



Universidad Ort Uruguay
Presentación de defensa de Obligatorio

**Métricas para la
detección y predicción
de fallas en discos
duros basado en Big
Data**

Diciembre 23, 2020

Resumen



Datos datos datos

Se calcula que en 2020 el negocio de los datos alcanzó los 411 Billones de dólares.



Futuro

Hay estimaciones que para el 2025 el negocio del storage alcance los 1300 Billones de dólares



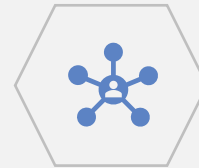
Confiabilidad

Hoy las empresas que trabajan dando servicios de datos necesitan unidades confiables



Breathing Data

Hoy los datos son esenciales para la vida moderna



Todo genera datos

Actualmente todo elemento de la vida genera datos que pueden o deben ser almacenados



Eficiencia en el data center

Hoy la diferenciación esta en cuan eficiente se puede ser un centro de datos

Objetivos

- Descargar datos generados en un Datacenter
- Analizar y pre procesar los datos
- Ingerir los datos con los que se trabajará
- Hacer análisis de estos datos
- Hacer predicciones a futuro



Problema

- El problema y los desafíos estaban en poder manejar más de 53 GB de datos crudos con millones de registros de información de salud de discos duros.
- Con ellos poder generar información útil y a tiempo así poder utilizar las herramientas de Big Data vistas en el curso

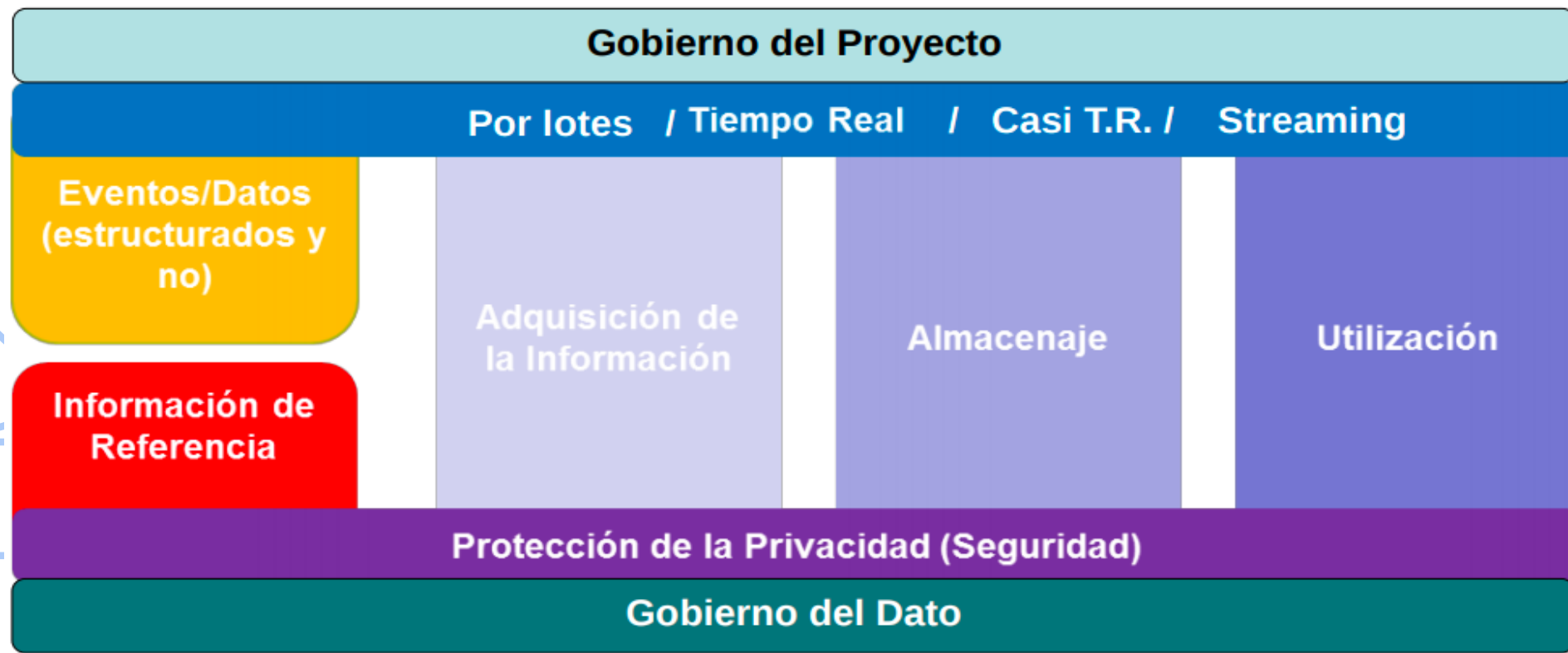


Justificación y Alcance

- El fin de este proyecto es generar conocimiento nuevo, que sea útil, resolviendo un problema de Big Data.
- Responder a un conjunto de preguntas planteadas sobre los datos de discos duros, realizar ingeniería de atributos, indexarlos.
Hacerlos disponibles para consultas en tablas y generar reportes que brinden insights útiles para mejorar eficiencia de un centro de datos.



Arquitectura



Herramientas

Source



Adquisición de la
Información



Almacenamiento y
procesamiento



Utilización



Análisis



Datos

+



Análisis

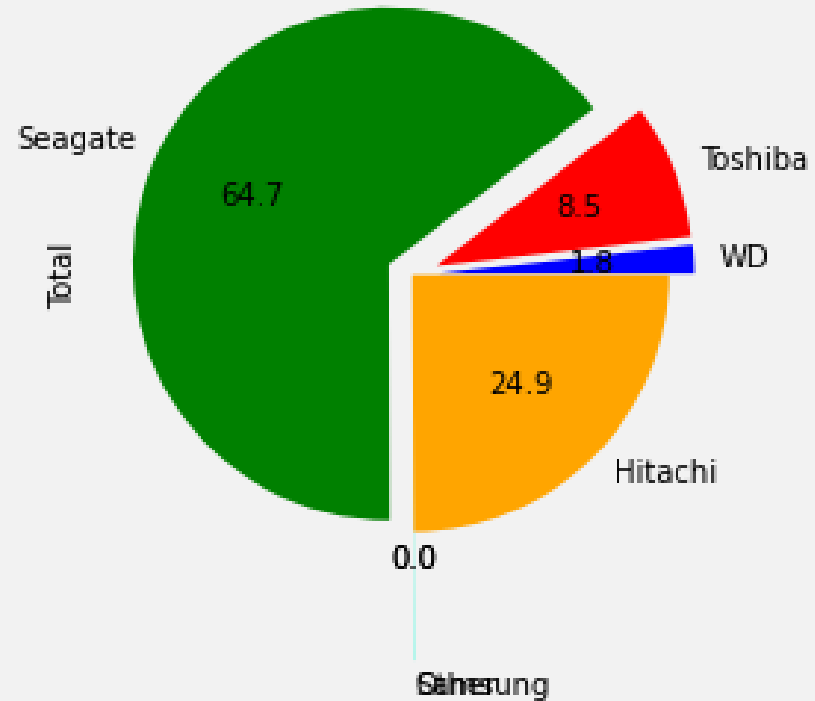
≡



Actionable Insights

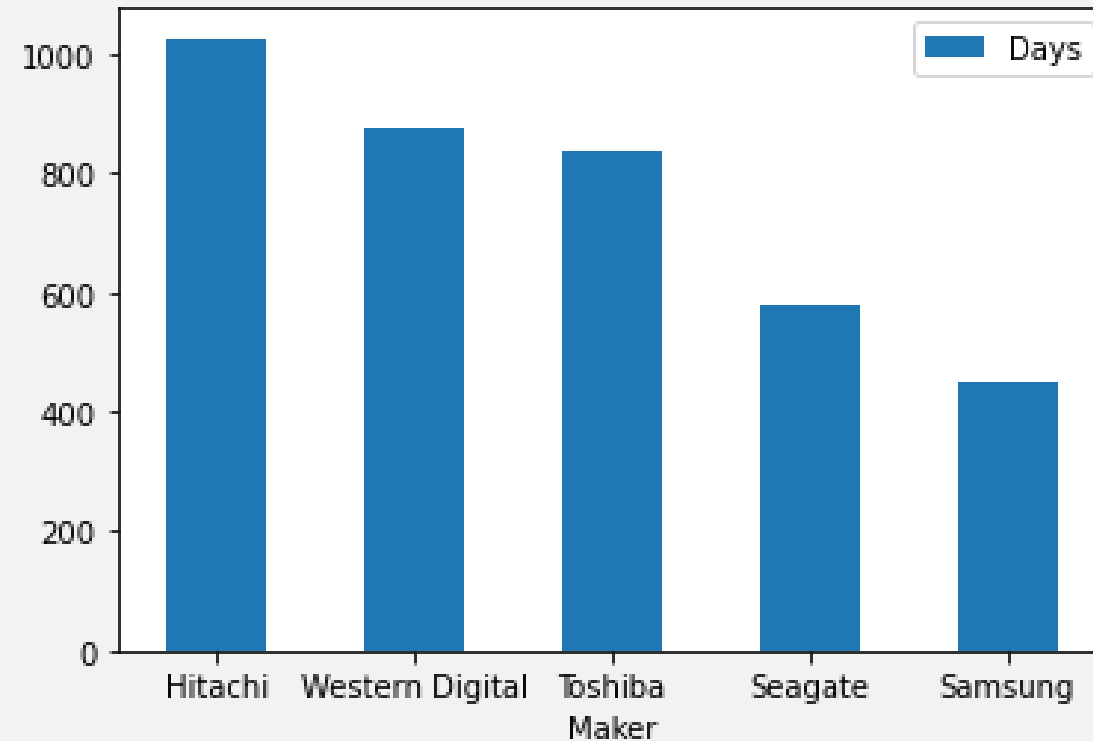
Total de discos que existieron en el datacenter agrupados por fabricante.

Maker	Total
Western Digital	4025
Toshiba	18844
Seagate	143144
Samsung	18
Other	100
Hitachi	55094



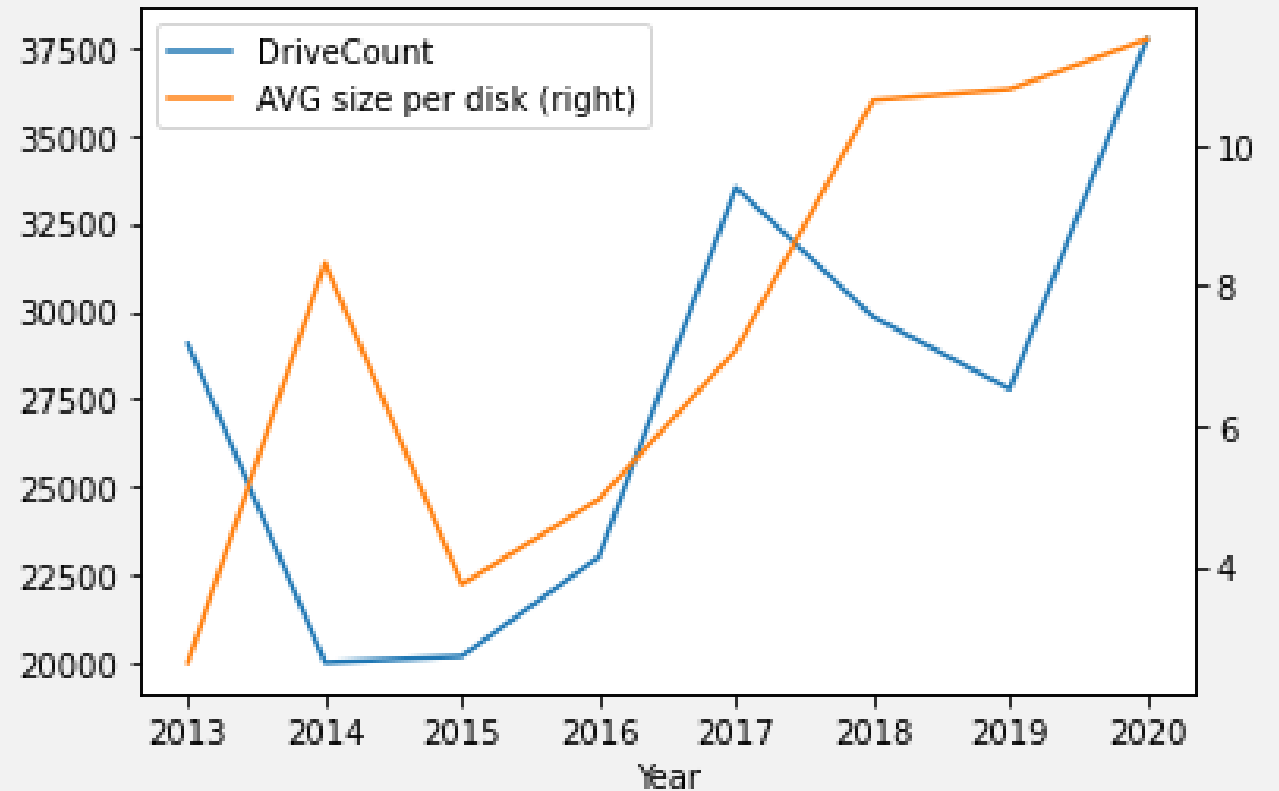
Cantidad promedio de días que un dura un disco duro por fabricante

Maker	Days
Hitachi	1027
Western Digital	876
Toshiba	835
Seagate	581
Samsung	449



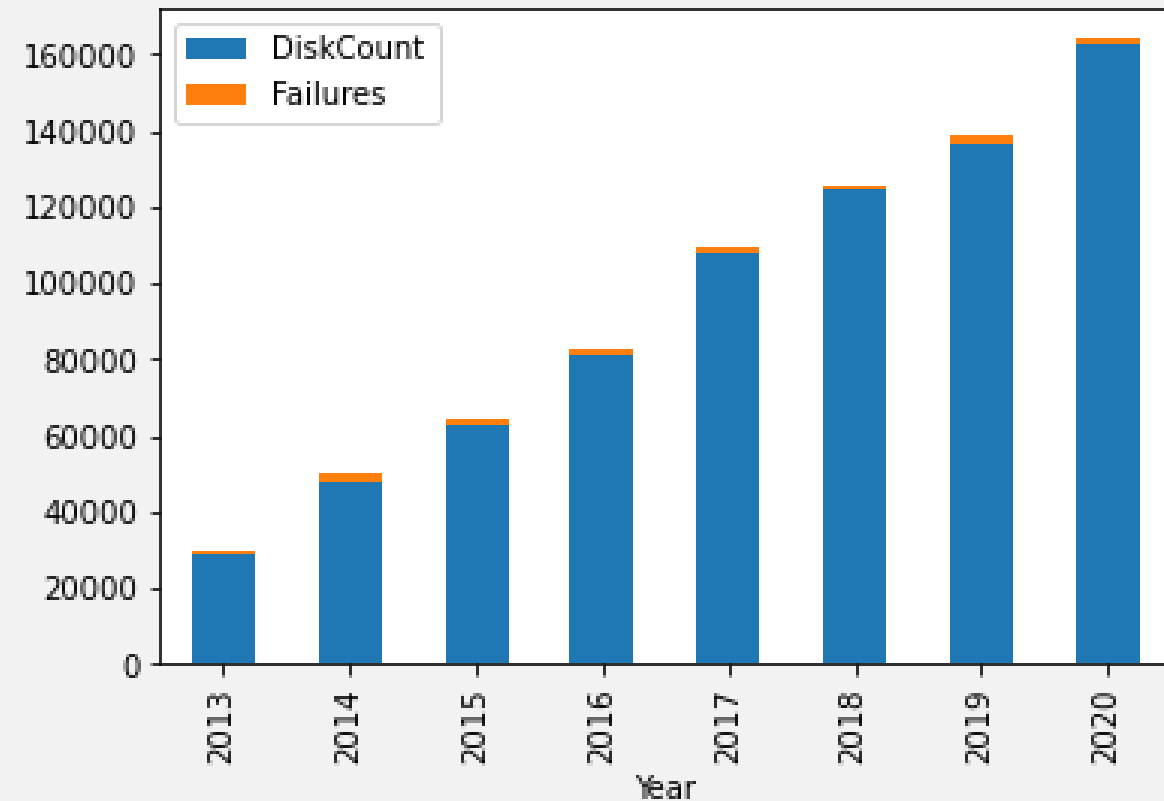
Cantidad de discos nuevos que ingresaron por año y tamaño aproximado de disco agregado

Year	DriveCount	SizeInTB	AVG size per disk
2013	29072	76776.20	2.64
2014	20035	166648.06	8.32
2015	20211	75692.28	3.75
2016	23030	114172.93	4.96
2017	33516	237143.00	7.08
2018	29849	317317.81	10.63
2019	27790	299931.30	10.79
2020	37723	433749.23	11.50



Cantidad de discos que fallaron vs discos en el data center discriminado por año

Year	DiskCount	Failures	Ratio
2013	29072	740	2.55
2014	47793	2206	4.62
2015	62898	1429	2.27
2016	81173	1431	1.76
2017	108282	1556	1.44
2018	124326	1381	1.11
2019	136568	2263	1.66
2020	162997	1093	0.67



Conclusiones

1



Los discos son cada vez mas confiables

Los discos no quedaron fuera de la evolucion tecnologica. Son cada vez mas confiables y mas densos

2



La demanda de datos es cada vez mayor

Se observa que el crecimiento en PBs es muy importante con el transcurso del tiempo

3



Machine Learning

Si bien machine learning no fue tema de este curso, con lo poco que se pudo ver en esta entrega práctica se ve el poder que tiene en la ayuda de la toma de decisiones

4



2.0 a 2.8 años

Se pudo determinar que la vida util de un disco duro es de 2 a 2.8 años, antes de que fallen. El fabricante más confiable es Hitachi pero el modelo mas confiable es WDC5000BPKT