Paper PP23

MODERN ANALYTICS WITHIN AN OPEN SOURCE FRAMEWORK

Sam Warden, d-wise, Geneva, Switzerland
Ali Dootson, d-wise, Manchester, UK

ABSTRACT

As sponsors increasingly look at providing open source technologies like R and Python for their biostats and programming teams, we provide an overview of the strategies to ensure this can be implemented and validated efficiently. Based on client collaborations and partnerships, we've created an innovation environment that tests and validates these technologies, maximizing analytic productivity within traditional Life Sciences confined systems.

INTRODUCTION

This presentation will look at how technology and analytics for clinical trial analysis are constantly evolving and dive into the latest innovative technology and methodologies for process optimization and change management to maximize productivity today, while future-proofing your analytic computing environment for tomorrow.

This provides an overview of the strategies to ensure the most relevant procedures are evaluated, implemented and validated efficiently and cost effectively.

TRENDS

Pharmaceutical companies have historically favoured a path of least risk with closed off the shelf packages for clinical analytics which profess to satisfy CFR21 part-11 regulation. However, these systems do not address the fact that a user would much prefer a local copy of PC SAS, hitting data on a file share, enabling them to work in a flexible and agile way. Users value freedom and speed of work over lock-in to an environment that does not meet their needs. It would be much preferable if an environment were designed for extensible and custom work applications (that can be modified and added over time) than to be locked into a closed environment. This insight implies that a more flexible infrastructure is needed, and new ways of thinking about what constitutes an application framework within a regulated environment.

This change in attitudes regarding compliant systems is happening (slowly) right now. Several companies are exploring rapid, agile development practices, and validating the applications as "black-box" at the end of the process. This allows a move away from purchase-install-validate to a develop-validate approach followed by a verify-deploy-support model. They are starting to view clinical trials as a form of "DevOps" pipeline where R code and data are kept in repositories and updates are managed by a automation tools .

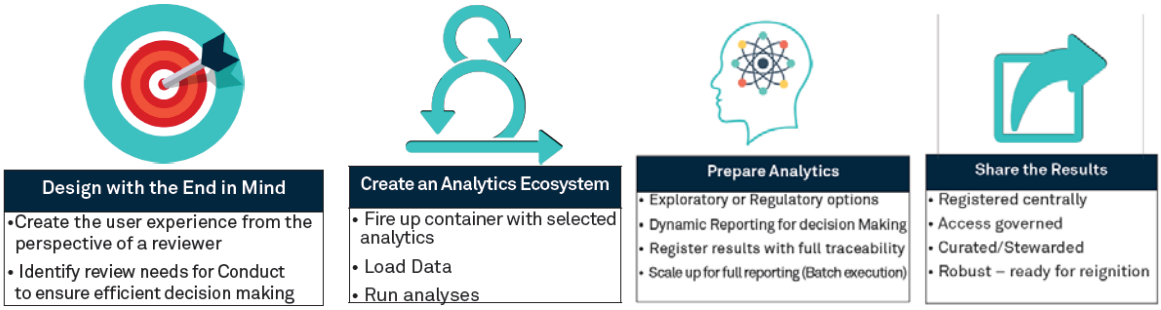
The transition to open source and away from a singular analytical platform for pharma R&D is already in motion and we are seeing that several fundamental technologies (e.g. Kubernetes, R, Github, etc.) being used are already open source and readily available. There are several barriers to adoption that include a concern about validation, lack of know-how in the tools and processes, preference at many large pharma for a buy vs. build approach; but these are not insurmountable.

The market evolution will mirror the adoption cycles such as Hadoop/Cloudera/Hortonworks and Linux/Redhat, albeit likely in a much slower arc in terms of taking the open source approach and making it validated, structured and stable for use in our highly regulated industry

CHARACTERISTICS

An Open Source framework for modern analytics is achievable in a performant and scalable infrastructure and d-wise have designed an architecture that enables this. Flexibility is built in from the very start understanding that no-one wants or needs to be tied into a single vendor for any component part. Innovation is encouraged and enabled through the framework design coupled with a strong agile process and governance to ensure that experimental or

exploratory features and functions can be transitioned into fully validated state in a supported and maintainable process.



DESIGN WITH THE END IN MIND

We need to turn the analysis and reporting paradigm on its head and instead start from the user experience of clinical submission reviewer's perspective. This will drive the information and presentations needed to support the evidence of a clinical trial or submission for the safety and efficacy in the population of interest. What is needed for the reviewer will then drive what data are captured, transformed, analysed and reported. Streamlining the process and cutting out redundant steps.

Identifying the review needs for study conduct to ensure efficient decision making is enabled throughout the course of the study with enrichment of analyses around decisions taken and the rationale behind it.

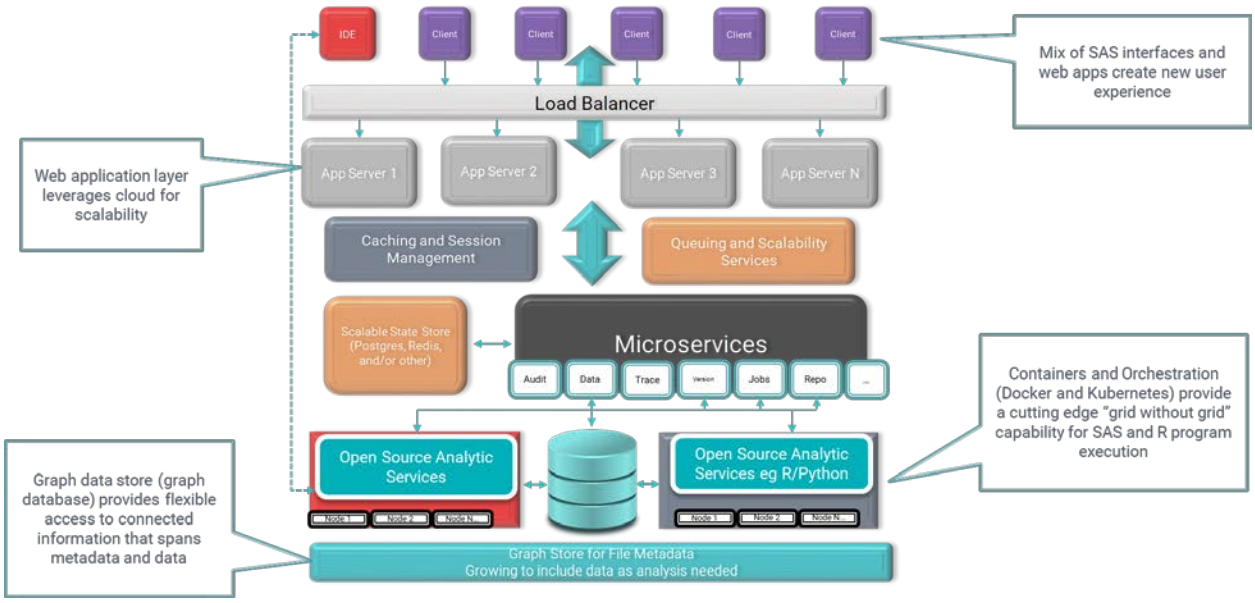
CREATE AN ANALYTICS ECOSYSTEM

They key to achieving this vision is to have a horizontally scalable, validated, cloud-agnostic deployment (i.e., we can deploy to multiple cloud environments, and enable the optional use of "hybrid/more than one" cloud environment(s),and migrate between environments easily) which has a set of APIs to deliver traditional clinical service capabilities; Auditing, reporting, user management, file system operations, task coordination, versioning, and others . These API's should be accessible to the end user through intuitive and user centric interfaces.

Storage of data must also be flexible, scalable and high performant to allow for easy access to analytical compute engines. Data should be stored in a format which is readable by multiple languages, is audit trailed and tagged at the value level for context, history, privacy/consent and validity. Whilst the requirements today are to produce SDTM and ADaM data in legacy SAS transport files, we can think of this as an output or rendition of data that are stored for operational purposes thus relieving organizations of the overheads of working with data which is not structured for re-use and analytics agility. Data feeds from capture should be enabled easily through the 'clinical' APIs describe above.

This ecosystem will enable the user to fire up a container with their selected analytics tools and languages, access or load the data with the valid view for their purpose (blinded/unblinded, fully consented for the purpose etc.) and then execute the analyses according to the plan.

The diagram below provide an example of an architecture that provides this flexibility in the data storage, API, and clinical 'tools' to support this approach.



PREPARE ANALYTICS

According to the purpose or intent of the analysis to be performed, a data scientist should be able to select the tools, functions or applications they need to explore and analyse the data. Having the ability to reproduce the journey or path as well as the results is helpful in all analytics, but mandatory for regulated decision support. A system which facilitates this in an agile way with a "switch" to record the state for reproduction is essential and possible.

Providing an application which guides the decision maker or reviewer in a "safe" and recorded manner is a concept under investigation in many pharmaceutical companies today. Technically this is simple to achieve, however requires education of the users not only in how to use the application supplied but also when to call for support or a statistically qualified opinion as well as having controls around how the output can be used.

SHARE RESULTS

Shifting from a paradigm of flat and comprehensive results which are difficult to navigate and decoupled from the journey of decision-making which study teams and investigators have made to determine the safety and efficacy of a medication to a more dynamic reporting concept is central to the modern analytics platform. Enabling a safe and secure way to encapsulate the data (in a view that is "fit for consumption") along with the navigation tools to explore, understand and infer its meaning can be provided with the containerized approach.

Enabling this in a platform where actions are registered for replay, access is managed, and expertise can be called upon helps us to move forward in the way that we leverage our clinical data assets.

CHALLENGES

Leverage Technology and Manage Change: Technology is not the most important component, but you need to know how to get the best from it. With greater flexibility and agility comes responsibility and education and trust are central to being successful. Data science now forms key modules in formal education programs, but teams need to agree on how they work with the data, when and how to seek expertise and support, as well as the necessary knowledge for using statistics and inference. A program of change management must accompany the transition to a modern analytics platform and is the only way that value will be secured.

Manage the Business Change: Maintain business as usual whilst we change through a solid transition and migration plan; enabling the old ways of working and analytics toolkit whilst moving to the new capabilities must be a foundational premise. This is why our architecture includes SAS and traditional data stores as well as open source and graph components.

Maintaining Control: Keeping control (Validation) is essential through the application of an agile and governed process. Automation of testing and strong frameworks around documentation are central themes. We propose a centrally curated service for this.

FUTURE

Unstructured data

The constraints of data which is stored and described in columns and rows is well known. Moving to a concept where data and meaning are stored and managed together will support the development of greater automation through the microservices approach we describe above. A flexible architecture built to support the evolution of both the data concept maturity and the automation approach is central to our proposal.

Process change

Failure to address the siloed nature of clinal data processing and analytics will prevent the industry from gaining full advantage of technological change. Teams need to look to the suppliers and receivers of their process to identify ways to streamline and automate, eliminating steps and interfaces in the process. A continuous improvement approach is recommended to ensure that the technology supports the change and full benefits are achieved.

Focus on collaboration

Thinking about how data will be shared and how teams can work together as they explore and analyse the data requires a change in mindset enabled through a collaborative toolkit. Designing trials with sharing in mind will ensure that you consider re-use and build the necessary controls and filters from the start.

Partnerships are key

Very little can be achieved by thinking that a single vendor, or a single team can or should deliver this as a “solution”. Defining a framework which leverages partnering concepts enables us to believe that this approach will change the way medicines are brought to market and even used in the future.

ACKNOWLEDGMENTS

Chris Olinger, Chief technology Officer at d-wise and Andrew Ratcliffe, Director, Technology Consulting at d-wise developed many of the concepts in this paper and we thank them for their generosity in allowing us to present them as a poster and paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Sam Warden
d-wise
Suite 7A, 53 Portland Street
Manchester, M1 3LD
sam.warden@d-wise.com
www.d-wise.com