



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

Diplomado en Big Data y Ciencia de Datos

Curso: *Ciencia de Datos y sus Aplicaciones*

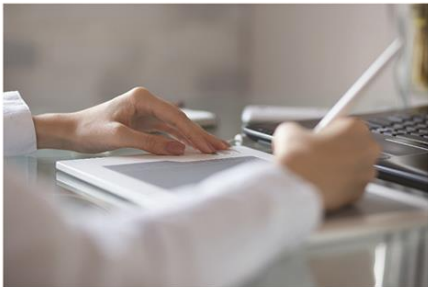
Educación Profesional
Escuela de Ingeniería UC

✉ regonzar@uc.cl

✉ rmunoz@uc.cl

✉ jcaiceo@uc.cl

Roberto González, Roberto Muñoz, Jaime Caiceo



Clase 08: Cierre

PROYECTO

Proyecto

- Hacer análisis de segmentación de clientes bancarios usando lenguaje R y métodos de clustering.
- Elegir un objetivo de negocio, segmentar a los clientes y proponer estrategias de marketing para uno de los segmentos.
- Recorrer el espacio de parámetros del algoritmo de clustering y anotar resultados.
- La base de datos contiene 45.200 registros y campos tales como fecha de nacimiento, actividad y saldo en cuenta.

Evaluación

- Criterios de evaluación

Aspectos a ser Evaluados	Ponderación en la Nota Final
Informe: Estructura, orden y claridad	30%
Informe: Descripción del problema, metodología aplicada, análisis de resultados y estrategia comercial	70%

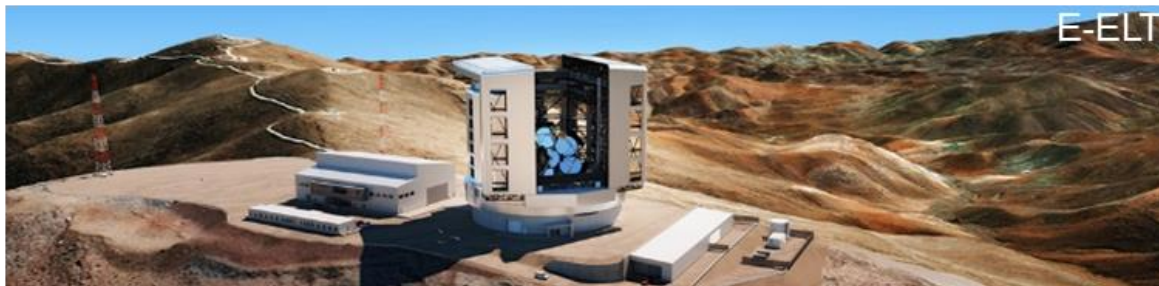
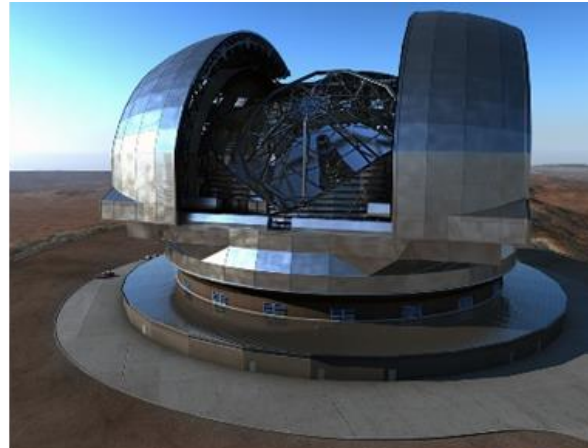
- Entrega de informe

Los informes deben enviarse hasta las 23:59 horas del martes 4 de Mayo a los profesores Jaime Caiceo, Roberto Muñoz y Roberto Gonzalez

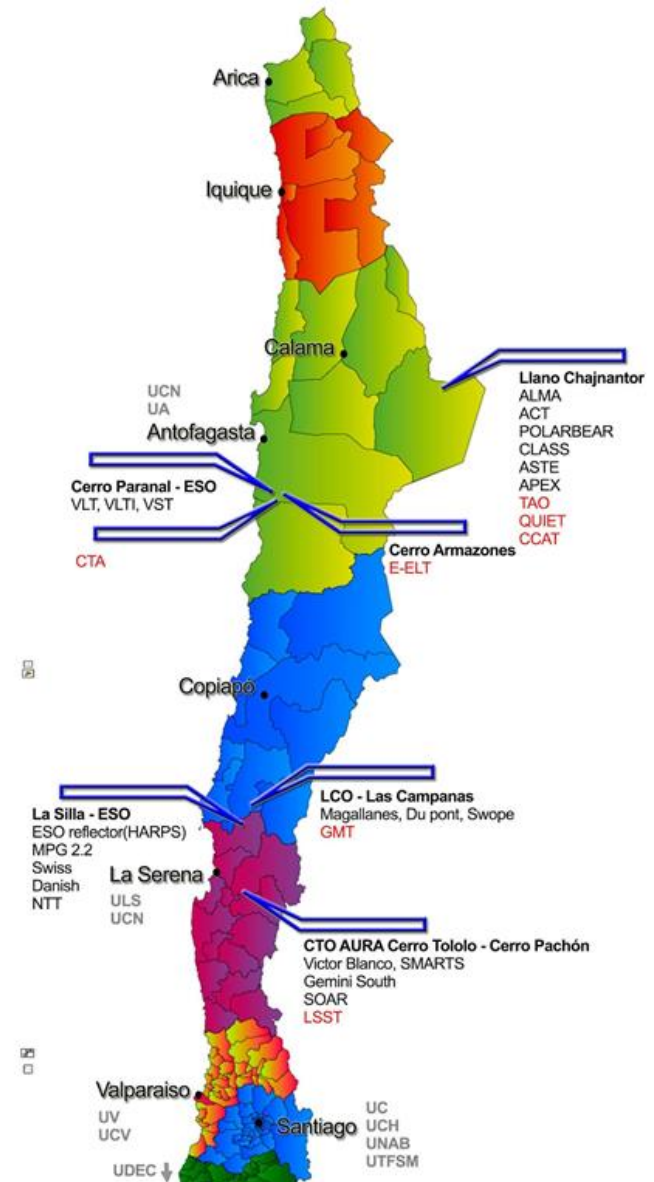
Clase 08: Cierre

ANALYTICS Y DATA SCIENCE EN CHILE

Observatorios en Chile



Chile albergará sobre el 55% del área colectora (reflectores) mundial, con la puesta en marcha de LSST, GMT y E-ELT.



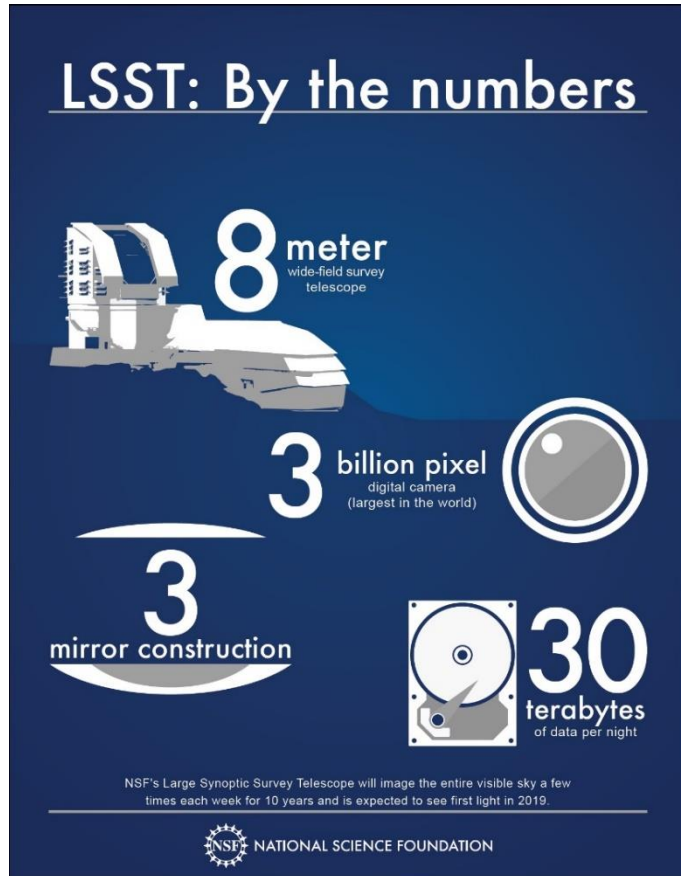
Astroinformática

<https://www.youtube.com/watch?v=XQclcxX6s9I>

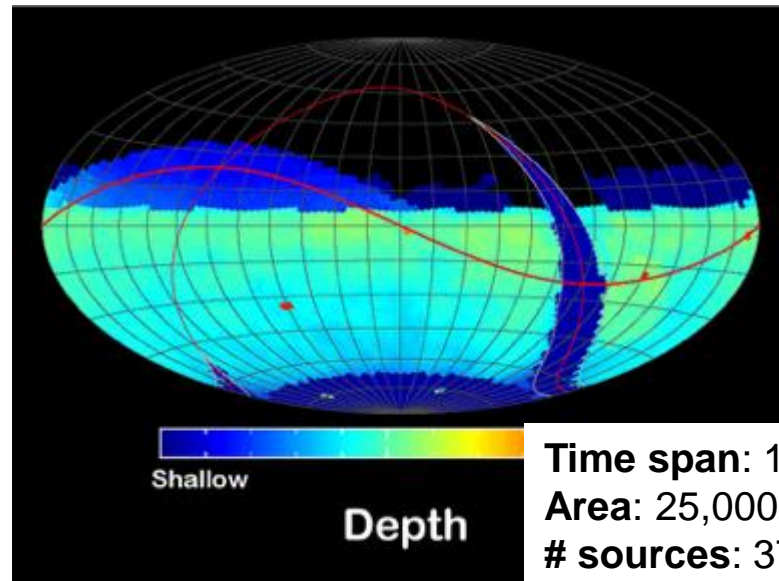
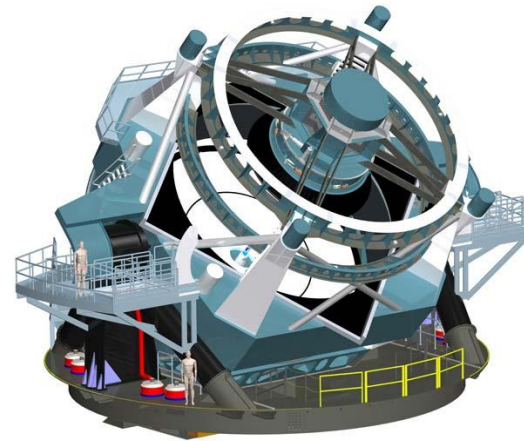


Large Synoptic Survey Telescope (Vera Rubin)

<https://www.lsst.org>



Observaciones comienzan
en octubre de 2022



Time span: 10 years
Area: 25,000 deg²
sources: 37,000,000,000

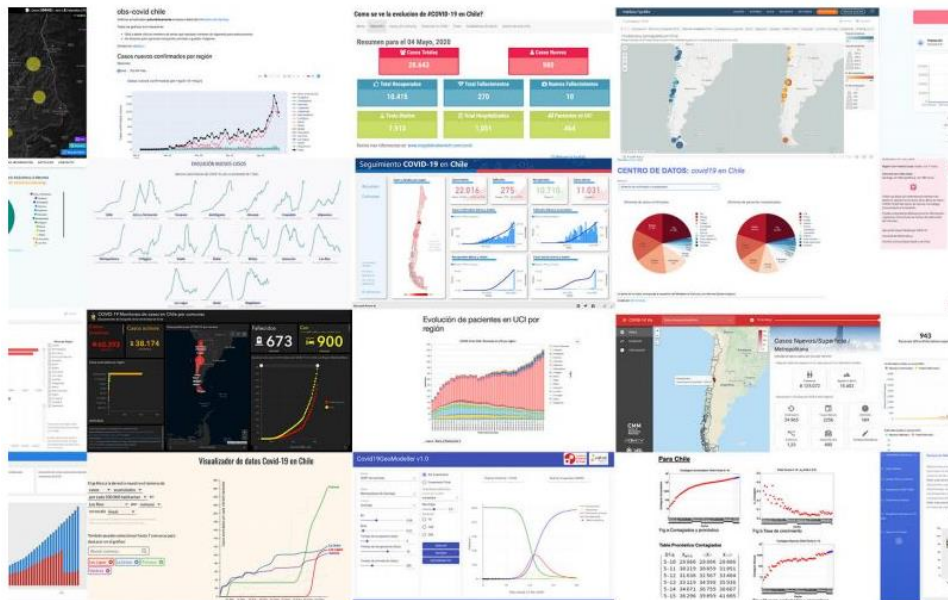
Data Observatory



<https://www.dataobservatory.net>

Organización sin fines de lucro con la misión de adquirir, procesar y almacenar datos generados por instituciones públicas y privadas

Data Observatory desarrolla e implementa componentes clave de la plataforma de datos COVID-19 de MinCiencia



Para 2020, la entidad tienen ambiciosos planes. Por un lado, busca sumar nuevos campos que se beneficien de su información, como también poder dar libre acceso a datos a la población general e incorporar más socios estratégicos, ya sea como inversionistas o bases de datos. MAGDALENA OVALLE V.



Volumen Datos: Proyecto SKA

- Square Kilometre Array (SKA). Australia y Sudáfrica
Observaciones comienzan el 2027



Créditos: SKA

Datacenter Microsoft

Nueva Región de Azure, 4 grandes ventajas al mercado chileno:

• **Residencia de Datos:** el hecho de que los edificios y sus servidores se encuentren dentro de las fronteras del país permite asegurar que los datos puedan permanecer en Chile, cumpliendo así las futuras normas que regulan el uso de datos personales y ciberseguridad en la región.

• **Baja Latencia:** La cercanía física con el datacenter permitirá reducir considerablemente los tiempos de conectividad y envío de información pudiendo desarrollar proyectos que incluso hoy son inviables, como serían análisis de datos en tiempo real (near real time), uso de IoT para gestión en línea, aplicaciones y/o servicios nativos en la nube para usuarios finales.

• **Seguridad:** La seguridad es fundamental para Azure. La infraestructura de Azure posee los más altos estándares de seguridad multicapa desarrollados por expertos en ciberseguridad que los supervisan activamente para proteger los recursos y los datos de personas y empresas que utilizan el servicio.

• **Continuidad Operacional:** Los datacenters de Azure están diseñados para brindar alta disponibilidad y tolerancia a fallos. La nueva región en Chile se diseñó para cumplir estos estándares, garantizando a los usuarios una continuidad en las operaciones que realicen en la nube.

EL MERCURIO

Vida • Ciencia • Tecnología

SANTIAGO DE CHILE, MIÉRCOLES 21 DE ABRIL DE 2021

Los detalles de una inversión anunciada en diciembre de 2020:

Así serán los tres *data centers* que Microsoft construirá al norte de Santiago

Las instalaciones buscan ofrecer mayor eficiencia a los usuarios; se alimentarán de energía renovable, tendrán una subestación eléctrica y serán los primeros en el mundo en tener una estación satelital terrestre.

Cómo funciona un centro de datos como el que se instalará en Chile

por su trayectoria, el compromiso con la transformación y por el acceso a energías limpias, entre otras variables", agrega.

"Chile tiene varias condiciones que lo hacen un lugar atractivo para las inversiones: cuenta con certeza jurídica, económica y una exigente agenda de energías renovables. Además, tiene una amplia red de conectividad, respaldada en el despliegue de carreteras de alta velocidad a través de los proyectos Fibra Óptica Nacional, Fibra Óptica Austral, más el proyecto del cable Transpacífico Humboldt", dice la subsecretaria de telecomunicaciones, Pamela Gidi, quien añade que Chile se ha consolidado como el hub digital de Latinoamérica.

Datos desde el cielo

Rademacher adelanta que Chile será uno de los primeros países en el mundo en contar con la tecnología Azure Orbital, una estación satelital terrestre que permitirá recibir, procesar y distribuir datos provenientes de satélites. "Es democratizar el acceso a los datos satelitales, y que puedas contratar los servicios y acceder a información a la cual es muy difícil hacerlo de otra forma", aclara. Esto puede ser usado para hacer estudios de cantidad de aguas con imágenes satelitales u obtener datos sobre condiciones ambientales, por poner solo un par de ejemplos.

Los tres *data centers* se alimentarán con energía proveniente de fuentes limpias, la que se comprará a proveedores. Rademacher cuenta que para mayor estabilidad energética, y no afectar a las empresas colindantes, la compañía piensa instalar una subestación eléctrica para atender sus propias necesidades.

Microsoft tiene estándares en la construcción de *data centers* que va replicando en los más de doscientos centros que tiene en 34 países. Ayer la empresa dio a conocer una visita virtual a uno de estos centros, cuyo entrada es estrictamente prohibida al público y que permitirá hacerse una idea de cómo será el de Chile (en <https://news.microsoft.com/stories/microsoft-datacenter-tour/html-version/es/>).

Energía y enfriamiento

Los *data centers* usan distintas energías para disipar el calor de los servidores. La más usada en los de Microsoft son las unidades de enfriamiento evaporativas de alto rendimiento.

Los *data centers* en Chile usarán energía renovable que comparan a distintos proveedores.

Tendrán su propia subestación de energía.

Respaldo de energía

Se usan generadores de energía propios. Cada generador puede alimentar hasta 3 mil hogares.

Baterías UPS para almacenar energía renovable en caso de microrroturas de energía.

Ya se comienzan a usar células de combustible de hidrógeno.

Azure Orbital

Recibe los datos generados por satélites para su procesamiento en Tierra.

Aí, instituciones o gobiernos pueden acceder con facilidad a imágenes o datos de diversa índole.

Chile será pionero en usar esta tecnología.

Se ubicará en la zona Norte de Santiago. Serán tres edificios independientes.

Sala de servidores

Un *data center* tiene varias salas cada una con cientos de servidores.

4 millones de servidores y 200 *data centers* en el mundo. Así las instalaciones de Chile forman parte de un entramado en que los datos van y vienen para atender con mayor eficiencia a los usuarios.

Lugar atractivo

En Latinoamérica habrá tres países con *data centers* de esta compañía: Brasil, México y Chile.

"Chile no compite con sus países vecinos por tener una infraestructura como esta, sino que fue elegido entre países de todo el mundo, lo que es más meritorio", aclara Rademacher. "Se le eligió por su trayectoria, el compromiso con la transformación y por el acceso a energías limpias, entre otras variables", agrega.

Los servidores del tipo blade (ultradensidad) se encuentran en racks diseñados bajo el nuevo concepto Proyecto Olympus.

Computadores de red (MDS) de alta velocidad permiten el intercambio de datos entre servidores.

Computadores de alta capacidad (DDF) comunican las distintas salas de servidores con la red principal de datos para salir al mundo por miles de redes de fibra óptica en las ciudades y bajo los océanos.

Si se pusieran los racks de un *data center* uno sobre otro se llegaría a 5,3 km de altura.

Fuente: Microsoft

EL MERCURIO

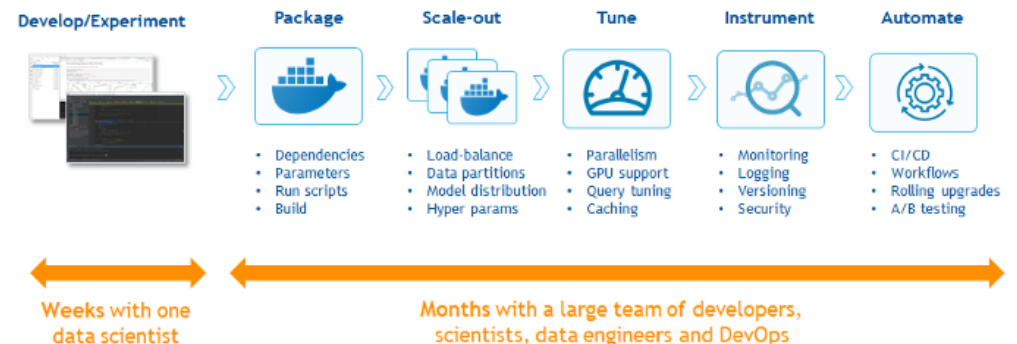
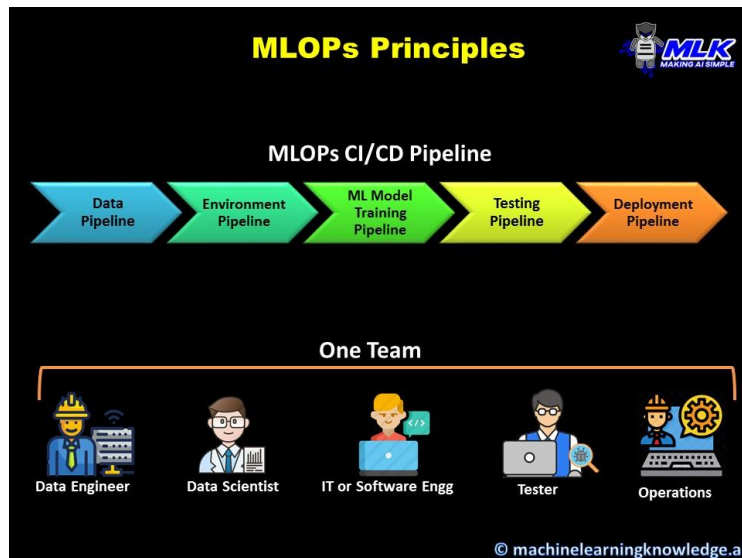
Clase 08: Cierre

TENDENCIAS DATA SCIENCE

Tendencias 2021

MLOPS – Machine Learning Ops.

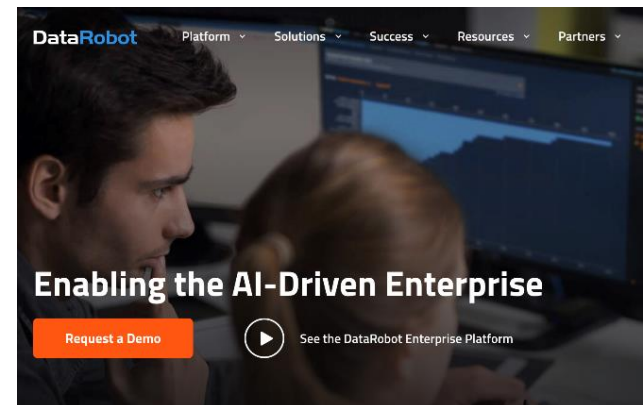
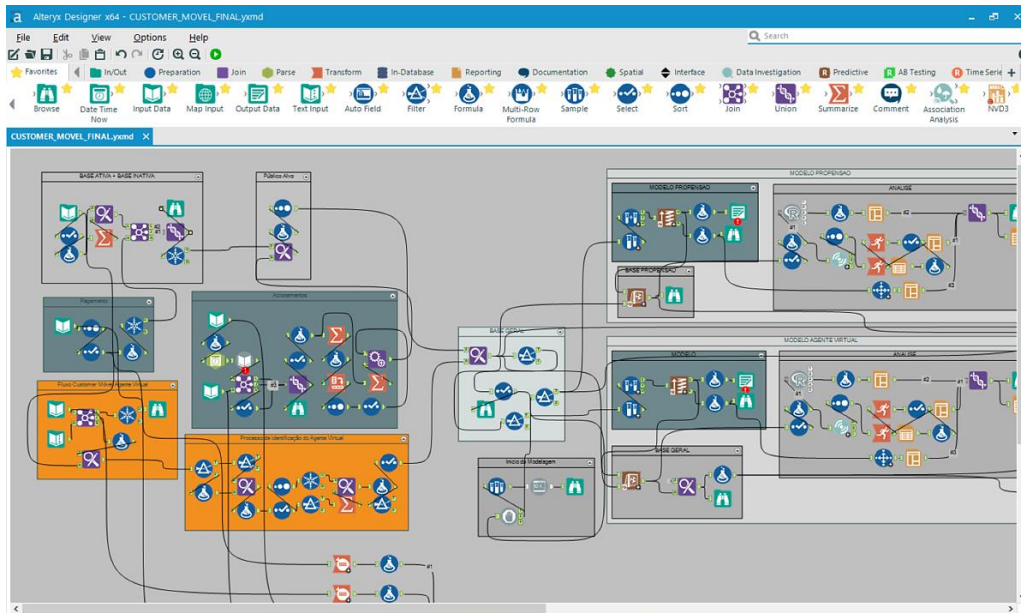
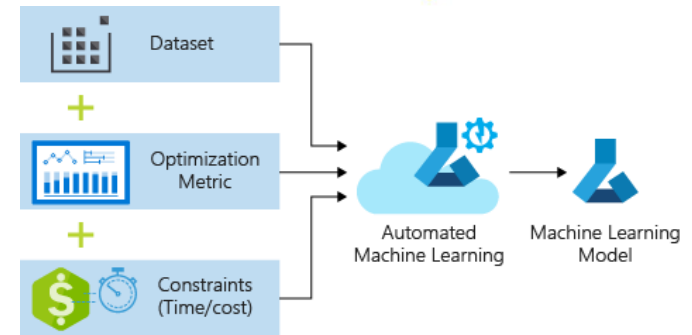
- Solo **22%** de los proyectos de ML son implementados en producción exitosamente.
- **39%** de los data scientist enfrenta grandes desafíos para poder administrar dependencias y ambientes en la etapa de implementación productiva.
- **38%** de data scientist admiten falta de conocimientos para implementar una solución productiva.
- **43%** de encuestados dicen encontrar dificultad para escalar su modelo y satisfacer las necesidades productivas de su organización. (Encuesta Anaconda 2020)
- DevOps pero para ML
- Continuous Integration/Delivery (CI/CD)



Tendencias 2021

Automatización de tareas en Data Science

- 1) Limpieza de datos
- 2) Ingeniería de features
- 3) Consolidación de plataformas como H2O



Tendencias 2021



INFORMACIÓN DE PRENSA

Durante este fin de semana, BancoEstado detectó en sus sistemas operativos un software malicioso. Apenas fue descubierto este problema, nuestros equipos de operaciones y de ciberseguridad se desplegaron para localizar, contener y solucionar esta situación.

Si bien algunas de nuestras plataformas podrían presentar algún tipo de interferencia, hasta el momento nuestros sistemas de cara a clientes (cajeros automáticos, CajaVecina, el sitio web personas y la App) no han sido afectados y se encuentran funcionando.

Solicitamos a nuestros clientes utilizar los canales digitales para las operaciones habituales. En el caso que requiera ir a una sucursal, informaremos por distintas vías la disponibilidad para mañana.

Pedimos las disculpas correspondientes a todos nuestros clientes por las molestias que esta situación, provocada por terceros, pueda acarrear.

Privacidad y seguridad de los datos

- 1) ¿A quién le pertenecen los datos?
¿Son suficientemente seguras las nubes?
- 2) Definir políticas de seguridad de datos en los productos y servicios que ofrecemos
- 3) Mejorar administración datos de clientes (SOC 2)
- 4) Trabajo Remoto ha invalidado seguridad de redes corporativas

Facebook no notificará a los 533 millones de usuarios expuestos en la base de datos online

Por Clare Duffy
12:48 ET(16:48 GMT) 10 Abril, 2021



UPDATED 13:14 EST / NOVEMBER 22 2019

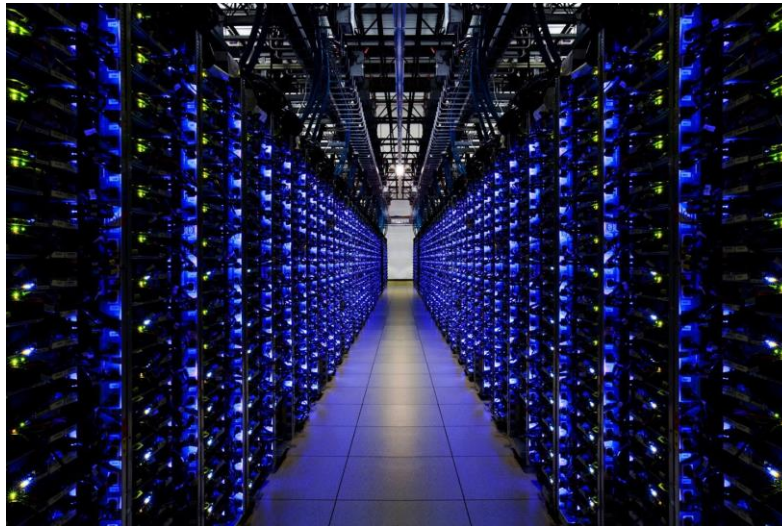


Account records for up to 1.2B people stolen in massive alleged data exposure

Tendencias 2021

Data science aumentada en la nube

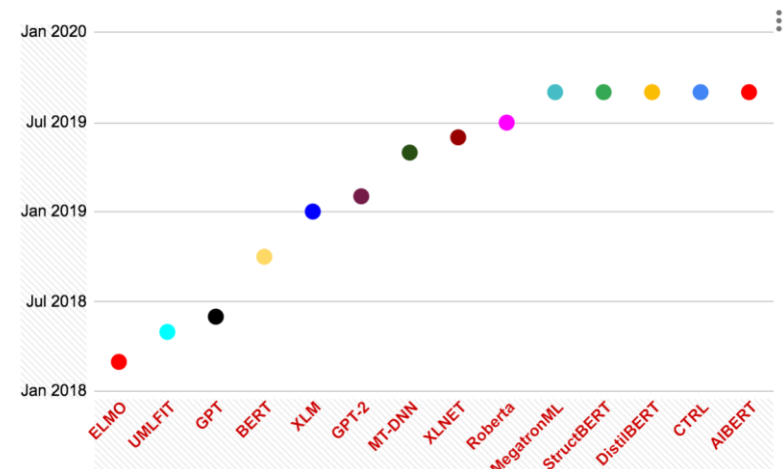
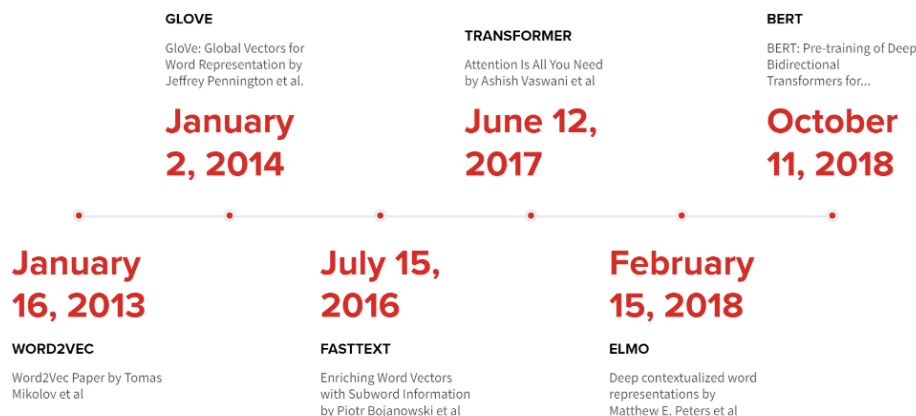
- 1) Volúmenes de datos en las organizaciones han crecido exponencialmente
- 2) Necesidad de almacenar y procesar big data
- 3) Data Science migra a la nube
- 4) Gran requerimiento de ingenieros y arquitectos de datos



Tendencias 2021

Procesamiento de lenguaje natural

- 1) Gran cantidad de información almacenada en archivos de texto, chats y logs
- 2) Nuevos modelos de Deep Learning han permitido avanzar en análisis semántico de texto

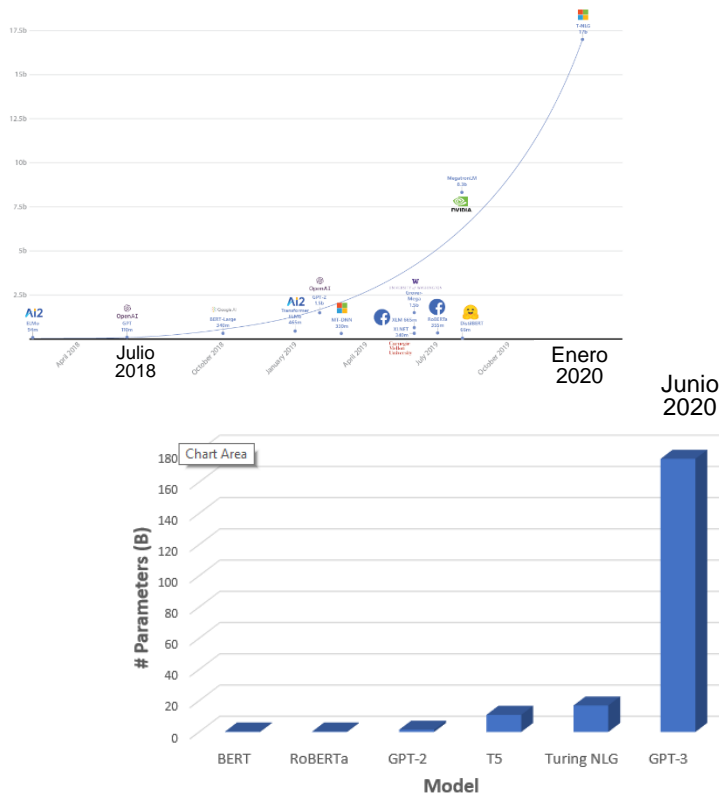


GPT-3

<https://openai.com/>



Modelo de NLP desarrollado por OpenAI(US\$4.6M training)



Describe a layout.

Just describe any layout you want, and it'll try to render below!

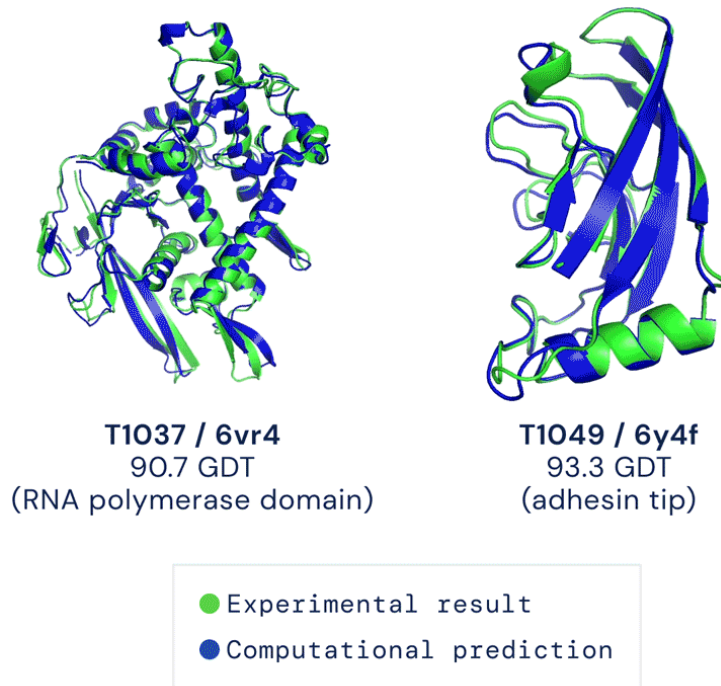
A div that contains 3 buttons each with a random color.

Generate

JSX Layout maker

Deepmind AlphaFold

- Modelo DL pliegue de proteínas

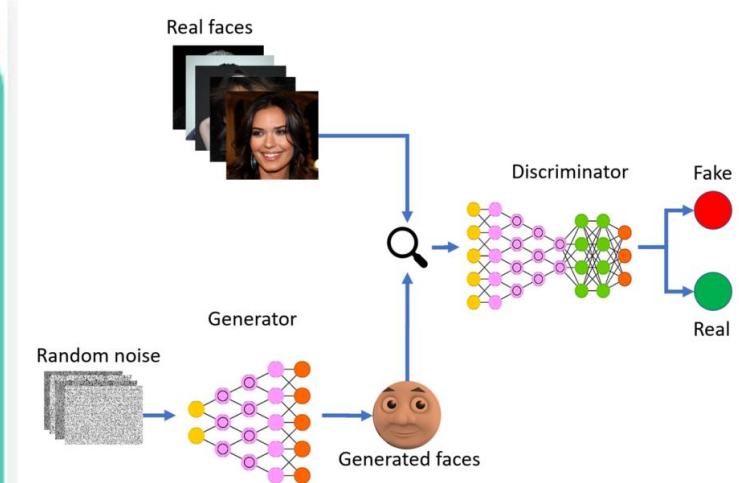
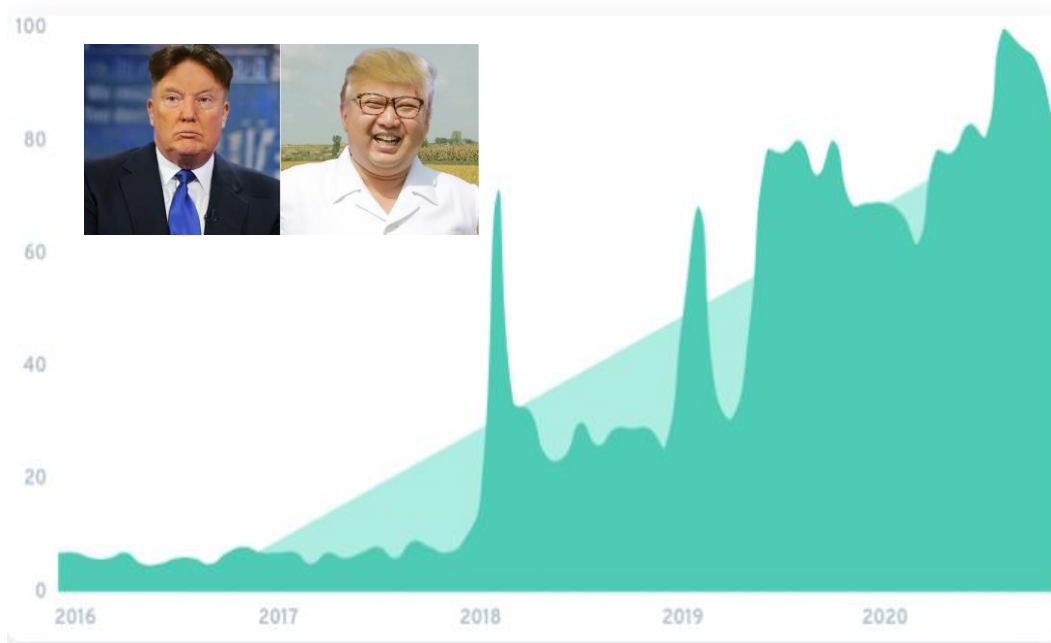


We have been stuck on this one problem – how do proteins fold up – for nearly 50 years. To see DeepMind produce a solution for this, having worked personally on this problem for so long and after so many stops and starts, wondering if we'd ever get there, is a very special moment.

PROFESSOR JOHN MOULT
CO-FOUNDER AND CHAIR OF CASP, UNIVERSITY OF MARYLAND

Contenido Deep Fake

- Explosión en generación de video y audio con GANS.
- Impacto en adquisición de datos.



COMPARTIR EXPERIENCIAS





ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

Diplomado en Big Data y Ciencia de Datos

Curso: *Ciencia de Datos y sus Aplicaciones*

Educación Profesional
Escuela de Ingeniería UC

✉ regonzar@uc.cl

✉ rmunoz@uc.cl

✉ jcaiceo@uc.cl

Roberto González, Roberto Muñoz, Jaime Caiceo

