一、卫生健康数据质控系统的组成部分：

1.质控对象管理子系统（数据库、共享文档、其它数据文件（如excel 等）。）

2.质控规则管理子系统（最大/最小值的异常值分析法、标准偏差异常值分析、正则表达式验证规则、SQL语句、指标计算分析比较）

3.质控任务管理子系统（质控规则或作业的具体执行）

4.质控报告管理子系统是（可视化显示输出、可共享交换的格式文档或接口、输出到数据库）。

5.系统及安全管理部分。

6.质量报告通知管理子系统（数据质量报告，通过邮件、短信等方式，告知第三方）。

二、初步设计思路

1.数据采集

支持接入多种数据库类型，按照数据库类型，可采用sqluldr,sqoop,datax等工具进行初始数据加载，采用ogg、dsg、logminer、物化视图日志等方式，获取后期增量数据。如果来源数据能够基于批次号来区分增量，同时不存在对历史数据的变更，也可以通过时间标识来进行增量数据采集。

数据文件，通过将数据文件内容，转换成数据库表方式进行处理。通过前台功能上传，明确数据文件字段与列对应关系。

2.规则管理

规则分为两类，一类为通用规则，一些常用且通用的规则，系统默认设置好，如：判断空或者非空、判断字段格式、判断身份证号码合法性等；另一类为自定义规则，按照实际的业务应用场景，自定义数据质量检测规则，如：费用明细是否等于费用合计，码值是否合法，数据取值范围等。

规则可以基于正则表达式、SQL语句方式或自定义算法函数来配置。

3.任务配置与调度

数据质检任务，设计基于流程图的方式进行配置，根据选择的数据表资源，再选择需要使用到的检验规则。按照流程，可以配置一个或多个检验规则，可以设计顺序执行，也可以通过依赖进行执行。

任务调度，基于Quartz开源作业调度框架，进行任务调度，支持手工调度和定时自动调度两种类型。

4.质检报告、可视化展示

平台支持导入自定义的质检报告模板，结合实际的质检分析数据，生成对应的质检报告，该报告支持下载、在线打印等操作。

可视化展示，可采用echarts、highcharts、three.js等开源技术，来对质检分析结果，进行可视化的展示。

数据输出到数据库，可采用sqlldr、datax等技术手段，实现将质检结果数据分发到数据库中。

5. 消息通知

在完成数据质量分析后，通过调用短信、邮件接口，发送短信和邮件通知。