Mejores prácticas para analítica a gran escala

Arca Continental

2022

Mejores prácticas para analítica a gran escala

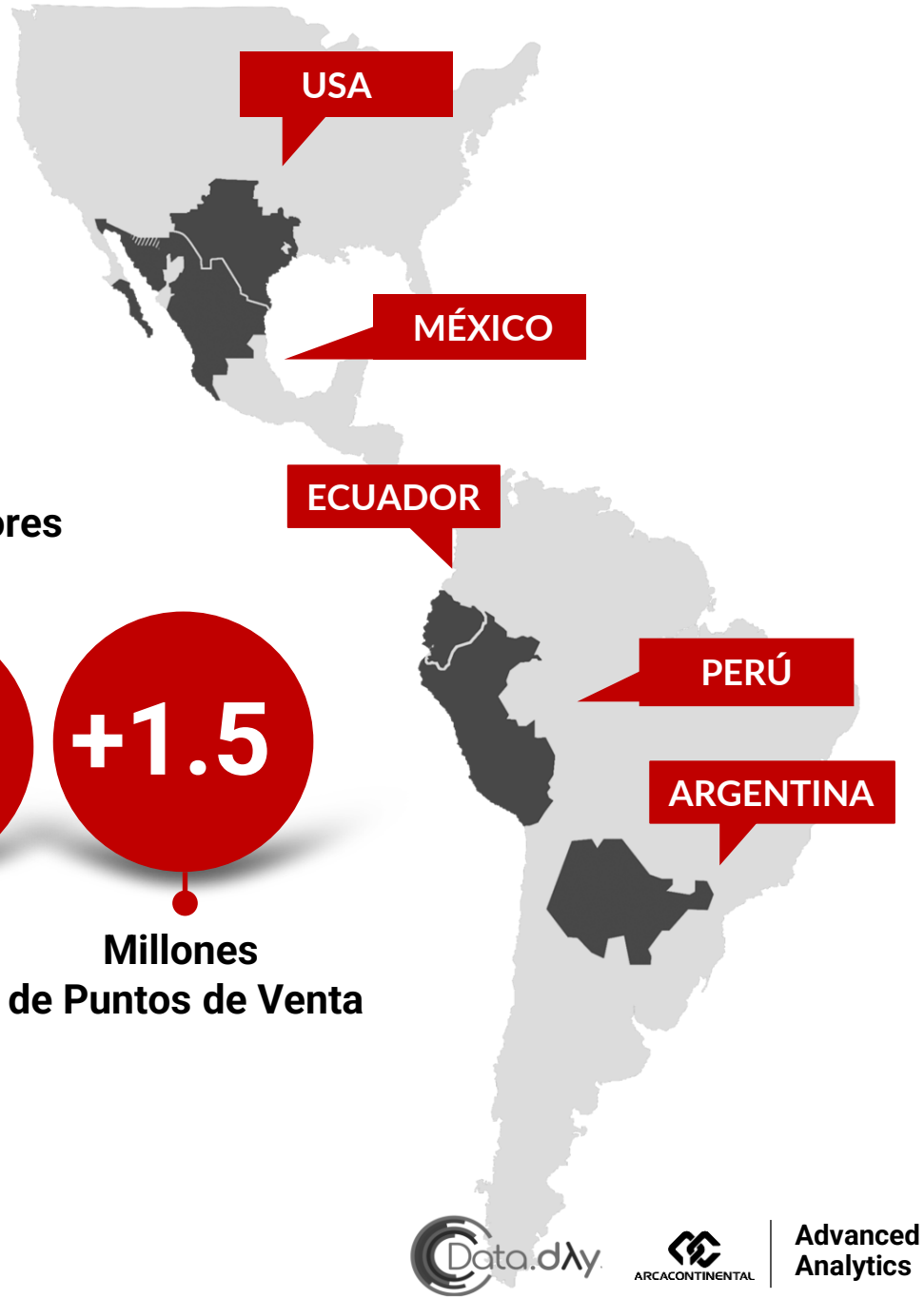
ARCACONTINENTAL | Coca-Cola

¿Quiénes somos?

César Villarreal
Jefe de Arquitectura de Datos

Iván Ramírez
Jefe de Ingenieria de Datos

Resumen de la compañía



Litros Vendidos (MM)

14.4

169

$ Ventas MMDP

Transacciones de Pedidos diarios MX (M)

2.5

364

Centros de Distribución

Mil Colaboradores

62

+1.5

Millones de Puntos de Venta

Data.dAy

ARCACONTINENTAL

Advanced Analytics

Timeline Analítica Avanzada en Arca Continental

2018

- Desarrollo Pedido Sugerido

10 mil clientes

2019

- Despliegue Pedido Sugerido
- Desarrollo Pedido Sugerido
- Colaboración MIT Next Best SKU
- Colaboración Tec de Monterrey Potencial de Venta

220 mil clientes

2020

- Piloto Pedido Sugerido
- Refactorización de Pedido Sugerido

230 mil clientes

2021

- Desarrollo y piloto Pedido Sugerido
- Despliegue Pedido Sugerido
- Piloto Portafolio por Cliente
- Potencial de Venta
- Trade Promotion Optimization CCSWB

420 mil clientes

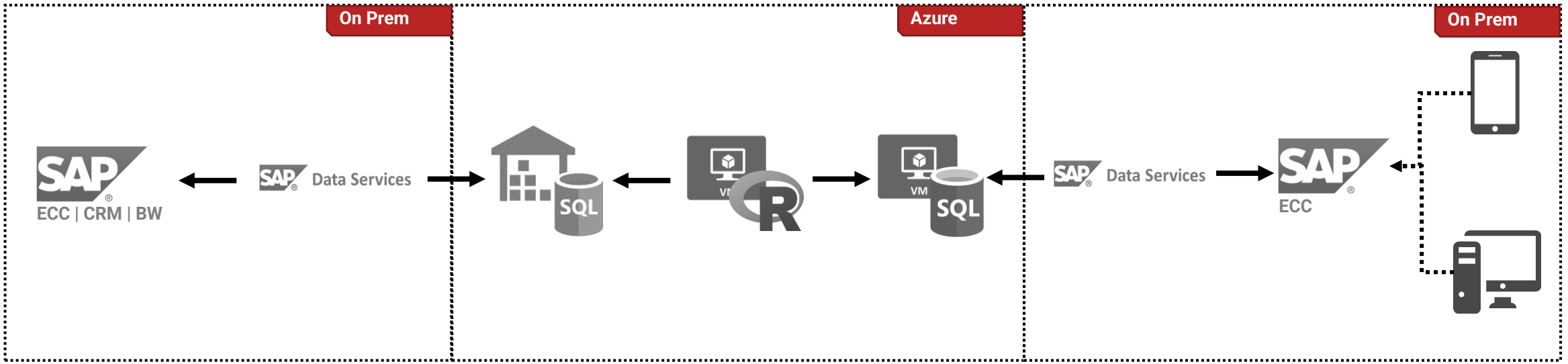
2022

- Desarrollo y piloto Pedido Sugerido
- Pedido Sugerido en vivo
- Despliegue Portafolio por Cliente
- Piloto Portafolio por Cliente
- Demand Forcast
- Piloto sistema de alertas

Data.dAy

ARCACONTINENTAL | **Advanced Analytics**

Implementación de Pedido Sugerido México - 2018



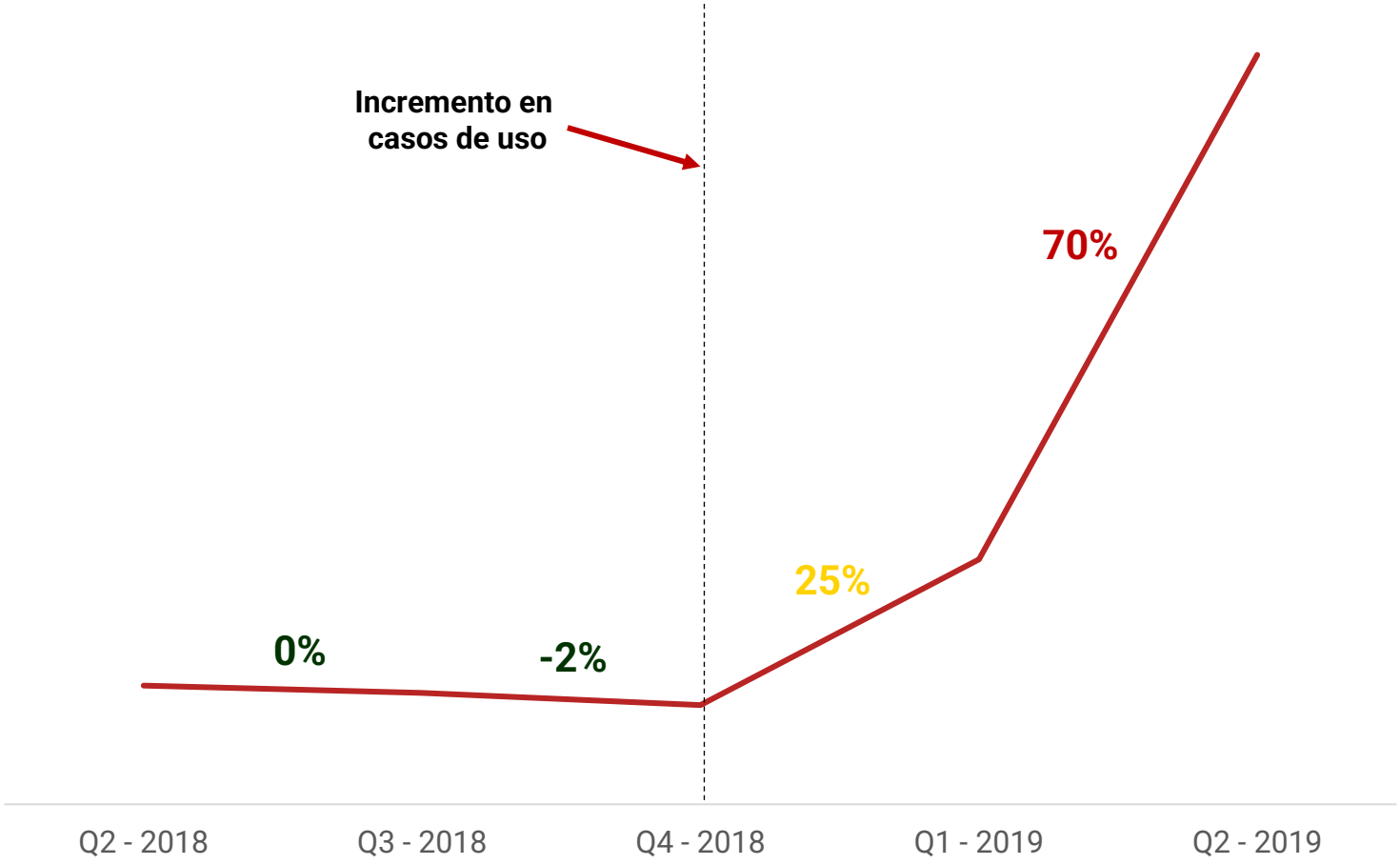
Características:

- Construido por proveedor externo
- No había separación entre limpieza de datos y feature engineering
- Incrementos en costos
- Falta de escalabilidad en la ejecución del modelo matemático
- Falta de versionamiento del proyecto

Qué aprendimos:

- Teníamos que buscar alternativas al stack tecnológico
- Necesitábamos dejar de ver la solución como un producto

Con el incremento de proyectos, los gastos en infraestructura comenzaron a crecer.



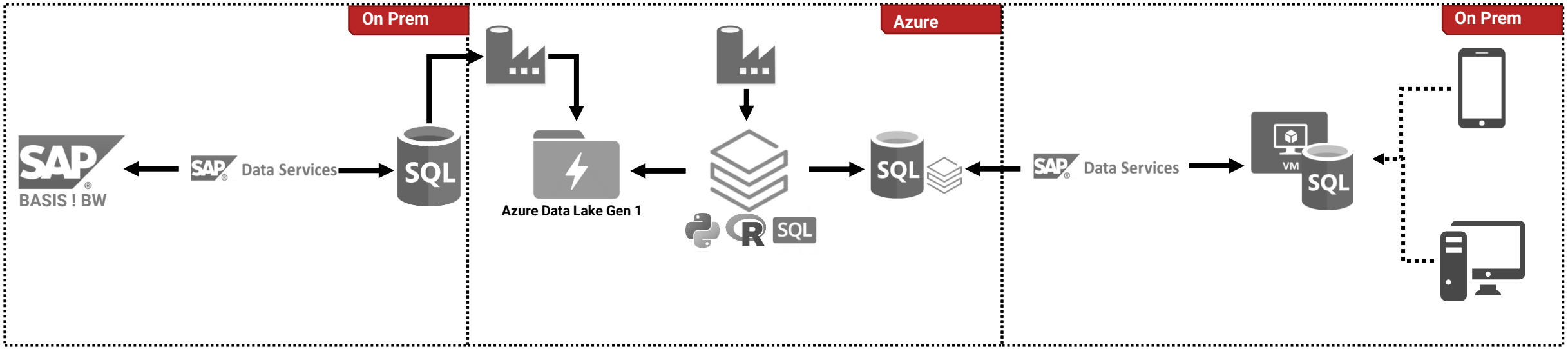
Tamaño del DWH

Personas

Procesos duplicados

Calidad de información

Implementación de Pedido Sugerido Perú - 2019



Características:

- Construido con otro proveedor externo
- Falta de lineamientos de programación + libertad de la herramienta tuvo desventajas.
- No había versionamiento de código
- Creación de BDs en cada ejecución del modelo

Qué aprendimos:

- Databricks era una herramienta con potencial, pero era necesario aprender a usarla correctamente
- Las bases de datos generadas por el modelo cada ejecución no era la mejor forma de continuar

Pruebas de cambio Stack Tecnológico

Azure Data Factory v2

- Integración con servicios de Azure
- Curva de aprendizaje corta para nuevos desarrolladores.
- Orquestación de procesos
- Permitía implementar CI/CD para el manejo de ambientes

Azure Data Lake Analytics

- Permitir el uso de USQL
- No requería recursos dedicados
- Potencia escalable.
- Compatible con Azure Data Lake Gen 1

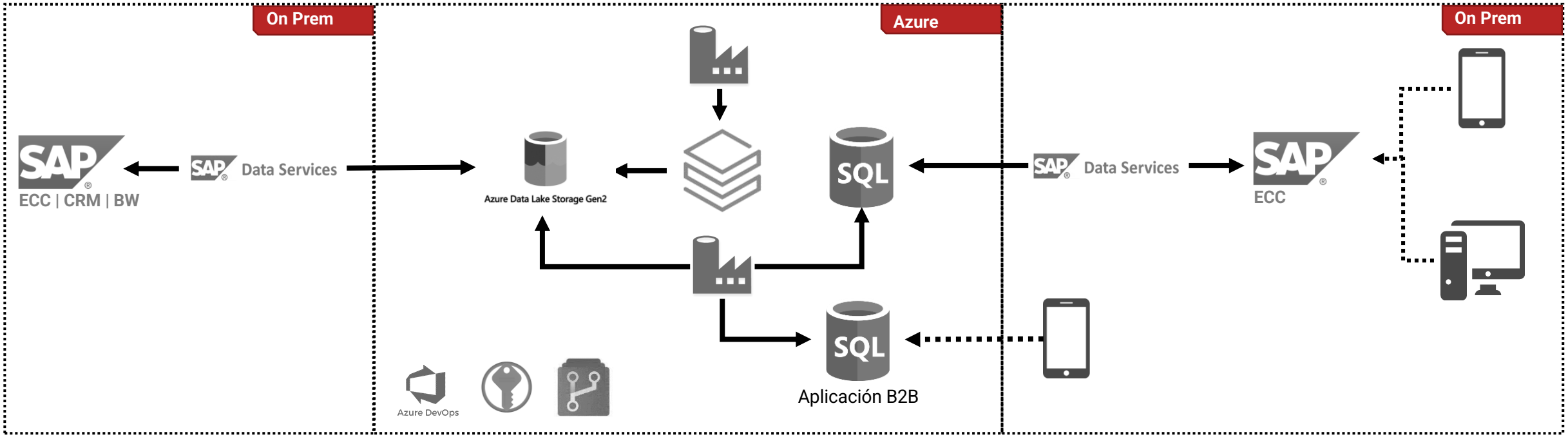
Azure Data Lake Gen1

- Permitió almacenar grandes volúmenes de datos a bajo costo
- Uso on-demand de la información almacenada
- Uso de herramientas alternativas para el procesamiento de los datos

Azure Databricks

- Configurable de manera sencilla
- Runtimes de clusters predeterminados con Apache Spark con opción de personalización
- Integración de procesos de Ciencia de Datos e Ingeniería de Datos

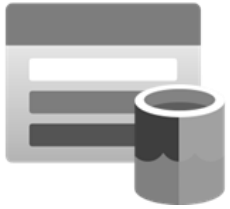
Refactorización de Pedido Sugerido – 2020



Características:

- Cambio a Azure Data Lake Gen 2
- Databricks como proceso de limpieza y modelos matemáticos
- Orquestación e ingesta con Azure Data Factory v2
- Implementación de DevOps + automatización de despliegues a distintos ambientes
- Replicado de forma global por operación

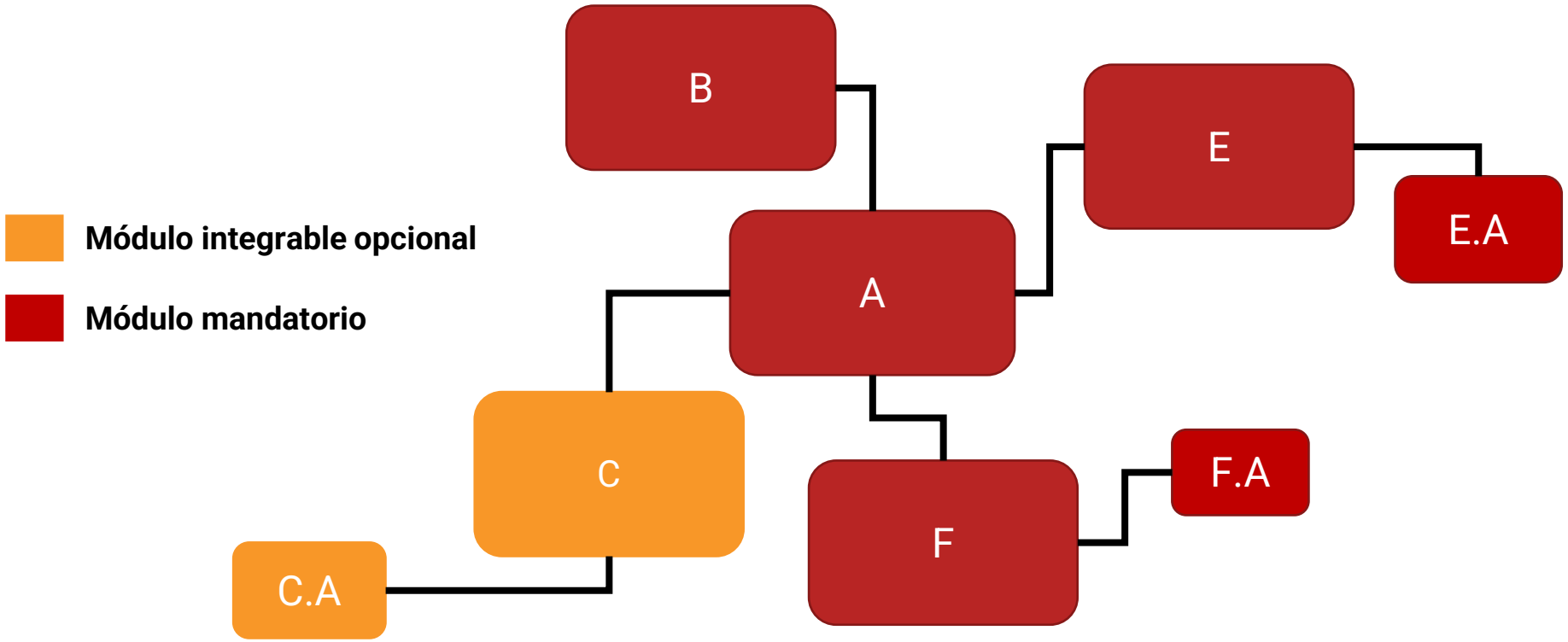
Cómo almacenamos nuestros datos



Estrategia de Zonas

Formatos	Nombre	Ejemplo
CSV, JSON, Parquet	RAW	**INTERNAL** • VENTAS **EXTERNAL** • CLIMA
DELTA LAKE	PROCESSED	**CURATED** • VENTAS **ANALYTICS** • PROYECTO 1
CSV, JSON, Parquet, DELTA LAKE	PLAYGROUND	**USER1** • ANALISIS 1 **USER2**

Estandarización del modelo de datos



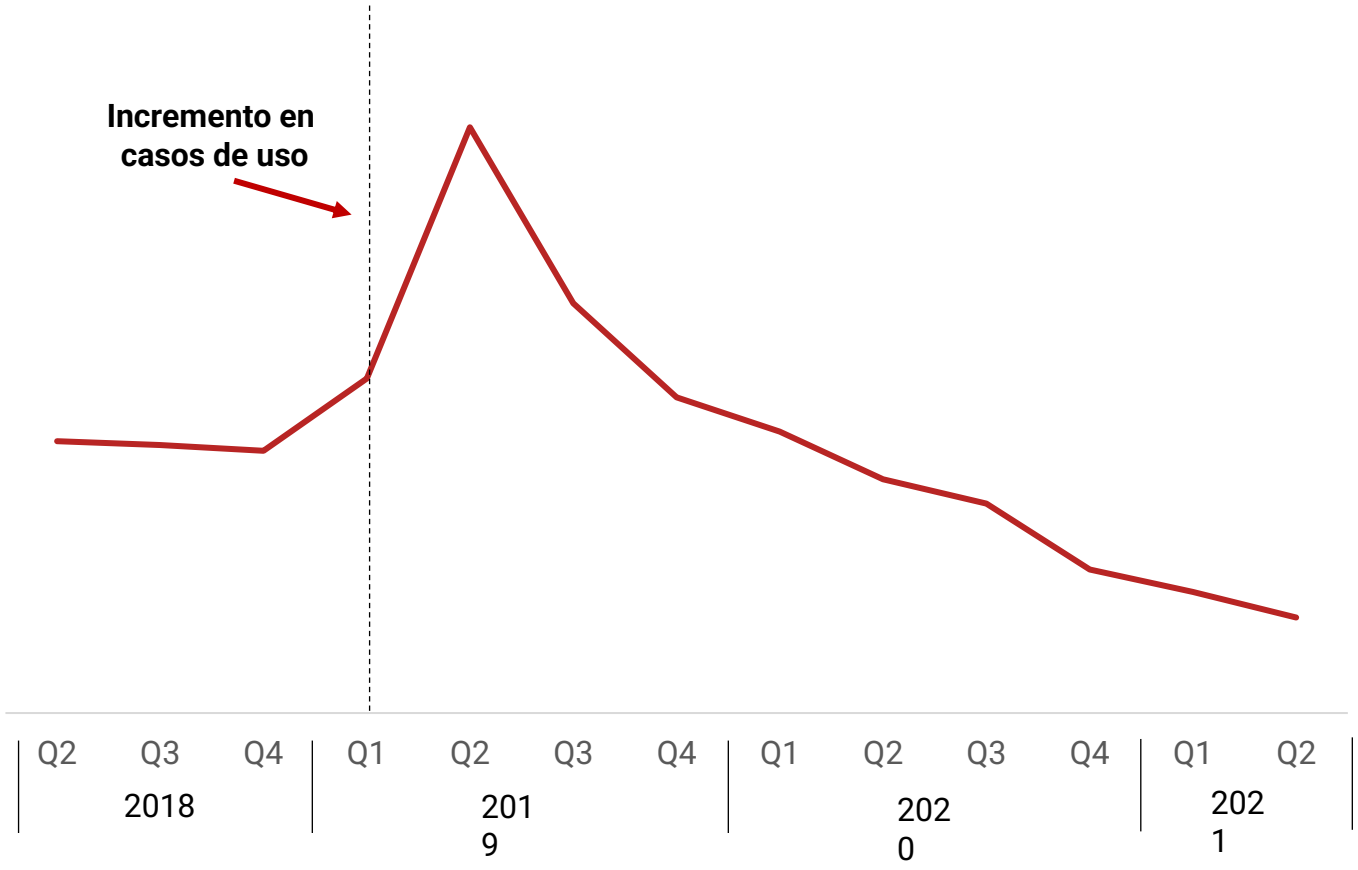
Qué hacemos:

- Conservar esquemas entre países para la solución
- Definir datos mínimos para la implementación de proyectos
- Complementar el modelo de datos una vez se tengan disponibles

Nos permitió:

- Implementar Pedido Sugerido en Ecuador y Argentina en menos de tres meses contra 1.5 años en Perú.

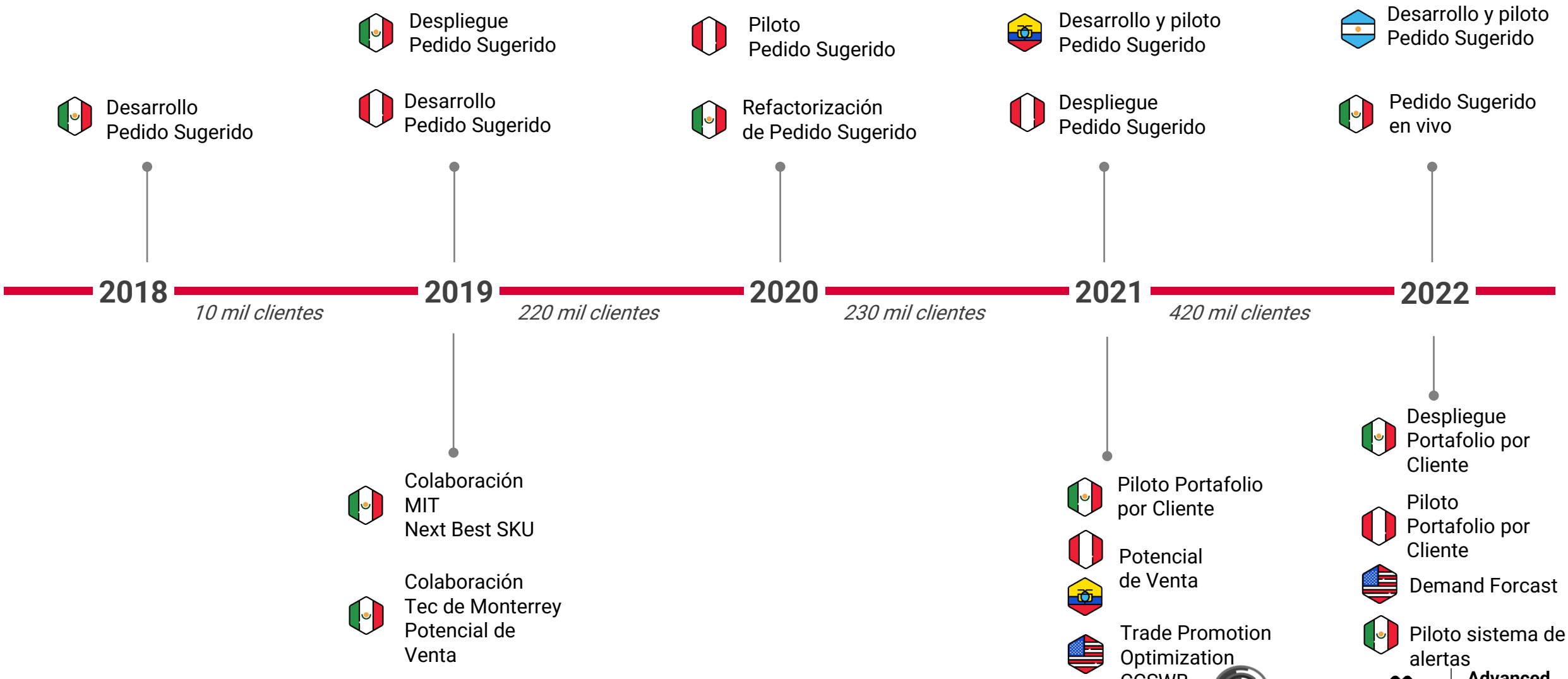
Con la solución actual para Analítica Avanzada hemos tenido **beneficios** considerables



Beneficios:

- Disminución de gastos
- Consistencia de información entre proyectos
- Agilidad de implementaciones
- Versionamiento de proyectos
- Facilidad de escalabilidad en los proyectos

Timeline Analítica Avanzada en Arca Continental



Data.dAy

ARCACONTINENTAL | **Advanced Analytics**

GRACIAS!

ARCACONTINENTAL