

2023年南湖HIT论坛

# 数据驱动医院运营管理

2023年11月11日 嘉兴市

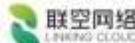


扫码观看视频直播

2023年南湖HIT论坛课件 版权归演讲人所有

主办： HIT专家网

承办：北京和思凯文化传媒有限公司

支持企业： 卫宁健康  联空网络  望海康信  ifmsoft  B-Soft 创业慧康  ClinBrain  美创  TRANSWARP 新华三  白峰科技 SmartBY Method  inspur 浪潮



# 助力医院构建“数据大脑”

秦晓宏

2023年11月

## 个人简介

**秦晓宏** 柯林布瑞公司联合创始人

《上海市医院信息集成平台建设与应用实践指南》 **第一副主编**

《广东省医疗数据中心建设规范》 **编委**

《新型冠状病毒肺炎患者科研基本数据集》 **编委**

《中医医院临床数据检索系统建设指南》 **编委**

《名老中医典型病案共享数据库建设指南》 **编委**

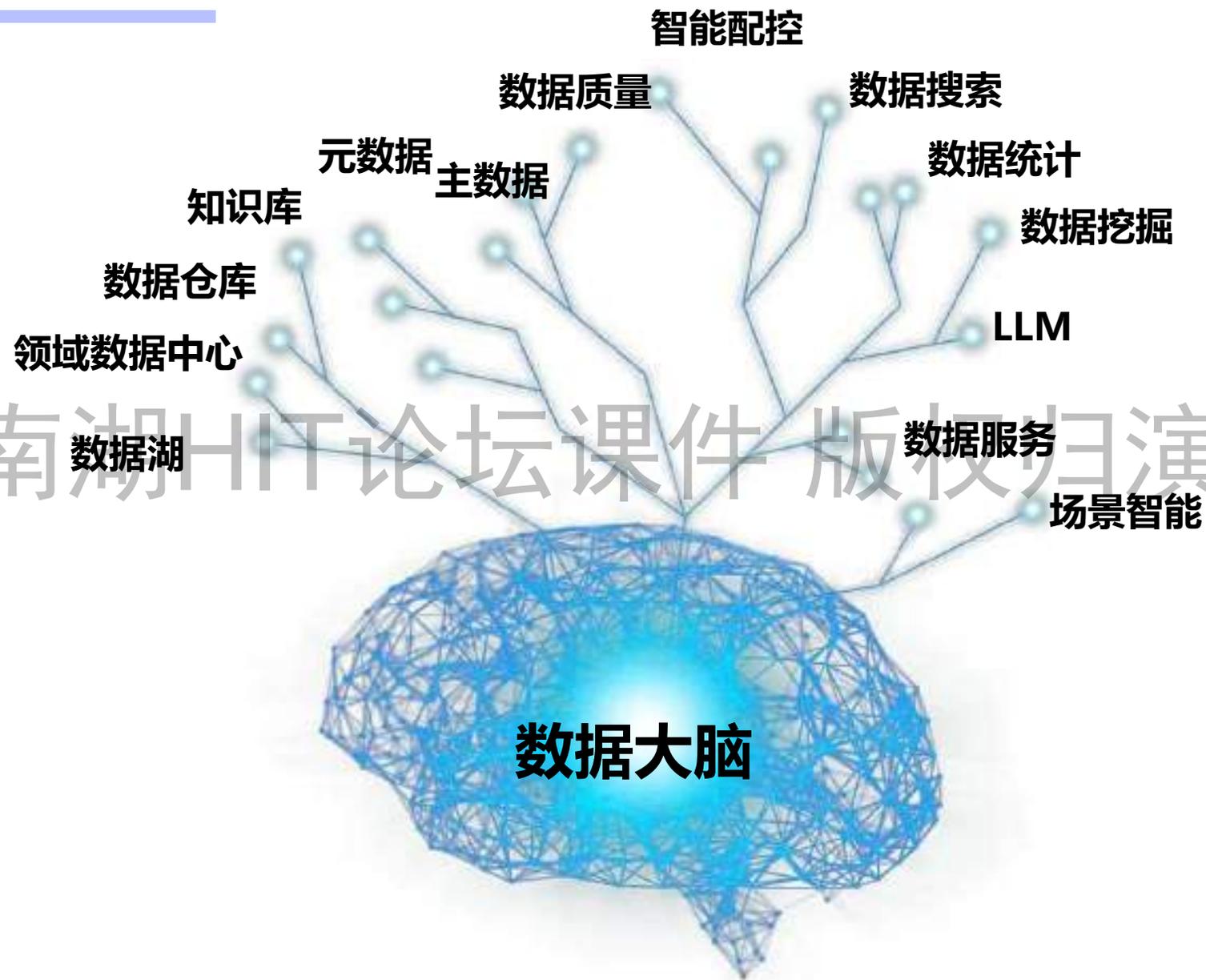
《大数据治理》大学教材 **编委**

**20项** 医疗大数据相关发明专利第一发明人

传统数据分析系统会因为医院数据的不断增长，导致数据整合困难，计算过程繁琐，计算性能低下等问题。

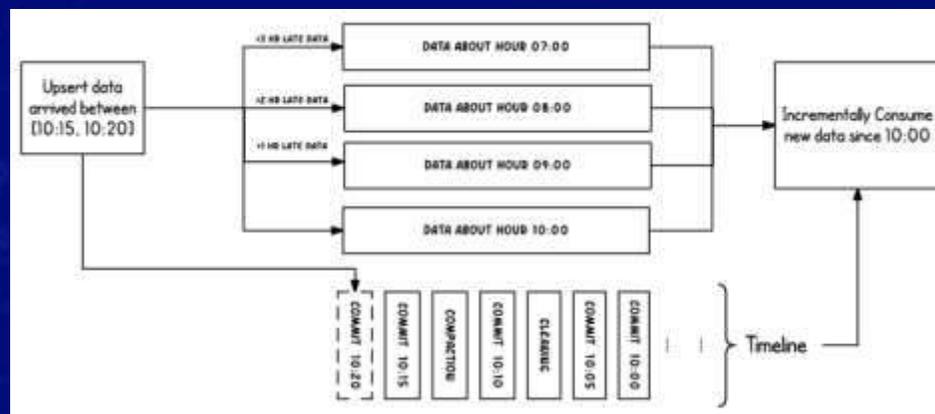
数据大脑正好相反，随着数据的不断增多，数据大脑的计算模型在大量数据的训练下变得更加智能、分析过程更加便捷、分析结果更加准确。

# 数据大脑



2023年南湖HIT论坛课件 版权归演讲人所有

数据湖是一个集中式的数据存储，以原始形式摄取和存储大量数据。进入数据湖后，数据便可以被加工处理并被用作各种分析需求的原材料。由于其开放、可扩展的架构，数据湖可以容纳来自任何来源的所有类型的数据，从结构化（数据库表、Excel 工作表）到半结构化（XML 文件、网页）再到非结构化（图像、音频文件、推文），所有这些都不会牺牲保真度





数据资产管理，是建立在元数据之上的。元数据相当于数据的户口本。以某张表为例，元数据描述了表名、表别名、表的所有者、数据存储的物理位置、主键、索引、表中字段、这张表与其他表之间的关系等等。所有这些信息的描述，就是这张表的元数据。

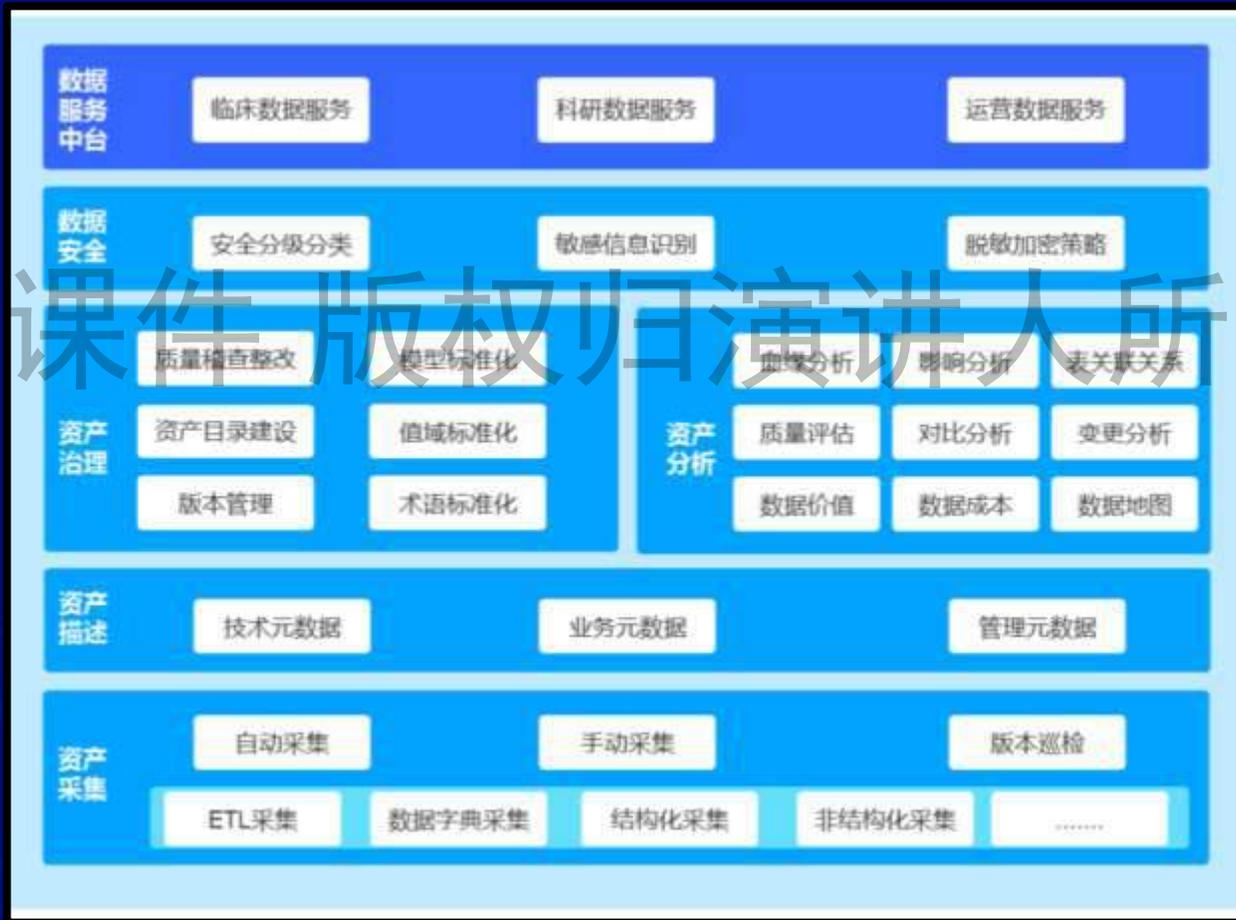
元数据在数据治理当中，就相当于所有数据的地图。它描述了我們有哪些数据，数据分布在哪里，这些数据分别是什么类型，数据之间有什么关系，哪些数据经常被引用，哪些数据无人查看。

在大数据平台中，元数据贯穿大数据平台数据流动的全过程，主要包括数据源元数据、数据加工处理过程元数据、数据主题库专题库元数据、服务层元数据、应用层元数据等。右图以一个数据中心为例，展示了元数据的分布范围：



# 元数据的位置

生产系统中的数据大多涵义不清晰，业务属性不明确，数据难以利用，难以发挥出数据的价值。所以医院需要构建元数据资产目录，参考DAMA指南中描述属性，将从业务、技术、管理三个角度来描述数据资产，帮助医院信息化人员和临床人员都能够知道拥有哪些数据、这些数据放在哪里、这些数据从哪里来到哪里去、分别由谁负责、其中每个数据代表什么含义、数据生命周期做了哪些内容、如何保证数据的安全性，以及数据的质量如何等等，从而让医院的数据摆脱黑盒管理。



# 元数据属性

**技术元数据：**技术元数据是结构化处理后的数据，方便计算机或数据库对数据进行识别、存储、传输和交换。包含库表结构、字段约束、数据模型、ETL 程序、SQL 程序、血缘图谱、表关联关系、数据质量等。

**业务元数据：**业务元数据描述数据的业务含义、业务规则等。明确业务元数据可以让人们更容易理解和使用业务元数据。元数据消除了数据二义性，人们对数据有一致的认知，避免“自说自话”，进而为数据分析和应用提供支撑。包含业务定义、映射实体物资、业务规则、业务术语等。

**管理元数据：**明确管理属性有利于将数据管理责任落实到部门和个人，是数据安全管理的基礎。包含数据所有者、数据质量定责、数据安全等级、备份、保留、创建日期、灾备恢复预案等。





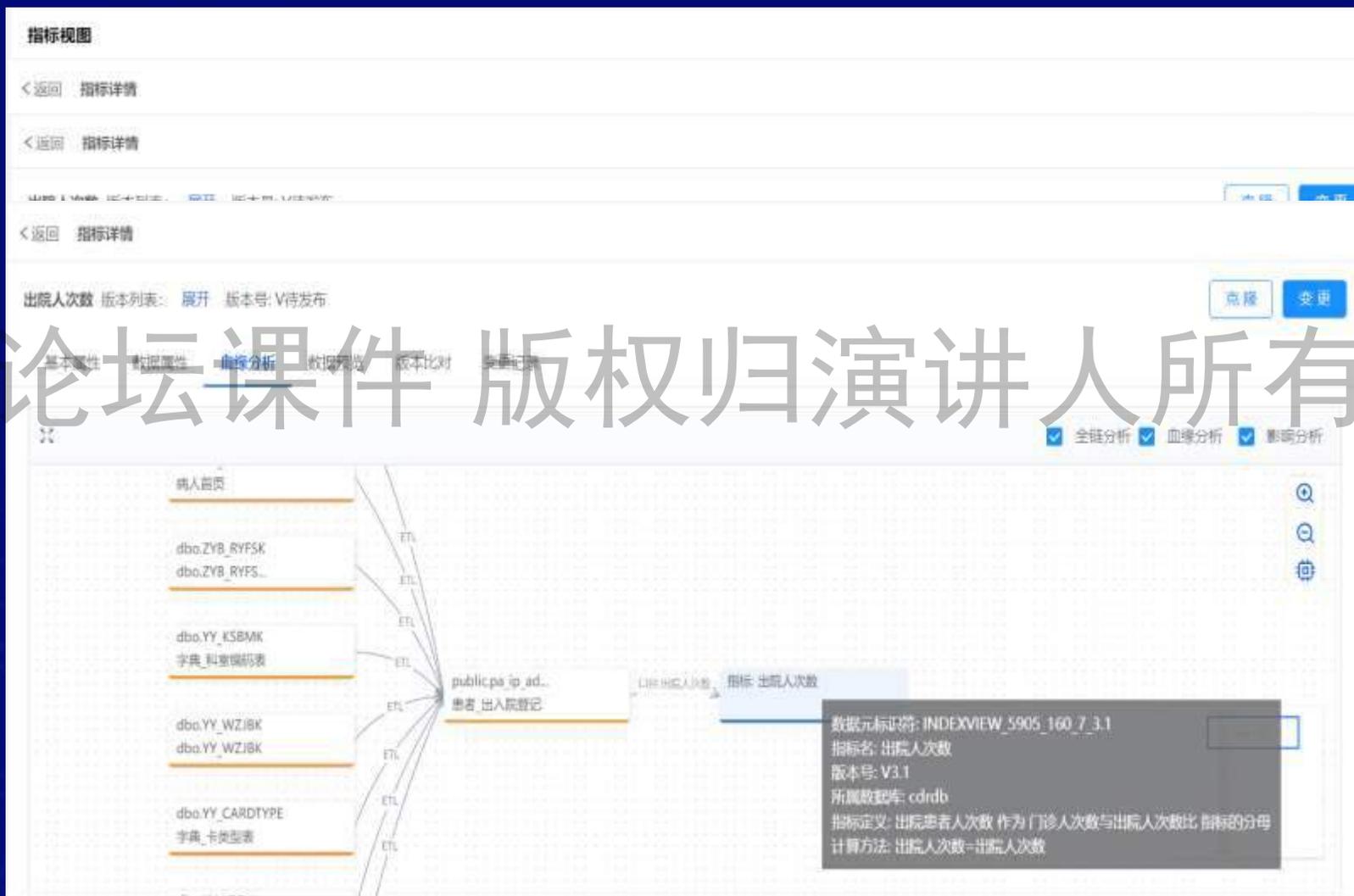
针对医院运营、质控、绩效、上报场景下的指标相关属性和口径描述，形成指标资产目录，并提供相关数据服务，目前已经积累三级医院等级评审、公立医院绩效考核、基础运营指标库三个体系，同时形成口径版本管理以更加全面的服务运管部门和领导组织。

指标遵循国考要求，对指标管理属性描述如下：

**基本属性：**指标名称、性质、定义、计算单位、计算方法、公式、指标导向、指标来源、指标口径说明等。

**数据属性：**取值来源、维度、度量、指导值、数据口径等。

**管理属性：**指标解释、应用场景、所属机构、状态、是否国家检测指标，有效期等。



# 业务地图:数据流与业务流的结合

业务地图，实现医院业务流与数据流的结合，通过业务流程图直观了解到医院的门诊、住院、体检等场景的业务闭环，以及各业务节点中的执行科室、业务规则和数据资产；

运管部门可以标记标签，关注自己运营管理中需要关注的业务节点，查看这些节点在底层体现在哪些数据资产上，并下钻查看这些数据资产的业务描述、技术参数、管理权限、数据血缘的来源和去处；

同时数据质量稽查模块对这些节点数据的一致性、完整性、规范性、逻辑性等质量维度做出综合评估，问题追溯至具体的明细记录，方便信息化部门和运管部门及时发现和整改业务缺陷和数据问题，最终实现通过数据流了解业务流，通过业务流掌握数据流，优化医院运营管理，发挥数据资产最大价值；

序	表名	中文表名	字段	字段中文名	字段类型	被检测记录	满足条件记录数	不满足条件记录数	分数	操作
1	ab_op_feeitem	费用门诊收费明..	cardkindcode	卡类型编码	卡类型代码	76543	76543	0	100	🔍
2	ab_op_feeitem	费用门诊收费明..	ybkindcode	医保类型代码	医保分类	76543	76543	0	100	🔍
3	ab_op_feeitem	费用门诊收费明..	regkindcode	挂号类型代码	挂号类型	76543	76543	0	100	🔍
4	ab_op_feeitem	费用门诊收费明..	sexcode	性别代码sys_code...	生理性别代码表	76543	76543	0	100	🔍

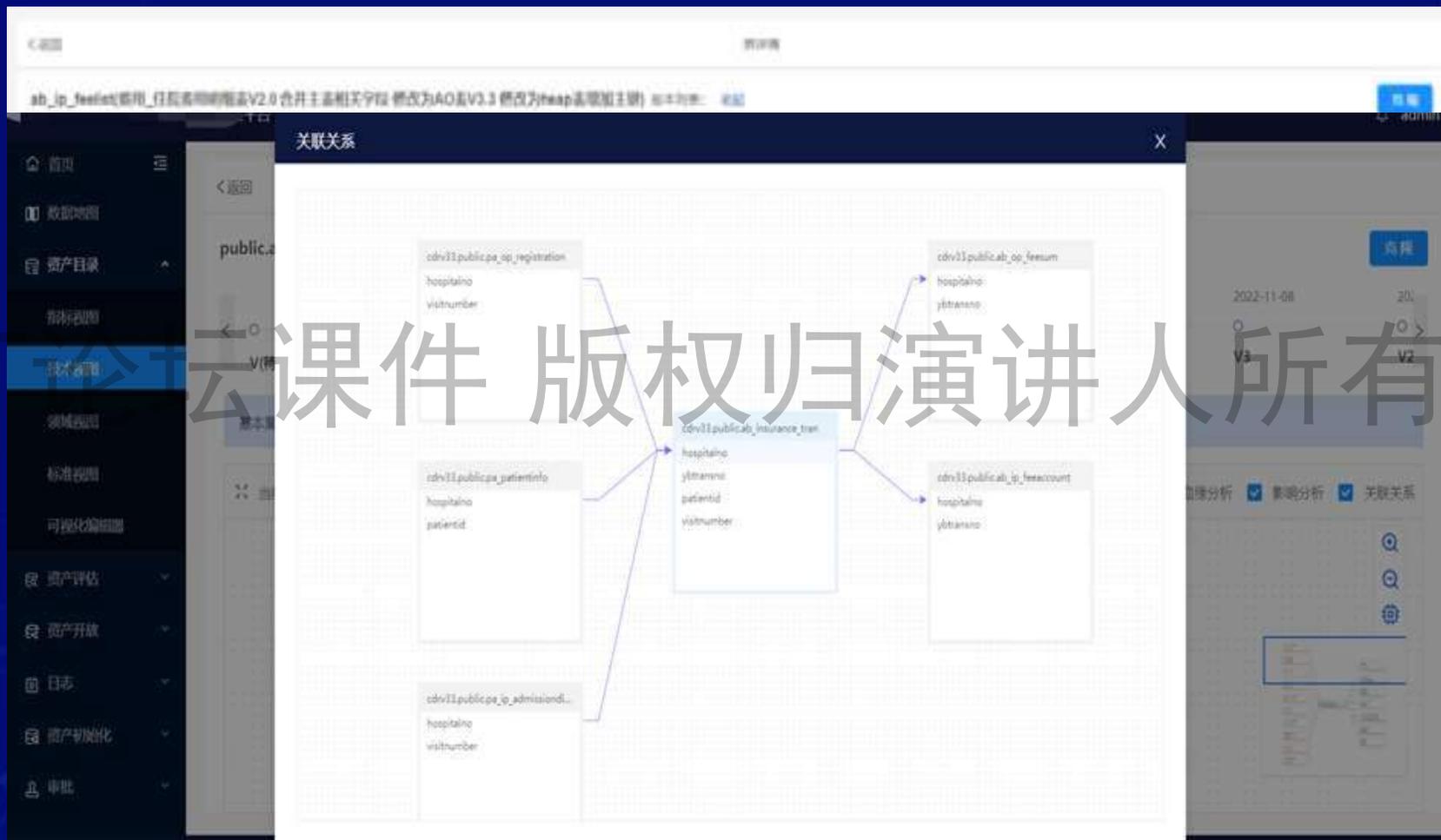
# 应用场景1：如何快速定位运营口径问题

**场景一：**举例某科室人员发现「月度运营分析」报表数据存在质量问题，于是向信息科提出异议，技术人员通过元数据血缘分析发现「月度运营分析」报表受到上游 ODS（数据湖）层四张不同的数据表的影响，从而快速定位问题的源头，低成本地解决问题。

## 解决方案：

**数据血缘分析**对于用户具有重要的价值，如：当在数据分析中发现问题数据的时候，可以依赖血缘关系，追根溯源，快速地定位到问题数据的来源和加工流程，减少分析的时间和难度。

除了纵向的血缘关系，我们还提供了横向的**表关联关系图谱**，充分表达了表之间的业务关系。



## 应用场景2：如何快速锁定模型变更带来的影响

**场景二：**某机构因业务系统升级，在业务表中修改了字段：“cardno” 长度由 8 修改为 64，需要分析本次升级对后续相关系统的影响。对元数据“cardno” 进行影响性分析，发现对下游数仓相关的表和ETL程序都有影响，信息科定位到影响之后，及时修改下游的相应程序和表结构，避免了问题的发生。由此可见，数据的影响性分析有利于快速锁定元数据变更带来的影响，将可能发生的问题提前消灭在萌芽之中。

### 解决方案：

**影响性分析**，它能分析出数据的下游流向。当系统进行升级改造的时候，如果修改了数据结构、ETL 程序等元数据信息，依赖数据的影响性分析，可以快速定位出元数据修改会影响到哪些下游系统，从而减少系统升级改造带来的风险。数据影响性分析和血缘分析正好相反，血缘分析指向数据的上游来源，影响性分析指向数据的下游。



# 应用场景3：一览院内数据资产价值

**场景三：**基于元数据资产目录，通过API共享中台开放数据服务，对外开放的API查询订阅的活跃程度由高到低，可以通过冷热度分析、广度分析让院内数据资产价值可视化，让院内的业务人员、管理人员能够清晰的看到数据的活跃程度，以便他们更好的驾驭数据，主持或激活僵死数据，从而为数据的自助式分析提供帮助。

## 解决方案：

**冷热度分析**主要是对数据表的被使用情况进行统计，如：表与ETL程序、表与分析应用、表与其他表的关系情况等，从访问频次和业务需求角度出发，进行数据冷热度分析，用图表的方式，展现表的重要性指数，对于用户有巨大的价值。

**广度分析**统计下游订阅该数据资源的订阅方数量，订阅方越多，影响力越大，以此可以评估影响分析。



## 监督微调 (SFT)

使用精选数据在预训练模型基础上做指令精调，以对齐指令意图。对抽取的prompt数据人工进行高质量回答，帮助模型更好地理解输入指令。

## RLHF 训练

利用 (PPO) 算法，根据 RW 模型的奖励反馈进一步微调 SFT 模型。最终生成医疗领域LLM。

1

## 增量预训练(PT)

基于transformer模型结构，使用的海量医疗数据包含电子病历、病理、检查数据，真实的医患问答等数据，注入模型医学领域知识。

2

3

## 奖励模型微调

人类对语言模型结果打分排序来训练一个独立的奖励模型，后续使语言模型能够更好地满足人类偏好。

4

# 基于LLM的运营新体验

The screenshot displays the ClinBrain LLM interface. On the left, a search history sidebar lists various queries such as '医院本年药占比是多少?' and '如何提高患者流转率?'. The central chat window shows a user query: '医院本年药占比是多少? 包括同比, ...'. The results page, titled '查看药占比指标数据', shows a search for '2023-01-01 - 2023-09-08' across '全部院区', '全部科室', and '全部医生'. The main result is '全院药品总收入占比 42.82%'. Below this, a section titled '查看药占比相关政策咨询' provides information based on the '三级医院评审标准' requirement that public hospitals' drug revenue should be less than 40% of total medical revenue. A bar chart shows the percentage of drug revenue for five consecutive years, with values consistently around 40%. The interface includes navigation buttons like '新建对话', '搜索历史记录', and '导出报告'.

查看药占比指标数据

2023-01-01 - 2023-09-08 全部院区 全部科室 全部医生

全院药品总收入占比  
42.82%

查看药占比相关政策咨询

根据《三级医院评审标准》要求, 公立医院药品收入占医疗总收入应小于40%。查看原文

按年份 按月份 2019 - 2023

全部院区 全部科室 全部医生

年份	药占比 (%)
2019	~38
2020	~39
2021	~39
2022	~40
2023	~41

导出报告

上海柯械市瑞信息技术有限公司 Copyright © 2023 ClinBrain. All Rights Reserved.

2023年南湖HIT论坛课件 版权归演讲人所有

- 建立医学院统一的行业标准库，涵盖ICD10、SNOMED CT中英文版、LOINC标准、药品ATC标准及各类常规的行业标准字典，为数据归一化建立基础。



2023年南湖HIT论坛课件 版权归演讲人所有

- 对接业务系统字典，建立符合国家标准的标准数据体系，通过提供唯一数据ID编码，将不同来源的同一数据进行映射和标化，同时对外发布标化服务。

The screenshot displays the '一体化基础平台' (Integrated Basic Platform) interface, specifically the '主数据管理' (Master Data Management) section. The interface is divided into several panels:

- 左侧导航栏 (Left Sidebar):** Contains menu items such as '模型初始化', '主数据维护', '主数据编辑', '内容审核', '主数据预览', '主数据映射', '主数据标注', '映射审核', and '映射关系总览'.
- 顶部 (Top):** Shows '一体化基础平台' and '主数据管理' tabs, along with a search icon and '管理员 退出' (Admin Logout).
- 中间目录 (Center Directory):** A tree view showing the hierarchy of data standards, including 'MDM\_院内手术\_V1.0' which is currently selected.
- 映射窗口 (Mapping Windows):** Two windows are open side-by-side, connected by a blue line. The left window is titled '【MDM\_院内手术\_V1.0】' and the right is '【ICD-9北京版\_V1.0】'. Both windows have columns for '字典编号' (Dictionary ID), '名称' (Name), and '操作记录' (Operation Record).
  - MDM\_院内手术\_V1.0 Table:**

字典编号	名称	操作记录
00.0100	头和颈部血...	回
00.0200x001	心脏血管病...	回
00.0300	其他治疗性...	回
00.0500x001	治疗性超声...	回
00.0901	高强度聚焦...	回
00.1000	化学治疗药...	回
00.1100	重组人免疫...	回
00.1101	重组蛋白输注	回
00.1200	吸入一氧化...	回
00.1300	东西注射法...	回
00.1301	人类血型物质...	回
00.1400	电击除颤仪...	回
00.1400x001	电击除颤仪...	回
  - ICD-9北京版\_V1.0 Table:**

字典编号	名称	操作记录
0.61011	经皮微创动...	回
0.62001	经皮微创动...	回
0.62004	经皮微创动...	回
0.62005	经皮微创动...	回
0.64007	经皮微创动...	回
0.64008	经皮微创动...	回
00.01001	头部血管治...	回
00.01002	颈部血管治...	回
00.02001	心脏治疗性...	回
00.03001	肾脏血管治...	回
00.10001	化学治疗药...	回
00.11001	输注重组蛋白	回
00.12001	一氧化氮疗法	回

2023年南湖HIT论坛课件 版权归演讲人所有

# 数据质量核查

- 参考数据核查PDCA方法论，对清洗之后的数据模型提供多种数据核查规则，实现多角度诊断数据问题，且出具核查报告，帮助倒逼系统优化



# 数据质量核查

- 质量规则定义后，通过核查任务生成不同规则的指标报告，对于有异常的数据能溯源到患者明细粒度。如‘医嘱表’的‘年龄’应该大于0，质量平台检测到出现126条年龄出现异常的数据，并通过问题稽查模块追溯到明细记录，发现年龄数据被赋值‘0.0’。

医疗信息数据治理及质量管理

异常数据

id	patientid	drugname	dosage	unit	unit	unit	unit	unit	unit	unit	unit	unit	unit	unit	unit
4201100	ONCE	None	None	None	身份证	None	0.0	nan	None	None	0378e05...	None	None	nan	
4201100	ONCE	None	None	None	身份证	None	0.0	nan	None	None	8f1ck1...	None	None	nan	
4201100	ONCE	None	None	None	身份证	None	0.0	nan	None	None	ab3zf76...	None	None	nan	
4201100	每日一次	None	None	None	身份证	None	0.0	nan	None	None	360b38...	None	None	40.0	
None	一日一次	None	None	ml	身份证	558	0.0	1014639...	None	1	ZXY230...	西药	None	1.0	
None	一日一次	None	None	mg	身份证	554	0.0	1014770...	None	1	ZXY230...	西药	None	1.0	
None	一日一次	None	None	ml	身份证	552	0.0	1015251...	None	1	ZXY230...	西药	None	1.0	
None	一日一次	None	None	ml	身份证	553	0.0	1014770...	None	1	ZXY230...	西药	None	1.0	
None	一日一次	None	None	mg	身份证	551	0.0	1014975...	None	1	ZXY230...	西药	None	2.0	

展示数据为部分抽取数据

2023 南湖HIT论坛课件 版权归演讲人所有

- 基于可视化指标拖拽引擎，实现院内指标的任意配置，支持多种图表，且配置后快速生成图表页面，满足各类指标系统的快速构建。



# 智能配挖- 自主创立智能化管理



数据挖掘指的是从大量的数据中，**提取隐藏在其中的、未知、潜在有用的价值的过程**，也指数据读取、数据清洗、数据变换、数据挖掘实施过程、模式评估、知识表示整个循环的过程，具体任务分为探索性描述任务、统计分析、预测任务（回归、分类）、关联分析、聚类分析、异常值检测等。

## 医疗应用场景



辅助临床科研



助力数字化建设



优化运营管理



推动医保改革

- 数据挖掘在数据量大、维度高、信息化建设完善的场景下可释放更大的价值



## 算法组件丰富

系统内置4大类、共**100多个组件**，满足日常数据清洗、数据分析、数据可视化、数据建模需求



## 预置分析模板

系统内置几十种运营指标计算和疾病分析的成熟算法模板，减少重复工作，供直接使用



## 简单易用

算法组件‘一键式拖拽’即可添加；模板‘一键式部署’即可调用；机器学习**超参自优化**；对接医院业务数据，**可视化筛选导入**



## 可视化展示

针对不同类型组件特制展示模板，包括数据可视化、分析结果可视化和模型评估结果和**报告结果展示**



## 个性化分析

对于复杂的分析和建模需求，系统支持python语言、SQL语言的自定义算法嵌入运行；可以项目形式，由专业团队**1对1沟通需求，制定分析流程**，提供分析服务



## 资源合理调度

系统基于k8s容器化部署，企业级负载均衡策略，具有高可用、弹性伸缩等特点，底层集成分布式计算引擎，**高效的利用资源**

系统通过对算法封装为组件，实现了从数据接入、数据预处理、数据探索、特征工程、模型构建、模型评估、模板部署到最终工程化应用的全流程解决方案，为医护人员提供自动化、智能化、一键式的数据分析、数据建模，降低学习门槛、缩短分析周期、提升建模效率。







事前

## 事前重“防”

- 防止首页、清单数据填报错误
- 防止主要诊断、手术选择错误
- 防止医保DRG/DIP入组错误
- 防止违规开具药品、耗材、诊疗项目



事中

## 事中重“治”

- 规范病案首页填写、质控流程
- 规范医保结算清单质控、上报流程
- 规范医嘱、用药审批流程
- 规范医生诊疗行为



事后

## 事后重“理”

- 梳理医院主要违规条目，并制定审核规则
- 对比各科室运营情况，寻找科室短板
- 梳理院内病种运营问题，实现精细化管理
- 分析回顾病人诊疗过程、费用结构等



■ 本科室    ■ 全院最优值    ■ 全院均值



## DIP科室画像

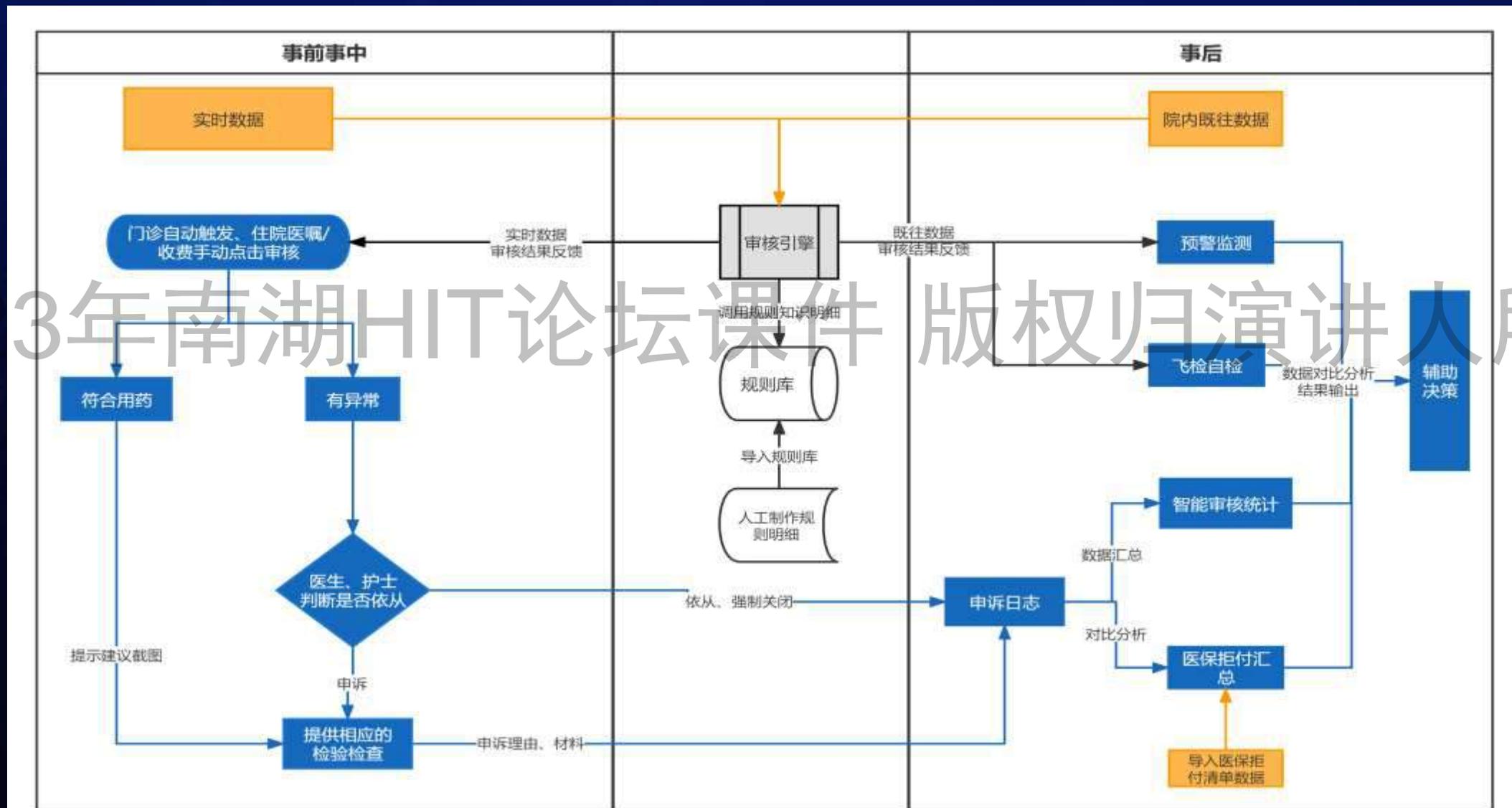
- 构建科室画像：选取DIP运营关键指标，从服务能力、服务效率、服务质量等方面，构建科室画像，直观反应科室优、弱勢；
- 找准定位：对比本科室与全院均值差异，分析本科室在全院服务能力、效率、质量各方面的定位，合理制定科室发展目标；
- 补齐短板：对比本科室与各指标全院最优值所在科室的差异，寻找科室短板及优势，集中精力攻关克难，补齐短板。

## 同一病种（诊断编码前四位）不同治疗方式对比分析

- 入组人数最高的治疗方式：关注次均结余，分析超支病例的主要超支原因，提高病种效益；
- 次均费用最高的治疗方式：关注该治疗方式下的费用结构，注重成本管控，降低病种费用；
- 新技术诊疗方式：关注诊疗技术提升，注重临床路径管理，打造品牌效应；



# 医保基金违规审核业务流程



❑ 人血白蛋白：限抢救、重症或因肝硬化、癌症引起胸腹水的患者，且白蛋白低于30g/L

## 限抢救且白蛋白低于30g/L

- 检验指标：白蛋白低于30g/L
- 费用记录：大、中、小抢救

## 限因肝硬化引起胸腹水 且白蛋白低于30g/L

## 限重症且白蛋白低于30g/L

- 检验指标：白蛋白低于30g/L
- 费用记录：重症相关费用，如呼吸机、重症监护
- ICD-10：重症相关诊断（如重症感染、HIV感染、重度烧伤）

## 限癌症引起胸腹水 且白蛋白低于30g/L

审核引擎

## 效果

1

### 合理提高医保结余

通过数据治理、病案首页质控、医保结算清单质控等，提高院内数据填报准确性，保证医保DRG/DIP合理入组，减少因数据缺失、错误导致的不合理支付，促进医保结余。

2

### 降低医保拒付风险

建立事前、事中、事后全流程审核模式，三道关卡保刚需。  
精准信息，精准提示，减少医生违规行为，降低医保扣费金额及风险。  
持续监测医保指标，医院管理者随时自查院内医保费用状态，及时调整管控策略。

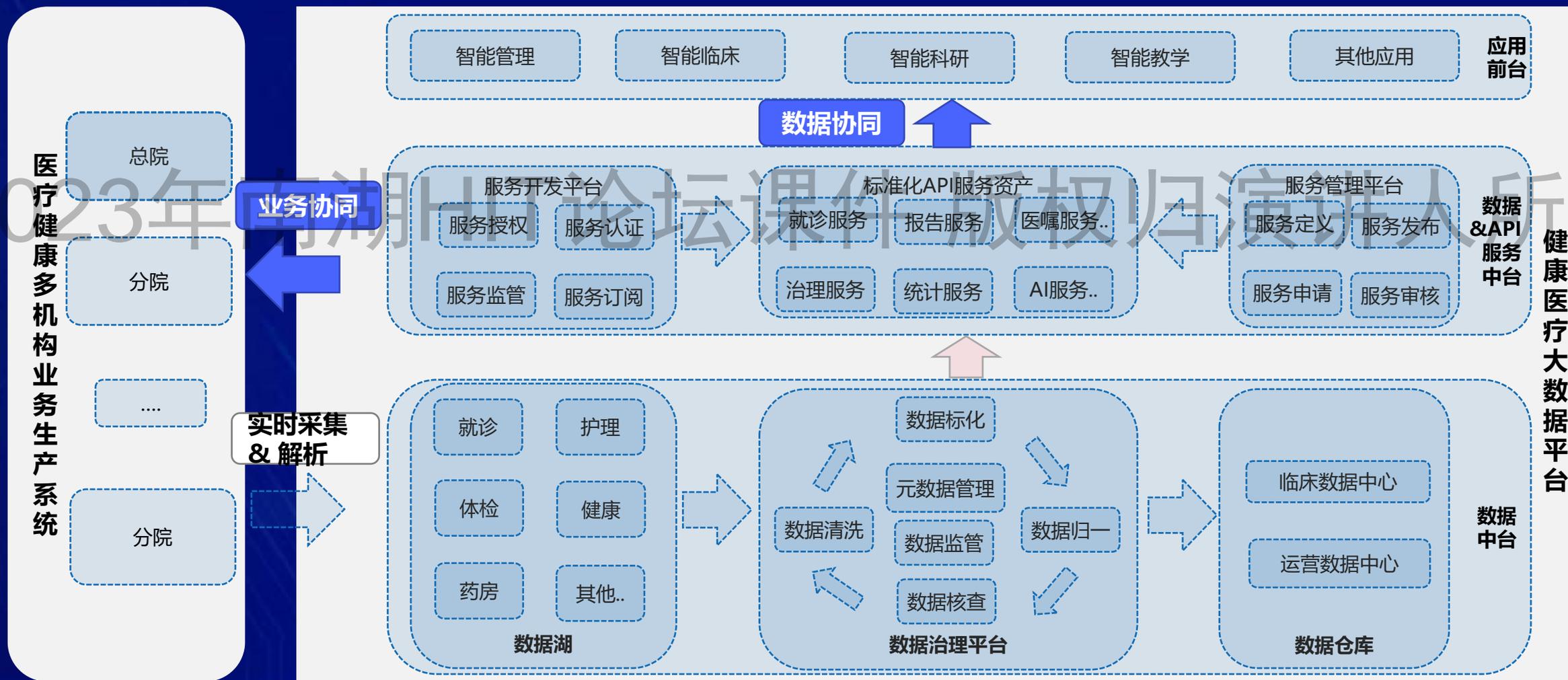
3

### 长效提升院内医保监管能力

建立医保诊疗监测模型，对重点风险行为、重点监控药品耗材、医疗风险场景和重点监控对象进行智能筛查分析，为医院管理决策提供参考。

# 业数协同

- 大数据平台利用高效的采集能力和完善的治理能力，构建了湖仓一体的数据中台，基于临床和运营数据中心对外定义访问服务并进行全生命周期管理，形成涵盖标准化API服务资产的服务中台，服务中台根据数据实时性，利用近实时服务满足院内管理、临床、科研等数据协同需求，利用实时服务满足多机构间各类实时报告查询等业务协同需求，从而实现医疗健康多机构“业数”协同平台的建设。



**数据大脑采用开放服务的方式，给医院各类业务系统提供基于数据提供、基本计算以及人工智能计算服务，同时数据量越多、指标越多、模式越多，数据大脑越智能。**

**数据大脑不再是单一的数据分析系统，而是作为各类业务的计算服务中心。**



秦晓宏@柯林布瑞  
上海徐汇



扫一扫上面的二维码图案，加我为朋友。

**ClinBrain**  
柯林布瑞

# 构建数据大脑

2023 柯林布瑞 - 医疗大脑

探索

发现

突破

2023年南桥论坛课件 版权归演讲人所有