

# Data Science Programming I

## Lab5 - Linux for Data Science

---

Yang Minqiang [yangmq@lzu.edu.cn](mailto:yangmq@lzu.edu.cn)

# Lab Session

- Shell/shellscript
  - **Bash**
  - Sh
  - Csh
  - Ksh
  - Zsh

# Lab Session

- 写个shellscript脚本lab5.sh, 输出以下结果:

基于yang.lzu.edu.cn/access.csv文件 (nginx日志)。

- 统计有哪些IP访问过我的nginx? 哪个IP访问最多?
- 统计有哪些种类的浏览器, 其比例分别是多少? (一个IP地址看做是一个终端)
- 客户端操作系统有哪些? 终端的操作系统占比分别是多少? (一个IP地址看做是一个终端)

# Lab on Class

- Lab5 : Data Science with Linux

- 1) 提交上一页ppt中问题的shellscript及结果截图
- 2) Linux有哪些工具可能会用于数据科学？分别用于什么场景、什么数据类型？（包括但不限于课堂介绍的，文字、表、图形式不限）
- 3) 举例说明上面工具的用法（请用**真实场景的数据**，给出terminal操作和输出日志。为了看起来方便，请将**命令部分**粗体显示）

**提交文件格式: lab5.pdf**

Deadline: March 27<sup>th</sup> 23:59