

Introducción al Modelado DataVault 2.0

Fundamentos del modelado de datos moderno

Rubén Hermoso Díez

Agenda

- ¿Qué es Data Vault?
- Componentes básicos
- Caso práctico
- Buenas prácticas
- Mini test de repaso

¿Por qué no siempre es suficiente con Kimball?

- Difícil adaptación a cambios frecuentes
- No se registran todas las versiones del dato
- Costoso mantener trazabilidad completa

Cliente

ID Cliente	Nombre Cliente	Email	Dirección
1	Juan Pérez	juan@example.com	Calle Falsa 123
2	María Gómez	maria@example.com	Calle Real 456

Pedido

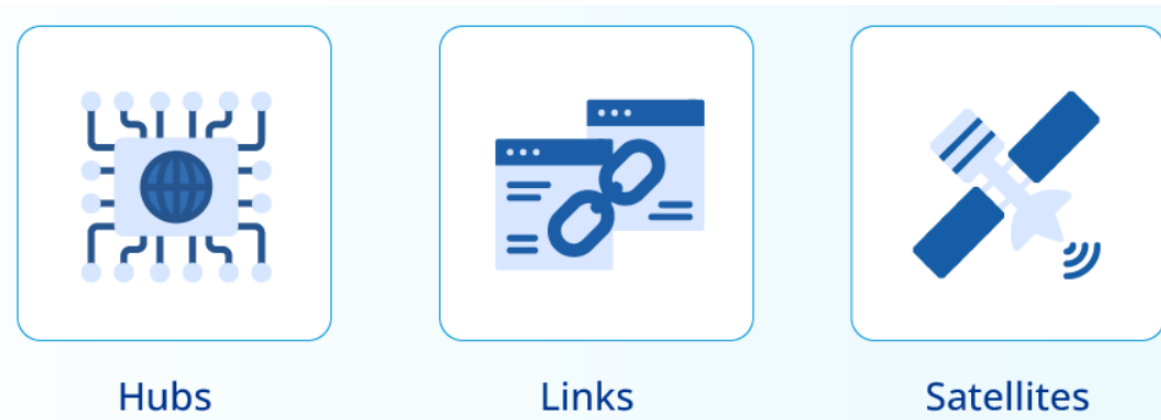
ID Pedido	ID Cliente	Fecha	Monto
1001	1	2025-05-01	250
1002	2	2025-05-02	400

Diferencias enfoque Kimball y DataVault

Aspecto	Kimball	Data Vault
Enfoque	Modelo dimensional (Hechos y Dimensiones), ideal para BI.	Modelo basado en historificación y trazabilidad (Hubs, Links, Satellites).
Flexibilidad ante cambios	Limitada , adecuado para datos estables.	Alta flexibilidad , adecuado para datos cambiantes.
Manejo de datos históricos	No conserva historial completo (solo datos actuales).	Preserva el historial completo de los datos.
Implementación	Fácil de implementar y entender.	Más complejo , requiere mayor esfuerzo de diseño y automatización.
Consultas y rendimiento	Alto rendimiento en consultas por su simplicidad.	Consultas más lentas debido a la reconstrucción del historial.
Mantenimiento	Bajo costo de mantenimiento en entornos estables.	Mayor costo de mantenimiento debido a la gestión de cambios y datos históricos.
Trazabilidad	Limitada , no permite auditoría completa de los cambios.	Alta trazabilidad y permite auditorías completas de los datos.
Adecuación	Ideal para entornos estables con datos poco cambiantes.	Ideal para entornos dinámicos y grandes volúmenes de datos cambiantes.
Automatización	No diseñado para automatización desde el inicio.	Diseñado para automatización , escalabilidad y facilidad de mantenimiento.

Qué es Data Vault

- Método de modelado orientado a la historificación y trazabilidad
- Separa claves, relaciones y atributos
- Pensado para grandes volúmenes y automatización

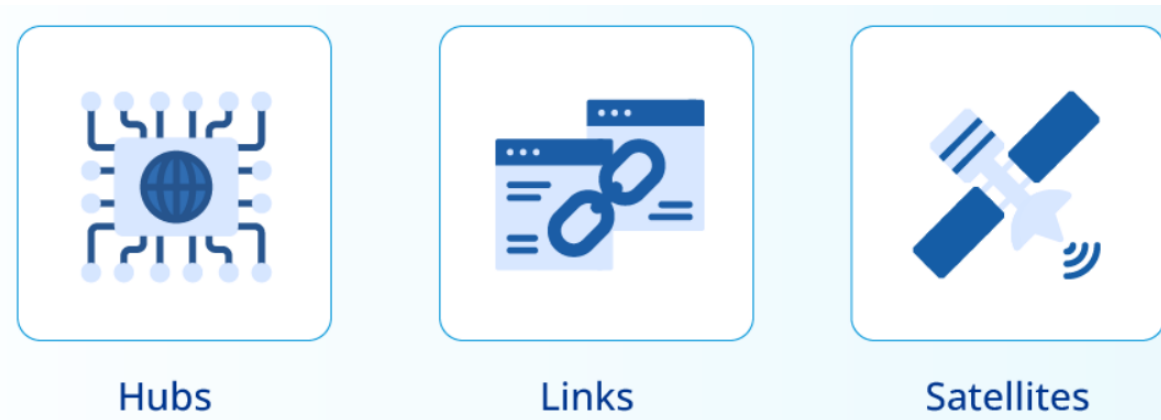


Comparativa de enfoques

Modelo	Enfoque	Ventaja clave
Kimball	Negocio	Sencillez y rendimiento
Inmon	Corporativo	Integración centralizada
Data Vault	Auditabilidad	Cambios y trazabilidad

Filosofía de Data Vault

- Lo que no cambia: business keys (Hubs)
- Lo que cambia: atributos (Satellites)
- Relaciones (Links) separadas para flexibilidad



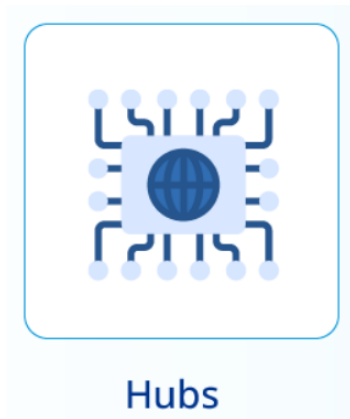
Componentes básicos

- Hubs: claves únicas de negocio
- Links: relaciones entre Hubs
- Satellites: información contextual e histórica



Hub

- Clave de negocio única (ej. ID Cliente)
- Los registros insertados no cambian con el tiempo
- Se crea un Hub por cada entidad principal involucrada en el proceso de negocio a modelar



HubCliente

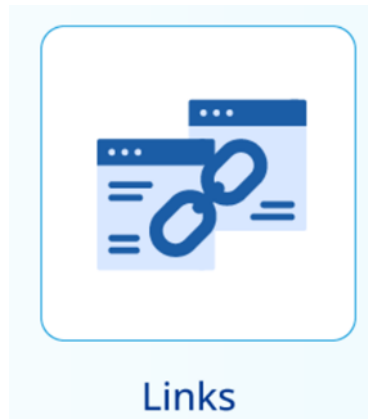
IDCliente	FechaCarga	Fuente
1	2025-01-01	Sistema A
2	2025-01-02	Sistema B
3	2025-01-03	Sistema A

HubPedido

IDPedido	FechaCarga	Fuente
1001	2025-01-01	Sistema A
1002	2025-01-02	Sistema B
1003	2025-01-03	Sistema A

Link

- Une dos o más Hubs
- Representa relaciones (ej. Pedido relaciona Cliente y Producto)
- También tiene clave única

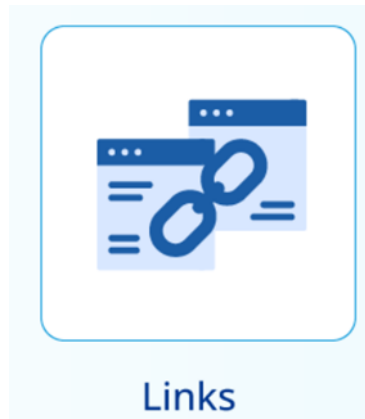


LinkCliente-
Pedido

IDPedidoCliente	IDCliente	IDPedido	FechaCarga	Fuente
1	1	1001	2025-01-01	Sistema A
2	2	1002	2025-01-02	Sistema B
3	3	1003	2025-01-03	Sistema A

Satelite

- Almacena atributos que cambian
- Siempre vinculado a un Hub o Link
- Incluye fecha de carga y fuente



SatCliente

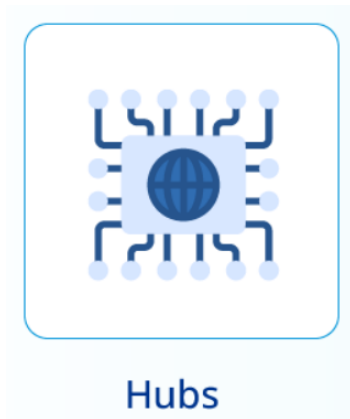
IDCliente	Nombre	Email	FechaCarga	Fuente
1	Juan Pérez	juan.perez@email.com	2025-01-01	Sistema A
2	Ana Gómez	ana.gomez@email.com	2025-01-02	Sistema B
3	Pedro Díaz	pedro.diaz@email.com	2025-01-03	Sistema A

SatPedido

IDPedido	Cantidad	FechaEntrega	FechaCarga	Fuente
1001	2	2025-01-10	2025-01-01	Sistema A
1002	5	2025-01-15	2025-01-02	Sistema B
1003	3	2025-01-12	2025-01-03	Sistema A

¿Y si el Hub Cambia?

- El **registro anterior** no se elimina. Se mantiene en el Satellite con la fecha de carga que tenía originalmente
- Mantenemos un histórico de cambios



HubCliente

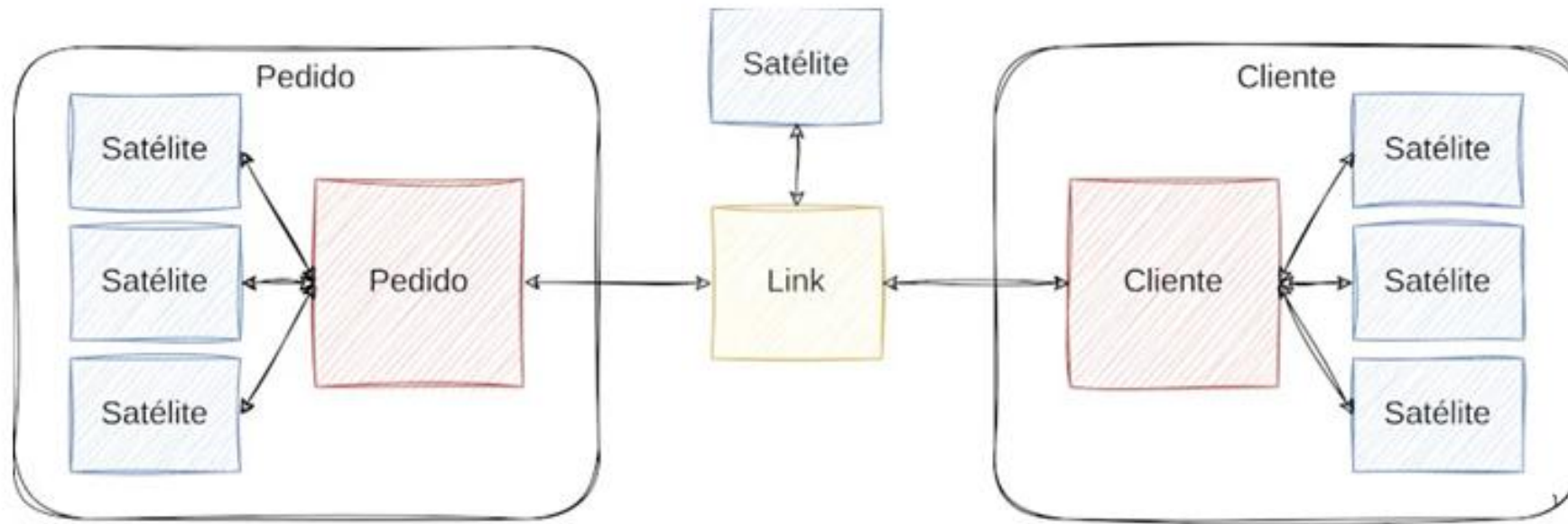
IDCliente	FechaCarga	Fuente
1	2025-01-01	Sistema A
2	2025-01-02	Sistema B
3	2025-01-03	Sistema A



IDCliente	Nombre	Email	FechaCarga	Fuente
1	Juan Pérez	juan.perez@email.com	2025-01-01	Sistema A
2	Ana Gómez	ana.gomez@email.com	2025-01-02	Sistema B
3	Pedro Díaz	pedro.diaz@email.com	2025-01-03	Sistema A
1	Juan García	juan.garcia@email.com	2025-02-01	Sistema A

Componentes básicos

- Hubs: claves únicas de negocio
- Links: relaciones entre Hubs
- Satellites: información contextual e histórica



Ejemplo: caso práctico Empresarial

- Hub = DNI de una persona
- Link = Registra cómo esa persona se **relaciona** con otras entidades
- Satelite = Detalles personales que cambian (dirección, empleo)



Buenas practicas

- Usar hash keys para identificar registros
- Siempre incluir fecha de carga y fuente
- Diseñar para automatización desde el inicio



Herramientas

- WhereScape, dbt, Talend, Azure Data Factory
- Diseñadas para generar y mantener modelos Data Vault

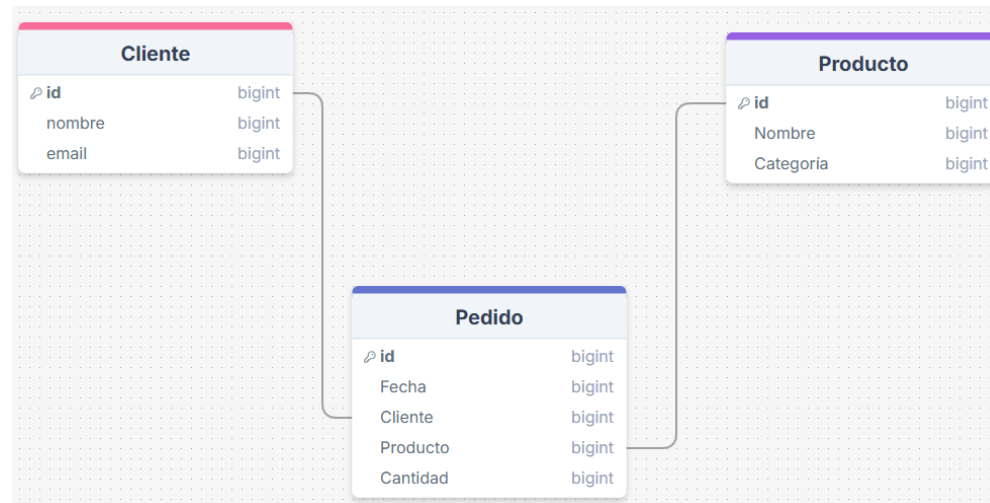


CASO PRACTICO

Caso práctico Datos de Ventas

Una empresa de logística necesita modelar sus datos bajo el estándar Data Vault 2.0. A partir de las siguientes tablas operacionales:

- Cliente (ID, nombre, email)
- Producto (ID, nombre, categoría)
- Pedido (ID, fecha, cliente, producto, cantidad)



Identificación de Hubs

- HubCliente → ID Cliente
- HubProducto → ID Producto
- HubPedido → ID Pedido (opcional)



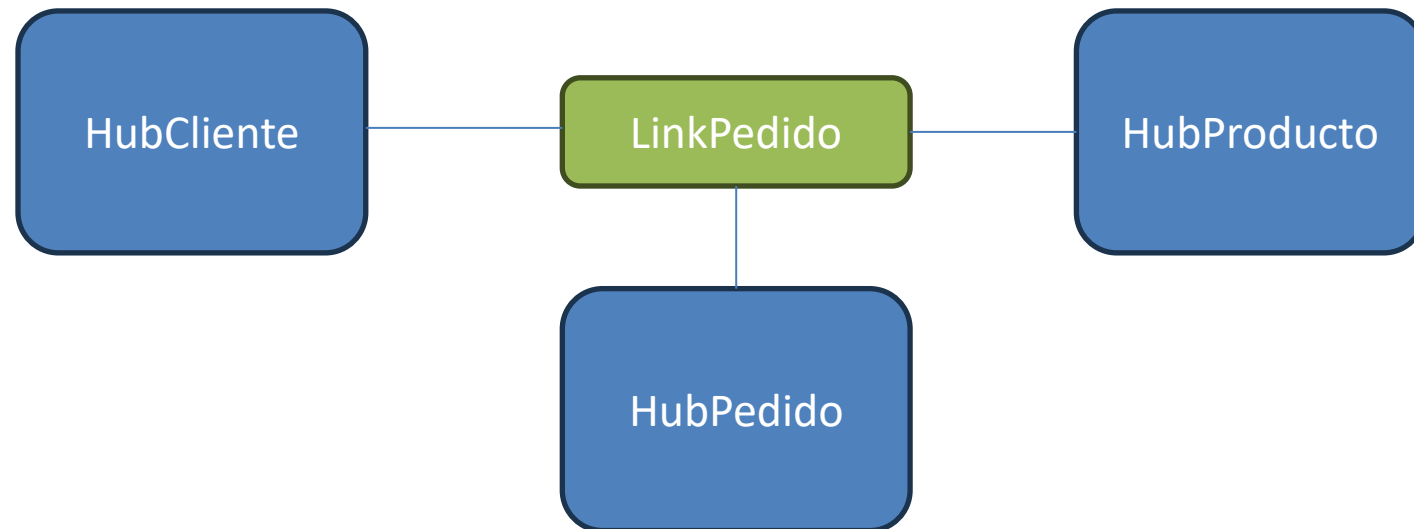
HubCliente

HubProducto

HubPedido

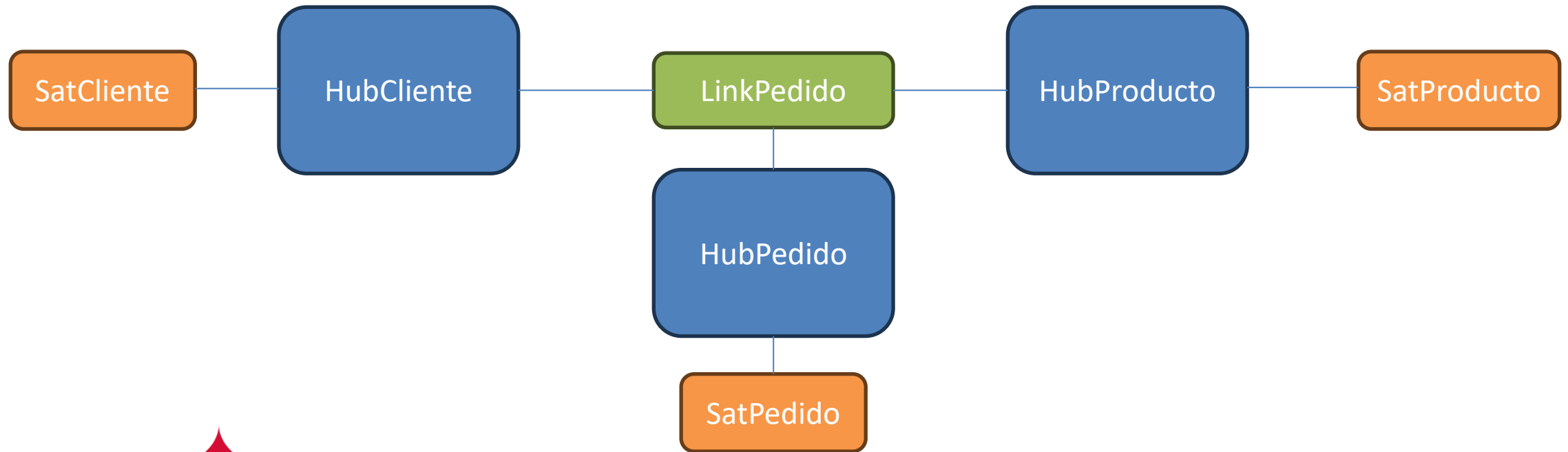
Identificación de Links

- LinkPedido une Cliente, Producto y Pedido



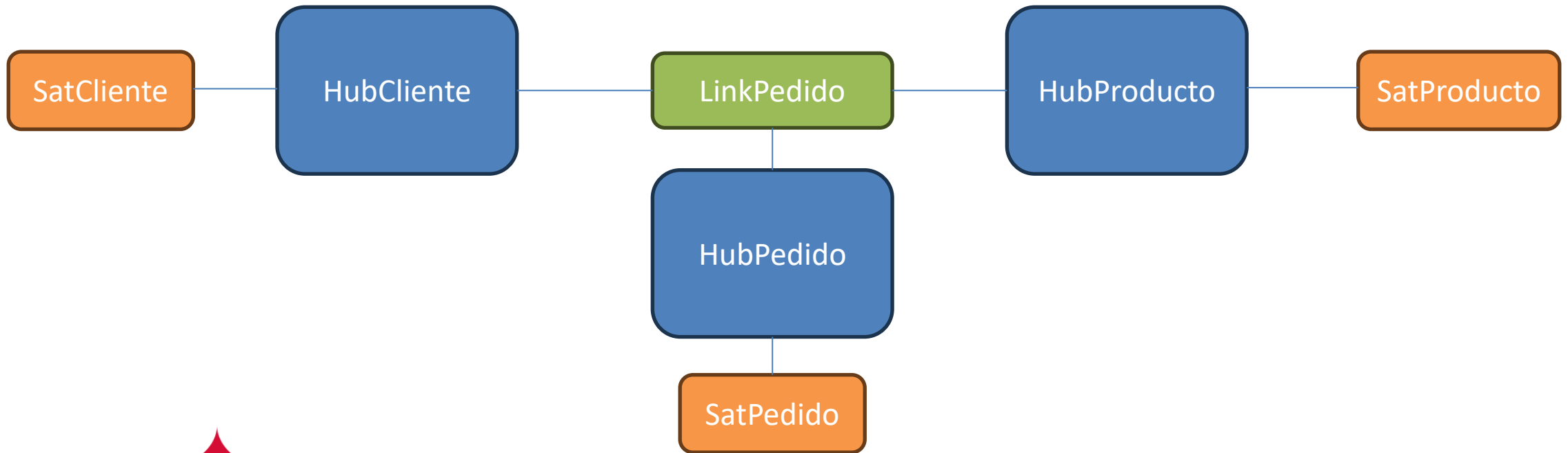
Diseño de Satelites

- SatCliente → nombre, email
- SatProducto → nombre, categoría
- SatPedido → fecha, cantidad



Buenas practicas

- Usar hash keys para identificar registros
- Siempre incluir fecha de carga y fuente
- Diseñar para automatización desde el inicio



Estoy preparado para el examen si...

- ☐ **Conozco los tres componentes de Data Vault:** Hubs, Links y Satelites.
- ☐ **Sé qué hace cada componente:**
 - ☐ Hubs: Claves de negocio únicas.
 - ☐ Links: Relaciones entre Hubs.
 - ☐ Satelites: Información histórica y atributos cambiantes.
- ☐ **Entiendo cómo Data Vault maneja los cambios en los datos:** Cada cambio se registra sin sobrescribir los datos anteriores.
- ☐ **Sé por qué Data Vault es más adecuado para entornos de datos cambiantes:** Permite la trazabilidad y la flexibilidad frente a los cambios frecuentes.
- ☐ **Entiendo la diferencia clave entre Data Vault y Kimball:** Data Vault es más escalable y mantiene el historial completo de los datos.



**Fundación
iberCaja**

