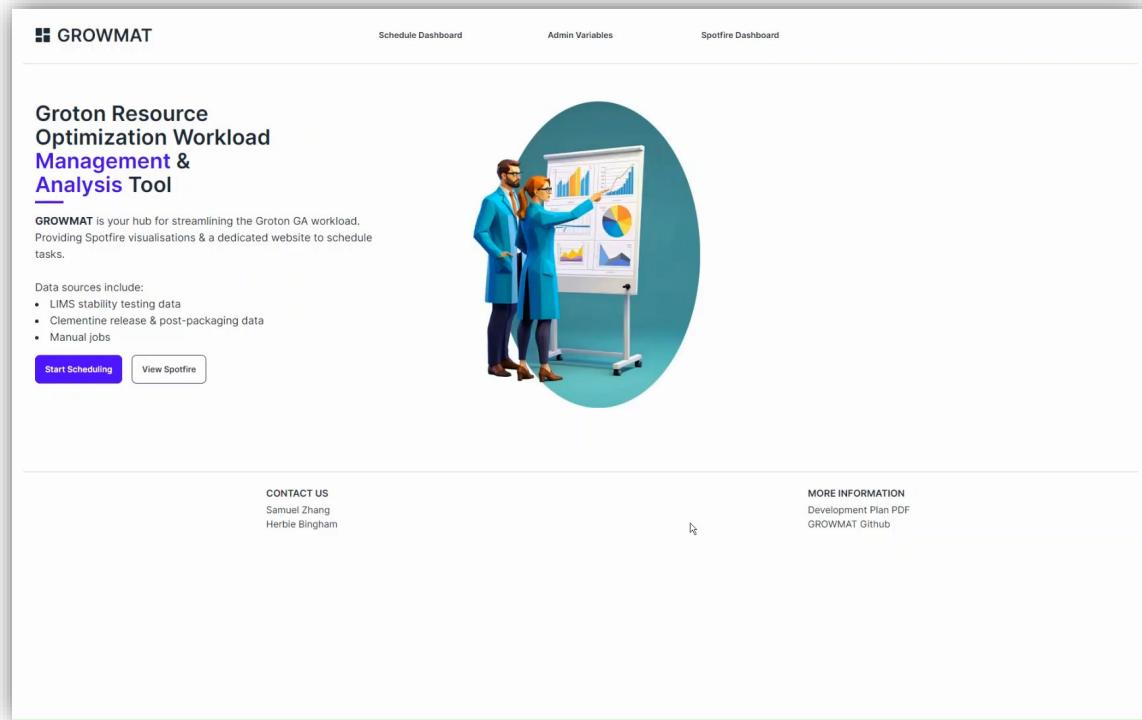
## GROWMAT

- GROWMAT Webapp Allows the user to quickly filter all scheduled jobs, search through them, and easily edit any job status, due dates or team assignment with a few clicks.
- Since all the calculations are done in the cloud, the user will not experience any slowdown unlike the Legacy tool.

# GROWMAT Workload Scheduler 2

## Adding Manual Jobs

### GROWMAT



### GROWMAT

- An easy way to add any additional manual jobs.
- The restrictive nature of the webapp and forced field validation is for data integrity. Only allowing users to change one field at a time and disallowing wildcard characters prevents users breaking the backend calculation.
- Standard user is unable to delete entries, further enforcing data integrity.

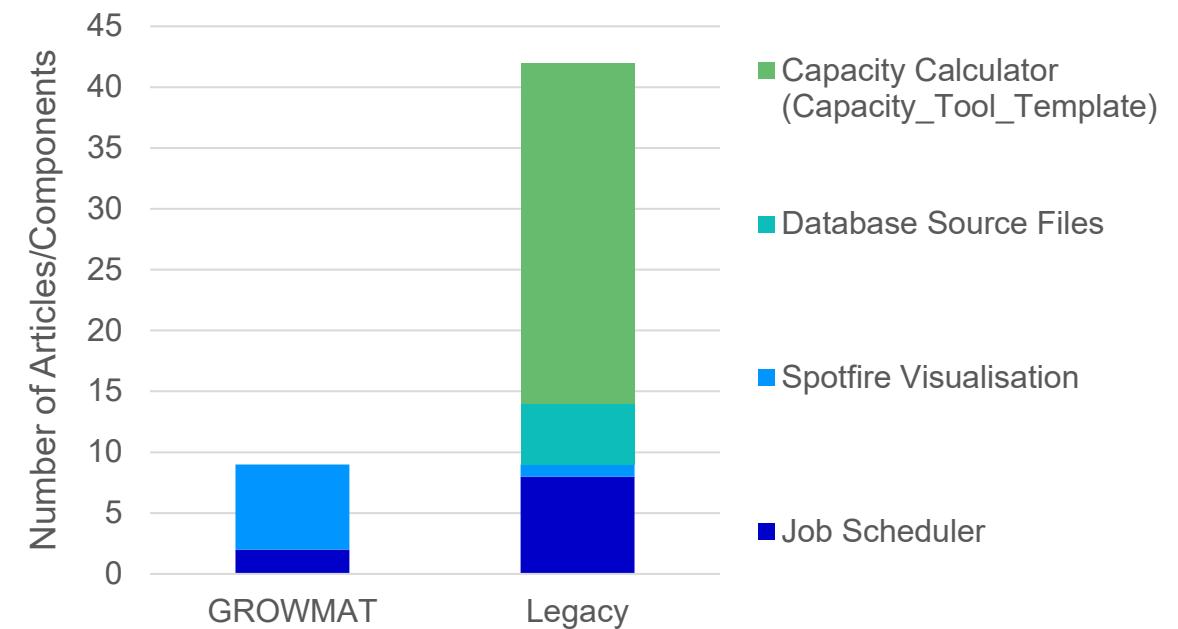
# User Experience

## Highlighting the Benefits of Reducing User Exposed Components

### Benefits of Reducing Number of Articles

- Complex projects have the inherit issue with more points of failure. To minimise the issue, most of GROWMAT's components are not exposed to the end user, reducing the risk of user breaking the system.
- All GROWMAT's user facing components are web based, allowing user to manage, visualise, and edit workload completely through the web.
- The backend of GROWMAT is cloud based and hidden from the user. Any specific requests such as recovering data, adding new features, deleting records should be done via request only to preserve data integrity and system stability.

### User Exposed Components (Lower is Better)



# GROWMAT Use Case Example 1

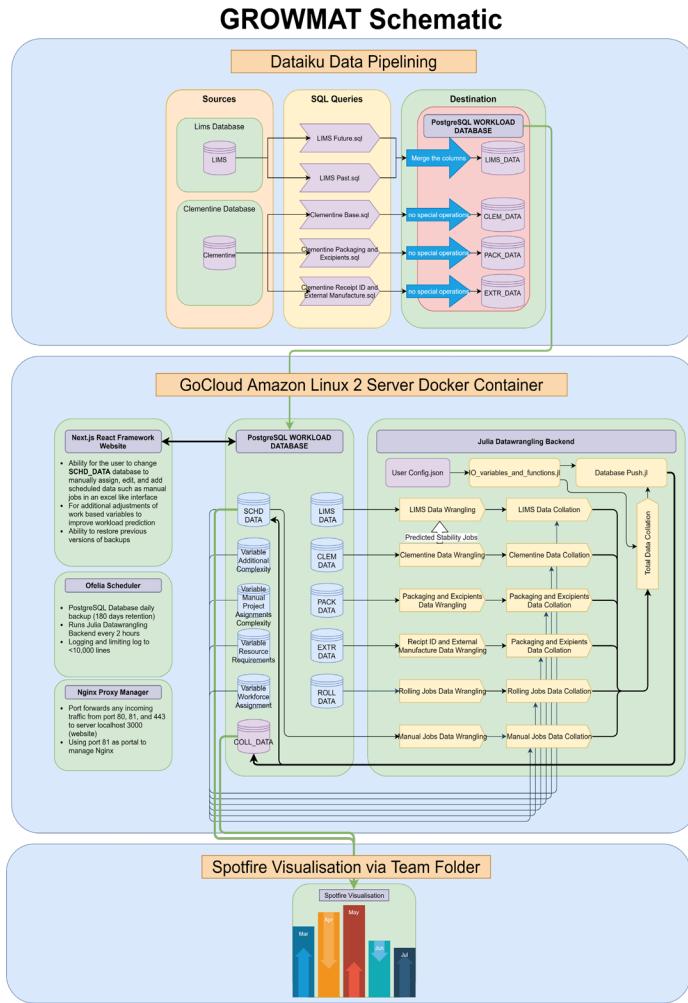
## GROWMAT

- A GA Team Leader felt that there are overwhelming work this month (Sep 2024) allocated to their team. They want to investigate what jobs are causing the issue and would like to reassign the task to another team.

## Demonstration

The screenshot shows two main parts of the GROWMAT system. On the left is the 'Welcome to GROWMAT Version alpha 1.0' page, which includes sections for Introduction, Model Assumptions and Variables, and Authors. It features a background image of a laboratory and a circular logo. On the right is a 'Spotfire' dashboard titled 'Date Shifted Workload Quick View'. The dashboard displays a bar chart of 'Sum(Calculated FTE Days/Duration)' over time, color-coded by 'Team Color + Job Type'. The chart shows a significant peak in workload around September 2024. The dashboard also includes a legend, a data table, and various navigation tabs at the bottom.

# Technical Details on GROWMAT's Backend



## Components of GROWMAT

- GROWMAT has 7 interlinked components functioning together, the data source comes from a **Dataiku Data Pipelining** schedule, the data is then pushed to a **GoCloud Amazon Linux 2 Server Docker Container** with 5 components that performs data storage (**PostgreSQL**), data wrangling (**Julia Backend**), task scheduling (**Ofelia**), and web server hosting (**Next.js on React Framework** and **Nginx Proxy Manager**). The output of which is then visualised using **Spotfire**.



All Photos by Unknown Authors are licensed under [CC BY-NC](#)

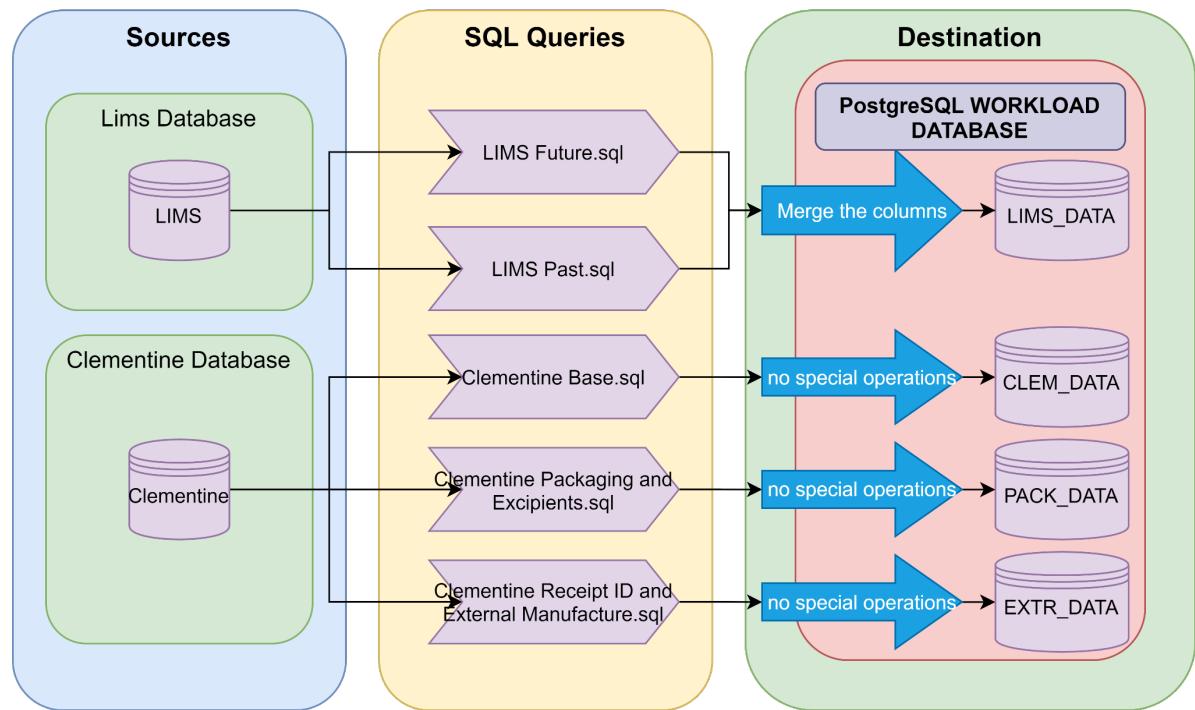
# Component 1: Dataiku Data Acquisition

Obtaining Data from Source Database Independent from GoCloud Server

## Dataiku Pipelining

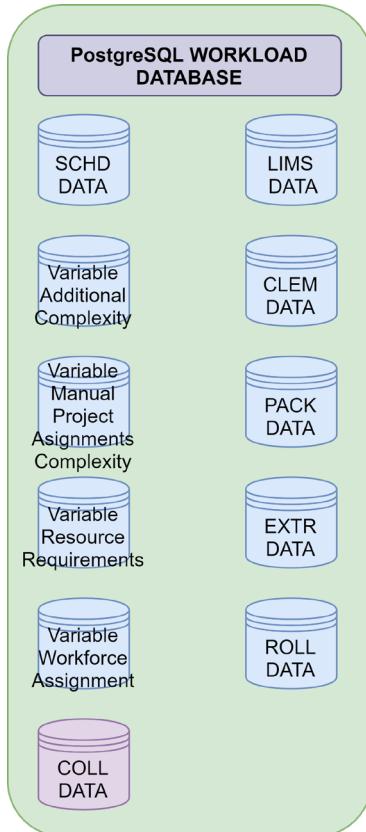
- Current and future project stability data are queried from the LIMS database, aggregated, and put into GROWMAT's own PostgreSQL database by a Dataiku Data Pipelining routine that runs daily.
- Project release data, packaging and excipients data, receipt ID and external manufacturing data are also put into GROWMAT with the same Dataiku routine.

## Dataiku Data Pipelining



# Component 2: PostgreSQL

## GROWMAT: Storing Databases Using Docker Containerised PostgreSQL



### PostgreSQL Database

- GROWMAT lives under the GoCloud Amazon Linux 2 server, within GROWMAT there are a few more components working together to make the entire service function. All components are containerised and deployed using Docker containers allowing easy management, deployment, and redeployment to other servers if needed.
- One of these containers acts as the data storage, it uses industry standard PostgreSQL to store 11 tables.
- Every column has data type, collation defined for data integrity. Every table has a unique key column.

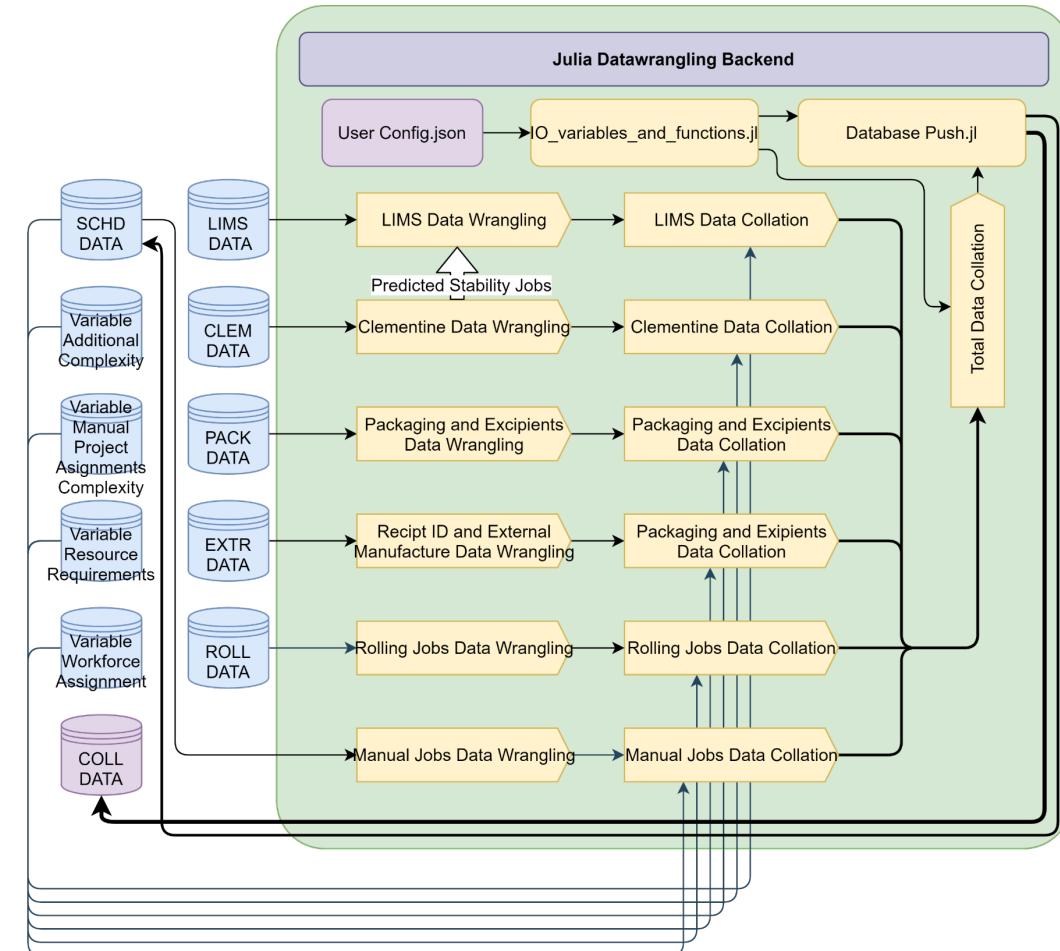
Data Table Name	Database Function
LIMS DATA	Project stability data
CLEM DATA	Project release data
PACK DATA	Project packaging, excipients data
EXTR DATA	Project external manufacturing, receipt ID data
ROLL DATA	Rolling jobs that repeats monthly
Variable Additional Complexity	User adjustable project counts economy of scales modifier
Variable Manual Project Assignments Complexity	Stores the list of all currently available projects and their individual complexity modifier
Variable Resource Requirements	User adjustable project types dependent FTE days and cycle time
Variable Workforce Assignment	Stores the current workforce (employees) FTE allocation and team assignments
COLL DATA	A full output database for Spotfire visualisation
SCHD DATA	A simplified output database for job assignment

# Component 3: Julia Backend

Background Scheduled Data Processing Using Docker Containerised Julia

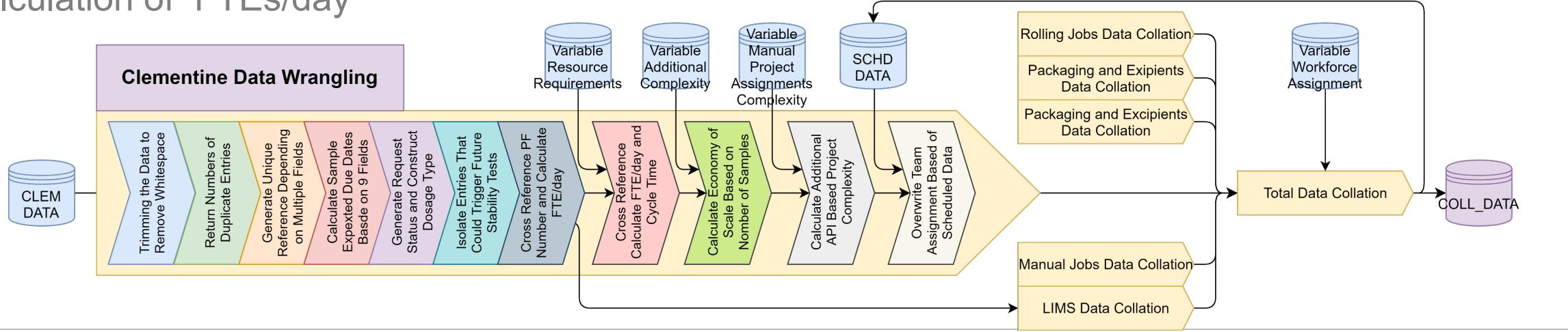
## Containerised Julia Data Wrangling Backend

- Julia is a high-level programming language like Python with the speed of C. Choosing Julia for data wrangling dramatically reduced the computational time needed.
- Data from 10 databases are fetched, calculated for FTE/Days depending on a variety of variables, and populated as **COLL DATA** for data visualisation and **SCHD DATA** for GROWMAT's workload scheduler.
- Julia service is containerised along with the PostgreSQL database, allowing seamless connection between the two without the need of configuration.



# Component 3: Julia Backend

## Calculation of FTEs/day



## Data Wrangling Pipeline (e.g. Clementine)

- Using Clementine databases as an example, raw data is fetched from **CLEM DATA**, formatted, removed duplicates, and assigned unique references. Expected sample due dates are calculated for each job depending on specific fields. Entries that might result in future stability tests are identified and sent to the process that performs LIMS data wrangling. The **FTE Days** and **Cycle Time** each type of job take is referenced from **Variable Resource Requirements**, additional economy of scale adjustment calculations are referenced and performed to the jobs' **FTEs/Day** from **Variable Additional Complexity**. **Variable Manual Project Assignments Complexity**, a PF number specific adjustment parameter is also referenced for any last-minute complexity multiplier for the **FTEs/Day** calculation. Scheduled data database **SCHD DATA** with user assigned Teams and **FTEs/day** then overwrites the automatic source. **Variable Workforce Assignment** calculates the current overall workload capacity. Wrangled data is then sent to **COLL DATA** and back into **SCHD DATA**.

# Component 4-5: Ofelia Task Scheduler and Nginx Proxy Manager

Other Docker Containerised Services Essential to GROWMAT's Functionality

## Component 4: Ofelia Task Scheduler

- Ofelia Task Scheduler schedules tasks to routinely run based on a timed schedule.

Name	Function	Schedule
Julia Backend	To process data	Every 2 hours
PostgreSQL Database	To back up the PostgreSQL Database	Daily with 180 days of retention
Log Trimmer	To keep the logfiles' size down	Every 3 Days

## Component 5: Nginx Proxy Manager

- For the website to be accessible to Pfizer internal network, Nginx Proxy Manager points the domain name <http://gacapacity.pfizer.com/> (only accessible via Pfizer intranet) to the web server within the docker container.



# Component 6: GROWMAT Web service

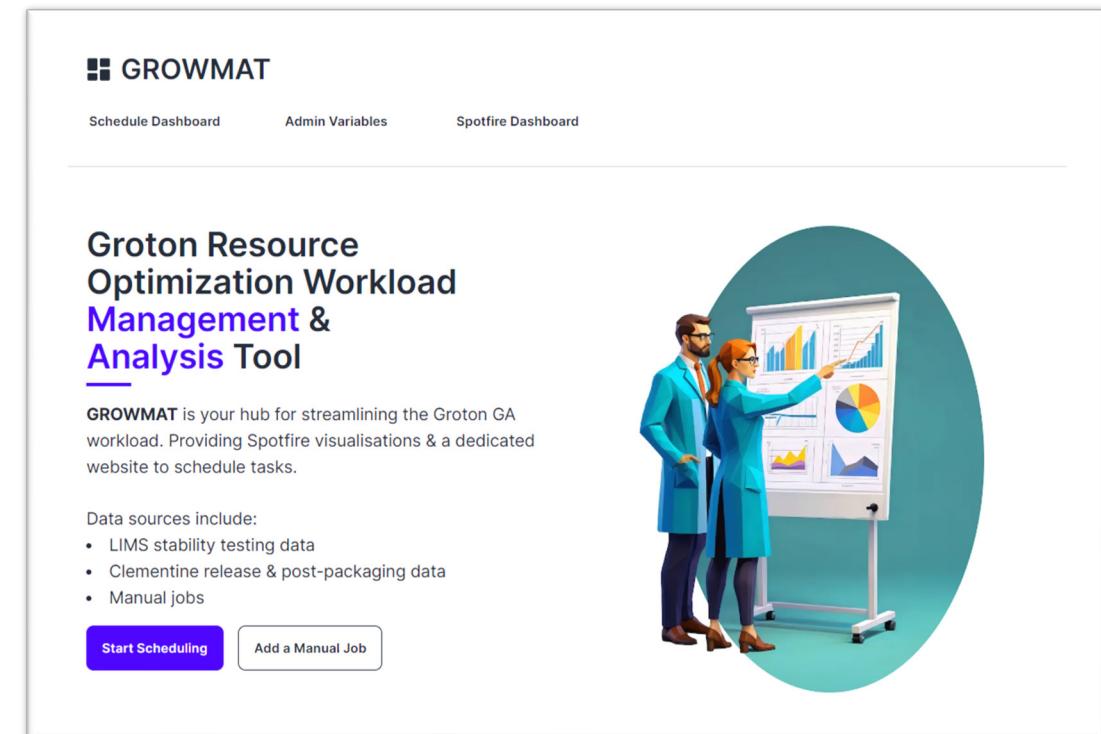
Docker Containerised Full Stack Web Server

## The Technology Stack

- Next.js is a full stack framework based on React, supported by Vercel
- It runs on top of Node.js
- Provides frontend and backend services, allowing for data fetching, caching, route handlers, service actions & more



<http://gacapacity.pfizer.com/>

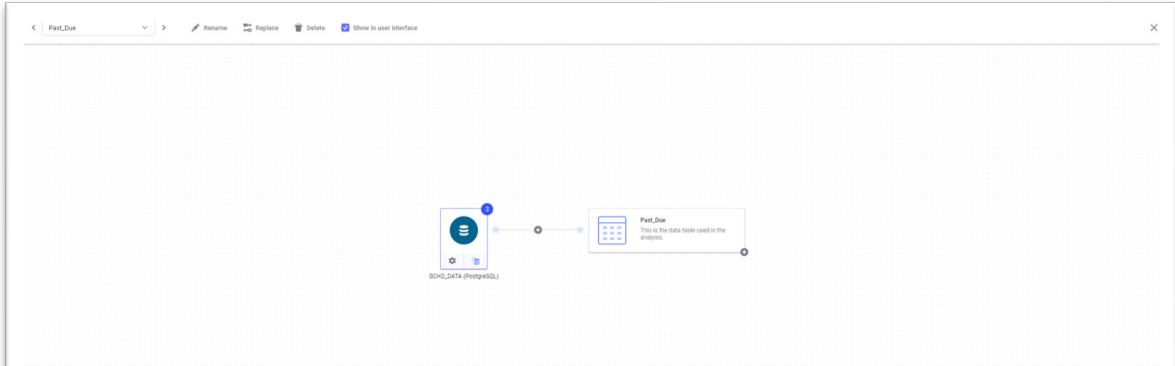


The screenshot shows the GROWMAT web application interface. At the top, there's a navigation bar with the GROWMAT logo, "Schedule Dashboard", "Admin Variables", and "Spotfire Dashboard". Below the header, the title "Groton Resource Optimization Workload Management & Analysis Tool" is displayed in a large, bold, blue font. A descriptive paragraph follows, stating: "GROWMAT is your hub for streamlining the Groton GA workload. Providing Spotfire visualisations & a dedicated website to schedule tasks." To the right of the text, there's a circular illustration of two people in lab coats looking at a whiteboard with various charts and graphs. At the bottom of the page are two buttons: "Start Scheduling" and "Add a Manual Job".

# Component 7: GROWMAT Spotfire Visualisation

Spotfire Visualisation Independent GoCloud Server

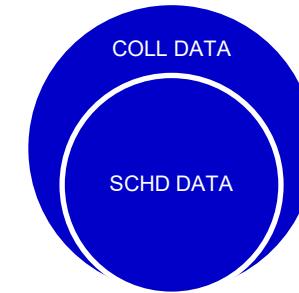
## Spotfire Data Canvas View



The screenshot shows the Spotfire Data Canvas interface. At the top, there's a toolbar with icons for Paste, Rename, Replace, Delete, and Show in user interface. Below the toolbar, a data flow diagram is displayed, showing a connection from a database icon labeled 'SCHD\_DATA (PostgreSQL)' to a data table icon labeled 'Past\_Due'. A tooltip for 'Past\_Due' states: 'This is the data table used in the analysis.' In the bottom half of the screen, there are two tabs: 'INFORMATION' and 'DATA'. The 'INFORMATION' tab shows a preview of the data with three rows highlighted: '1. Filter rows', '2. Filter rows', and '3. Filter rows'. The 'DATA' tab displays a detailed table with numerous columns, including Job Details, Job Reference, Sample Average, Test due date, Sample Duration, Manual Due Date, Lifecycle, Analyst, Reviewer, Notes, is urgent, QTS record, Job Status, Number of Singleleaf, FTE Days, Team, Age of Job, Product, and Mfg. The table contains many rows of data, mostly 'Complete' status entries.

## Spotfire Visualisation

- The Spotfire Visualisation fetches from the GROWMAT's PostgreSQL database tables **COLL DATA** and **SCHD DATA**. The only difference between the two databases is **COLL DATA** contains 53 columns, and **SCHD DATA** contains all the same rows as **COLL DATA** but with only the necessary fields essential for job scheduling. With the two databases, the end user can build any visual or apply any custom filter without affecting the database.



# Permissions Table

A Comprehensive Table Detailing the Permissions and Tasks Assigned to Each Role

## Standard User Permissions

1. View, add manual jobs, and edit (not delete) existing jobs in GROWMAT Scheduler
2. View Spotfire Visualisations
3. Download Schedule for user's own analysis

## Elevated User Permissions

1. All the permissions of the Standard User
2. Granted database elevated permission user account, access databases using DBeaver
3. Edit locked fields on the GROWMAT Scheduler, and delete entries under **COLL DATA** and **SCHD DATA**
4. Edit, add, and delete specific economy of scale multiplier under **Variable Additional Complexity**
5. Edit, add, and delete additional PF numbers under the **Variable Manual Project Assignments Complexity** data table
6. Edit, add, and delete project type specific FTE Days and cycle time under **Variable Resource Requirements**
7. Edit, add, and delete employee FTE availability within teams under **Variable Workforce Assignment**

## Admin Permissions

1. All the permissions of the Elevated User
2. Read, write, delete, create database, grant roles options for all database items
3. Able to roll back database recovery up to 180 days
4. Able to roll back the server as disaster recovery plan
5. Able to change any code base and redeploy for any other server

# Acknowledgements and Conflicts of Interest

GROWMAT

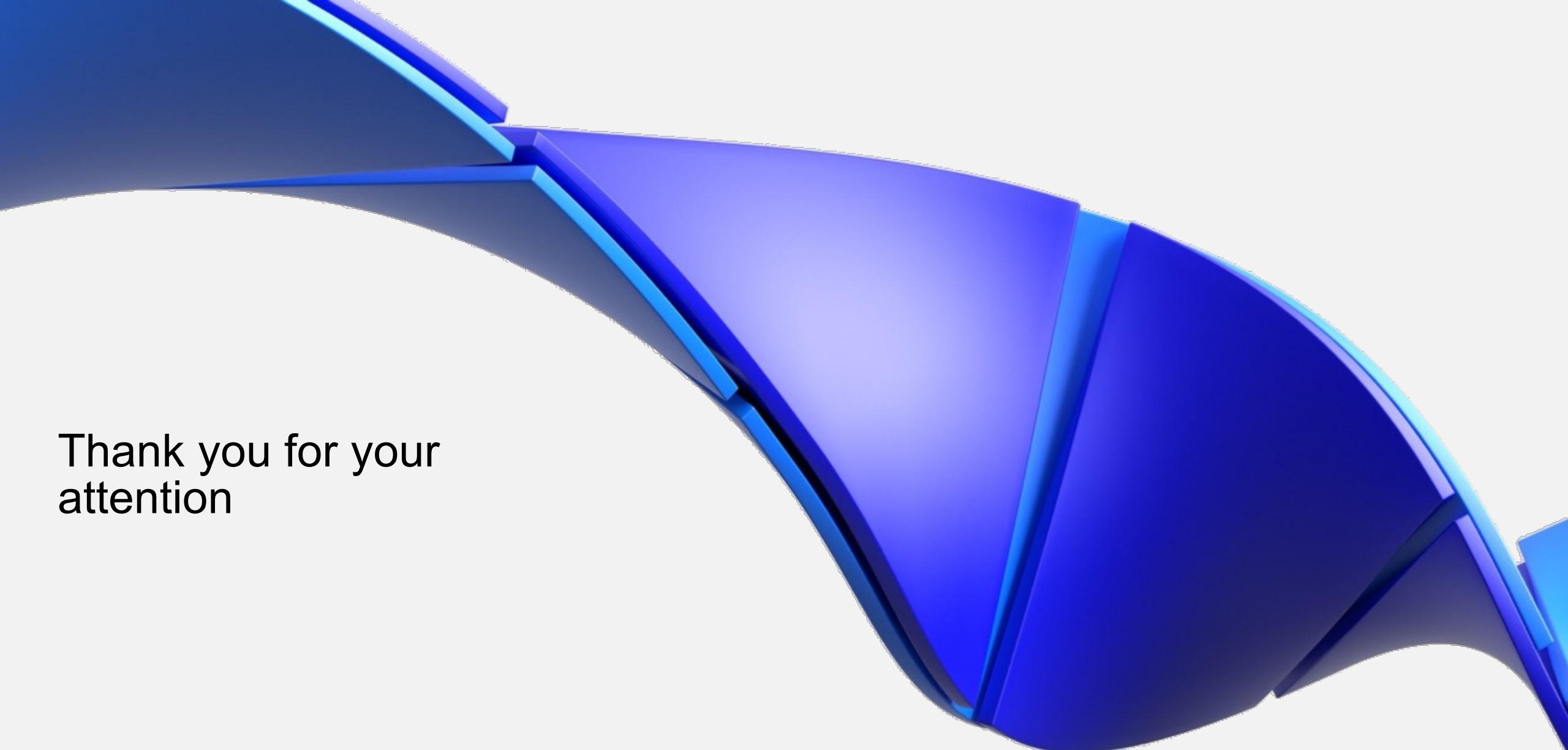
---

## Acknowledgements

- *Zhang, Samuel*
- Bingham, Herbie
- Matthew Summers
- Du, Bo
- Carrick, Paul
- Vargas, Jose
- Gomersall, Stephen
- Chang, Qing
- Glover, Darren
- Dyer, Dana Marie

## Conflicts of Interest

- None



Thank you for your  
attention