

一种面向软件配置参数管理的数据处理方法

李朝阳, 张小锋, 周 波, 徐同乔
(中国西昌卫星发射中心, 四川 西昌 615000)

摘 要:针对国产平台下应用软件运维效率低下的难题,提出一种面向软件配置参数管理的 Word/Excel 数据清洗处理方法。该方法通过综合应用 pandas, paramiko 以及 python-docx 等多种 python3 开源库,构建了以原始数据文件处理为核心,融合远程控制,文件传输及日志记录等多种功能于一体的国产应用软件配置文件生成模型。基于该模型在国产麒麟操作系统下采用 python3 语言编程实现了一套面向多种业务应用系统的配置参数文件快速处理软件。该软件在典型应用环境中的部署应用实践表明,提出的面向软件配置参数管理的数据处理方法结构清晰,设计合理,可操作性强,具有良好的可扩展性和应用前景。

关键词: Word; Excel; paramiko; 配置管理; 麒麟操作系统

中图分类号: TP311.11

文献标识码: A

文章编号: 1673-629X(2021)0083-04

A Data Processing Method for Software Configuration Parameter Management

LI Zhao-yang, ZHANG Xiao-feng, ZHOU Bo, XU Tong-qiao
(China Xichang Satellite Launch Center, Xichang 615000, China)

Abstract: Aiming at the problem of low efficiency of application software operation and maintenance under domestic platforms, we propose a Word/Excel data cleaning processing method oriented to software configuration parameter management. This method uses a variety of python3 open source libraries such as pandas, paramiko and python-docx to construct a domestic application software configuration file generation model that takes raw word or excel data files processing as the core and integrates remote control, file transfer, and log recording. Based on this model and the domestic Kylin operation system, we use python3 programming language to accomplish a configuration parameter rapid processing software for multiple application systems. The deployment and application practice of the software in a typical application environment shows that the proposed data processing method for software configuration parameter management has a clear structure, reasonable design, strong operability, and good scalability and application prospects.

Key words: Word; Excel; paramiko; configuration management; Kylin operating system

0 引 言

随着互联网安全态势进一步严峻,许多重要领域的 IT 基础设施的高可用性和高安全性面临更大风险和挑战。于 2006 年启动的核高基项目作为最重要的举措之一,其对政府、国防、金融、电力及保密等领域的信息化建设和应用影响深远,最为显著的便是很多部门都采用中标麒麟可信操作系统及其配套的国产硬件。基于麒麟系统构建的应用软件于过去近十年以来已经形成规模,但其繁杂配置管理任务给运维人员带来巨大挑战。从技术层面分析上述运维困局,一方面基于麒麟平台构建的应用系统的运行稳定性非常依赖于繁杂的配置文件,手工编写及修改则出错几率增加;另一方面配置文件所需各类参数通常包含在各个部门

提交的 Word 文件、Excel 文件及文本文件中,手工逐个转换耗费大量人力资源且正确性难以保证。鉴于此,文中针对国产麒麟平台下的应用系统配置管理任务,提出一种基于开源 python 库的 Word/Excel 数据文件清洗处理应用方法,为更进一步构建自动化智能化的运维系统提供基础服务。

1 技术途径

1.1 基于 pandas 的 excel 数据处理方法

pandas^[1] (python data analysis library) 是基于 numpy 的一个工具集,专为解决数据分析任务而创建。Pandas 纳入了大量库和一些标准数据模型,为大型数据集提供高效操作工具。同时, pandas 是 python 的一

收稿日期: 2021-03-20

基金项目: 中国西昌卫星发射中心科研项目 (XSLCTX202002160006)

作者简介: 李朝阳 (1981-), 男, 硕士, 工程师, 通信作者, 研究方向为计算机软件与应用。

个数据分析包,最初被作为金融数据分析工具而研发,其在时间序列数据分析方面表现卓越,pandas 的主要组件是 Series 和 Dataframe。文中主要以 pandas 为工具对各部门提交的 excel 数据进行清洗和处理,从中提炼出各类配置管理任务所需参数。

1.2 基于 python-docx 的 word 数据处理方法

Python-docx^[2]是一个对 word 文件进行创建和更新操作的 python 库。其最主要的类为 Document,通过一个文档对象实现对段落、标题、格式及其表格对象的创建和更新。文中主要通过 python-docx 实现对表格中的参数自动提取,在麒麟操作系统上与之配合使用的 word 处理套件为永中 office。

1.3 基于 paramiko 的远控主机管理方法

paramiko^[3]是 python 中的一个非常著名的开源 SSHV2 项目,基于 python2.7 和 python3.4+ 开发,其最早的开发者是 Jeff Forcier。paramiko 同时支持 SSH 服

务端和客户端,源码可以从 Github 上下载。paramiko 基于 Cryptography 库完整实现 SSH 的网络概念。文中一方面利用 paramiko 的 SSHClient 对象在远程麒麟主机上执行 shell,另一方面则利用其支持的 sftp 功能将经数据分析所得的配置文件上传远程麒麟服务器端的指定位置。

2 典型结构设计

图 1 为典型结构设计示意图。从操作流程而言包括三个阶段,即用户配置输入阶段、数据清洗处理阶段及结果输出应用阶段^[4]。

用户配置输入阶段涵盖程序运行所需系统参数配置和相关 Word/Excel/TXT 原始数据文件准备两方面工作。前者配置内容包括数据库连接信息、生成配置文件路径信息、远端麒麟主机控制信息及 SSH 登录认证信息等。

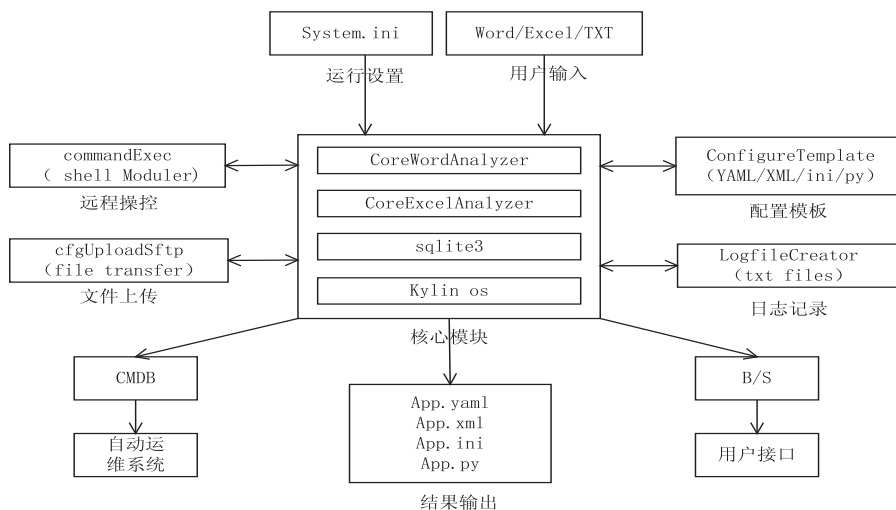


图 1 典型结构设计示意图

数据清洗处理阶段包括核心模块、远程操控、配置模板、文件上传及日志记录五个模块,其中核心模块是整个程序的核心部件,其余四个模块皆围绕核心模块进行工作。核心模块由构建于麒麟操作系统和 SQLite3 数据库之上的 CoreWordAnalyzer 和 CoreExcelAnalyzer 两个子模块,分别实现对永中 Word 数据文件及永中 Excel 数据文件进行清洗处理^[5]。远程操控模块 commandExec 则基于 paramiko 实现,主要完成以 SSH 方式登录远程麒麟服务器并执行相应 shell 命令,为配置文件更新操作做好准备。配置模板模块 ConfigureTemplate 针对麒麟系统上的不同应用服务,提供与具体应用软件相对应的配置文件模板。模板中的参数将由数据分析处理结果进行替代,并最终产生实际应用需求的配置文件。根据应用类型的不同,产生的配置文件格式包括 yaml、xml、ini 及 py 等。文件上传模块基于 paramiko 库的 sftp 类实现,主要完成将

生成配置文件自动上传特定应用系统之上,以提高其自动化运维效率。日志记录模块 LogFileCreator 主要实现对数据清洗处理阶段中各关键环节产生的文本信息进行实时跟踪记录,以便故障排除所用^[6]。

结果输出应用阶段包括用于生产所用的各类配置文件、CMDB 应用及 B/S 用户接口。其中 CMDB 简称配置管理数据库,作为一个信息中心自动化运维的关键和基础,其信息的完备性和有效性对运维工作影响重大,文中所提及的 word 及 excel 原始数据文件的分析处理结果一般是各部门提交的应用系统运维参数,利用此分析结果有效充实 CMDB 将进一步提高信息中心自动化运维系统的执行效率^[7]。

3 典型实例实现

3.1 核心模块实现

文中所述面向配置管理的 Word/Excel 数据清洗

处理程序基于 python3 开发,其编程实现的重点是图 1 所示的数据清洗处理阶段。鉴于篇幅所限,文中以伪码方式重点阐述 CoreExcelAnalyzer 的实现细节。CoreWordAnalyzer 与 CoreExcelAnalyzer 的实现不同仅体现在对 pandas 和 python-docx 库的应用处理上。基于 pandas 实现该数据清洗处理程序的 CoreExcelAnalyzerHandler 核心类,其方法主要包括公共函数和功能函数两部分^[8],实际应用中,依下面伪码所示定义顺序依次调用 CoreExcelAnalyzerHandler 类的八个功能函数,即可从 Excel 原始数据文件中提炼配置参数并结合配置模板生成应用系统所需配置文件^[9]。

```
class CoreExcelAnalyzerHandler(object):
    #基于 excel 数据文件的配置文件生成器
    def __init__(self, m_db_name,
                 m_table_name, m_fexcel_name):
        self.db_name = m_db_name
        self.db_path = self.get_abspath\
            ("db", self.db_name)
        self.table_name = m_table_name
        self.fexcel_name = m_fexcel_name
        self.fexcel_path = self.\
            get_abspath("excel", self.fexcel_name)
        self.insert_data = None
        #=====公共函数=====
    def get_abspath(self, flag, filename):
        #获得各种绝对路径信息
        pass
    def fill_classFlag(self, series_obj):
        #为 Series 对象增加分类标识符
        pass
    def fill_total_row(df, rep_str):
        #处理 NAN 值
        pass
    def switch_template(self, template_flag):
        #根据标识符,选择模板
        pass
        #=====功能函数=====
    def create_db_for_zhys(self):
        #为配置管理信息创建数据库
        if os.path.exists(self.db_path):
            os.remove(self.db_path)
        conn = sqlite3.connect(self.db_path)
        cur = conn.cursor()
        return cur
    def delete_database(self):
        #删除数据库
        pass
    def create_table_based_excel(self):
        #创建配置管理信息数据表
        pass
```

```
def clean_data_from_excel_file(self):
    #借助 pandas 对 excel 进行清洗处理
    io = self.fexcel_path
    df = pd.read_excel(io,
                      sheet_name=0, skiprows=[0, 1])
    df = df.fillna(method="ffill")
    df = df.replace(r"\n", "", regex=True)
    s = df["location"]
    df["myindex"] = range(1, len(df) + 1)
    df["number"] = df["myindex"].groupby\
        (df["super_depart"]).rank()
    return df
def insert_data_to_db(self, df):
    #把 pandas 中的数据构造造成可插入到 DB 的数据
    pass
def create_dictdata_for_phone(self):
    "从数据库中取出数据,构造成字典"
    pass
def create_cfgfile_for_phone(self,
                             cfg_name, template_flag):
    "根据电话信息产生配置文件"
    pass
def create_cfgfile_for_seater\
    (self, cfg_name, template_flag):
    "根据座次信息产生配置文件"
    pass
```

3.2 文件上传模块实现

文件上传模块 CfgUploadSftp 主要基于 paramiko 实现,其关键步骤如下面伪码所示。

```
defcfg_upload_sftp():
    try:
        tobj = Transport((host_ip, int(port)))
        tobj.banner_timeout = timeout
        tobj.connect(username=user, password=password)
        sftp_obj = SFTPCClient.from_transport(tobj)
        sftp_obj.put(local_path_leader_file,
                    remote_path_leader_file)
        sftp_obj.put(local_path_phone_file,
                    remote_path_phone_file)
        tobj.close()
        return True
    except Exception as e:
        return False
```

4 应用实践

目前,各领域基于国产平台研发越来越多的应用软件系统,大多基于 QT、Python 开发。针对上述系统运行状态转换快,配置管理繁杂等难题,文中给出图 2 所示的一个典型面向配置管理的 Word/Excel 数据文件清洗处理程序应用场景。

从图 2 可以看出,以 WE(Word/Excel)数据处理程序为中心,Word/Excel 数据文件作为输入,产生格式为 YAML、XML、ini 等不同配置文件;通过 SSH 方

式将各类配置文件自动发送相应远端国产应用系统,以便随任务进程推进而及时、准确实现应用软件状态转换和操作管理,从而提高工作效率,节省人力资源。

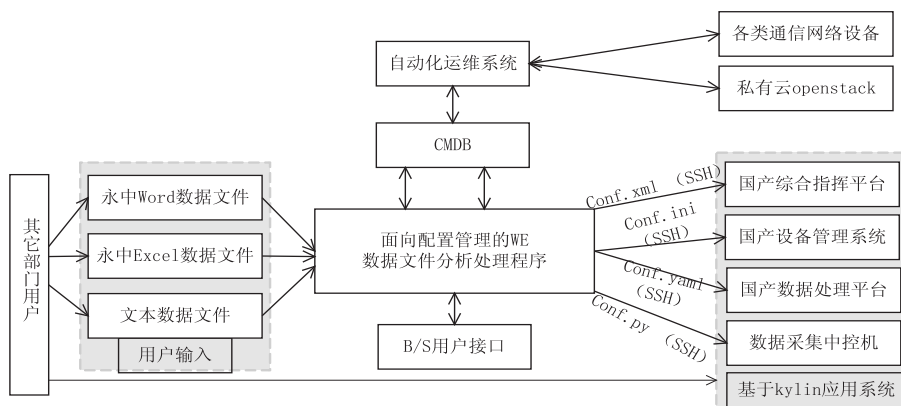


图 2 典型应用示意图

5 结束语

随着国产麒麟操作系统在政府、国防及金融等领域的深入部署应用,构建于其上的大量类 Linux 系统应用软件配置管理繁杂且出错率高的问题日益严峻。为此,提出一种国产平台下面向配置管理的 Word/Excel 数据清洗处理方法。该方法通过综合应用 paramiko、python-docx、pandas 等开源 python 库,构建了一个以 Word/Excel/txt 原始数据文件处理为核心,融合远程控制,文件传输及日志记录为辅助的配置管理文件生成框架。文章采用 python3 语言对该框架描绘的内容在麒麟操作系统上进行了实现,并对实现程序在典型应用环境中安装部署应用。应用实践表明,该方法结构清晰,设计合理,可操作性强,具有良好的可扩展性和应用前景。

参考文献:

[1] MCKINNEY W. Pandas: powerful Python data analysis toolkit Release 1.2.0 [OL]. (2020-11-26) [2021-01-01]. <https://pandas.pydata.org/docs/>.

[2] CANNY S. Python-docx 0.8.10 documentation [EB/OL]. (2019-06-11) [2021-01-01]. <https://python-docx.readthedocs.io/en/latest/>.

[3] FORCIER J. Paramiko's API documentation [EB/OL]. (2020-09-12) [2021-01-01]. <http://docs.paramiko.org/en/stable/>.

[4] MCKINNEY W. Python for data analysis: data wrangling with Pandas, NumPy, and IPython [M]. 2nd ed. Farnham, UK: O'Reilly Media, 2012: 128-135.

[5] MOLIN S. Hands-on data analysis with pandas [M]. Birmingham, UK: Packt Publishing, 2019: 321-326.

[6] 攸 阳, 杨素梅. 软件配置管理在军用软件中的应用 [J]. 软件, 2018, 39(6): 85-88.

[7] 敬 军, 郝福珍, 徐东华. 国产自主平台下可信执行环境的设计与实现 [J]. 计算机工程与设计, 2012, 33(10): 3801-3806.

[8] 封富君, 姚俊萍, 李新社, 等. 大数据环境下的数据清洗框架研究 [J]. 软件, 2017, 38(12): 193-196.

[9] 骆 蓉, 黄 俊, 黎茂锋, 等. 基于 Word 模板的复杂文档快速生成方法 [J]. 计算机应用与软件, 2020, 37(10): 57-63.