

 腾讯医疗健康

基于大数据技术 医疗数据中台应用与实践

腾讯医疗数据中台 谢彬
2021年7月

CHIMA2021会议资料
未经授权请勿转载
www.chima.org.cn

有没有同样的烦恼?

数据多样性

多源异构 融合内外数据

数据标准不统一

业务口径 管理口径不统一

需求多、接口多

流程再造 数据上报 IOT

处理速度

计算处理速度 科研&决策速度

数据安全

想用数据，用不踏实

数字化转型需要一个新平台驱动

信息化现状

- His+EMR+Lis+PACS+手麻+财务+人力+成本核算+后勤.....
- 互联网医院+互联网服务+科室科研+AI+院内IOT+大数据.....
- 公共卫生垂直系统+基层+急救协同.....

应用能力

赋能

| | | | | | |
|------|--------|------|--------|--------|--------|
| 业务重塑 | 医学知识图谱 | 医患标签 | 多维增长分析 | 数据实时上报 | 疾病预测监测 |
| | AI预测模型 | 规则引擎 | 应用协同服务 | 新冠监测预警 | 精准推荐管理 |

对外服务

管理

| | | |
|------|------|------|
| 统一出口 | 统一指标 | 统一服务 |
| 资产管理 | 安全审计 | 隐私脱敏 |

大数据云计算

提速

| | | | |
|-----------|------|------|------|
| 数据开发 | 数据治理 | 数据质控 | 数据资产 |
| 大数据引擎 | | | |
| 实时计算/离线计算 | | | |

医学数据标准

提质

| | |
|--------|--------|
| 医疗数据标准 | 医学术语标准 |
| 自然语言处理 | 医学质控规则 |

全域数据源

| | | | |
|----|----|----|----|
| 医院 | 公卫 | 区卫 | 疾控 |
| 公安 | 民政 | 交通 | 其他 |

腾讯数据中台整体架构

腾讯医疗健康



实时高效 Real-time

 流批一体数据处理

 高效快速运算

用户导向 User-oriented

 一体式可视化

 用户设计体验

安全可信 Security

 资产透明管理

 全链路安全管控

开放共赢 Together

 中台能力开放

 生态合作共赢

Real-time 实时同步，高效运算

腾讯医疗健康

流处理：监测预警及健康信息 5s更新



批处理：全市数据10G/天

千万级千条质控规则&NLP算法 30min 治理
传统模式3小时以上

亿级数据 2s查询

User-oriented 用户导向——一体化可视化全流程设计

腾讯医疗健康

✓ETL流程-模型构建

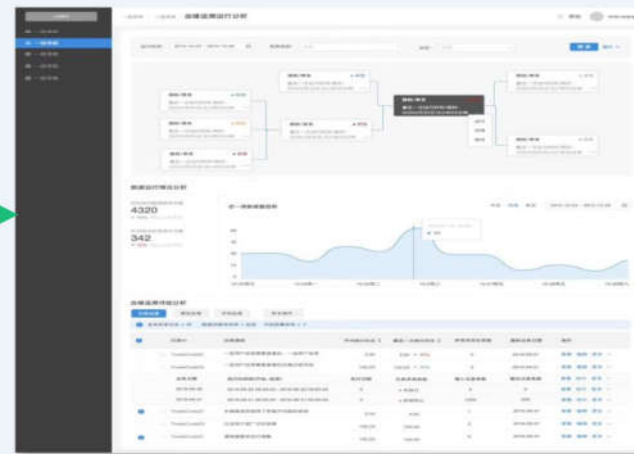
| 字段英文名 | 字段中文名 | 字段类型 | 字段长度 | 字段描述 | 主键 | 单位 | 字段标识 | 值域范围 | 操作 |
|-----------------|--------|----------|------|-----------|----|----|------|------|----|
| id_time | 业务时间 | datetime | 100 | 记录字段创建... | 否 | 秒 | | | |
| import_time | 数据导入时间 | datetime | 100 | 记录字段创建... | 否 | 秒 | | | |
| upload_batch_id | 数据上传批次 | varchar | 100 | 记录字段创建... | 否 | | | | |
| name_id | 姓名ID | varchar | 100 | 记录字段创建... | 否 | | | | |
| height | 身高 | int | 100 | 记录字段创建... | 否 | 厘米 | | | |
| weight | 体重 | int | 100 | 记录字段创建... | 否 | 公斤 | | | |
| sex | 性别 | int | 100 | 记录字段创建... | 否 | | | | |

| 分区名称 | 字段英文名 | 分区类型 | 分区数据格式 | 操作 |
|------|---------|------|-----------------------|----|
| 一级 | id_time | 时间分区 | 按月 | |
| 二级 | id | 范围分区 | 1-100,100-200,200-400 | |
| 三级 | name_id | 哈希分区 | YYYYMMDD | |

✓数据质控

| 字段名称 | 数据类型 | 值域范围 | 内容合理性 | 操作 |
|------|------|---------|----------|----|
| 姓名 | 字符串 | 1-100 | 是否包含特殊字符 | |
| 年龄 | 整数 | 0-150 | 是否在合理范围内 | |
| 身高 | 浮点数 | 1.5-2.5 | 是否在合理范围内 | |
| 体重 | 浮点数 | 40-150 | 是否在合理范围内 | |
| 性别 | 枚举 | 男,女 | 是否有效 | |

✓治理运维



自动化+可视化

根据数据标准自动创建多层级数据仓库模型、数据质控规则的能力

根据医疗数据加工流程的特性，数据中台提供了自动创建数据同步、数据质控 workflows 的能力

5层数据质控保障

数据一致性质控：业务系统->数据中台数据不丢失
字段级质控：字段的关联，值域，范围，内容合理性
标准化质控：是否完成数据标准化，标准更新
数据趋势质控：一段时间数据突然减少或是增多提醒
数据溯源质控：根据数据血缘定位数据源头问题

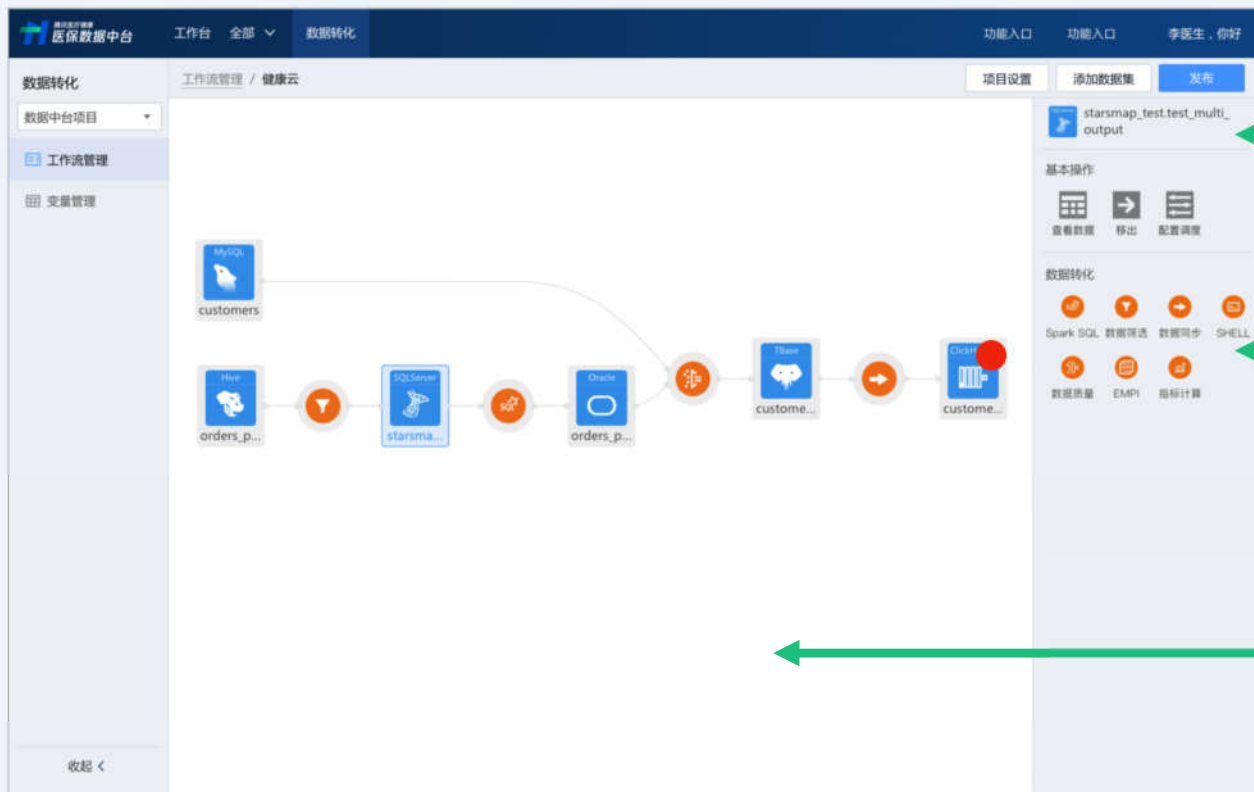
一站式+易溯源

由于数据中台的数据加工在一个模块中向导式、一站式的完成
可以从数据的任意一个节点向前、向后溯源。例如从指标、数据服务向前可以很容易的溯源到数据源头的业务表。为全链路的数据治理提供保障。

CHIMA2021会议资料
未经授权请勿转载
www.chima.org.cn

User-oriented 用户导向——引导式操作体验

腾讯医疗健康



页面操作流程简单便捷，用户操作引导充分

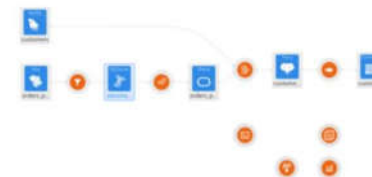
高效数据处理

集成数据处理常见操作，拖拽生成处理流程



所见即所得

复杂数据加工流程可视化展现，结果实时预览



数据DevOps平台实现了一站式，可视化的数据治理编排能力
通过图形实现复杂数据生产治理流程的编写，大幅度提升了数据开发与管理人员的操作效率

Security 安全可信——资产透明管理

腾讯医疗健康

患者医保使用行为分析？

诊断术语标准有多少？ 质控规则有哪些？

当前发烧患者多少？ 医生行为数据分析？

诊断术语标准有多少？ 检查类目有多少？

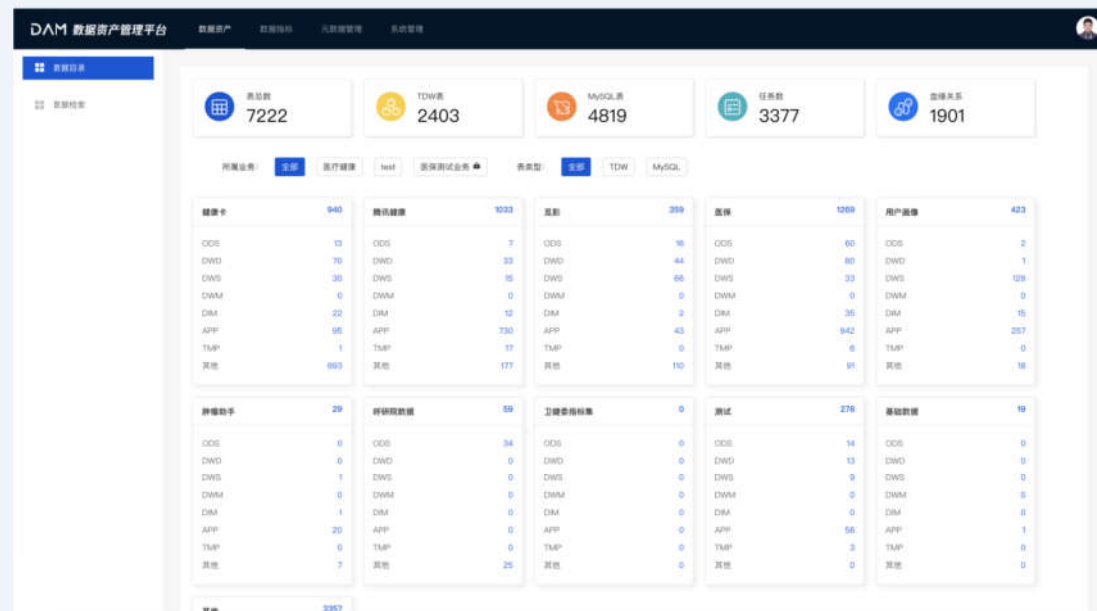
当前患者数量有多少？ 标准数据集当前什么版本？

病历数量？ 门诊急诊？ 住院多少？

运营分析指标有哪些？ 数据治理哪些内容？

当前医疗数据域有哪些？

公卫协同服务有哪些？



数据资产全景

资产全链路管理

数据资产目录

医患行为标签

医疗业务指标

元数据管理

质控规则

血缘链路

逻辑定义

CHIMA2021会议资料
未经授权请勿转载
www.chima.org.cn

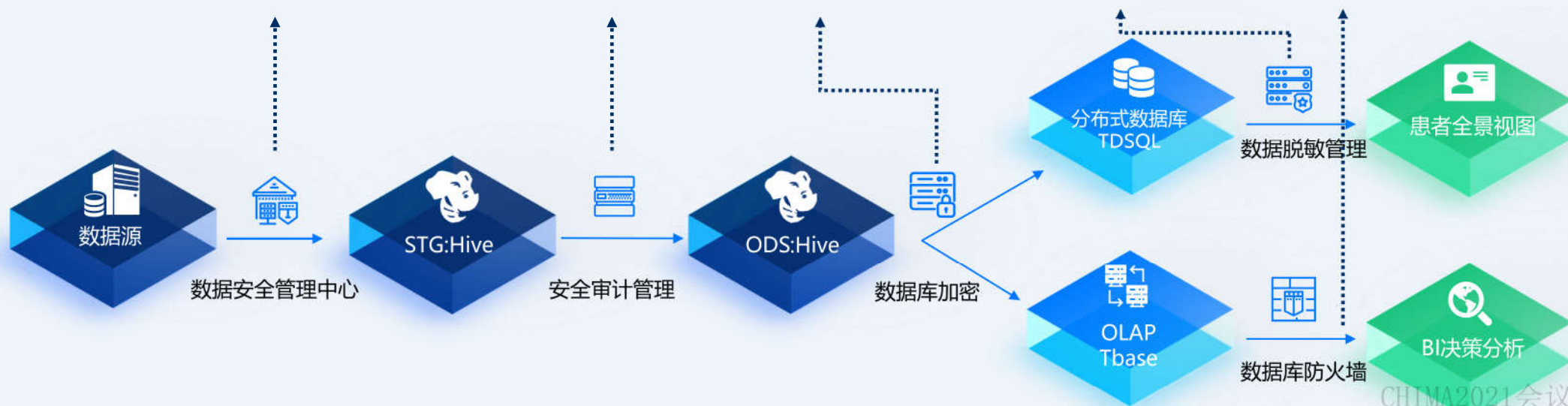
Security 安全可信——全链路安全管控

腾讯医疗健康

医疗资源目录

- 医疗数据标准集
- 元数据管理
- 主数据管理
- 医疗清洗&质控则
- 医学术语标准
- 医疗NLP算法
- 业务数据模型
- 医疗业务指标
- 临床标签分类
- 医疗协同服务
- 查询注册服务
- 定制医疗服务
- BI分析模板
- 医学知识规则
- 疾病预测模型等

- 数据采集风险
- 数据来源不明
- 流向未知
- 数据治理/运维风险
- 违规操作
- 越权访问
- 数据开发风险
- 未脱敏数据
- 违规访问
- 数据共享风险
- 敏感数据泄露
- 越权使用风险
- 数据使用风险
- 第三方及业务调用泄露风险



CHIMA2021会议资料
未经授权请勿转载
www.chima.org.cn

中台应用 症候群监测

腾讯医疗健康

数据汇集->新冠标准->治理->质控

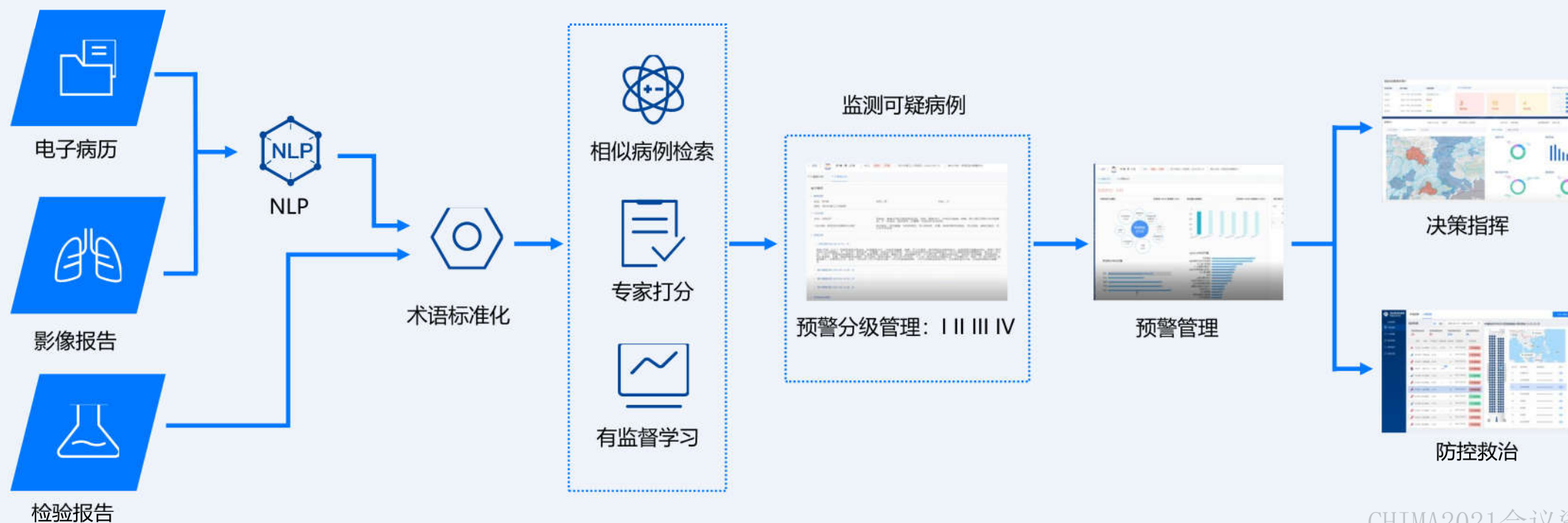
NLP+AI模型

症状提取->规则评分->画像监测

流批一体

监测预警->决策指挥->防控救治

实时预警



中台应用 企业微信链接

腾讯医疗健康



临床数据预警



临床数据查询



运营数据查询

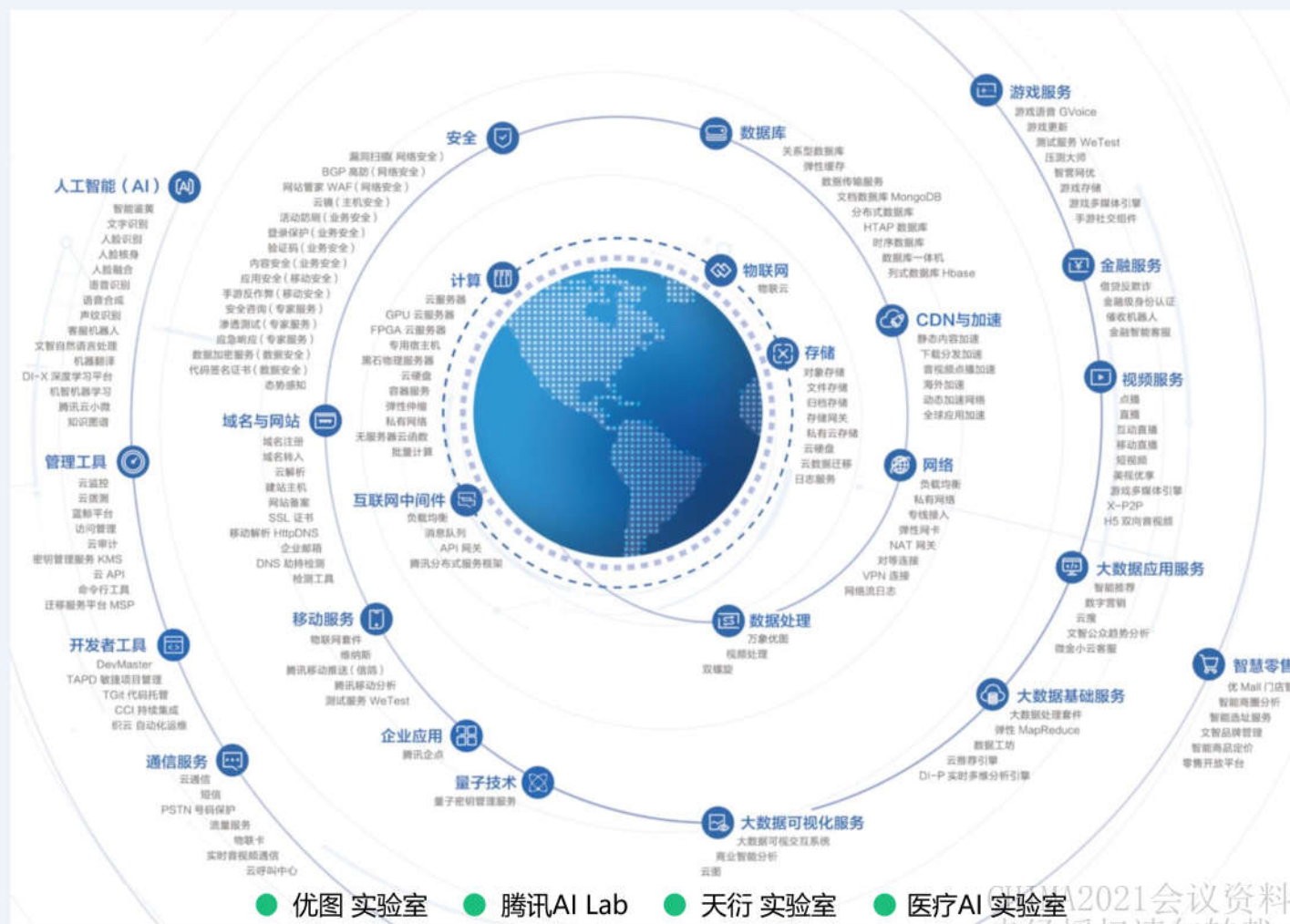
中台应用 医学术语、图谱和AI模型的积累

知识图谱术语覆盖内容

- 50万医学实体，200万关系，术语网络收录疾病、症状、药物、检查、检验、手术、食物、治疗、穴位等。
- 症状疾病部分，支持ICD10国际四位编码（9200条），ICD10国标六位编码（32000条）；影像学检查，支持RadLex3.0（45000条）；手术与操作部分，支持ICD9-CM3（9000条）；药品，支持CFDA药品库3.0（35万）

医疗NLP引擎

- 覆盖12类常见的医疗NLP算法，包括实体识别、意图识别、实体链接、query纠错、语义搜索等，平均准确率高达95%
- 诊断术语标准化引擎：映射至医保ICD10准确率97%
- 手术术语标准化引擎：映射至医保ICD9-CM3准确率95%
- 药品归一化模型：能识别药品商品名、通用名等。
- 医疗文本结构化引擎：调用量1亿+/天



技术支撑开放

大数据运算能力

大数据引擎运算与OLAP库开放

数据通道开放

数据开发->数据质控->数据服务通道开放

数据标准开放

信息数据标准开放

数据集, 值域, 质控规则开放

术语标准开放

医学术语, 后结构化服务开放

应用对接开放

数据工具使用开放

人群标签, 规则引擎, 机器学习工具开放

业务应用引入开放

应用系统引入开放, 提供数据服务

省市级卫健委、医院案例

数据信息

| 便民公众号 | | 平台门户 | |
|--------------|---------|----------------|----------------|
| 基础管理平台 | 业务应用和监管 | 标准 规范 体系 | 安全 保障 体系 |
| 数据协同服务 | 大数据应用 | | |
| 数据治理(采集治理质控) | | | |
| 测评服务 | | | |

卫健委



医院



T

Tethnology

科技助医

海量数据处理能力
全面医学知识库

R

Real-time

实时高效

流批一体引擎
实时高效运算

U

User-oriented

用户导向

体式可视化
引导式操作体验

S

Security

安全可靠

资产透明管理
全链路安全管控

T

Together

开放共赢

数据能力开放
生态合作共赢



腾讯医疗健康