

## 建立全诊疗数据的数据仓库分项报价

序号	具体工作内容	单价 (万元)	工时 (人*天)
1	构建数仓技术架构：至下而上构建(ODS, DWD, DWS, ADS)医学数据仓库的四层架构。	6	30
2	数据源连接：识别多模态数据源，如文本、关系数据库、非关系数据库。并支持不同数据源转化后入库数据仓库进行存储。	6	30
3	数据抽取(Extract)：利用 ETL 定义数据抽取规则，支持根据时间戳、数据变更标识等条件，支持定期或实时地从各数据源中抽取数据。	4	20
4	数据及文本解析：以文件形式存在的数据，如 CSV、XML 使用文件读取技术后转化为脚本数据进行入库	4	20
5	数据清洗模型-缺失值处理清洗：抽取到数仓中的缺失值数据，采用填充法、均值、中位数、众数填充数值型缺失值；基于决策树的多重填补法来预测缺失值。	4	20
6	数据清洗模型-错误值处理：如不符合逻辑的日期、超出范围的数值等。对于错误值，根据业务规则进行修正或删除。	4	20
7	数据去重：使用基于数据库的 DISTINCT 关键字或 ETL 工具中的去重组件，去除重复的记录。	4	20
8	数据标准化与归一化：将 ODS 中数据进行标准化或归一化处理，使其具有相同的均值和方差，以便于后续的分析建模。	3	15
9	维度建模：采用星型模型，雪花模型设计数据仓库的架构。以患者维度、时间维度、疾病维度等为核心，构建事实表和维度表。	3	15
10	数据粒度确定：根据分析需求确定数据的粒度。根据不同维度指标，构建由细到粗的粒度存储模型。	3	15



11	开发指标模型实现对数据的多维分析：通过动态配置维度和指标的方式，快速生成各种报表和可视化图表，进行钻取、切片、切块等操作。	2	10
12	访问控制：建立严格的访问控制机制，使用身份验证和授权技术，确保只有授权用户能够访问和操作数据仓库中的数据。	1	5
13	数据加密：对敏感的医学数据，如患者的身份证号码、联系方式、疾病诊断结果等，在存储和传输过程中进行加密处理。使用对称加密算法（如 AES）对数据进行加密。	1	5
14	元数据定义与存储：定义和存储数据仓库中的元数据，包括数据源信息、数据模型定义、数据转换规则、数据字典等，在数据仓库中来管理和维护元数据。	1	5
15	质量监控与评估：定期对数据质量进行监控和评估，检测数据是否符合质量规则，及时发现和解决数据质量问题。	1.9	9.5
16	构建各层数据链路并打通，集成 ETL（数据抽取与加载）。	1.9	9.5
17	<b>合计（万元）</b>	<b>49.8</b>	<b>249</b>



深圳市中科童颜生物科技有限公司

2025年1月10日

