

# 现代化临床试验数据科学平台搭建的考量

SAS生命科学技术总监---顾天伦



# Aganda



**临床试验生物统计数字化过程中的挑战**

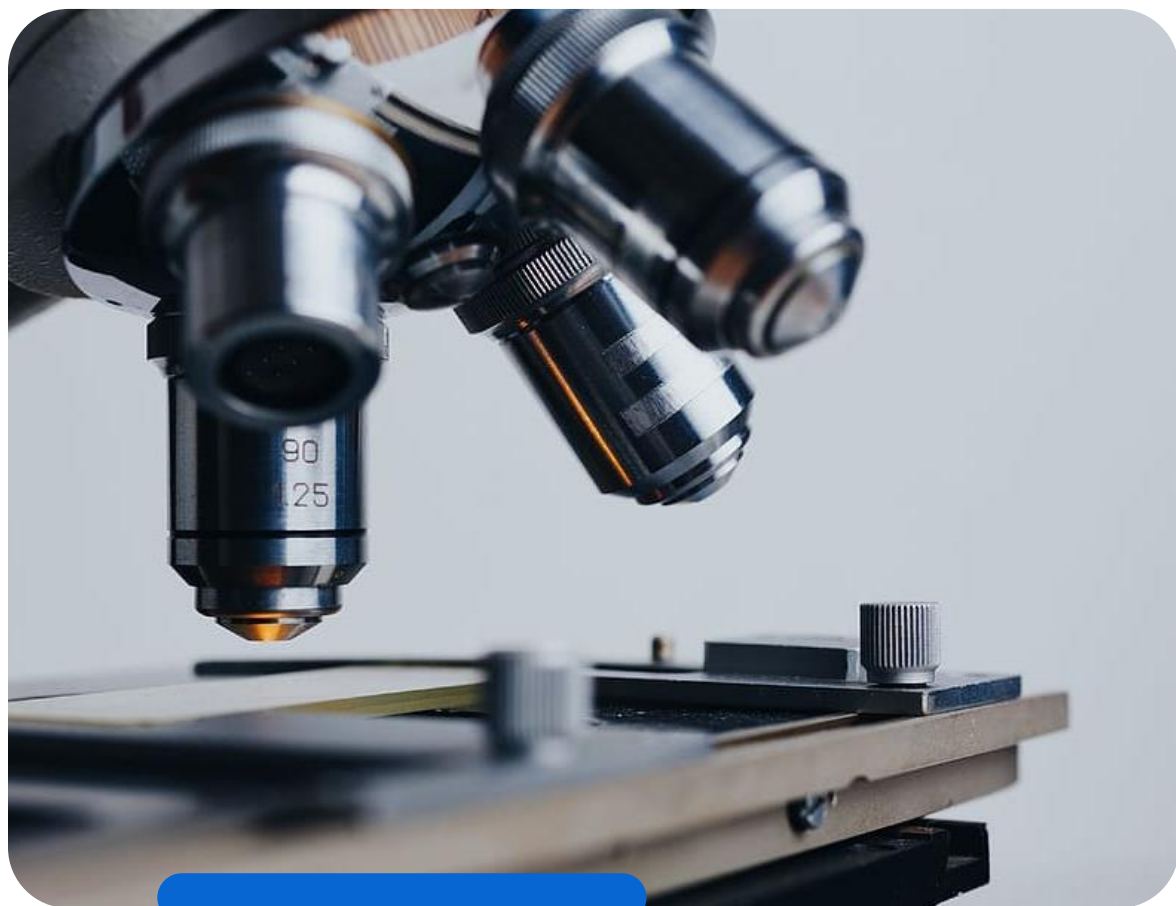
**数据科学在临床试验全生命周期中推广**

**SAS临床试验数据科学平台的设计理念**

**未来生物统计技术平台的展望**

# Top Trends in Life Sciences

## Reshaping Life Sciences in 2024



1

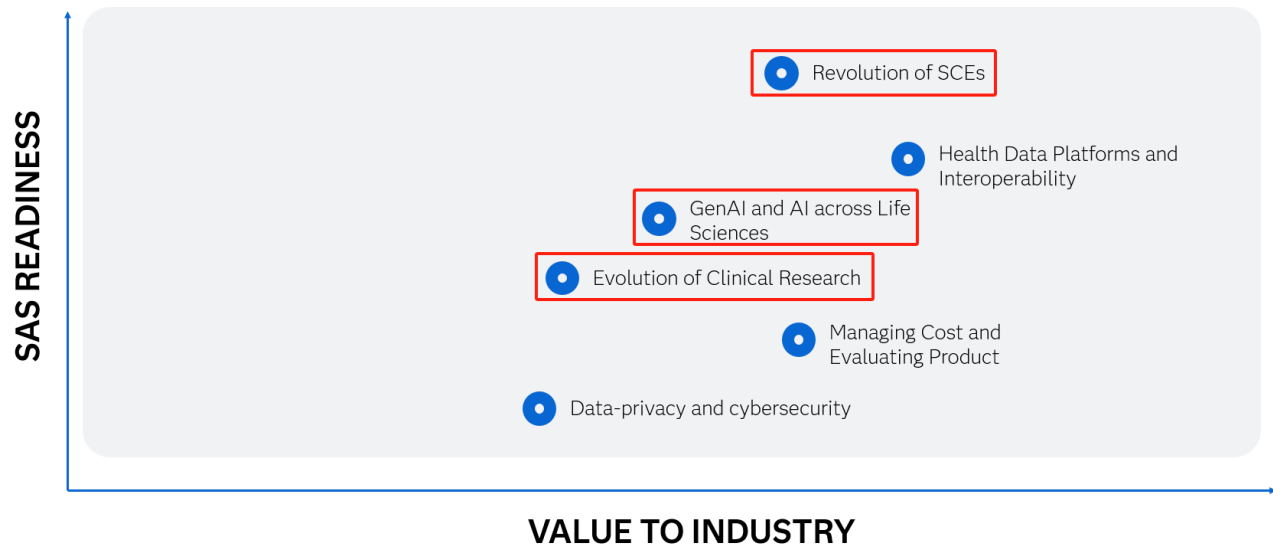
Generative AI and AI across Life Sciences including patient engagement, protocol design, drug delivery, monitoring, manufacturing, and more

2

A Revolution of Statistical Compute Environments powered by Big Data and cloud technologies

3

The Next Evolution of Clinical Research driven by scientific advancements and new technologies, and accelerated by the use of synthetic data and real-world evidence



# PhUSE SCE White Paper – Key Requirements

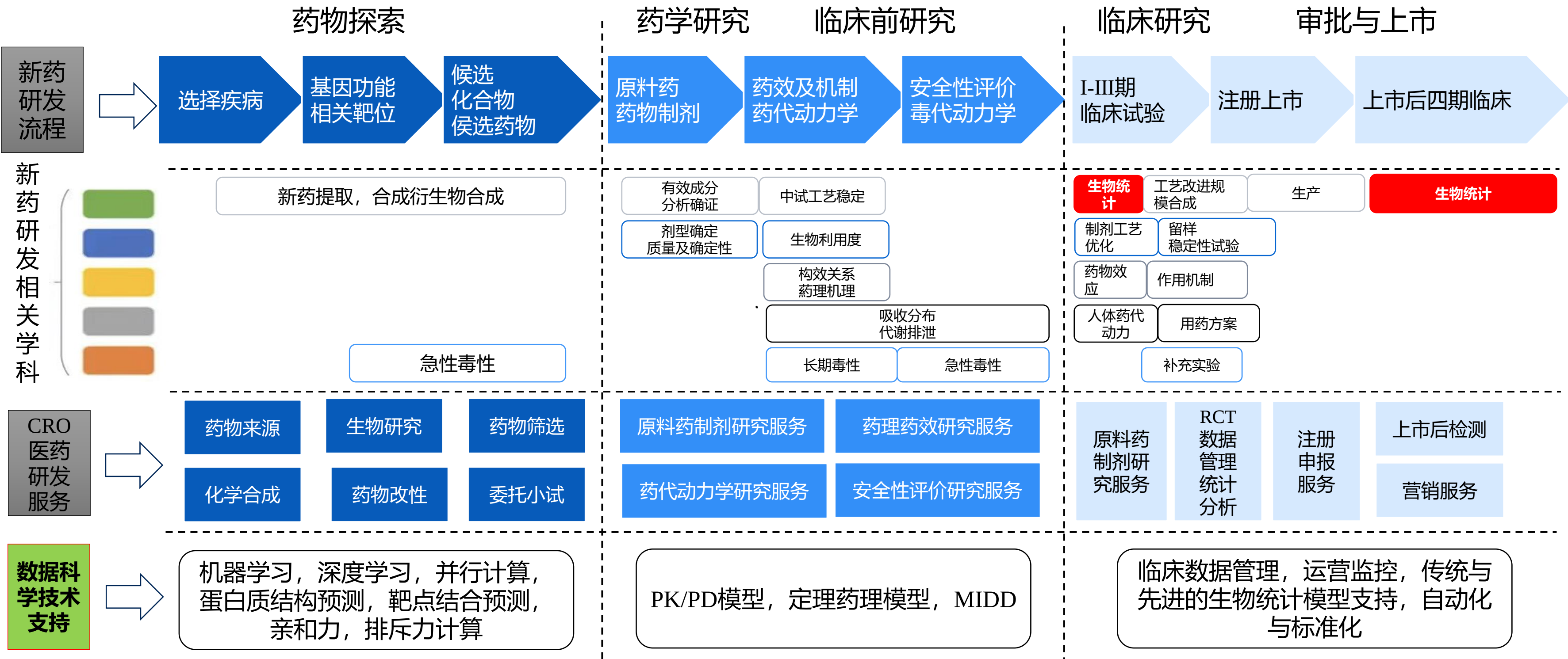


For simplicity, the analytics environment will be referred to as the Statistical Computing Environment (SCE) throughout this paper. To construct a fully functional SCE that meets evolving business need whilst maintaining regulatory compliance, it must:

- Be an application that is intuitive and easy to use
- Involve modern and scalable forward focused technology. Adopt technology that is reliably used for big data
- Facilitate multiple programming languages
- Effectively integrate with a variety of systems and tools to allow seamless access to data
- Successfully manage programming workflow e.g., version control, traceability, reproducibility
- Allow effective outsourcing and collaboration
- Include automated functionality, AI and machine learning to replace repetitive and manual-heavy tasks
- Include machine learning to facilitate decision making
- Enable high compute power (high performance computing capability) to enable statistical methods that require complex simulations to be run in minutes
- Allow the capture and reporting of essential business-related metrics to help track value generation over time, as well as identify possible opportunity areas to drive further value
- Enable search services and dashboards with drill down functionality
- Provide capability to anonymize data and documents
- Seamlessly manage access provisioning and review
- Easily integrate with sponsor data structure ecosystem
- Support data and document reproducibility and archiving
- Use metadata to enable the search and re-use of code
- Be a high availability cloud-based computing platform

[https://phuse.s3.eu-central-1.amazonaws.com/Events/2021/US+Connect+2021/SCE\\_White\\_Paper\\_Final\\_15Dec2021.pdf](https://phuse.s3.eu-central-1.amazonaws.com/Events/2021/US+Connect+2021/SCE_White_Paper_Final_15Dec2021.pdf)

# 新药研发的全流程分析



# 临床试验过程中相关组织和机构的数字化挑战

## 患者招募难

- 患者招募多依赖机构
- 患者招募的准确性低
- 临床入组速度慢

## 数据来源复杂

- 数据来源繁多
- DCT增加收集源头
- 人工整合效率低下

## 企业临研项目分散

- 自己管理+外包CRO
- 数字化系统组合多样
- 项目文档存储位置分散

## 海量数据如何利用

- 累积数据挖掘
- 分析模型构建
- 风险反馈决策

## 数据标准不统一

- 数据库结构不一
- 非结构化数据难利用
- 传输标准需要统一

## 无数据集成平台

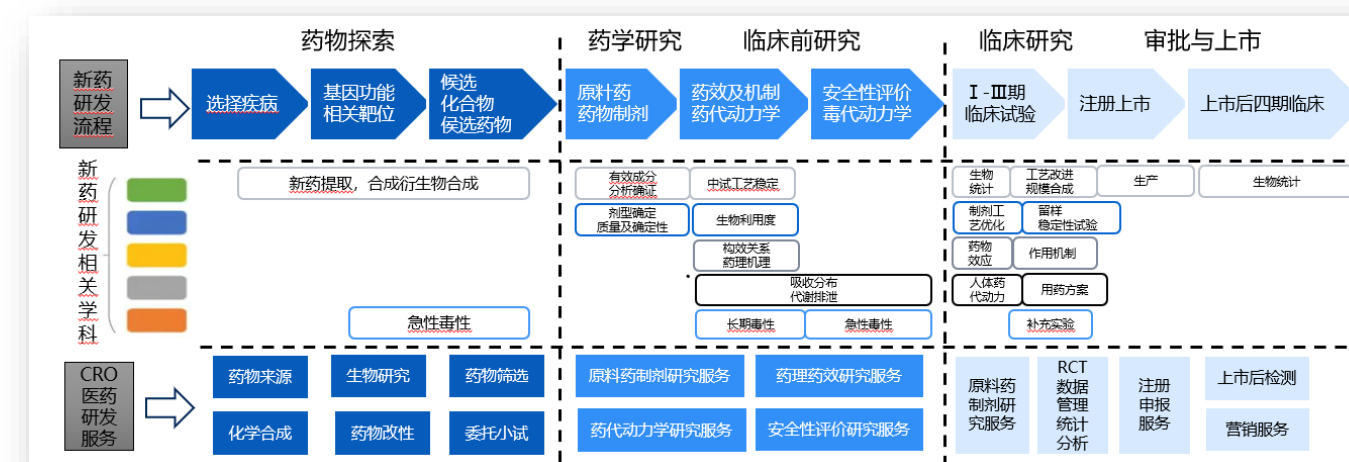
- 无集成信息化平台
- 数据副本留存散乱
- 多种格式数据传输

## 数据人工采集耗时长

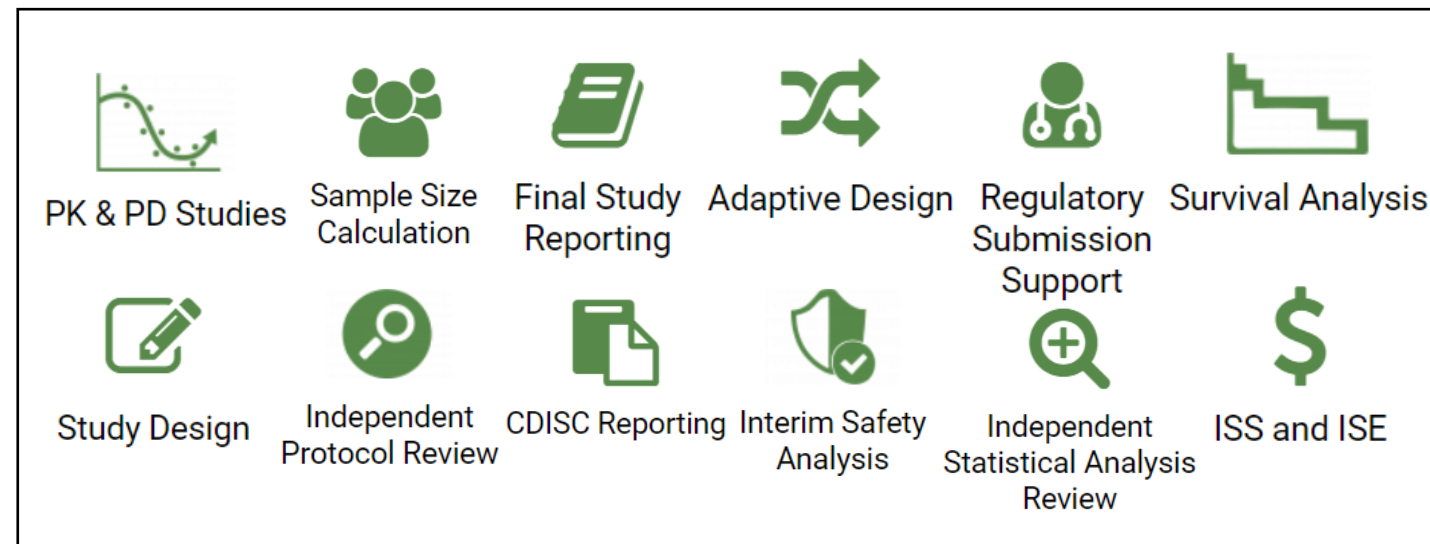
- 临研数据靠人采集
- 准确性需多角色审阅
- 系统割裂导致重复采集

## 缺少可行性评估

- 数据是否包含研究所需要的全部影响因素?
- 缺失数据的发现与处理

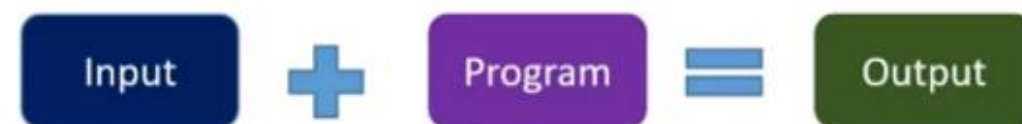
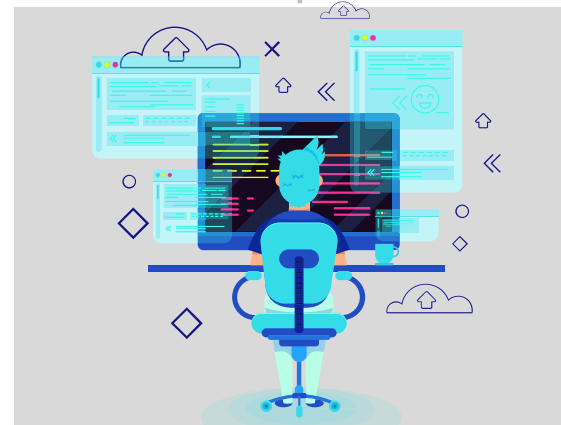


# 现阶段临床试验生物统计平台的痛点



## Traditional Analytics Methodology

- Static data
- Fixed analysis plan
- Small sample size
- Based on hypothesis
- Interpretability of the model



Traditional Software Model

## 新方法/新算法的挑战

- 效率挑战：人工决策VS基于数据预测结局
- AI自适应能力，根据数据、结构和参数进行迭代演进
- 传统RCT应用通过硬编码技术实现研究目标

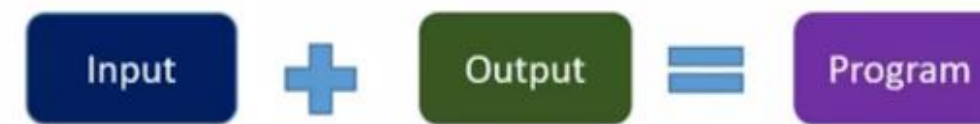


## 工程化/系统化的挑战

- 自动化：用户级/项目级的全流程自动化能力的搭建
- 生物统计研究结果的跨部门，跨团队的共享，以及向监管递交的敏捷性
- 通用应用程序的发布与分享（医学/DM/PV的便捷使用）
- Rule-Based & Model-Based Decision

## Advanced Analytics Methodology

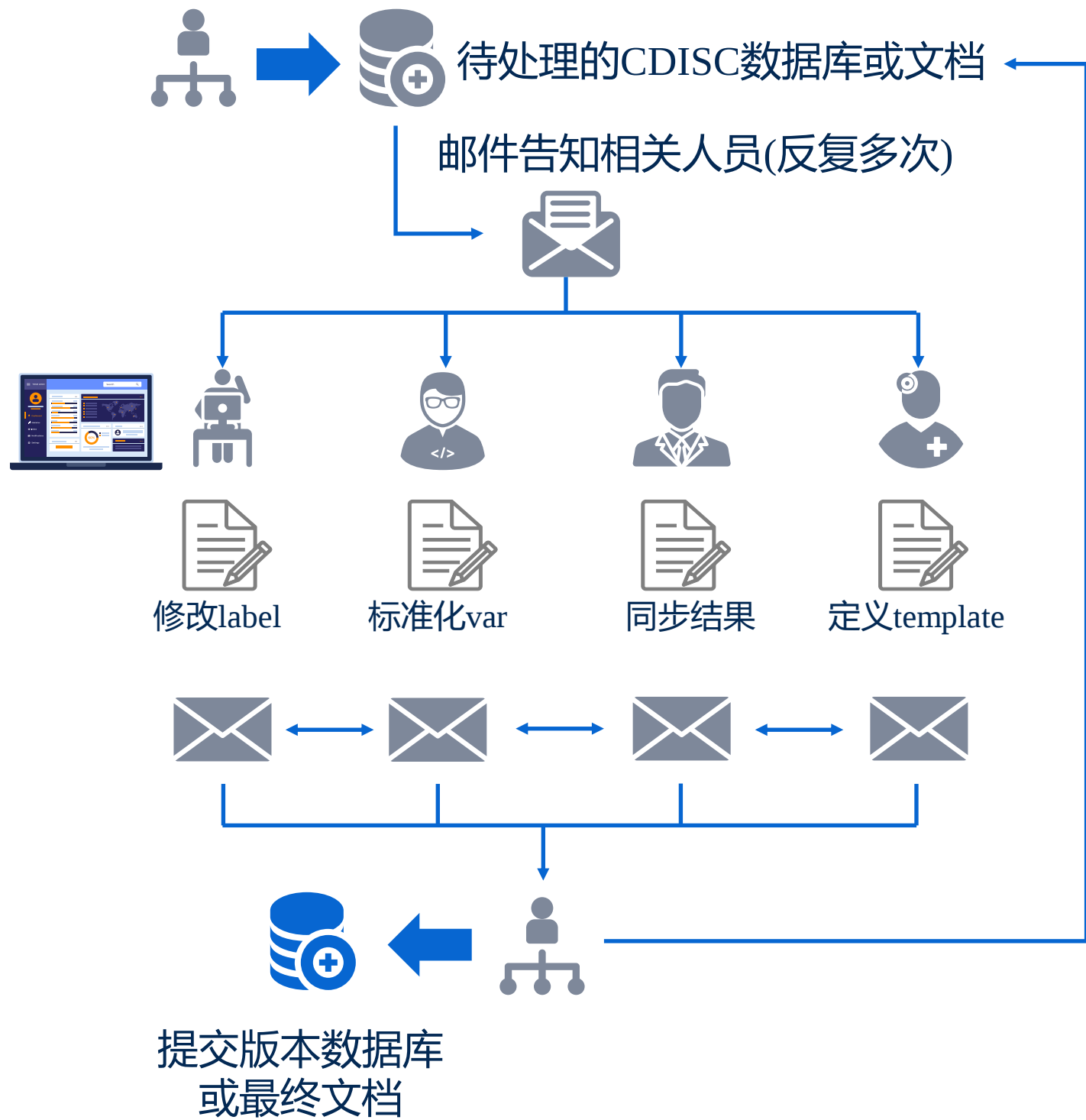
- Dynamic data
- Fitting the optimal model
- Big data from different sources
- Data-driven
- Focus on model accuracy and relevance



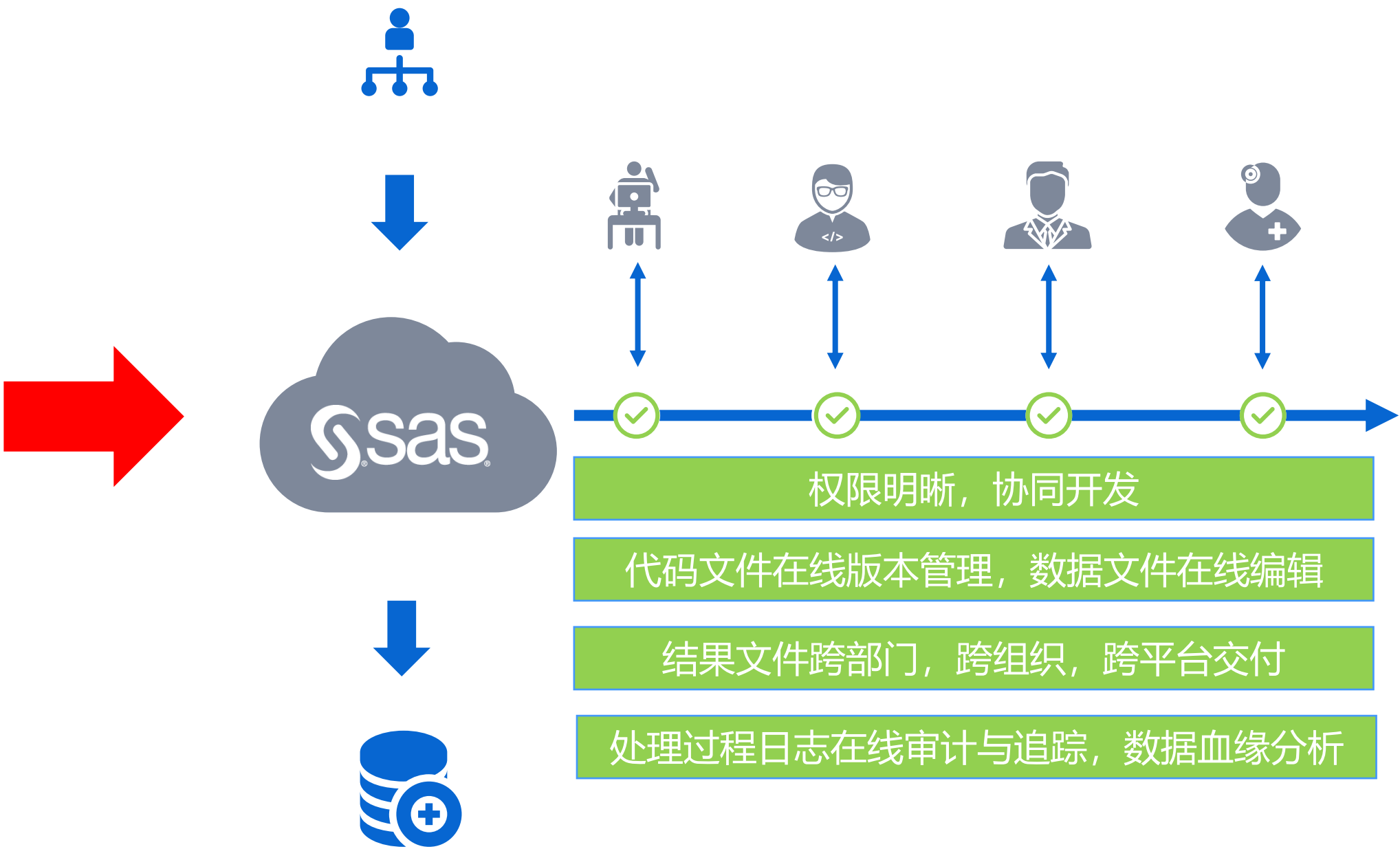
AI/ML Based Models

# 现阶段临床试验数据及文档管理的痛点

传统文件系统的管理



现代化文件系统的管理



# 现阶段临床试验运营核查数据的痛点

## 试验过程数据

1. 受试者数据是临床研究的核心内容,数据管理员通过整体比较各访视数据更易甄别其中异常数据点
2. 传统临床试验, 数据管理员需逐页翻阅病例报告表 (CRF) 或浏览录入数据库才能获得受试者全部数据内容
3. 采用eCRF仍需逐个表单查阅才能获取受试者整体数据
4. 受试者状态在整个试验过程中数据管理员、申办方和研究者无法快速实时汇总查看

## 疗效数据

1. 临床试验研究目的不同其疗效指标各不相同, 需要对某些关键指标的重点追踪分析
2. 关键指标的重点关注逻辑的差异, eg单一指标的前后比较, 正常值的上下限范围, 临床试验结果趋势

## 安全性数据

1. 在临床试验过程中,药物的安全性数据和疗效数据同等重要,安全性数据通常包括实验室安全性指标及不良事件等
2. 对实验室安全性指标通常以文字或列表的形式进行描述
3. 对于不良事件 (AE) 的分析通常只列出AE记录并描述AE发生率,很难发现隐含在这些数据背后的药物安全性信号

# 临床数据管理和数据分析流程

研究计划

创建协议文件

批准或者签署

研究设计

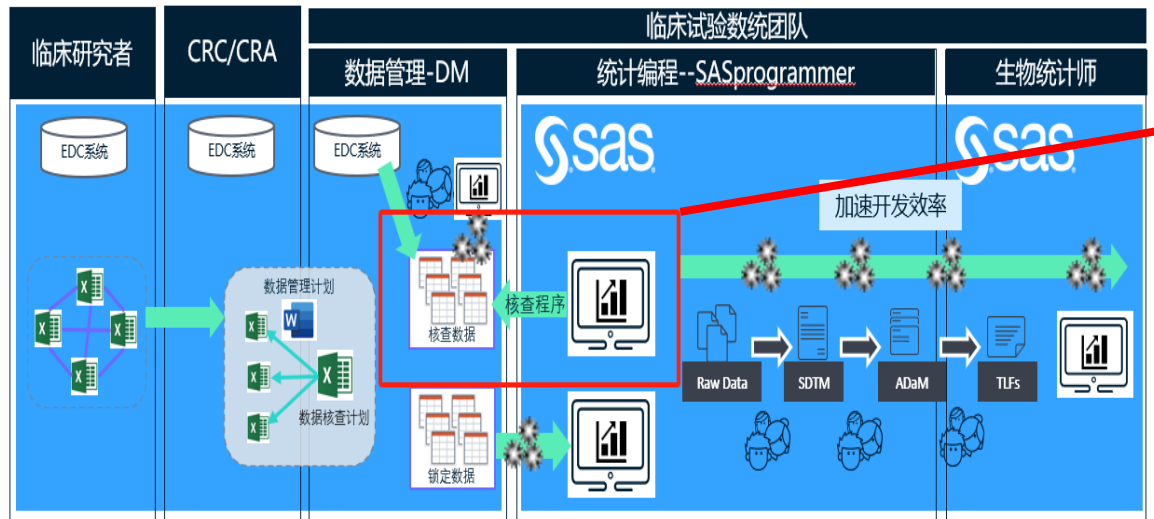
DB Setup

CRF表格设计

编辑检查

验证

数据管理/审核

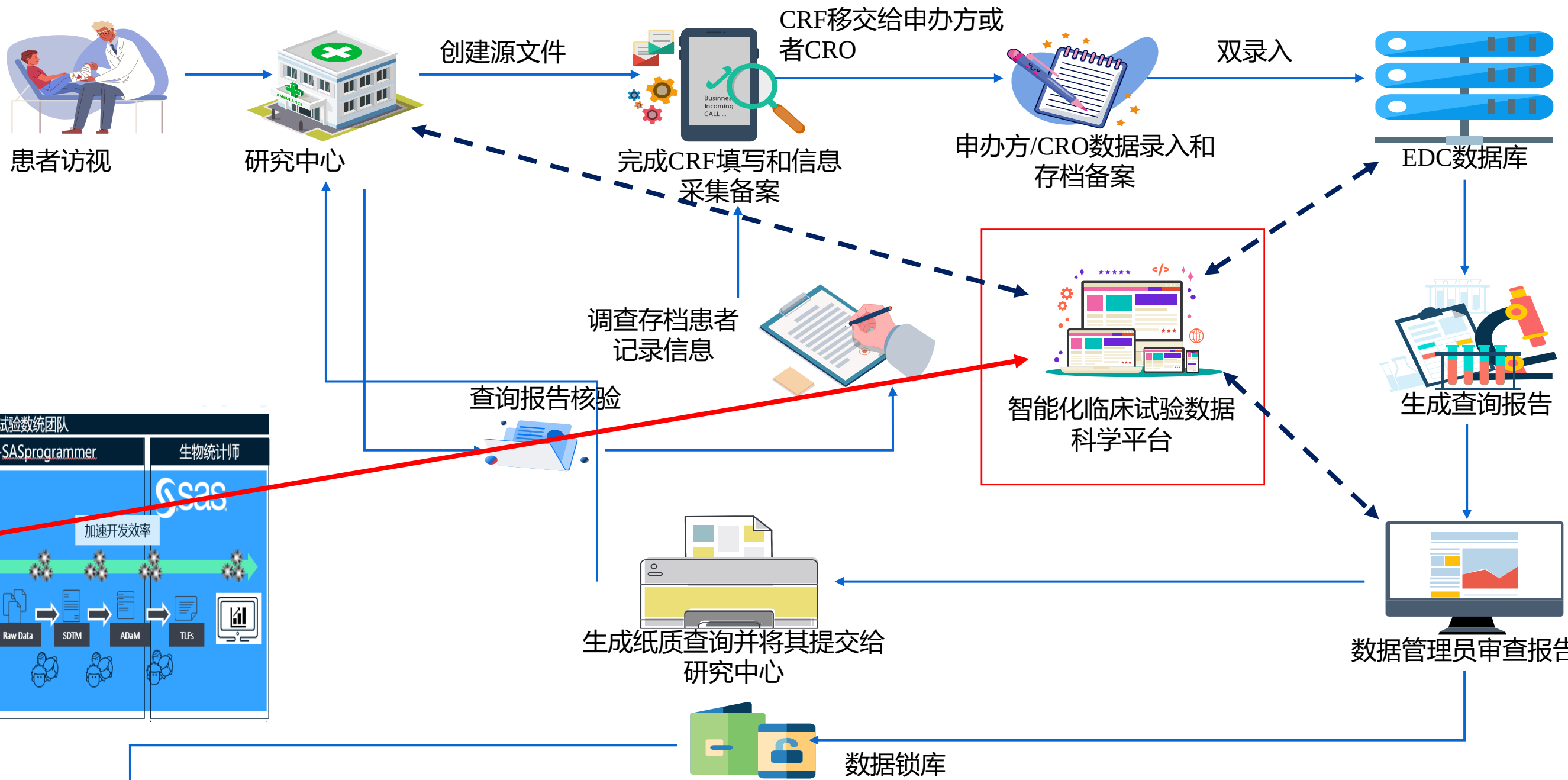


统计分析

统计分析计划

统计分析编程

统计报告医学写作



# Aganda



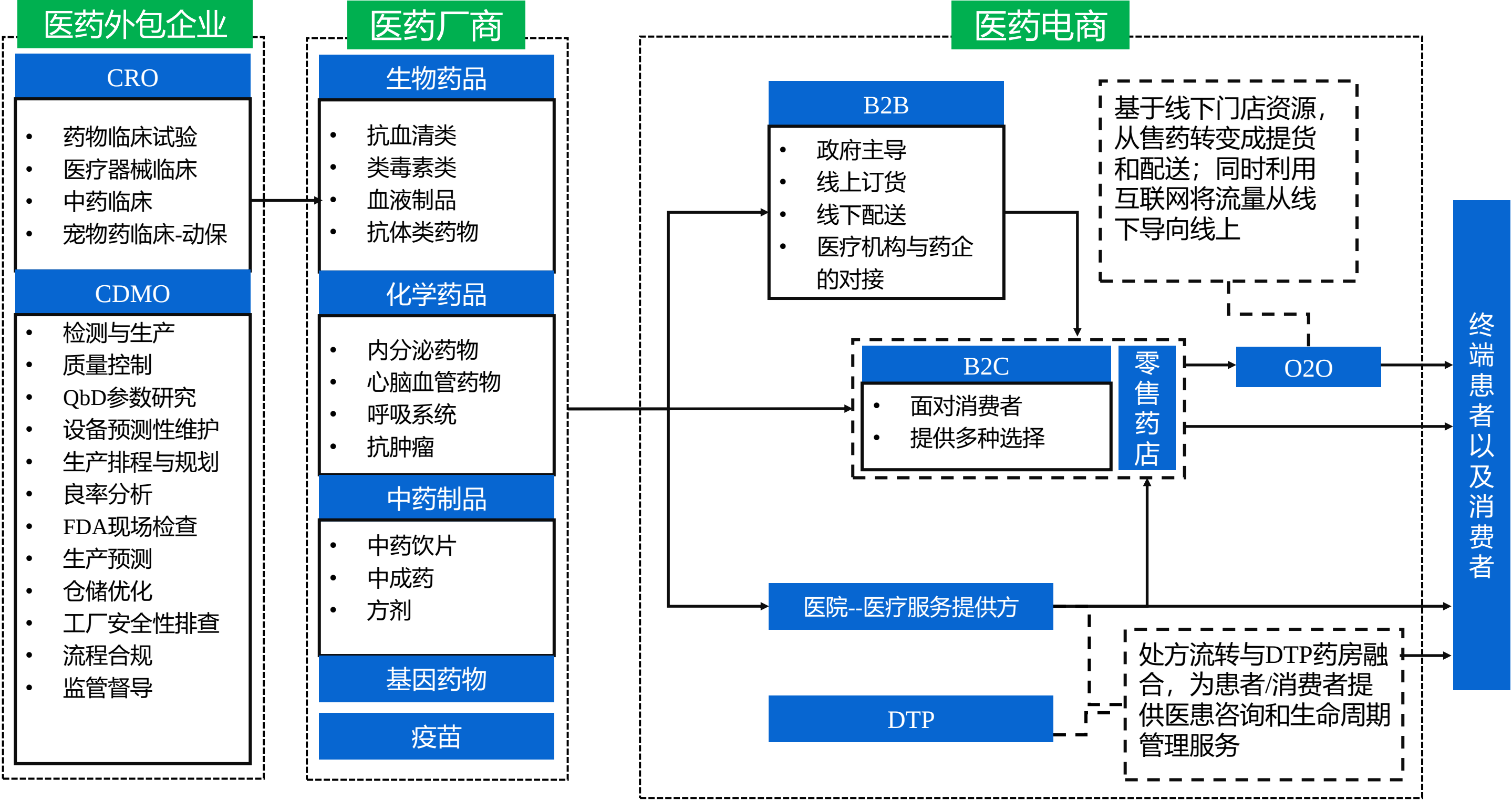
**临床试验生物统计数字化过程中的挑战**

**数据科学在临床试验全生命周期中推广**

**SAS临床试验数据科学平台的设计理念**

**未来生物统计技术平台的展望**

# 医药产业链各环节的实体关系与应用场景



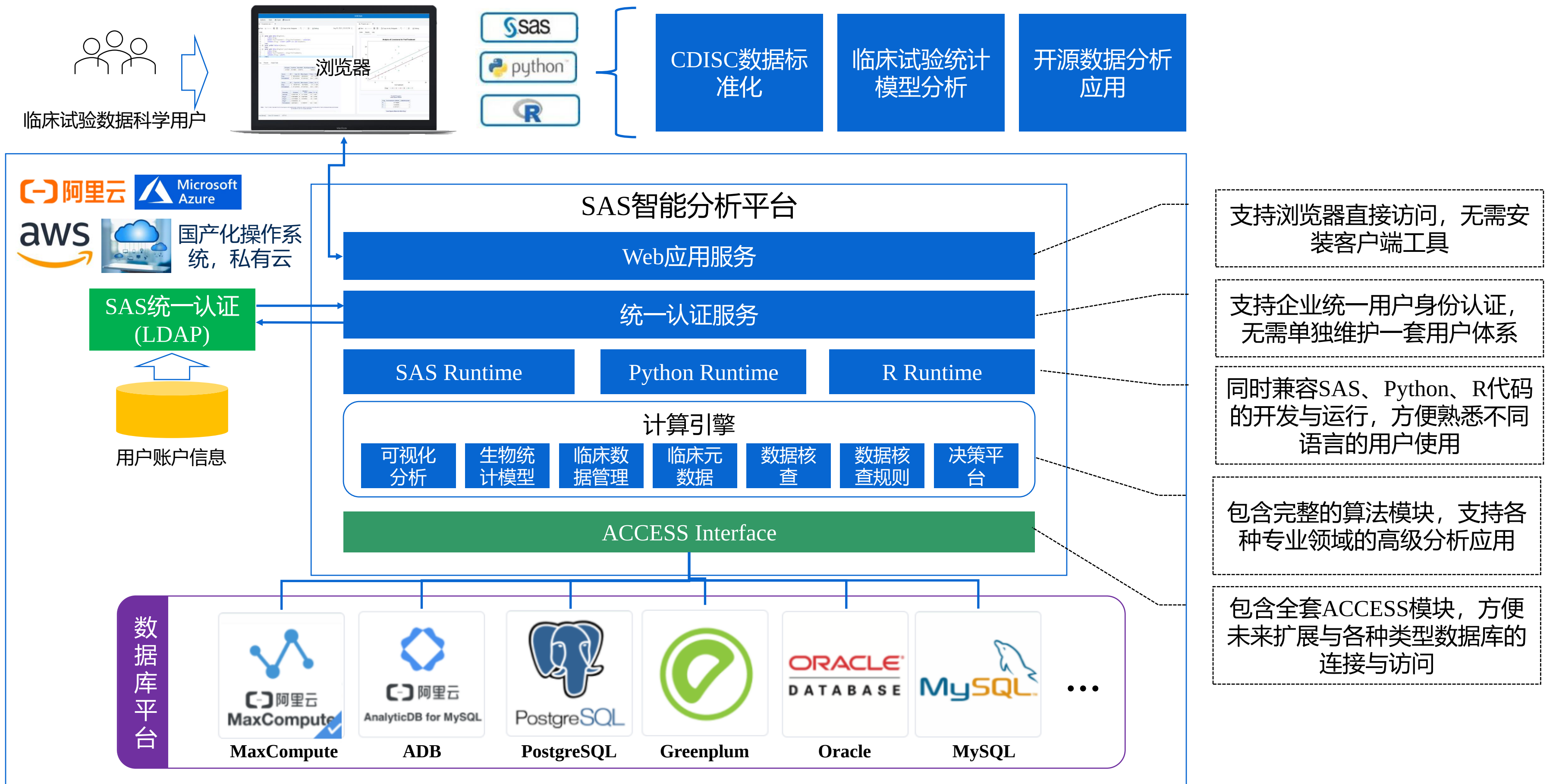
- SAS药械研发领域应用**
- ✓ 支持药械安全性验证
  - ✓ 支持药械有效性验证
  - ✓ 支持药械研发上市
  - ✓ 支持药械变更适应症
  - ✓ 临床试验数据标准化处理
  - ✓ 支持传统医学走向现代化
- SAS药械商业化与医保支付应用**
- ✓ 支持药物经济学定量评价
  - ✓ 支持药厂基于销售数据的经营现状与预测分析 (SFE)
  - ✓ 支持医保支付端DRG欺诈发现
  - ✓ 支持B2B政府对医药电商价值的定量评估
  - ✓ 支持零售药店销售数据监控
  - ✓ 支持O2O提货数据分析

SAS MQC Solution

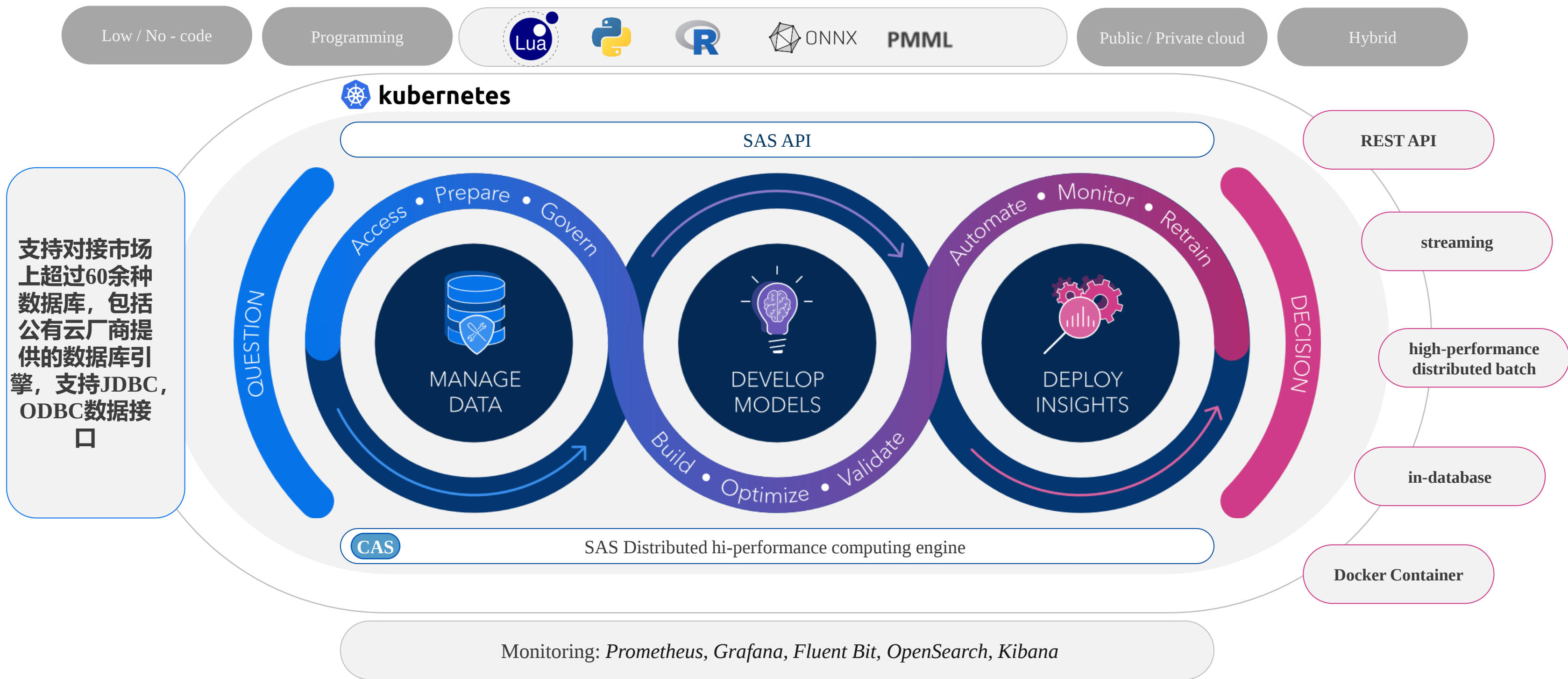
SAS RCT Solution

SAS 药械商业化—医保支付类应用场景

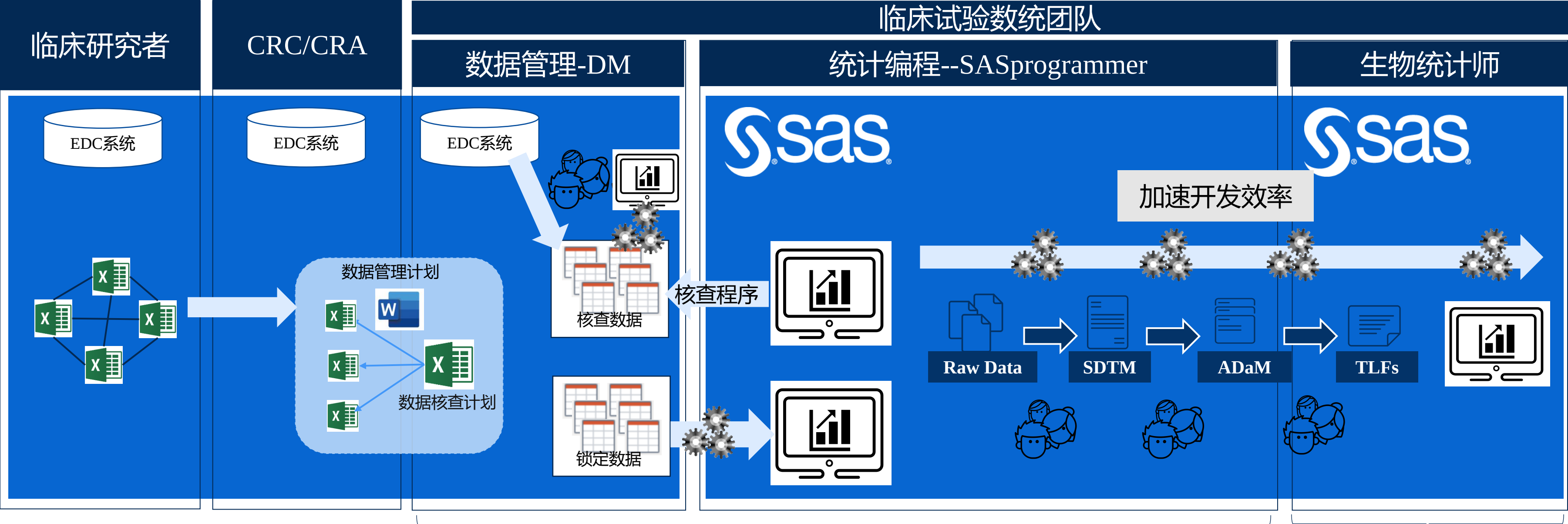
# 临床试验数据分析平台系统逻辑架构图



# SAS现代化临床试验数据科学平台设计理念



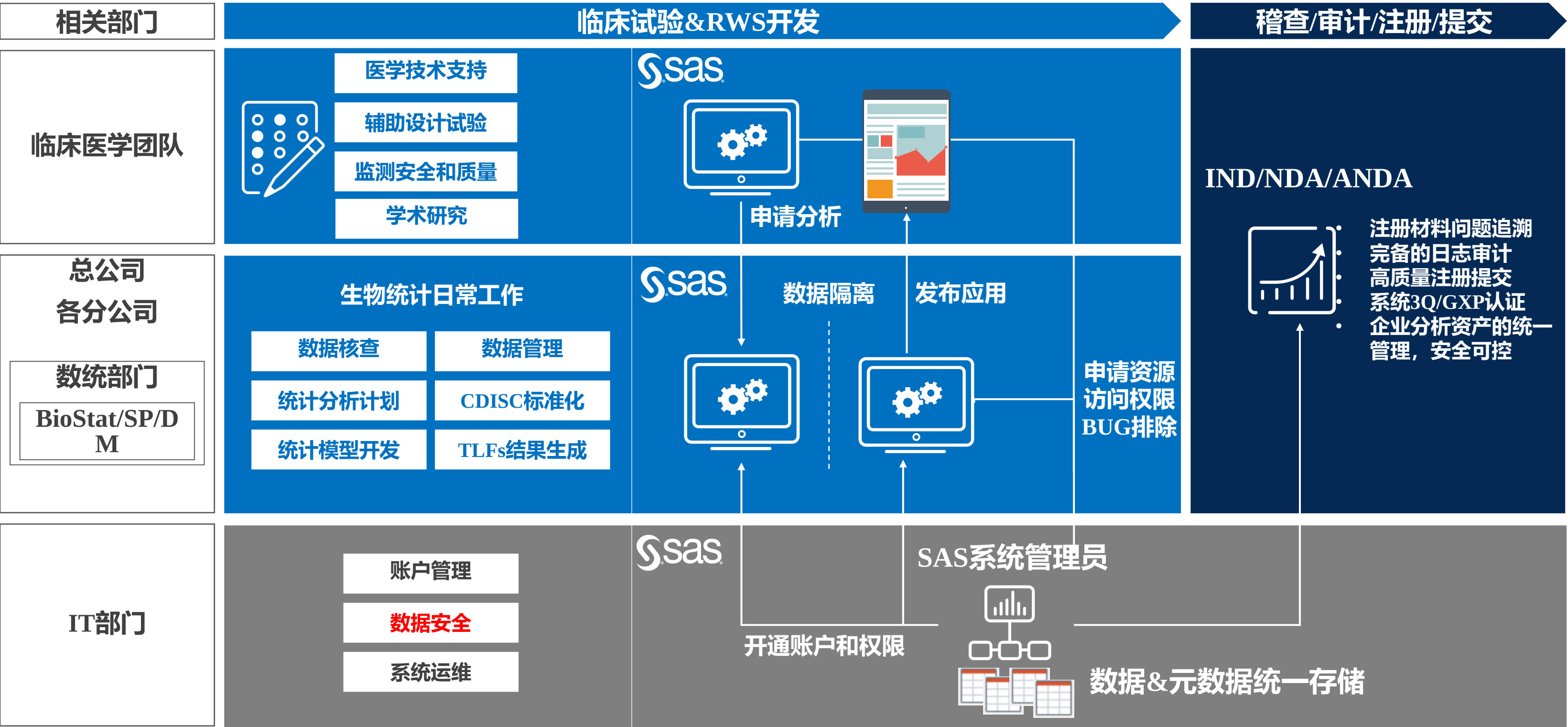
# 现代化临床试验的技术考量---加速研发效率



- DM团队数据核查和质控负担重---劳动密集型
- SAS程序员为DM团队编写数据核查脚本任务繁重，调用机制复杂
- CDISC数据管理和元数据管理缺乏系统化标准操作流程
- 缺乏SAS宏程序知识产权管理

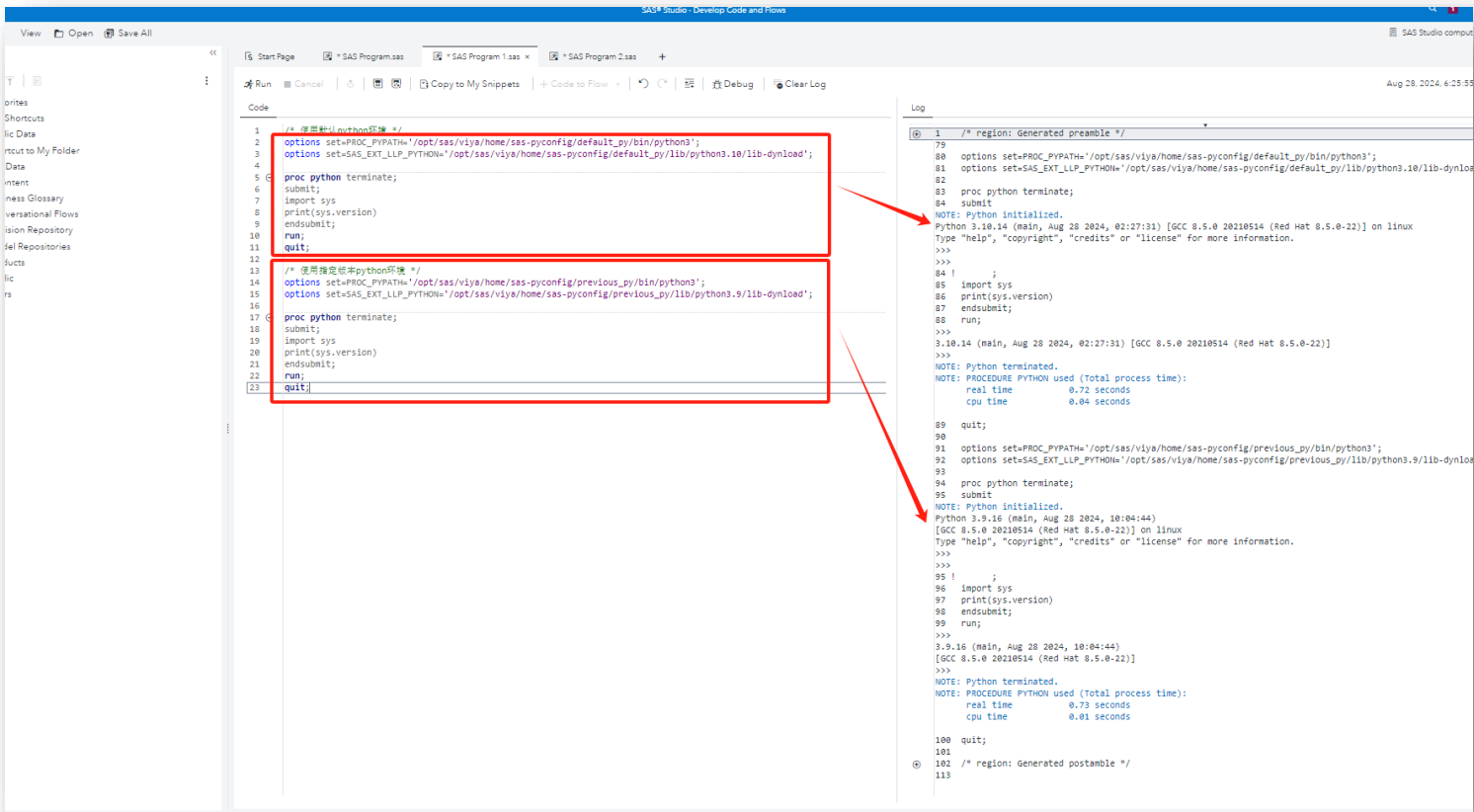
- SAP计划编写和分发
- 统计结果实时查看
- 三期适应性模型开发

# 现代化临床试验的技术考量---生物统计应用发布

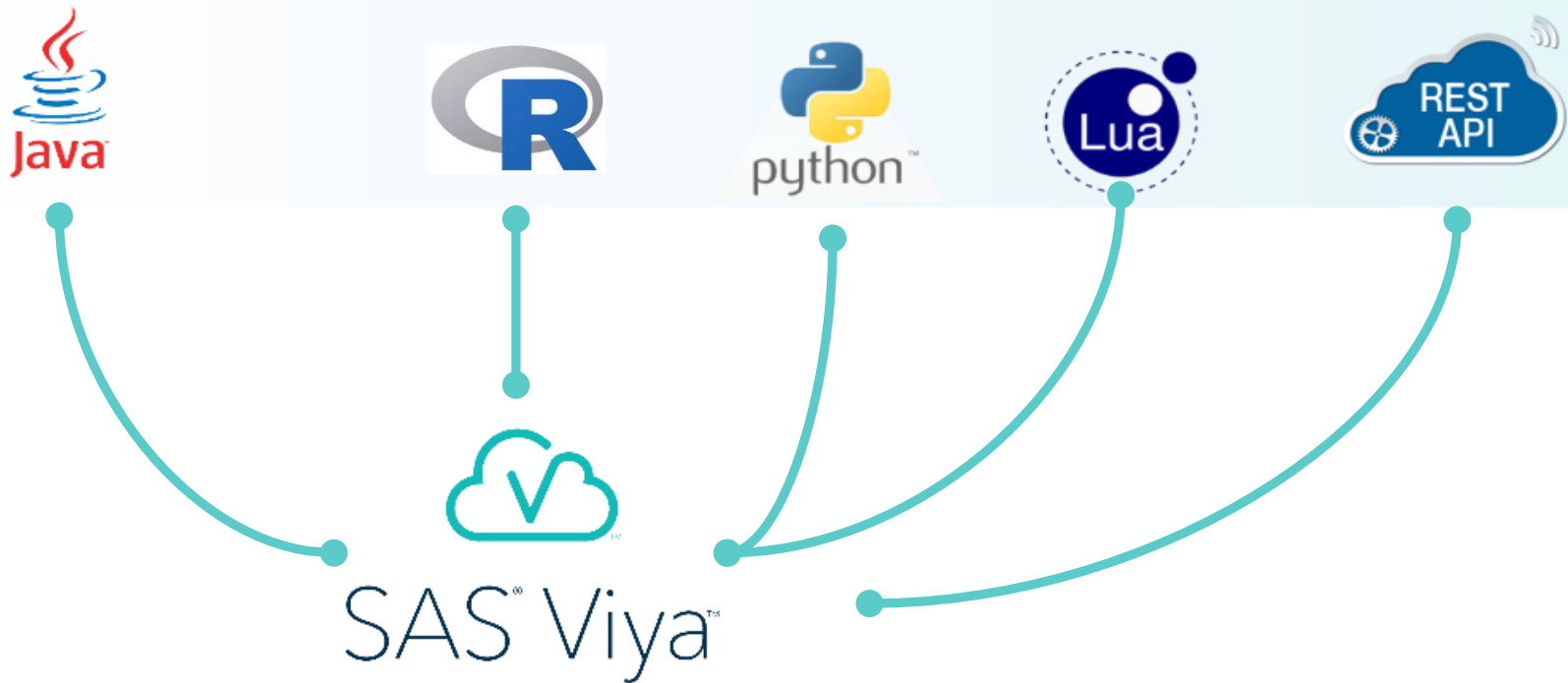
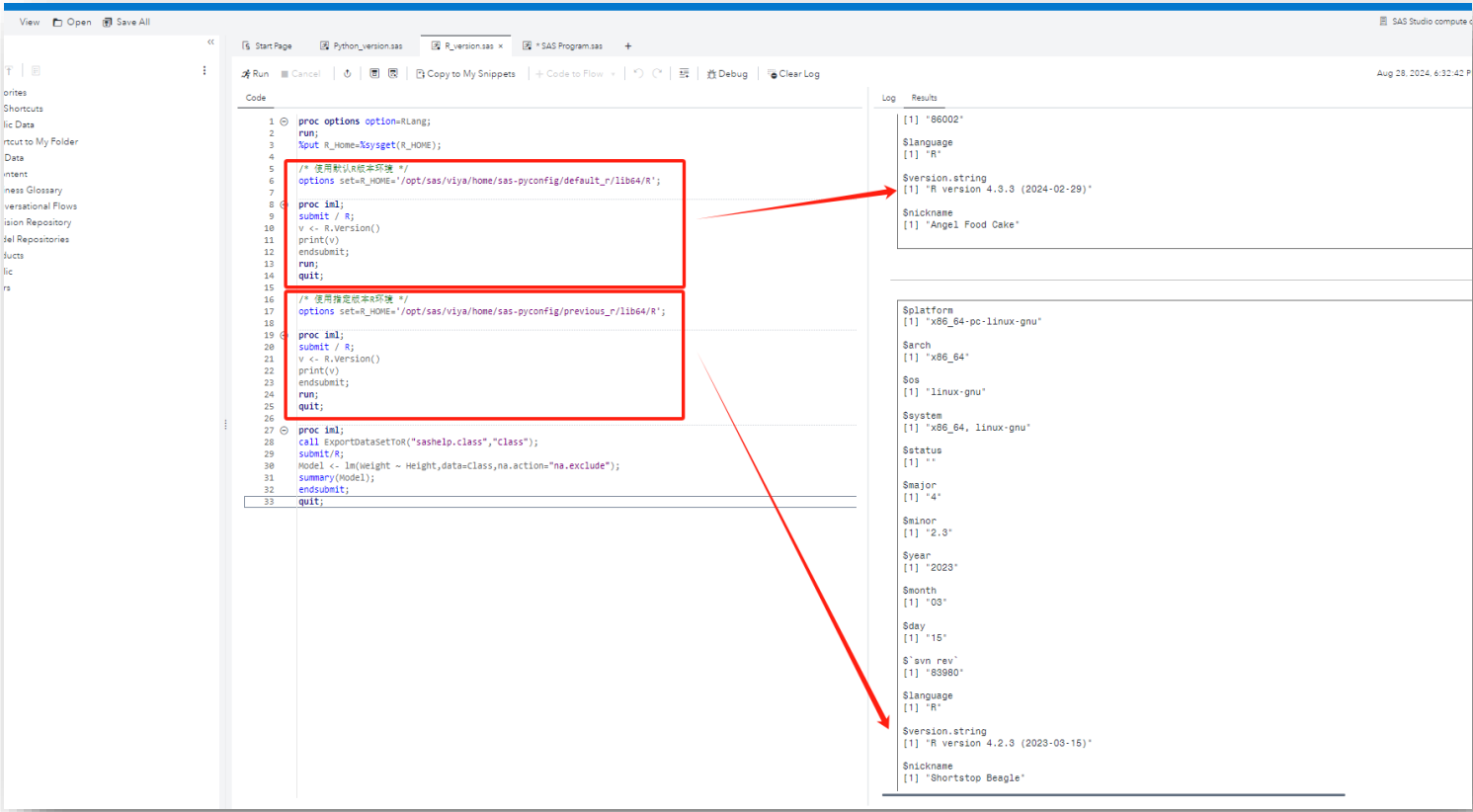


# 现代化临床试验的技术考量---开源工具的版本管理

## 多版本python环境集成



## 多版本R环境集成



# 临床试验数据管理的统一存储库

- 临床试验统一数据仓库

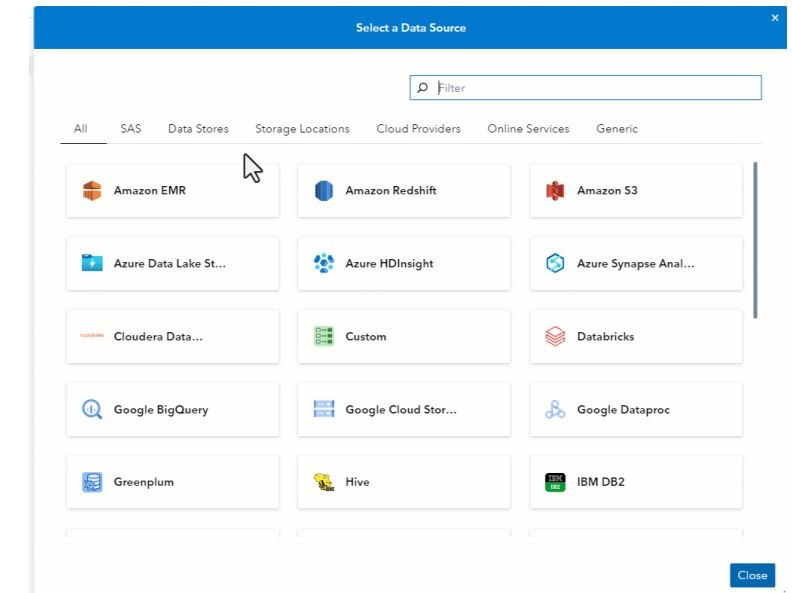
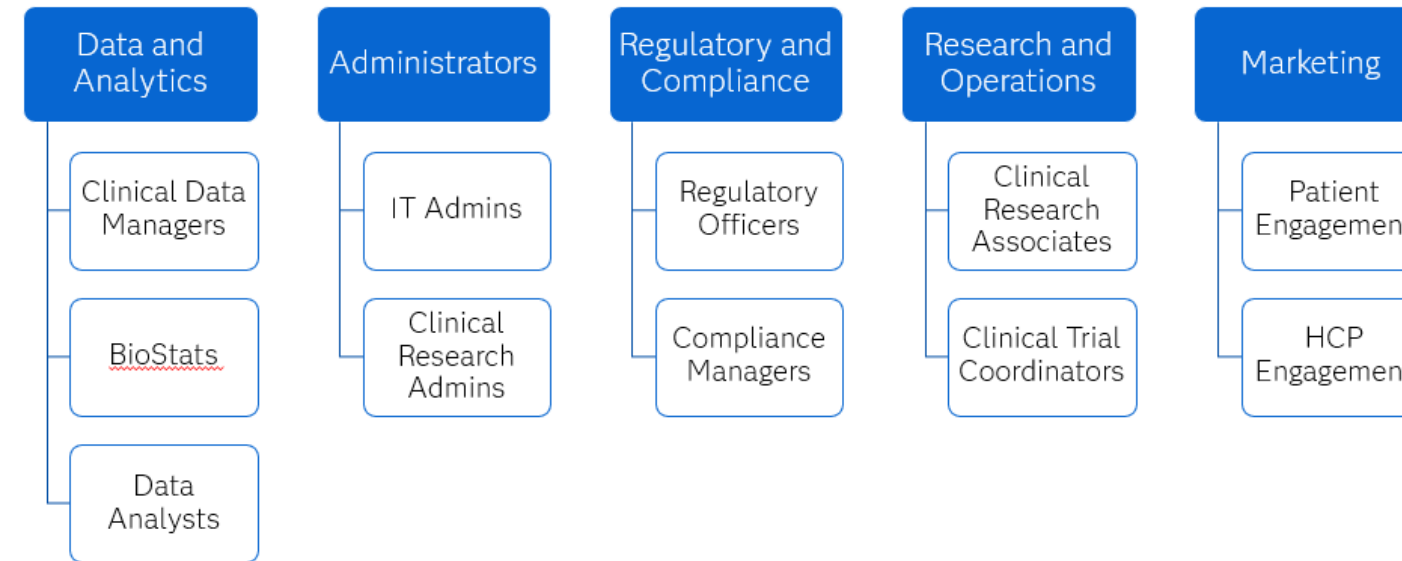
- 用户组 & 访问权限
- 用户角色 & 权限赋予
- 临床文档的版本管理
- 电子签名
- 在线数据回收站
- 文档检索
  - 文件内容搜索
  - 可在对话框中的 "文件规则" 选项下搜索文件内容

- 管理员监控与管理

- 操作审计和日志审计

- 可搜索您的操作

- 动作与操作状态管理



| Principal           | Access Level | ReadInfo | Select | LimitedPromote | Promote | CreateTable | DropTable | DeleteSource | Insert | Update | Delete | AlterTable | AlterCaslib | ManageAccess |
|---------------------|--------------|----------|--------|----------------|---------|-------------|-----------|--------------|--------|--------|--------|------------|-------------|--------------|
| Authenticated Users | Read         | ✓        | ✓      | ✗              | ✗       | ✗           | ✗         | ✗            | ✗      | ✗      | ✗      | ✗          | ✗           | ✗            |
| Samiul Haque        | Write        | ✓        | ✓      | ✓              | ✓       | ✓           | ✓         | ✓            | ✓      | ✓      | ✓      | ✓          | ✗           | ✗            |
| Data Builders       | Write        | ✓        | ✓      | ✓              | ✓       | ✓           | ✓         | ✓            | ✓      | ✓      | ✓      | ✓          | ✗           | ✗            |



Dashboard icons: Home, Key, Clock, Add

System Message

请大家保持文件夹树结构清晰

- 临床试验项目下不允许存放不相关的文件
- SAS代码的版本管理需要由所属者进行处理
- 如果出现任何系统问题请及时联系管理员 (tianlun.gu@sas.com)

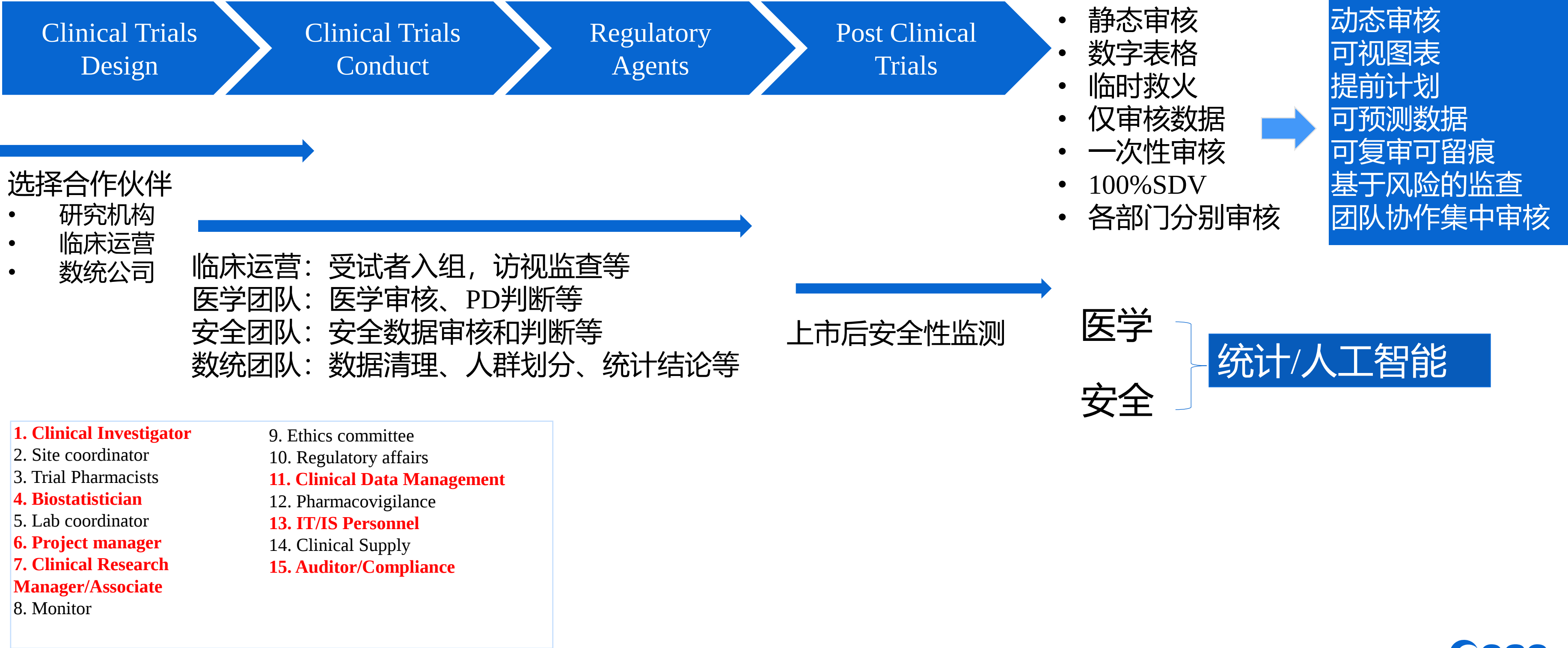
Repository Utilization

Total space used on 2024年9月02日 10:25:20: 0.00 MB

Utilization (MB)

Date

# 现代化临床试验运营数据审核交互式可视化技术



# 临床试验数据管理与核查分析的可视化技术

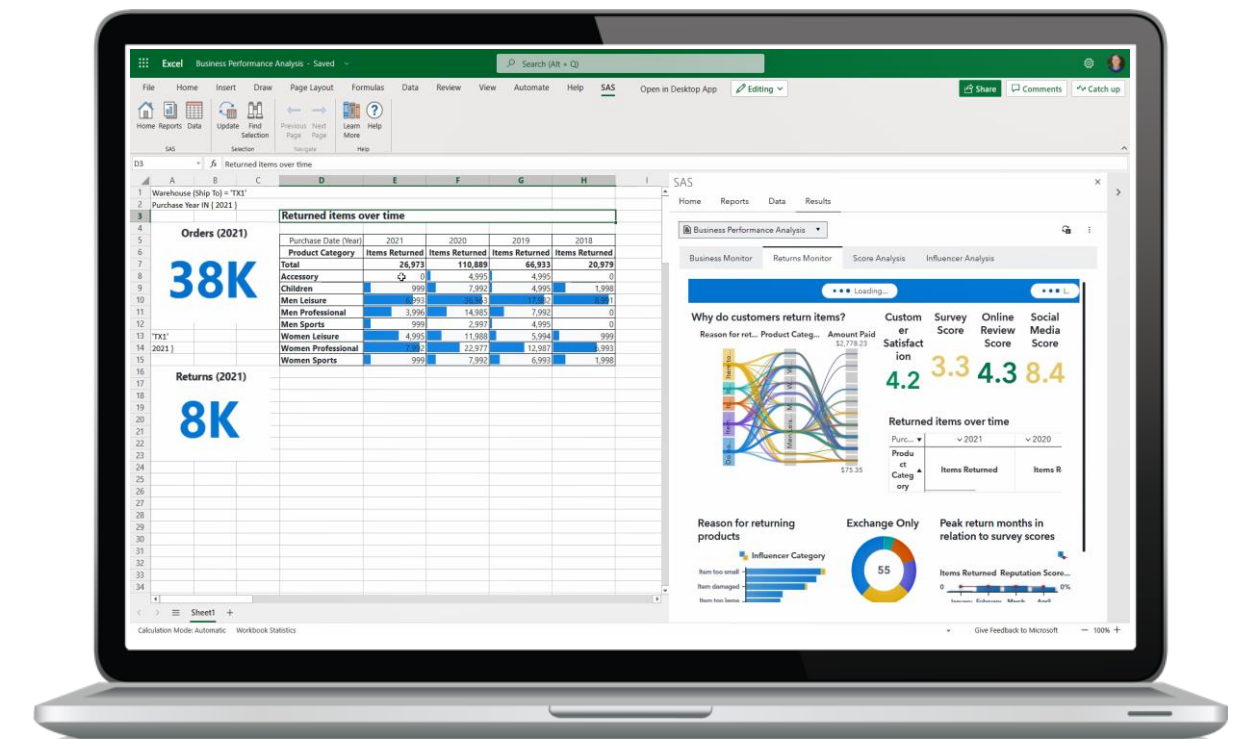
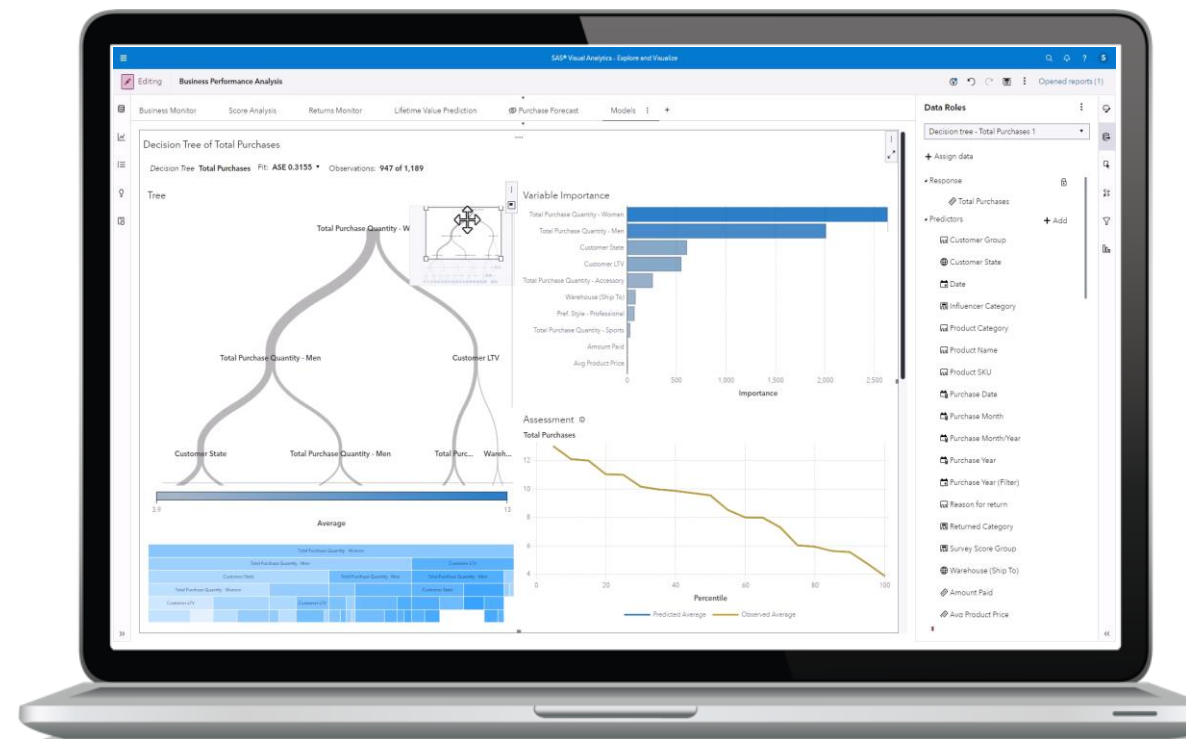
- 列出临床试验数据常用模块: DM/IC/IE/MH/LB/PE/VS/AE/CM/DS
- 逻辑核查类型
  - 窗口和范围核查
  - 交叉逻辑关系核查
  - 缺失和不完整行数据核查
  - 外源数据核查
  - 其他核查方法

目的：及早发现药物安全信号！



- 基于风险的质量管理
- 医学监察可视化审阅
- 药物警戒信号可视化
- 研究数据库实时可视化
- 数据管理可视化
- 真实世界数据交互可视化
- 统计结果TFL可视化

可视化分析技术是充分利用人们的视觉感知能力，对数据进行呈现以放大人们对数据含义的认知



## SAS Visual Analytics in SAS Viya Clinical Studies Visualization

## Demographic Analysis



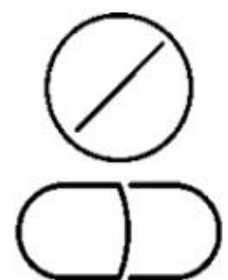
### Adverse Event Analysis



## Patient Profile



### Concomitant Medication Analysis



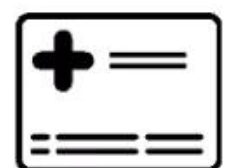
## Lab Analysis



## Hy's law DILI Analysis



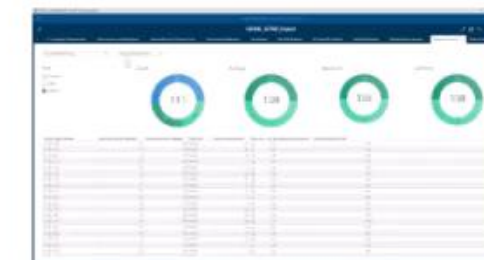
### Medical History Analysis



### Vital Signs Analysis



## Exposure Analysis



# Aganda



**临床试验生物统计数字化过程中的挑战**

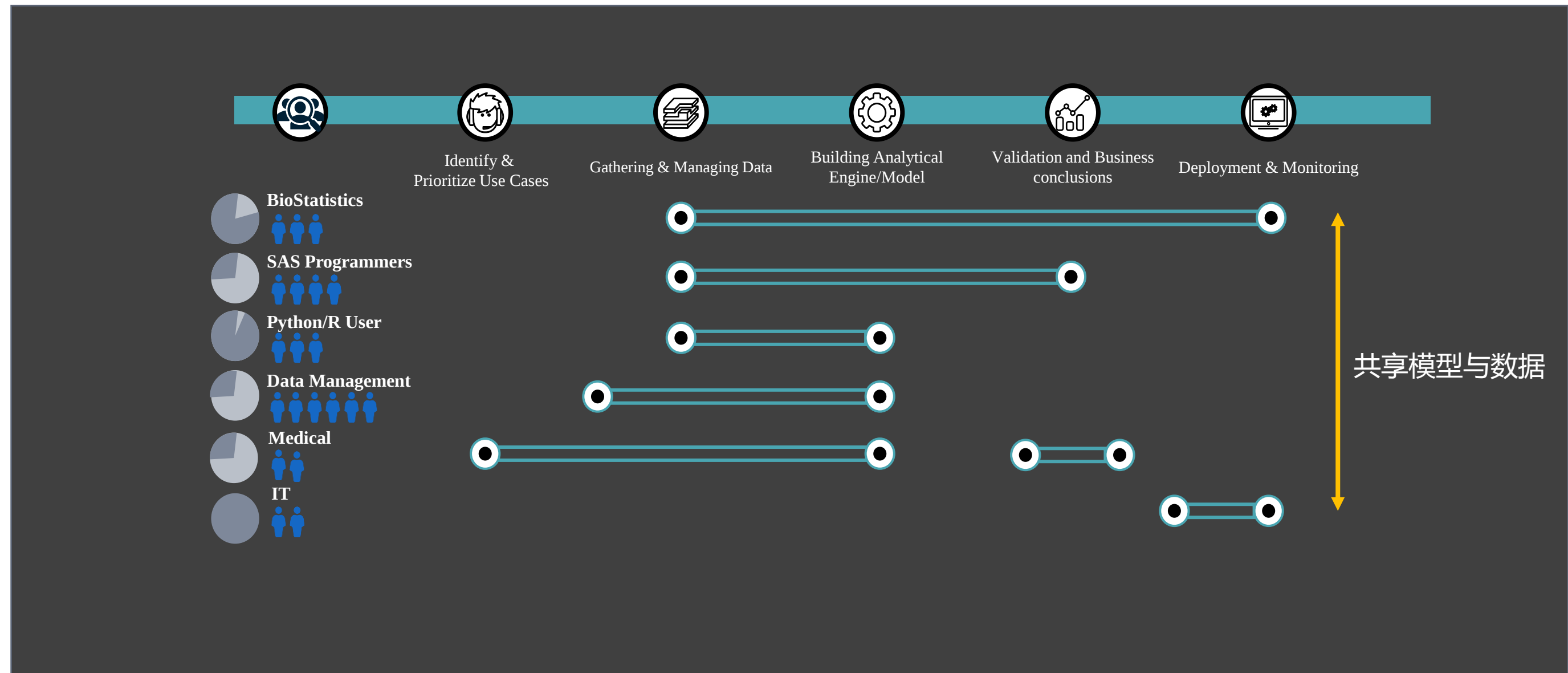
**数据科学在临床试验全生命周期中应用**

**SAS临床试验数据科学平台的设计理念**

**未来生物统计技术平台的展望**

# 未来数据科学技术平台的展望

- 临床试验数据互联互通：数据流转畅通无阻，数据标准业界趋同，数据访问安全可控
- 生物统计应用无处不在：分析工具因人而异，分析场景业务驱动，分析应用灵活轻量
- 人工智能自动化辅助：无缝衔接试验设计与试验实施，自动监控不合规操作，人工智能模型可审计可解释



# Thank YOU

