

Observabilidad de datos-para la fiabilidad de los datos empresariales

Presenter's name

Presenter contact information

¿Porqué?

...Complejidad y Estructura

A full-page background image of an underwater scene. On the left, a large, dense school of silver fish, possibly jackfish, swims towards the right. On the right, a scuba diver in a blue wetsuit and mask swims away from the viewer, leaving a trail of bubbles. The bottom of the frame shows a rocky seabed covered in green algae or coral. The water is a deep, clear blue.

Observabilidad de datos

La base para
información
fiable.

Sin embargo, muchas organizaciones consideran que **sus datos son poco confiables**

82%

de las compañías dicen que las preocupaciones sobre la calidad de los datos representan una barrera para sus proyectos de integración de datos.

80%

de los ejecutivos de empresas encuestados no confían en sus datos.

73%

de los ejecutivos de empresas encuestados no están satisfechos con la calidad de sus datos.

Y esta información **poco confiable** es costosa desde una perspectiva de tiempo y financiera.

6%

Impacta en los ingresos anuales debido a datos inexactos de los clientes.

40%

Proporción promedio de tiempo que los equipos de datos dedican a la limpieza, integración y preparación de datos.

\$15M

es el costo promedio de datos incorrectos.

Por lo tanto, muchas organizaciones no pueden aprovechar sus datos para la innovación de procesos.

61%

de las organizaciones son incapaces de aprovechar los datos para crear una ventaja competitiva y sostenible.

63%

de las organizaciones son incapaces de usar sus propios datos para crear nuevos modelos de negocios.

63%

de las organizaciones no son capaces de crear un nuevo servicio o producto.

La Importancia de Data Observability



Kyle Schmidt

Engineer @Meta | Working
on the Metaverse |
Systemantics | Systems
Thinking

[View full profile](#)



Kyle Schmidt • 2nd

Engineer @Meta | Working on the Metaverse | Systeman...

4d •

Gestión de incidentes de datos

Los sistemas de datos pueden fallar por muchas razones y no existe un enfoque único para entender cómo o por qué. El siguiente flujo de clasificación es muy común:

- 1) hacer ping frenéticamente a la ingeniera de datos senior que ha estado en el equipo durante más de 4 años y pedirle ayuda urgente
- 2) si no está disponible, pasar horas depurando la canalización comprobando al azar miles de mesas.

A medida que los sistemas de datos se distribuyen cada vez más y las empresas incorporan cada vez más datos, la posibilidad de error no hace más que aumentar. Si bien no existe mucha literatura sobre cómo los equipos de datos pueden manejar la gestión de incidentes, ya existen excelentes recursos y mejores prácticas que podemos aprovechar del desarrollo de software.

La **Importancia** de la Observabilidad de Datos.



“Mientras tanto, la dirección de **Unity espera una pérdida de ingresos de 110 millones de dólares en 2022** por este único error”.

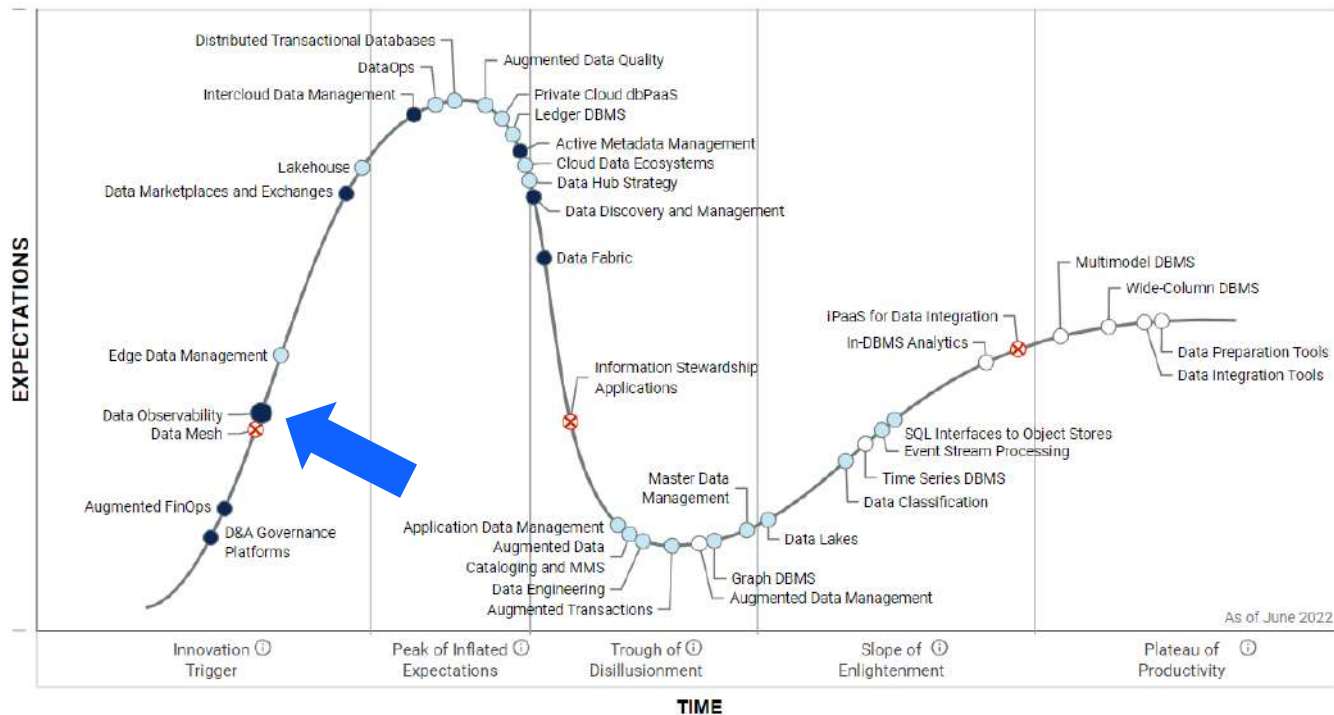
– John Quast, The Motley Fool, <https://www.nasdaq.com/articles/can-unity-fast-track-its-recovery-with-this-1-surprising-move>

”Hemos perdido el valor de entrenamiento de datos, en cierta parte el responsable es el consumo de información de **poca calidad** ... Esperamos que la recuperación pase por pasos secuenciales, reconstrucción de datos, capacitación y mejora del modelo, y luego recuperación de ingresos”.

– John Riccitiello, Unity CEO, Q2 2022 Earnings Report

¿Qué?

Hype Cycle para Data Management, 2022

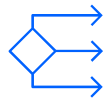


Source: Gartner, Hype Cycle for Data Management, 2022, Donald Feinberg, Philip Russom, Nina Showell, 30 June 2022

Gartner no respalda a ningún proveedor, producto o servicio descrito en sus publicaciones de investigación y no recomienda a los usuarios de tecnología que seleccionen sólo aquellos proveedores con las calificaciones más altas u otra designación. Las publicaciones de investigación de Gartner consisten en opiniones de la organización de investigación de Gartner y no deben interpretarse como declaraciones de hechos. Gartner renuncia a toda garantía, expresa o implícita, con respecto a esta investigación, incluida cualquier garantía de comerciabilidad o idoneidad para un propósito particular.



Databand.ai
permite
Observabilidad y
confiabilidad
desde el inicio
hasta el final de
tu proceso

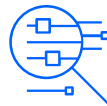


Full-stack, enterprise
aplicacion en
observabilidad.

INSTANA

an IBM Company

Una solución de observabilidad
que de inicio a fin proveen
soluciones atraves de data y
machine learning.



Observabilidad de
datos para bases de
datos dinamicas y
estaticas.

Databand.ai

an IBM Company

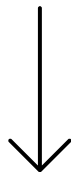


Observabilidad del modelo
para una inteligencia
artificial confiable y de
confianza a lo largo de su
ciclo de vida.

Watson® Studio
en IBM Cloud Pak®
para-Data

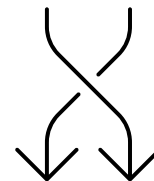
Monitoreo de datos vs Observabilidad de datos

“El pipeline **fallo...**”



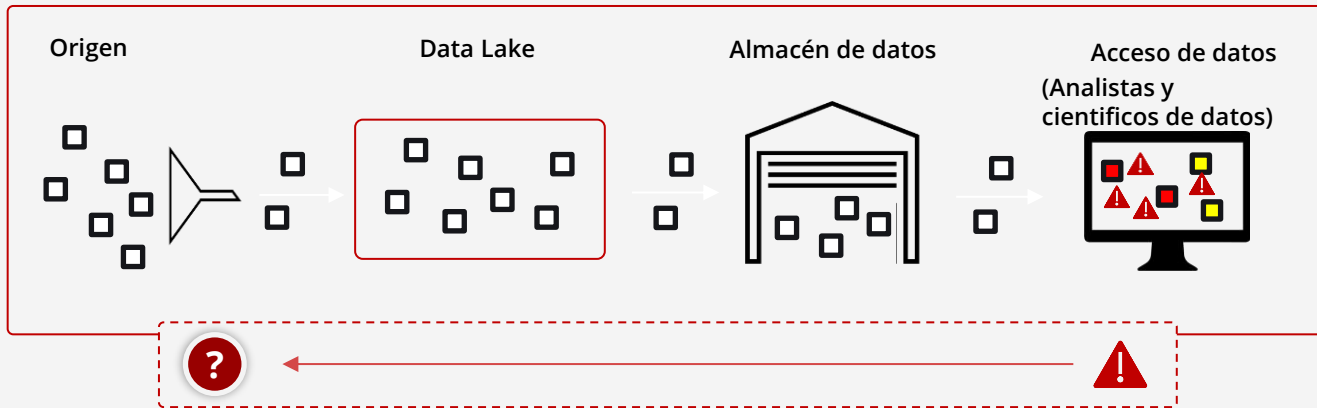
Monitoreo de datos	Observabilidad de datos
Vista limitada	Vista Holística
¿Qué paso?	¿Que? ¿Dónde? ¿Por qué? Paso
Sistema aislado	Sistemas Dependientes
Detección estática	Detección dinámica
Reactivo	Proactivo

“El pipeline fallo,
aquí esta la **causa
raiz** y como va a
afectar a las
dependencias XYZ...”



Problema

Los ingenieros de datos son **reactivos** a la información



Muchos errores de calidad de datos son **ignorados**

Los problemas de la plataforma salen a la luz solamente cuando **los consumidores reportan el error.**

Después de que los errores son reportados **no se resuelven rápido**

The root causes



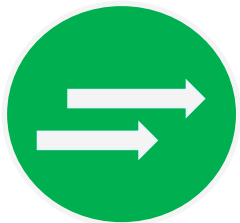
Cadena de herramientas fragmentada



Volumen de datos



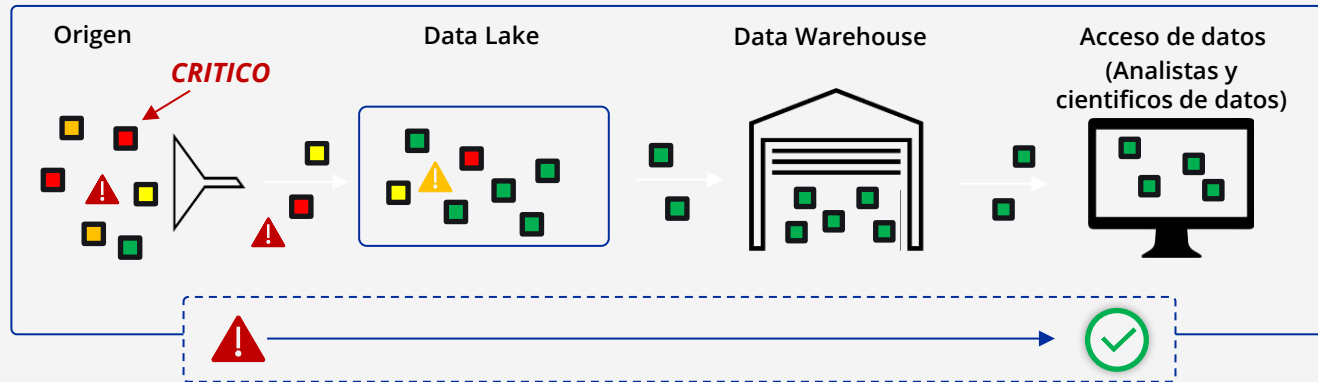
Inundación de ruido.



Nuestra
solución se
enfoca en
observar
información
en
movimiento

Solución

Observabilidad de datos proactiva: **Shift + boton izquierdo** y resuelve problemas desde el origen.



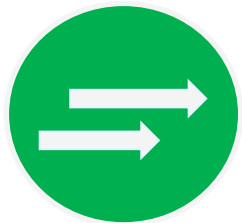
Mejoras de **MTTD**

Encuentra errores de código en tiempo real, tan pronto haya leído todo el código

Mejoras de **MTTR**

Identifica instantáneamente la causa de los errores de Código.

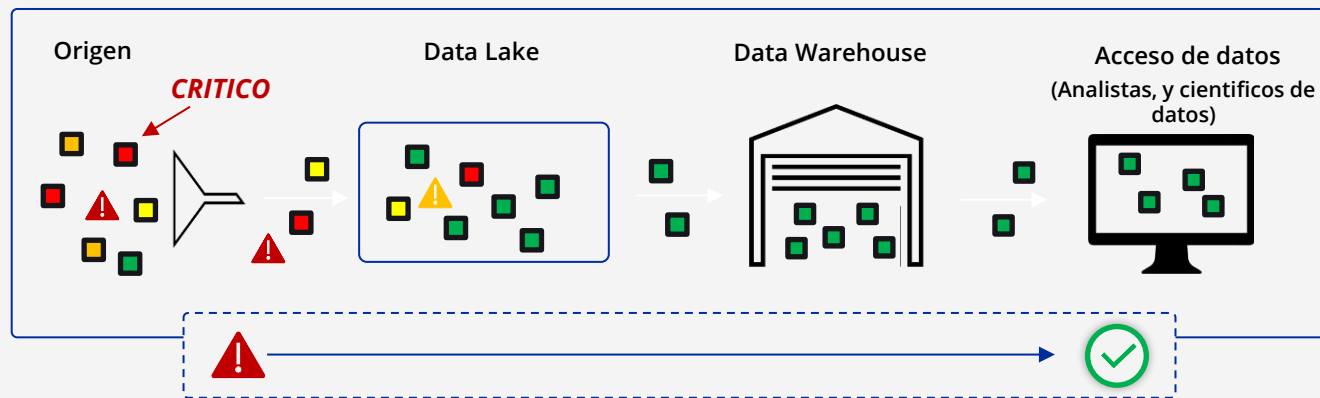
Mejoras en la **calidad** de producción de datos. *Mejora la confianza y la satisfacción del consumidor*



Databand se enfoca en los datos en movimiento

Solution

Proactive Data Observability: *Shift izquierdo* y resuelve problemas en la fuente



Observa el pipeline de datos en los **proceso de calidad**

- Estatus
- Desempeño
- Latencia

Observa la **calidad de datos y su fiabilidad**

- Esquema de cambios
- Forma de datos
- Frescura de datos

Observación de **Linaje**

- *Analisis de impacto*
- *Descubrimiento de interconexiones*
- *Identificación de causa raíz.*

Observabilidad de datos: un nuevo caso de uso IBM's data fabric

Conductor

Reto

Solución

Iniciativas de eficiencia

Demasiado tiempo perdido en depurar pipelines y costos fuera de control

Reduzca el tiempo de depuración y mejore la utilización de los recursos.



“Databand tells **you right away** when a data process is underperforming. Now **we don't have to wait** for a pipeline to fail to know if we'll have missing or incomplete data downstream.

- Ben, Director of Data Engineering at HBO Max



“Databand helps us **detect data quality issues faster** so we can meet our data SLAs.

Without Databand, we didn't know we had problems **until two or three days later.**

- Fithrah, Data Engineering Lead at Shipper

Desarrollo de nuevos productos de datos

Falta de confianza en los datos e incapacidad para mantener el release del producto de datos de bloque de SLA

Crear confianza en los SLA, desbloquear nuevas áreas de crecimiento.



“Databand automatically monitors the health of our Airflow pipelines so we spend **less time debugging** and **more time building ML models.**

Before Databand, **60% of our pipelines** had at least one data incident. **Now less than 1% of pipelines have incidents.** This resulted in a **3X increase** in our customers since we can now manage our ML deep learning models at scale.”

- Tzooof Hemed, AI-Engineering Team Leader at Trax Retail

¿Cómo?

1

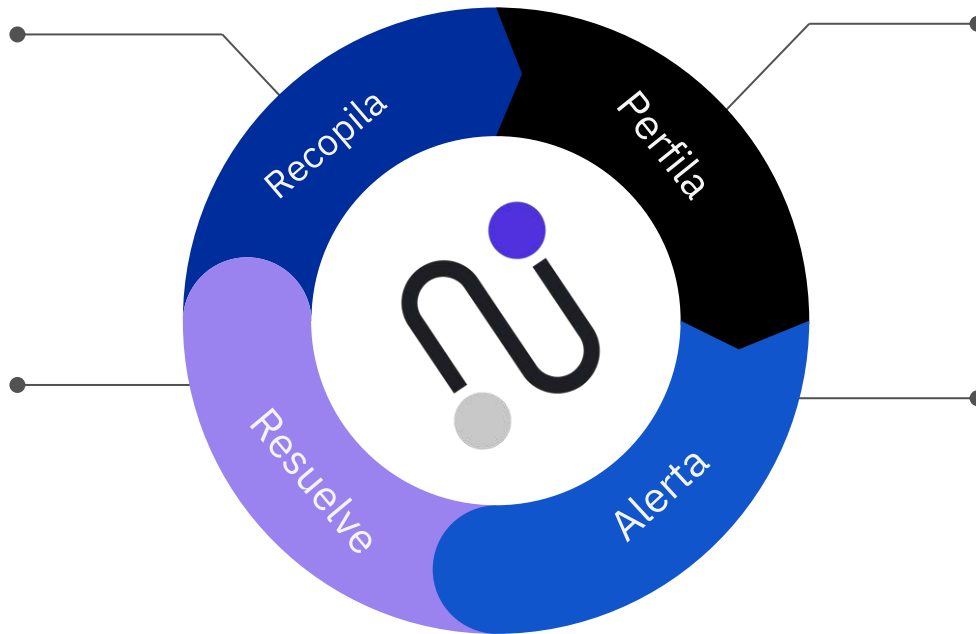
Recopilar automáticamente metadatos

de todas las soluciones clave en la pila de datos moderna.

4

Resuelva a través de la clasificación

Cree flujos de trabajo inteligentes para solucionar problemas de calidad de los datos y mantener los acuerdos de nivel de servicio en marcha.



2

Construir línea base histórica

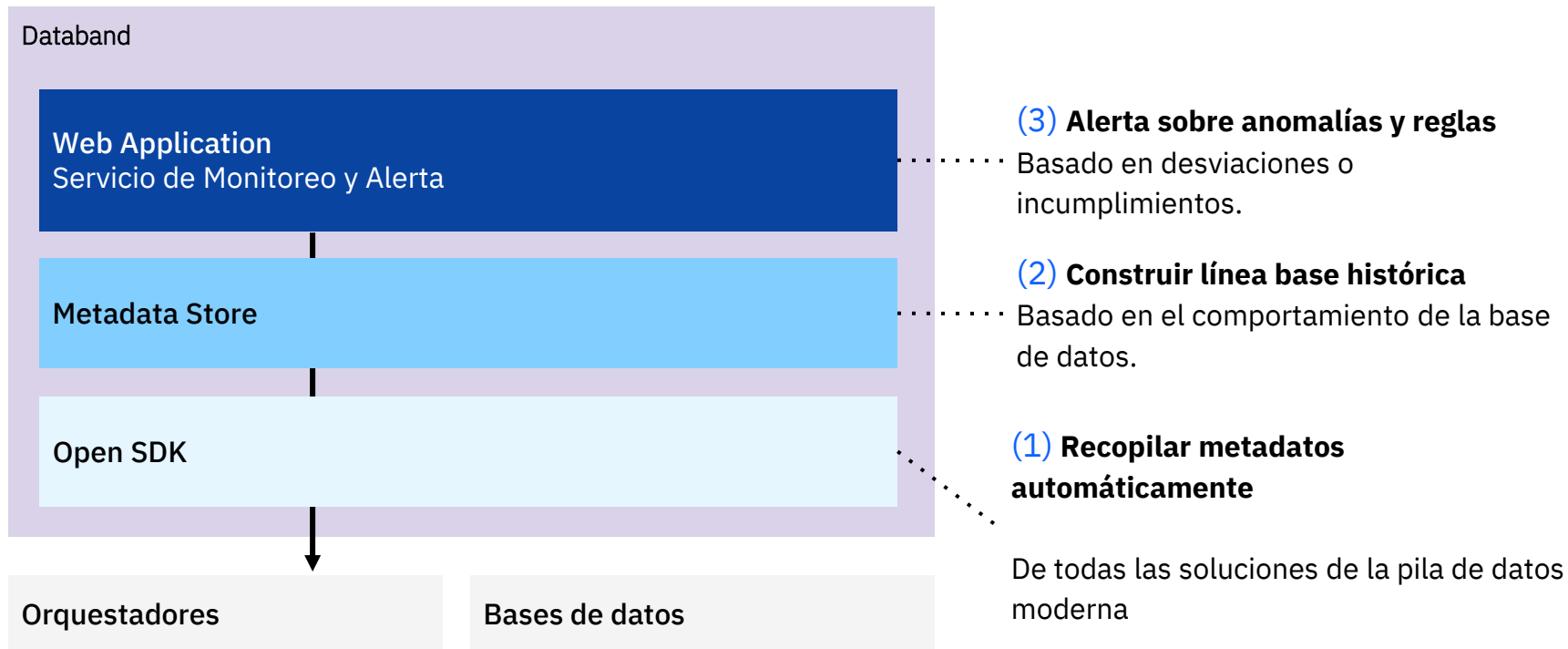
Basado en un comportamiento de interconexión de datos común.

3

Alerta sobre anomalías y reglas

Basado en desviaciones o infracciones.

¿Cómo funciona?



Diferenciadores clave

Alta variedad de metadatos críticos

Databand recopila todas las fuentes de información de incidentes clave

Covertura a través de tecnologías

La base de datos rastrea los problemas del **data lake** a través del warehouse

Alertas en tiempo real

Databand aumenta las alertas de incidentes en tiempo real, proporcionando el tiempo más rápido para la detección





Presenter's name

Presenter's contact information