2024-3-28

BMC 服务器故障预测与诊断平台

概要设计

高天润 刘开元 王永琪

哈尔滨工业大学
容错与移动计算研究中心

目录

1 比赛要求

1.1 赛题背景

BMC（基板管理控制器）是服务器的控制管理单元，是用于监控和管理服务器的专用控制器，其负责服务器的资产信息显示、硬件监控、散热调控、系统配置、远程监控、日志收集、故障诊断、系统维护等重要功能。BMC 可以提供 IPMI、SNMP、RedFish、WEB 等接口，满足多种系统集成需求，可以对服务器进行精细监控，有效提升服务器的管理效率，降低其运营和维护成本。服务器状态监控管理是 BMC 的一个重要功能，针对服务器场景的高可靠性与稳定性的要求，BMC 需要对服务器各部件的状态进行监控，如 CPU、内存、硬盘等，收集运行状态信息、故障信息，进行分析、统计、提示，进而进行故障预测，这是 BMC 技术发展的方向和必然趋势。

1.2 赛题任务

本次赛题以 BMC 技术为基础，针对服务器典型场景，设计一个故障采集、诊断及预测平台，作品包含但不限于软件应用、硬件实现、软硬件结合的整体解决方案等。参赛方可在系统移植、应用开发两个领域内，任选一项或多项参赛。

2 作品特点

根据本次赛题要求，我们团队选择应用开发，在 Ubuntu22.04 系统上，针对服务器场景，部署了服务器故障与诊断平台，实现数据采集与呈现，故障分析，故障预测三个模块，并给出相应的实现和解决方案。平台以 Web 形式呈现，以 React 前端加 Django 后端为主体架构，后端负责调用模型，模仿真实服务器提供模拟数据，前端负责数据的展示以及检测结果的呈现。

总体架构如下图 2- 1 所示：

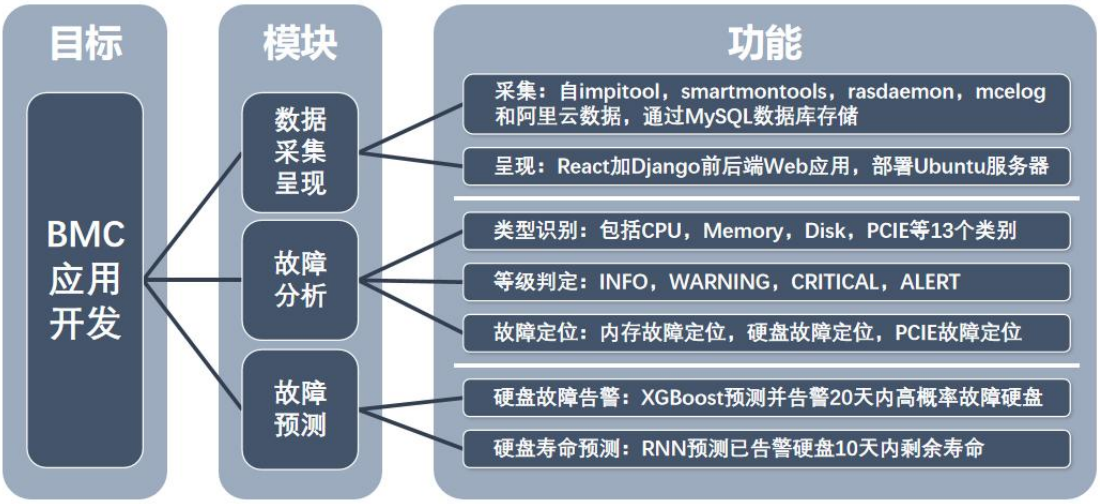

图 2- 1 故障预测与诊断平台总体架构

在数据采集与呈现模块中，平台使用多种 RAS 检测工具收集故障数据并分析故障格式，生成样本容量更大并具有更多故障的模拟数据存储于后端数据库中，前端使用 Web 形式呈现数据以及检测结果。

在故障分析模块中，平台以日志分析为主体，主要实现了故障类型识别、故障等级判定和故障定位三个功能。

在故障预测模块中，平台实现了判断硬盘是否可能发生故障的硬

盘故障预测和判定告警硬盘剩余寿命的硬盘寿命预测。

3 设计思路

整体软件架构示意图如图 3-1 所示，我们实现了从真实的服务器采集数据，交由后端存储和分析，并最终由前端呈现给用户这一完整的系统结构。

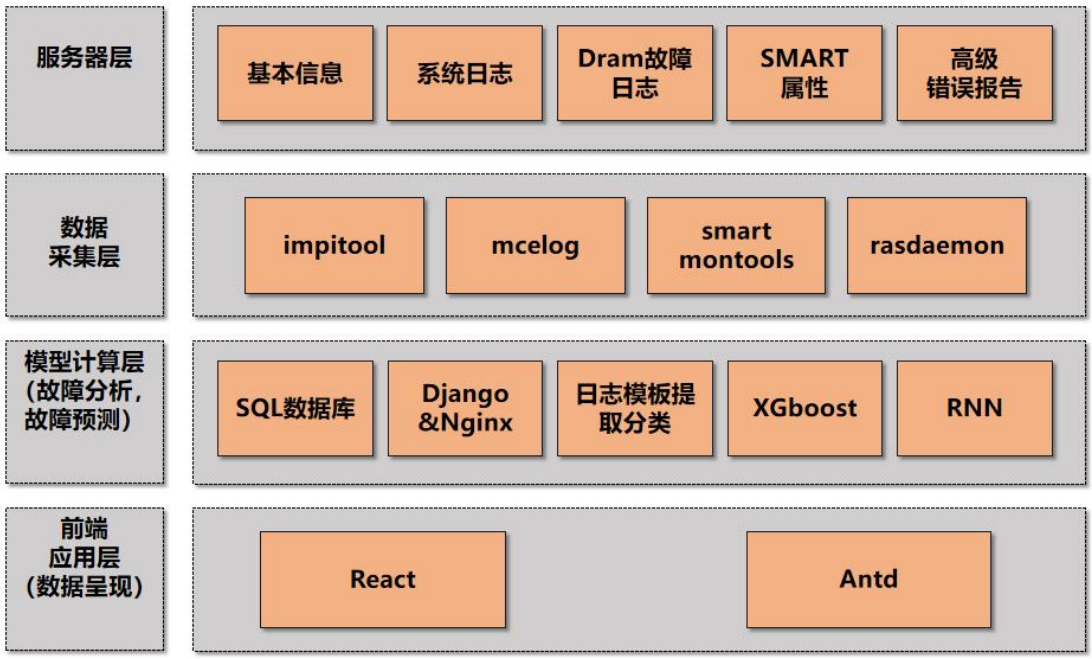

图 3- 1 软件系统总体架构

3.1 数据采集与呈现模块

对于数据采集与呈现模块，平台收集 5 种类型的数据，分别是服务器基本信息、服务器系统日志、Dram 故障日志、硬盘 SMART 属性、PCIE 高级错误报告，其中服务器基本信息由 ipmitool 收集，用来呈现服务器的总体状态，服务器系统日志、Dram 故障日志、硬盘

SMART 属性， PCIE 高级错误报告分别由 ipmitool， mcelog，smartmontools，rasdaemon 这 4 个 RAS 检测工具收集，用于后续的故障分析和故障预测功能。

由于检测工具收集的真实服务器数据样本过少且错误发生频率较低，平台根据工具收集的数据格式，产生相同格式的模拟数据存储在后端数据库，最终由 Web 前端做后续功能的展示。

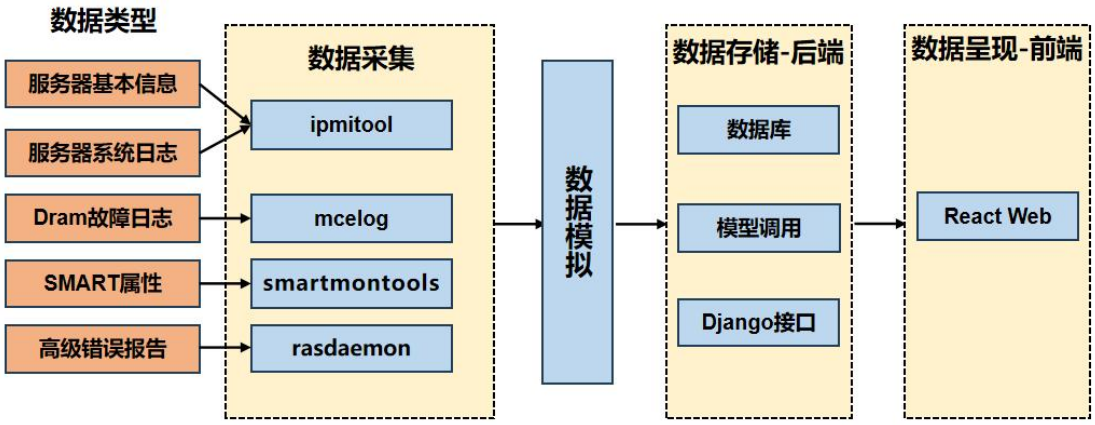

图 3- 2 数据采集与呈现模块架构

3.2 故障分析模块

故障分析模块的主体数据是服务器系统日志数据，根据服务器系统日志的内容，判断出服务器故障等级和故障类型，其中故障等级判定的思路为通过 Drain3 日志解析树提取出原始日志的所有模板并对模板进行分析，根据模板含义将每个模板归类为 INFO、WARNING、CRITICAL 和 ALERT 四个级别。

对于故障类型识别，首先将原始纯文本的日志数据根据单位时间内模板事件发生的次数转为特征向量，并将向量输入全连接网络进行预分类，在预分类的基础上，根据关键字实现最终的 12 分类：CPU、

PSU、Memory、Disk、PCIE、FAN、INTRUSION、OS、STATUS、ACPI、Boot、LAN 和 Other。

平台将故障等级大于 INFO 的故障判定为错误，对 Memory、Disk 和 PCIE 三种类型的错误，平台根据 Dram 故障日志、硬盘 SMART 属性、PCIE 高级错误报告给出的定位信息，对每一条错误日志给出相应的定位。

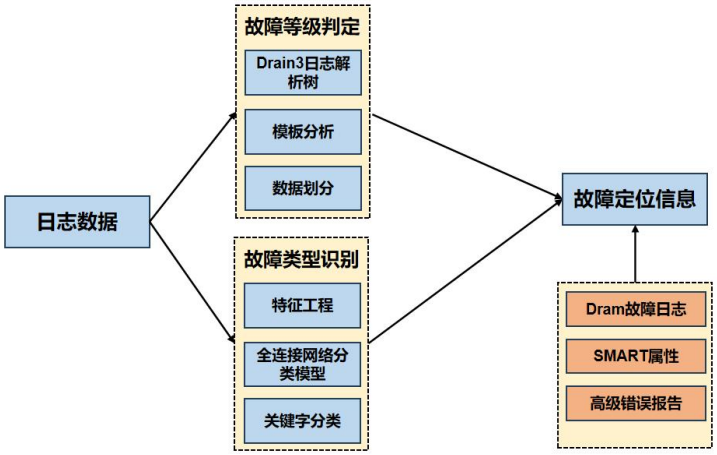

图 3- 3 故障分析模块架构

3.3 故障预测模块

故障预测模块的主体数据是服务器硬盘的 SMART 数据，根据每天生成的 SMART 信息判断硬盘是否会在不久的将来发生故障，如果有较大可能性发生故障（即告警），则预测硬盘的剩余寿命。

对于告警预测部分，我们采用 XGBoost 模型作为基础建立。我们使用 Backblaze 的数据集，挑选某一类型的硬盘中的损坏硬盘，将它们每日的 SMART 数据清洗后，每个硬盘选择设定时间窗口大小的损坏前日志作为正样本，其余数据为负样本，我们还随机抽取等量于损坏硬盘数量的正常硬盘，并以同样的方法，但将数据均标记为数据

集的负样本，迭代训练，并选择效果最好的 XGboost 模型作为告警模型。

对于寿命预测部分，我们采用 RNN 模型作为基础建立。对 XGBoost 模型中的正样本数据，按 9：1 的比例划分训练集和测试集，输入 RNN 中进行训练，并挑选测试集 Loss 最小的作为寿命预测模型。

在服务器上，模型均处于静态，将对应日期的 SMART 属性输入整个模型即可获得告警结果和预测结果。

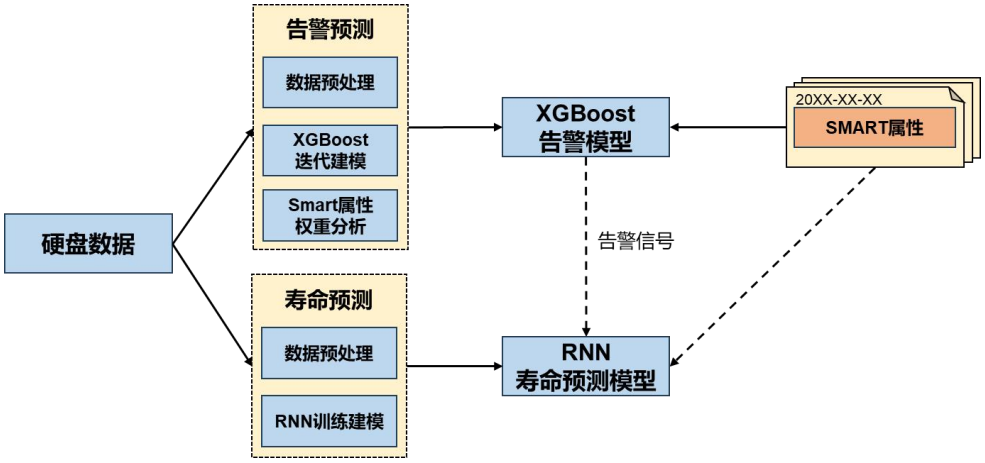

图 3- 4 故障预测模块架构