**سرفصل‌های پیشنهادی دوره کلان‌داده**

در دوره پیشنهادی کلان‌داده، مطالب زیر به صورت عملی، معرفی و مرور خواهند شد :

* تاریخچه، مفاهیم پایه و آشنایی با ***اکوسیستم کلان‌داده*** . (۱ جلسه)
* راه‌اندازی و کار با یک کلاستر ***هدوپ*** (کار عملی با HDFS/MapReduce و Hive) (۳ جلسه)
* آشنایی مقدماتی با ***کافکا*** به عنوان گذرگاه اصلی تبادل داده در سامانه‌های معاصر اطلاعاتی دنیا. (۲ جلسه)
* کار با ***اسپارک*** به عنوان ابزار توزیع شده پردازش داده‌ها (۳ جلسه)
* معرفی و کار عملی با ***ایرفلو*** (airflow)به عنوان یک ابزار مدیریت کارهای زمان‌مند و انجام ETL های روزانه (۲ جلسه)
* معرفی و آشنایی با ***Trino/PrestoDB*** به عنوان یک ابزار همه منظوره اجرای SQL بر روی انواع منابع داده در شبکه . (۱ جلسه)
* معرفی و کار عملی با ***Clickhouse*** برای ذخیره داده‌های تحلیلی و اجرای کوئری‌های پیچیده بر روی حجم عظیم داده‌. (۲ جلسه)
* ذخیره و نمایش داده‌های متنی به کمک ***الستیک سرچ*** و کیبانا (۱ جلسه )

**پیش‌نیازها**

* **داکر** : برای راه‌اندازی کلاسترهای مختلف مورد نیاز (هدوپ/اسپارک/کافکا/ایرفلو)
* **خط فرمان لینوکس** : با توجه به اینکه ایمیج‌های مورد استفاده، مبتنی بر لینوکس هستند، آشنایی مقدماتی با خط فرمان لینوکس، موجب تسریع فرآیند آموزش خواهد شد.
* **پایتون** : زبان برنامه نویسی مورد استفاده در این دوره پایتون است.

**نحوه ارزیابی**

ارزیابی درس به صورت زیر انجام خواهد شد :

* مینی پروژه هدوپ و ایجاد یک دیتالیک ساده با HDFS
* مینی پروژه کافکا و اسپارک
* مینی پروژه ایرفلو
* پروژه پایانی درس : کلیک‌هوس + الستیک سرچ + کیبانا