

GESTIÓN DE PROYECTOS DE CIENCIA DE DATOS						
Detección de anomalías de tráfico en servidores web						
E23S-18014: MITSIU ALEJANDRO CARREÑO SARABIA						
Criterio	Nivel 1 Deficiente	Nivel 2 Aceptable	Nivel 3 Bueno	Nivel 4 Excelente	Evaluación	Comentarios
Relevancia	Los datos no son pertinentes para los objetivos del proyecto	Parte de los datos es relevante, pero no toda la información necesaria está presente	La mayoría de los datos es relevante y útil para los objetivos del proyecto	Todos los datos son completamente pertinentes y adecuados para los objetivos del proyecto	4 - Excelente	Los datos ofrecen datos únicos que expresan el contexto a partir del cual se generó la conexión y cómo la manejó el servidor
Exactitud	Los datos contienen errores significativos	Algunos datos son precisos, pero hay errores ocasionales	Los datos son mayoritariamente precisos, con pocos errores	Los datos son completamente precisos y reflejan la realidad sin errores.	4- Excelente	Los datos representan fielmente los datos que recibe el servidor, y se entienden las limitaciones al existir mecanismos y tecnologías de ofuscación
Compleitud	Los datos están incompletos, faltan muchas piezas clave	Algunos datos faltan, pero no impiden significativamente el análisis	La mayoría de los datos está completa, con pocas piezas faltantes	Todos los datos están completos y abarcan todas las dimensiones necesarias	4- Excelente	Los datos al ser generados de manera automática como bitácora de un servidor web, ofrecen toda la información relevante.
Consistencia	Los datos presentan numerosas inconsistencias y contradicciones	Hay algunas inconsistencias, pero la mayoría de los datos son coherentes	Los datos son mayoritariamente consistentes, con mínimas discrepancias	Los datos son completamente coherentes y uniformes en todos los aspectos	4- Excelente	Los datos tienen una granularidad alta, lo que significa que un solo renglón representa una conexión completa, además de seguir el modelo RESTful
Actualidad	Los datos están desactualizados y no son pertinentes al periodo de estudio	Los datos son algo desactualizados, pero aún son parcialmente útiles	La mayoría de los datos está actualizada y es pertinente para el estudio	Todos los datos están completamente actualizados y son pertinentes para el periodo de estudio	N/A	
"Estructura	La estructura que presentan los datos ha cambiado	La estructura de los datos ha cambiado parcialmente	La mayoría de la estructura de los datos se preserva y es pertinente para el estudio	La estructura de todos los datos es completamente relevante y sigue en uso	4- Excelente	La estructura de los datos a pesar de ser de hace un año sigue exactamente igual por lo que tiene valor para el estudio
Accesibilidad	Los datos no están disponibles o son muy difíciles de acceder	Los datos están disponibles, pero el acceso es complicado y lento	Los datos son accesibles con algunos obstáculos menores	Los datos son fácilmente accesibles y disponibles cuando se necesitan	3- Bueno	Los datos son difíciles de acceder por diseño, al formar parte de los datos sensibles de la empresa, además tienen una duración limitada de 15 días (configurable) para controlar la demanda de almacenamiento

Comprensibilidad	La documentación de los datos es confusa o inexistente	La documentación está presente pero es incompleta o difícil de entender	La mayoría de la documentación es clara y comprensible	Toda la documentación es clara, completa y fácil de entender	2- Aceptable	La documentación indica las posibles configuraciones a aplicar a los datos (duración, campos, formato, ubicación en disco) pero no ofrece una explicación directa del dato por lo que se necesita experiencia en el campo y sus tecnicismos
Fiabilidad	Las fuentes de datos no son confiables	Algunas fuentes de datos son confiables, pero otras no lo son	La mayoría de las fuentes de datos son confiables	Todas las fuentes de datos son altamente confiables y verificables	4- Excelente	Los datos provienen de una única fuente, que es el servidor web mismo, y de una única aplicación, el proxy que se encarga de manejar el tráfico, por lo que son altamente confiables
Seguridad	Los datos no están protegidos adecuadamente y pueden ser vulnerables	La protección de datos es consistente, con algunas áreas vulnerables	Los datos están mayoritariamente protegidos, con pocas vulnerabilidades	Los datos están completamente protegidos y cumplen con todas las normativas de seguridad y privacidad	4- Excelente	Los datos están únicamente alojados en el servidor en cuestión, protegidos con diversas medidas de seguridad entre las que destacan, conexión encriptada y acceso únicamente mediante contraseña
Rango temporal	Los datos son antiguos, de un periodo temporal corto y el comportamiento que describen pudo haber cambiado	Los datos son moderadamente recientes, de un periodo temporal corto y parte del comportamiento que describen aún se mantiene	Los datos son recientes o comprenden un rango temporal hasta el pasado reciente y generalmente describen comportamientos recientes	Los datos son completamente recientes y comprenden un rango de tiempo extenso que permite analizar tendencias nuevas y pasadas	2- Aceptable	Los datos comprenden un periodo de 15 días del año pasado, por lo que permiten realizar un análisis de comportamiento generalizado, pero es posible que las tendencias hayan cambiado moderadamente