

مهندس داده (Data Engineer)

- ✓ زمان لازم برای انجام: ۳ روز
- ✓ داوطلبان گرامی لطفاً پروژه زیر را با دقت خوانده و مستندات مرحله اول شامل توضیح کامل راه حل خود را در قالب فایل word ارسال کنید.
- ✓ مستندات مرحله دوم شامل توضیح کامل مراحل نصب Hadoop و نحوه قراردادن فایل ها در hdfs در قالب فایل word ارسال کنید.
- ✓ فایل یا فایل‌های مربوط به اسکریپت های مرحله اول را به صورت کامل کامنت گذاری کنید و همراه با فایل‌های مستندات ارسال کنید.
- ✓ حجم فایل ارسالی بایستی کمتر از ۱۵ مگابایت باشد و پاسخ را در یک فایل zip. به نام خودتان قرارداده و بارگذاری نمایید.

هدف پروژه

هدف از این پروژه، ایجاد یک Linux shell script است که فرایند log rotation را در محیط لینوکس شبیه‌سازی کند. این اسکریپت باید فایل‌ی را که در اختیارتان قرار می‌گیرد، از انتها به صورت خط به خط بخواند و بر اساس معیارهای ذکر شده در فایل‌های جداگانه سازماندهی کند.

قسمت اول:

- فایل file.log باید از انتها به صورت خط به خط خوانده و پردازش شود.
- فایل file.log باید بر اساس نوع لاگ (INFO,WARNING,ERROR) در فایل‌های جداگانه دسته‌بندی شود.
- نحوه نام‌گذاری فایل‌ها به صورت زیر باشد:

log-error.log	log-warning.log	log-info.log
log-error-1.log	log-warning-1.log	log-info-1.log
log-error-2.log	log-warning-2.log	log-info-2.log
log-error-3.log		log-info-3.log
log-error-4.log		

- نحوه قرارگیری لاگ‌ها در فایل‌ها به این صورت باشد که فایل log-error.log شامل جدیدترین لاگ‌ها و فایل log-error-n.log (برای مثال log-error-۲.log) شامل قدیمی‌ترین لاگ‌ها باشد.
- حجم هیچ فایل‌ی نباید از ۱ مگابایت بیشتر شود.

- زمانی که حجم فایل به حد نصاب مشخص شده رسید، یک فایل جدید ایجاد کنید و محتوای فایل را به فایل جدید منتقل کنید (بر اساس نام‌گذاری تعیین‌شده، فایل‌های جدید را نام‌گذاری کنید) تا فایل جاری (مثلاً log-error.log) خالی شود و آماده دریافت لاگ‌های جدید شود.

- مثالی از ساختار فایل

```
2023-12-31 20:55:28 | WARNING | Missing parameter in request
2023-12-31 20:55:53 | ERROR | Uncaught exception
2023-12-31 20:57:03 | INFO | Configuration file loaded
2023-12-31 20:58:41 | ERROR | Database connection failed
2023-12-31 21:00:14 | ERROR | Null pointer dereference
2023-12-31 21:00:48 | WARNING | Disk space running low
2023-12-31 21:01:47 | ERROR | Uncaught exception
2023-12-31 21:04:52 | INFO | New device connected
2023-12-31 21:07:41 | WARNING | Potential security breach
```

قسمت دوم:

- هدوپ نسخه 3.3.4 از این [لینک](#) دریافت و نصب کنید.
- یک دایرکتوری با نام log در hdfs بسازید.
- تمام فایل‌هایی که در مرحله قبل ایجاد شده را در دایرکتوری ایجاد شده در hdfs قرار دهید.
- دستور `hdfs dfs -ls/log` را در سیستم خود اجرا کنید. فایل‌های مورد نظر باید در این دایرکتوری قابل مشاهده باشد. اسکرین‌شات خروجی این دستور را به مستندات مرحله دو اضافه کنید.