

Formulario de aprobación de curso de posgrado/educación permanente

Asignatura: Calidad e Integración de Datos

(Si el nombre contiene siglas deberán ser aclaradas)

Modalidad:

(posgrado, educación permanente o ambas)

Posgrado

Educación permanente

Profesor de la asignatura ¹: Dra. Flavia Serra, grado 3, Instituto de Computación

(título, nombre, grado o cargo, instituto o institución)

Profesor Responsable Local ¹:

(título, nombre, grado, instituto)

Otros docentes de la Facultad: Dra. Adriana Marotta, grado 4, Instituto de Computación

(título, nombre, grado, instituto)

Docentes fuera de Facultad:

(título, nombre, cargo, institución, país)

¹ Agregar CV si el curso se dicta por primera vez.

(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

[Si es curso de posgrado]

Programa(s) de posgrado: Diploma de Especialización en Sistemas de Información y tecnologías de gestión de datos, Maestría en Sistemas de Información y tecnologías de gestión de datos, Diploma de Especialización en Ciencia de Datos, Maestría en Ciencia de Datos Aplicada.

Instituto o unidad: Instituto de Computación

Departamento o área:

Horas Presenciales: 45

(se deberán discriminar las horas en el ítem Metodología de enseñanza)

Nº de Créditos: 6

[Exclusivamente para curso de posgrado]

(de acuerdo a la definición de la UdelaR, un crédito equivale a 15 horas de dedicación del estudiante según se detalla en el ítem Metodología de enseñanza)

Público objetivo: Estudiantes de posgrado interesados en el área de Calidad de Datos.

Cupo mínimo: 6 estudiantes.

Objetivos: Presentar a los estudiantes fundamentos sobre calidad de datos e integración de datos. Proveerles conocimientos, herramientas prácticas y tecnologías que les permitan enfrentar el problema de la evaluación de calidad de datos, limpieza y prevención de errores. Presentar los problemas y soluciones posibles relativos a la integración de datos.

Conocimientos previos exigidos: Conocimientos de Bases de Datos y Programación

Conocimientos previos recomendados:

Metodología de enseñanza:

(comprende una descripción de la metodología de enseñanza y de las horas dedicadas por el estudiante a la asignatura, distribuidas en horas presenciales -de clase práctica, teórico, laboratorio, consulta, etc.- y no presenciales de trabajo personal del estudiante)

Descripción de la metodología:

Se presentarán en clase los conceptos teóricos y sus aplicaciones prácticas mediante ejemplos. Además, se estudiarán trabajos recientes de investigación en el área. Promediando el curso se presentará a los estudiantes un trabajo obligatorio, el cual irán resolviendo en clase con apoyo de los docentes, y en parte como trabajo domiciliario.

- Horas clase (teórico-práctico): 30
- Horas clase (práctico):
- Horas clase (laboratorio): 10
- Horas consulta: 5
- Horas evaluación:
 - Subtotal horas presenciales: 45

- Horas estudio:
- Horas resolución ejercicios/prácticos:
- Horas proyecto final/monografía: 45
 - Total de horas de dedicación del estudiante: 90

Forma de evaluación:

- Realización de un proyecto donde se aplican las técnicas vistas en clase.
- Defensa oral del trabajo realizado.

Temario:

- Introducción
 - Conceptos de calidad de datos
 - Problemas y consecuencias de la mala calidad
 - Necesidades en distintas áreas de aplicación
 - Gestión de calidad de datos
- Dimensiones y métricas de calidad
 - Multi-dimensionalidad de la calidad
 - Metamodelo de calidad
 - Estudio de un conjunto de dimensiones y métricas
 - Agregaciones según datos y según aspectos de calidad
- Evaluación de calidad de datos
 - Data Profiling
 - Modelo de calidad
 - Requerimientos de calidad
 - Metadatos de calidad
- Mantenimiento de la calidad de datos
 - Integración de datos
 - Limpieza de datos
 - Prevención
 - Monitoreo y roles

Bibliografía:

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

- Carlo Batini, Monica Scannapieco. Data and Information Quality. Springer, 2016. ISBN: 978-3-319-24104-3.

- Peter Christen. Data Matching - Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Springer, 2012. ISBN: 978-3-642-31163-5.

- X. L. Dong y D. Srivastava. Big Data Integration. Morgan & Claypool Publishers, 2015. ISBN: 978-1-62705-224-5

Datos del curso

Fecha de inicio y finalización: Del 19/05/25 al 27/06/25

Horario y Salón: Lunes, Miércoles y Viernes de 18 a 21 hs.

Arancel: \$30.000

[Si la modalidad no corresponde indique "no corresponde". Si el curso contempla otorgar becas, indíquelo]

Arancel para estudiantes inscriptos en la modalidad posgrado: \$30.000

Arancel para estudiantes inscriptos en la modalidad educación permanente: \$30.000
