



## Brief Description (Elective Course)

### Big Data (IM 6673)

In this course we explore key data analysis and management techniques, which applied to massive datasets are the cornerstone that enables real-time decision making in distributed environments, business intelligence in the Web, and scientific discovery at large scale. In particular, we examine the map-reduce parallel computing paradigm and associated technologies such as distributed file systems, Hadoop Ecosystem, NO-SQL databases, and stream computing engines. Additionally, we review machine learning methods that make possible the efficient analysis of large volumes of data in near real time.

### دام ٦٦٧٣ (البيانات الضخمة)

في هذا المقرر، نستكشف تقنيات تحليل وإدارة البيانات الرئيسية، والتي تُعد تطبيقها على مجموعات البيانات الضخمة حجر الأساس الذي يتيح اتخاذ القرارات في الوقت الفعلي في البيئات الموزعة، وكفاءة الأعمال في الويب، والاكتشاف العلمي على نطاق واسع. وعلى وجه الخصوص، ندرس نموذج الحوسبة المتوازية المختزلة والتقنيات المرتبطة بها مثل أنظمة الملفات الموزعة، ونظام Hadoop، وقواعد البيانات NO-SQL، ومحركات الحوسبة المتدفقة. بالإضافة إلى ذلك، نستعرض أساليب تعلم الآلة التي تجعل التحليل الفعال لكميات كبيرة من البيانات ممكناً في الوقت الفعلي تقريباً.