DataOps, MLOps, AIOps… Why now?

Luis Cajachahua
Senior Data Scientist
Center of Excellence, Teradata Americas

Julio 2020

Problemas vs Oportunidades

+Años

Time-to-market para implementar un data product en cualquier organización, debido a distintas barreras y a la complejidad. Los modelos demoran tanto en llegar a implementarse que se vuelven obsoletos antes.

15 %

Sólo el 15% de los proyectos Big Data llegan a producción (Gartner). Las barreras generalmente están relacionadas a conflictos organizacionales, falta de agilidad y de alineamento con el negocio.

+6 Meses

Para implementar un modelo o data product en producción. De no ser utilizado un data product **perderá su valor completamente en un período breve.**

El contexto ha cambiado

Antes

- Modelos y Algoritmos simples/tradicionales
- Implementación en SQL
- Datos Estructurados
- Data Marts de Modelos o ABTs
- Amplios ciclos de vida para modelos
- Roles establecidos
- Herramientas maduras

Ahora

- Modelos y Algoritmos sofisticados
- Implementación en diversas plataformas
- Datos Estructurados, No estructurados, Semi-estructurados, etc.
- Features Stores, Librerías de Datos
- Ciclos de vida cortos para modelos
- Múltiples roles interactuando
- Muchas herramientas que aparecen y cambian continuamente

DevOps: El inicio de todo

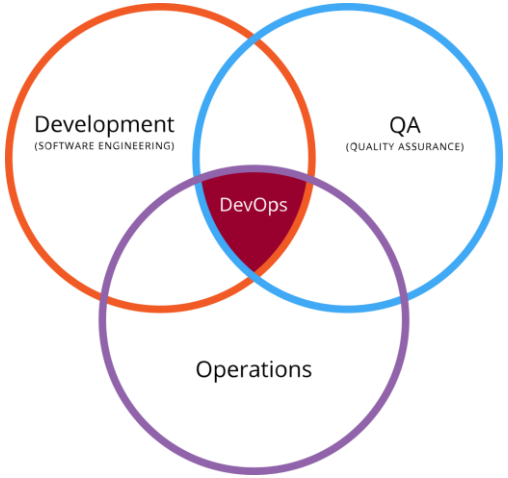
Developers & **Op**erations trabajando juntos en equipo

Una **cultura** que efatiza la **colaboración** y **comunicación** entre Desarrolladores y TI

Automatizando el proceso de entrega de Software

Los procesos de Building, testing & releasing se dan de manera más rápida, **frecuente** y **confiable**

Fuente: https://en.wikipedia.org/wiki/DevOps



teradata.

Herramientas DevOps a través del Proceso

4300 Biscayne Blvd
Suite 203 Miami,
Florida 33137

Phone: +1 (305)515-5324
partnernetwork@certijoin.com
www.certijoin.com

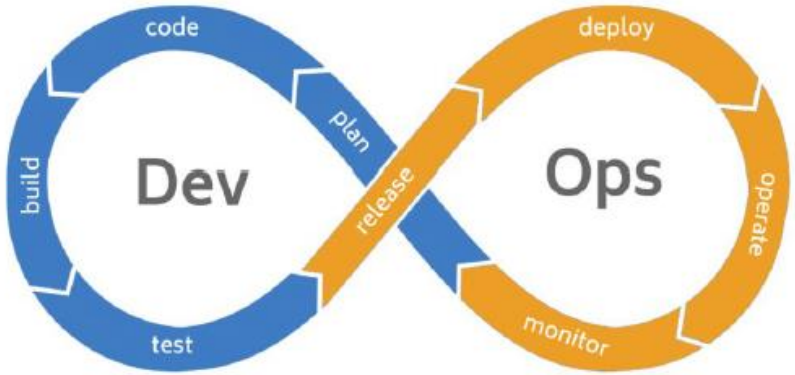
PERO

No se trata sólo de utilizar herramientas DevOps para desarrollar e implementar modelos

Encuentra las diferencias

Hay claras diferencias entre:

- Desarrollar Software/Apps
- Desarrollar BI
- Desarrollar Soluciones Analíticas



Data Science Project Lifecycle

The data science life cycle is described by several different analytical models, including Microsoft's Team Data Science Process (TDSP). Several companies are following this popular approach (Fig.1) to structuring data science projects.

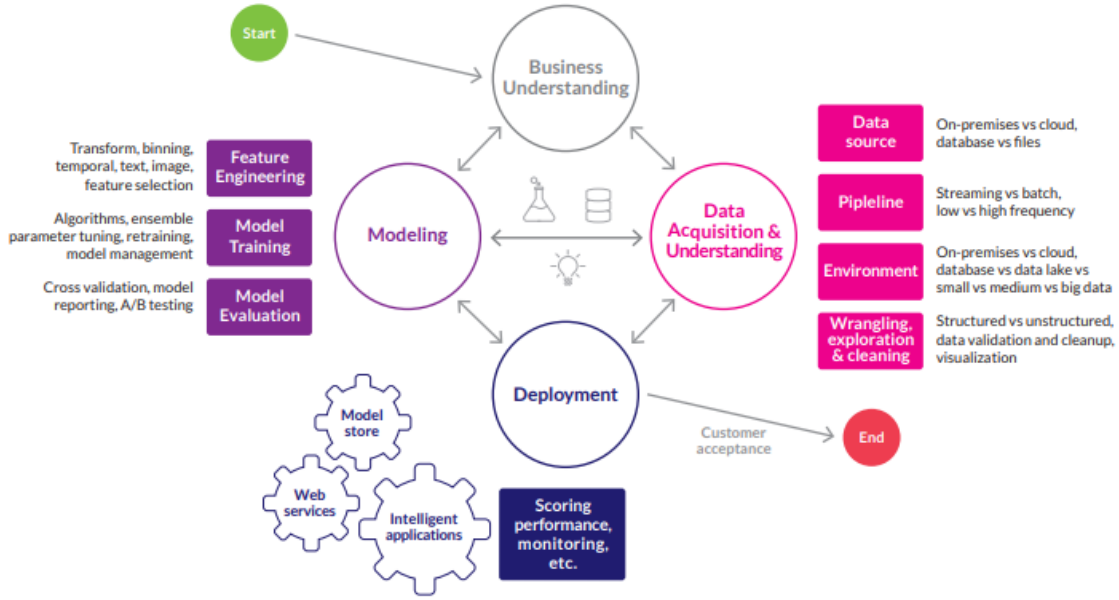
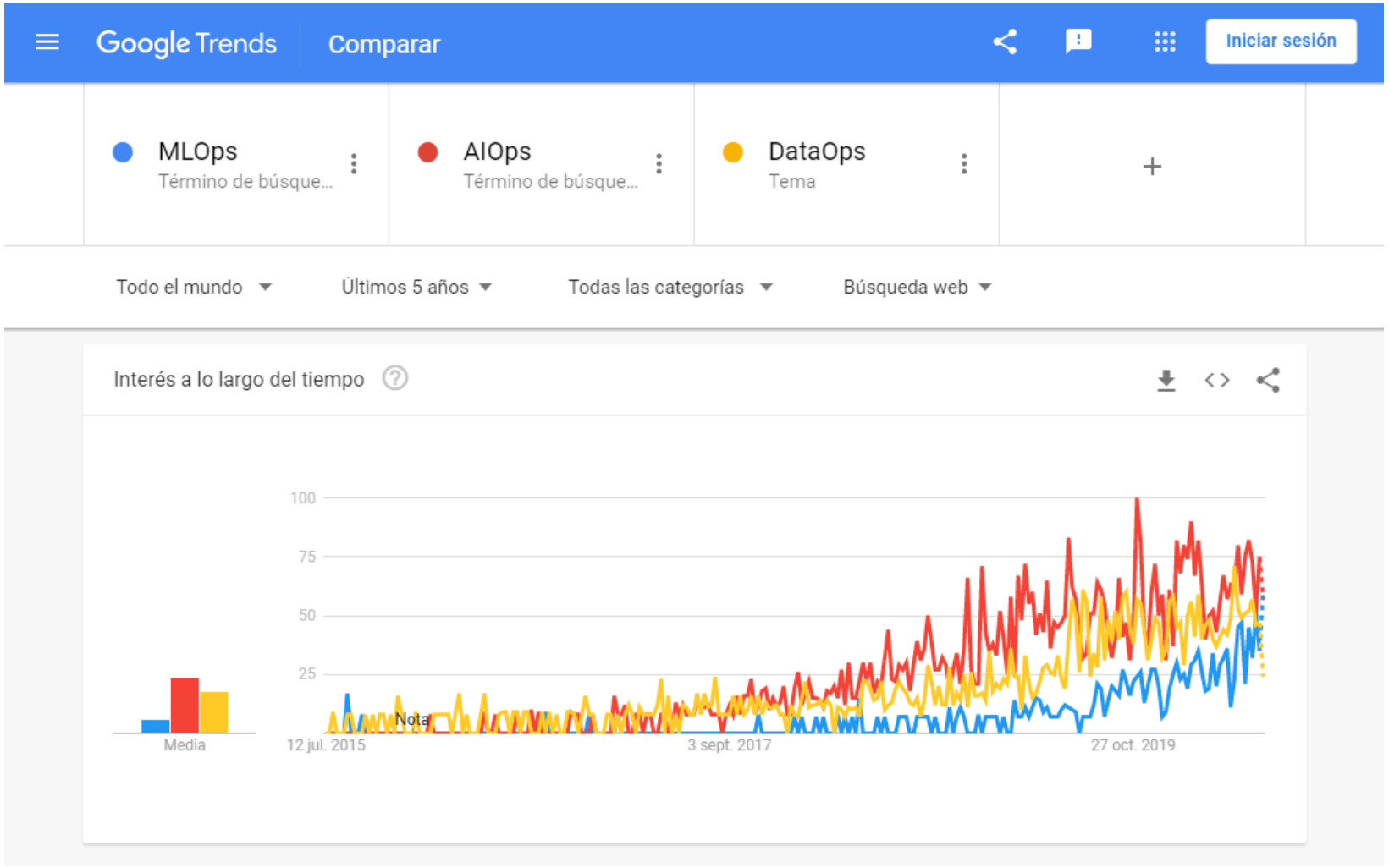


Figure 1. Data science lifecycle

teradata.

DataOps, MLOps, AIOps

- **Hecho:** Los tres términos se utilizan de forma intercambiable
- **Opinión:** DataOps debería estar más enfocado en armado de pipelines, MLOps debería estar más relacionado con Modelos y AIOps debería estar relacionado con automatización.



teradata.

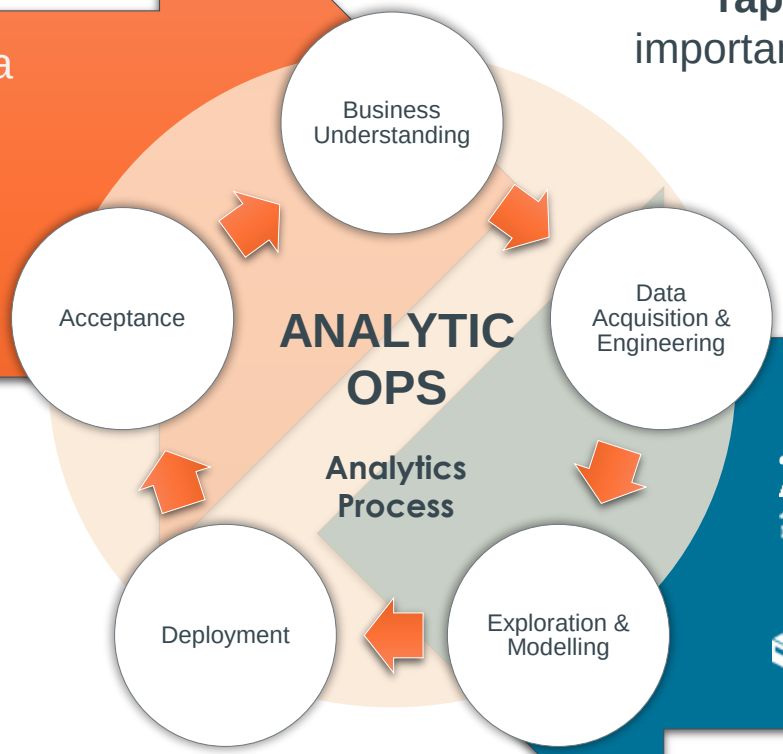
¿Y en qué consiste?

AnalyticOps es **un framework que posibilita la operationalización de modelos en una forma escalable,** tal que las organizaciones puedan llevar los casos analíticos **a producción rápidamente,** cumpliendo con requerimientos importantes como **gobierno, calidad** y **seguridad**

ANALYTICS

- flexibility
- exploration
- discovery
- modelling
- blue-sky ideation
- external data
- iteration
- data-mining
- statistics
- value-driven

Data Science Practices



OPERATIONS

- security
- governance
- compliance
- curation
- deployment
- maintenance
- integration
- testing
- engineering
- process-driven

Software Engineering Practices

teradata.

Los tres pilares de AnalyticOps

Desarrollo

Facilitar la integración de fuentes y exploración de datos

Plantillas reusables de extracción y transformación de datos

Desarrollo centralizado, seguro y auditable

Configuración del proceso de modelado (scheduling, reportes y aprobaciones)

Verificación de modelos y dependencias antes de la automatización

Automatización

Ejecución de pruebas unitarias de los modelos

Ejecutar rutinas de validación y performance de los modelos

Ejecutar evaluacion de modelos tipo champion / challenger

Crear indicadores para aprobar los pases a producción

Almacenar metadata de los modelos y de las métricas de performance

Consumo

Despliegue simplificado de modelos

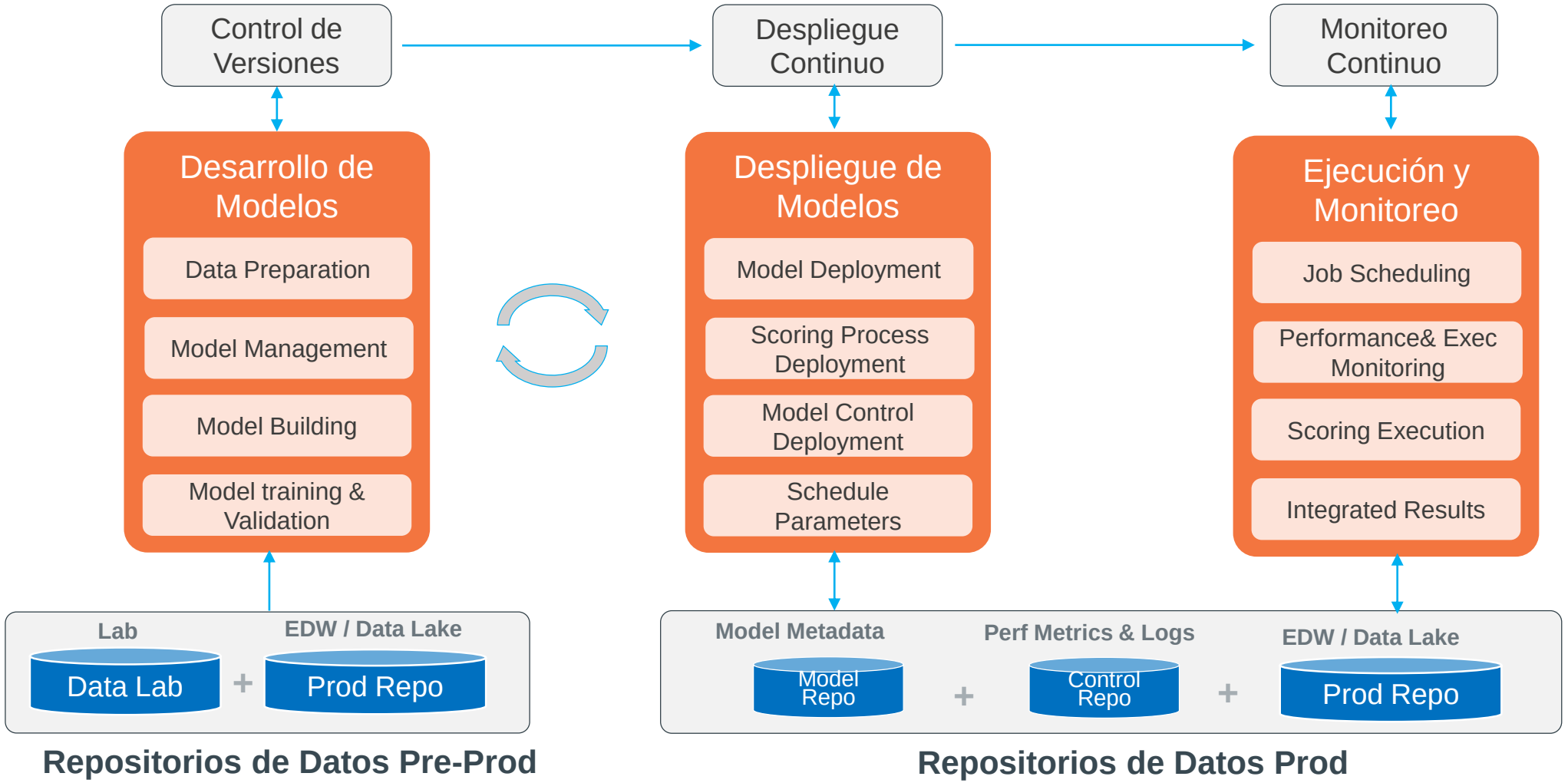
Almacenamiento de logs para el monitoreo y mejora continua

Integrar los modelos entre los procesos de negocio

Aplicar estrategias graduales o de A/B testing para elegir entre modelos alternativos

Despliegue de artefactos en el ambiente de ejecución (motor de scoring, web service, etc.)

Componentes de Arquitectura de AnalyticOps



teradata.

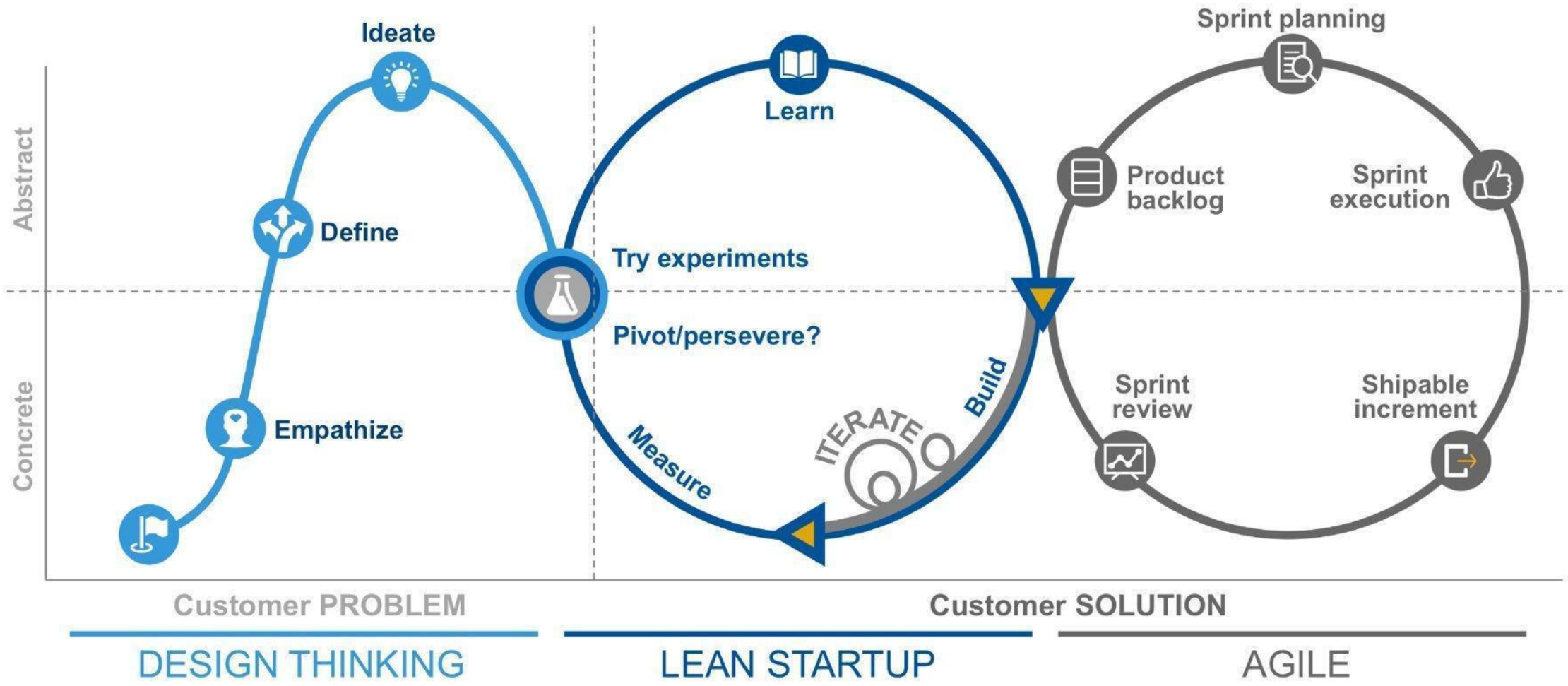
Y RECUERDA QUE:

Lo más importante no está en las herramientas, sino en la manera de hacer las cosas

Finalmente, ¿Qué se necesita?

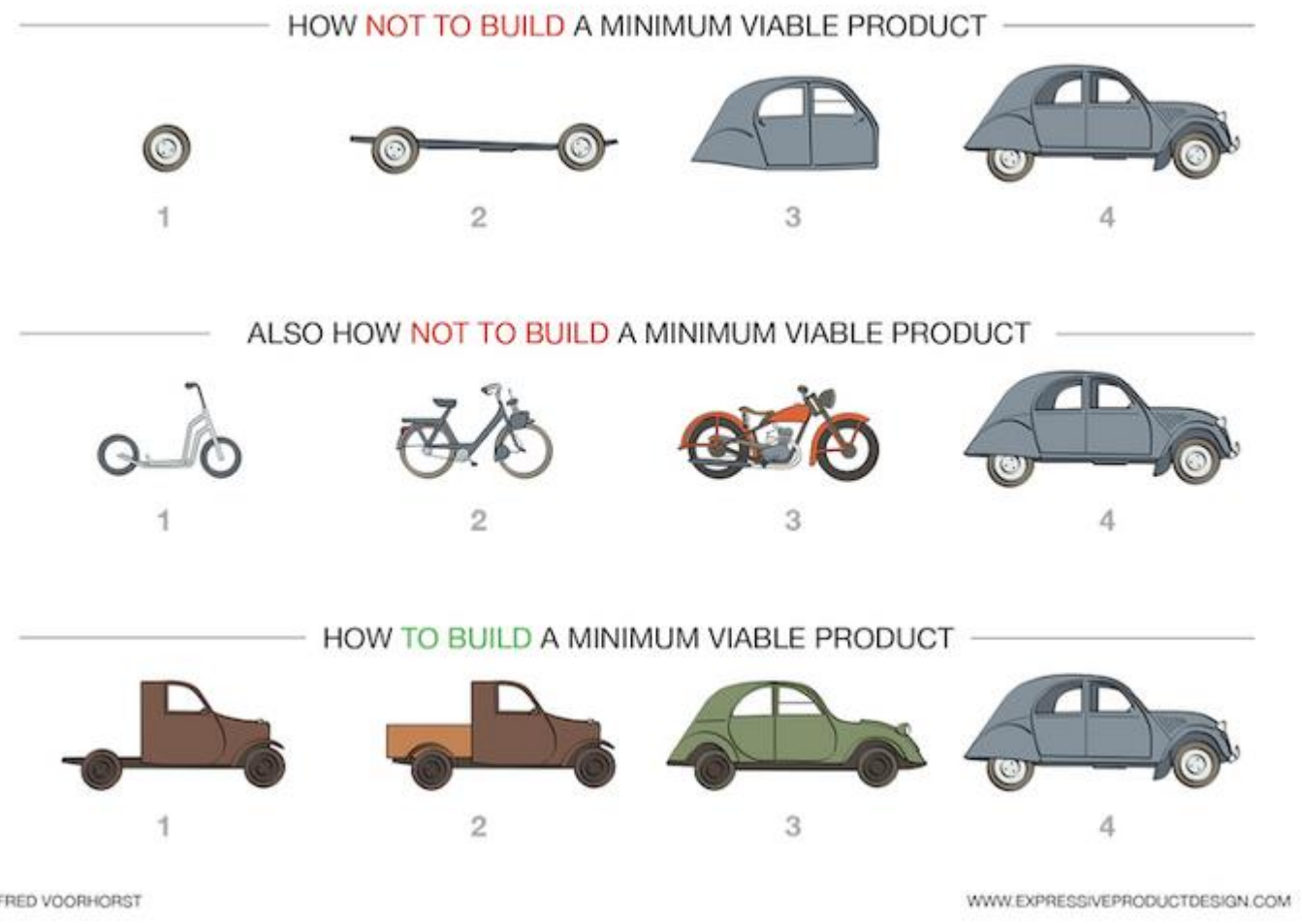
- Agilidad (Cultura)
- Gobierno de Datos + Gobierno de Modelos/Analytics
- Ops (Metodologías + Frameworks + Herramientas)
- Cambiar paradigmas:
  - Modelos como APIs
  - Federated Learning
  - Champion Challenger
  - Uplift Modeling
  - Edge Computing

Combine Design Thinking, Lean Startup and Agile



Gartner

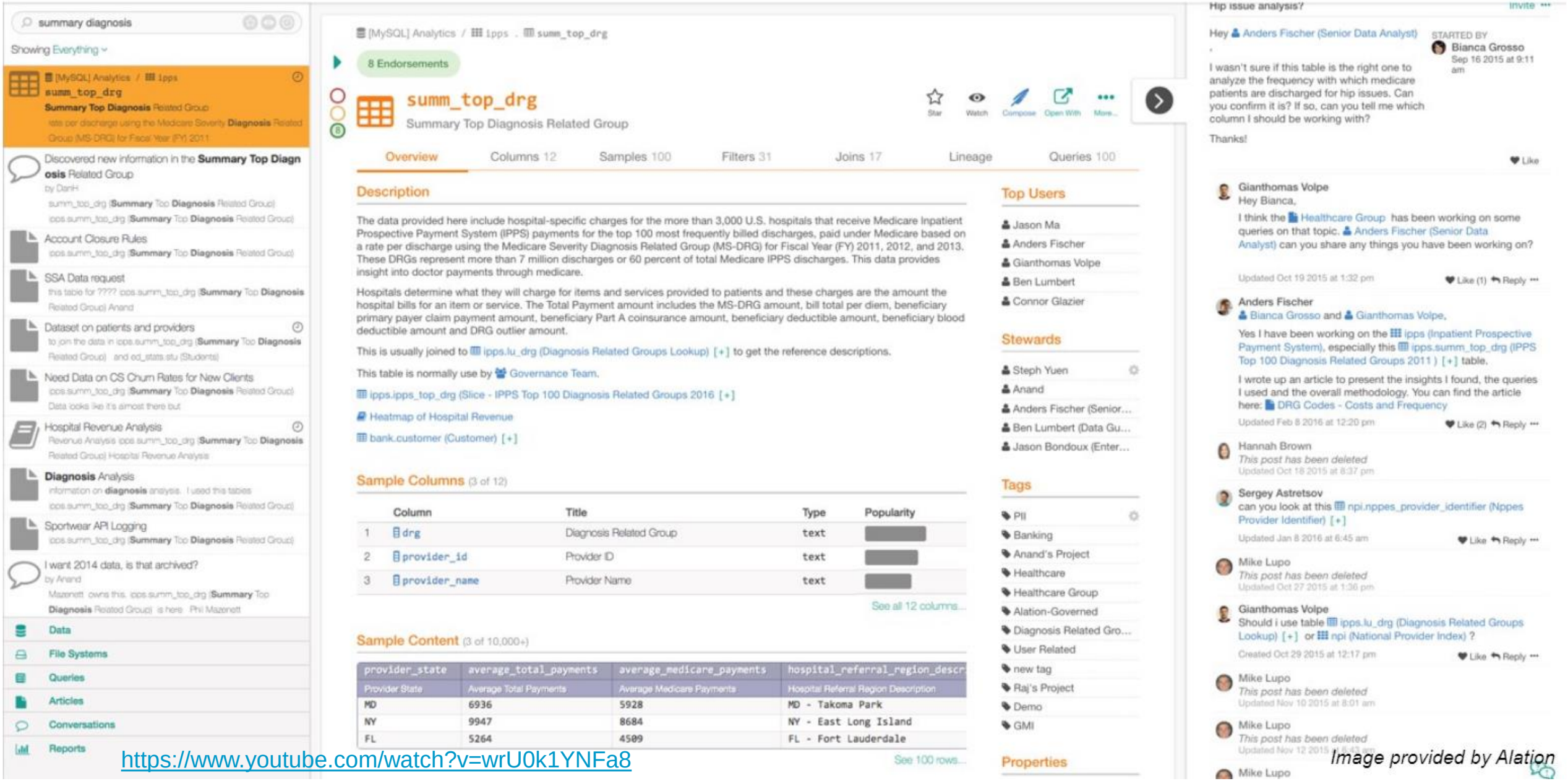
¿Cómo armar un MVP?



"An MVP is the simplest version of a product which already brings some clearly defined value and benefits to the user, can be tested and pushed forward to production. It contains compact functionality and quality design. It is reliable and easy to use."

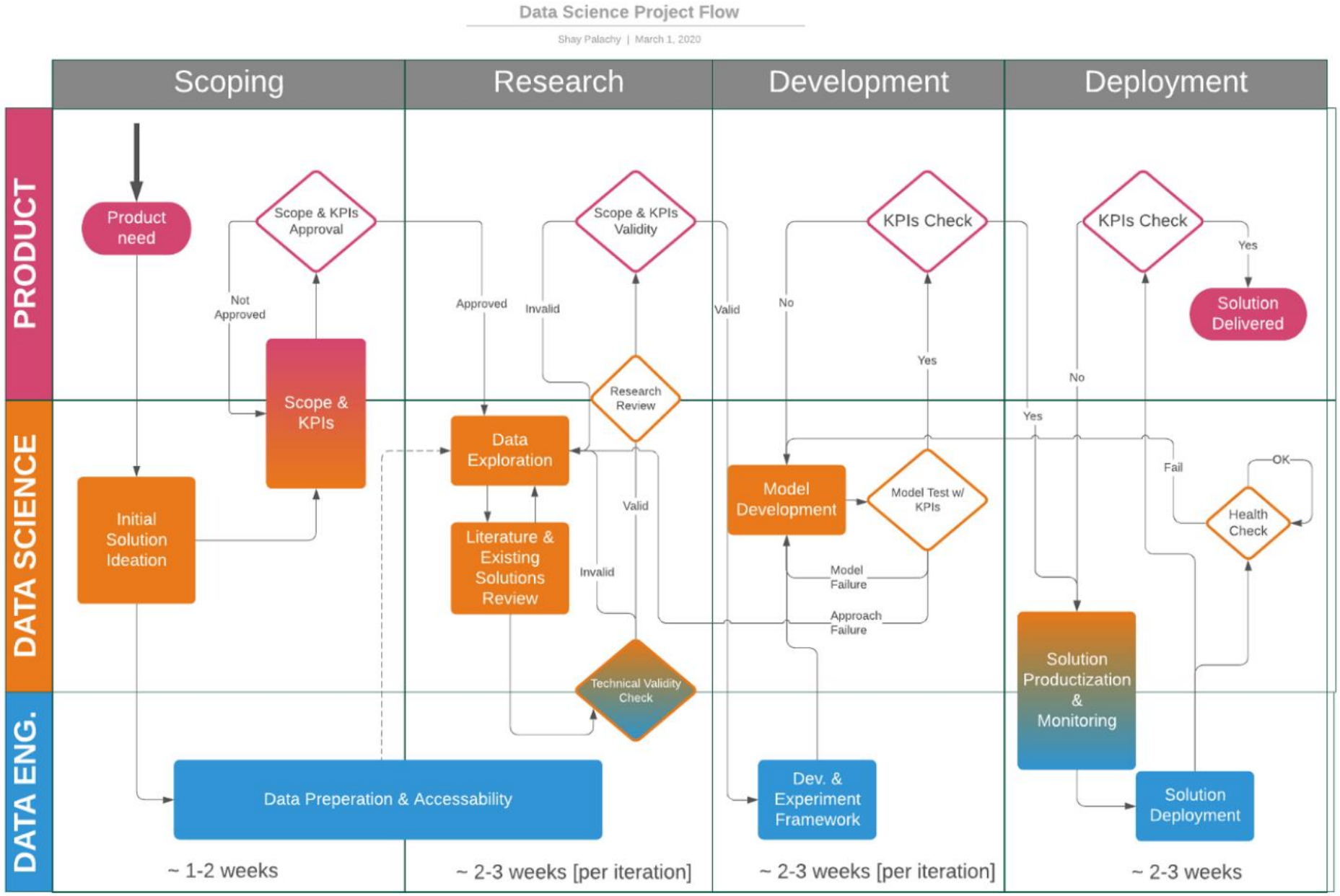
https://pixelfield.co.uk/blog/mvp-what-is-it-and-why-is-it-crucial-for-your-business/

Enterprise Data Catalog



https://www.youtube.com/watch?v=wrU0k1YNFa8

Image provided by Alation


Data Science Project Flow
Shay Palachy | March 1, 2020
Scoping
Research
Development
Deployment
PRODUCT
DATA SCIENCE
DATA ENG.
Product need
Scope & KPIs Approval
Not Approved
Scope & KPIs
Initial Solution Ideation
Approved
Scope & KPIs Validity
Invalid
Research Review
Data Exploration
Literature & Existing Solutions Review
Valid
Invalid
Technical Validity Check
Valid
KPIs Check
No
Yes
Model Development
Model Test w/ KPIs
Model Failure
Approach Failure
KPIs Check
Yes
Solution Delivered
No
Yes
Fail
OK
Health Check
Solution Productization & Monitoring
Solution Deployment
Data Preperation & Accessability
Dev. & Experiment Framework
~ 1-2 weeks
~ 2-3 weeks [per iteration]
~ 2-3 weeks [per iteration]
~ 2-3 weeks

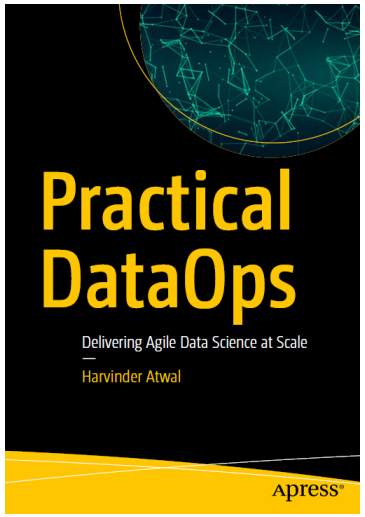
teradata.

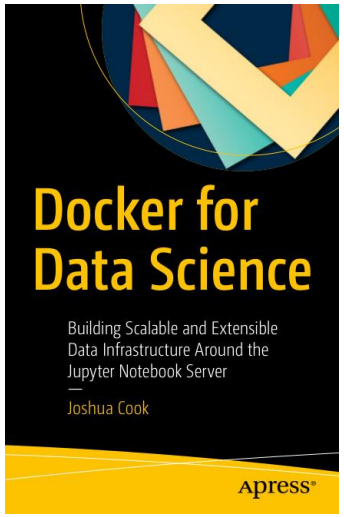
Beneficios Esperados

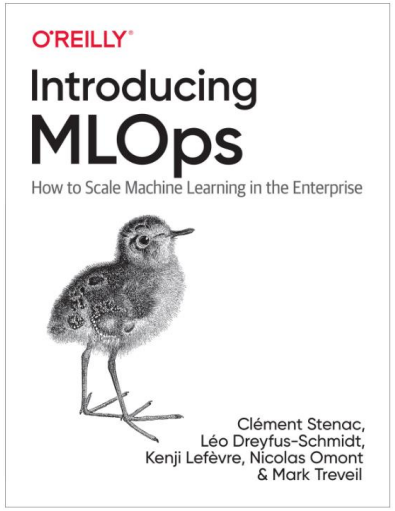
Antes
Proyectos largos – desarrollar sólo un modelo toma muchos meses
Despliegue manual – volátil, sujeto a errores, pérdidas y difícil de auditar
Falta de gobierno – a lo largo de todo el proceso analítico
Silos – Tecnología, Analytics y el Negocio trabajando en silos
Pobre confiabilidad y escalabilidad – Dificultades para hacer un proceso escalable y robusto

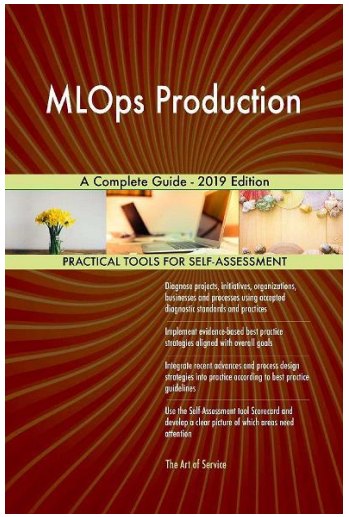
Después
Acelerar el time-to-market – desarrollo de modelos en menor tiempo
Automatización – en la construcción, testeo y pase a producción de modelos
Incremento de Gobernabilidad – mayor control y auditabilidad de los procesos
Trabajo colaborativo – todas las áreas alineadas para optimizar el despliegue
Infraestructura escalable – la solution capaz de evolucionar hacia un acceptable nivel de madurez

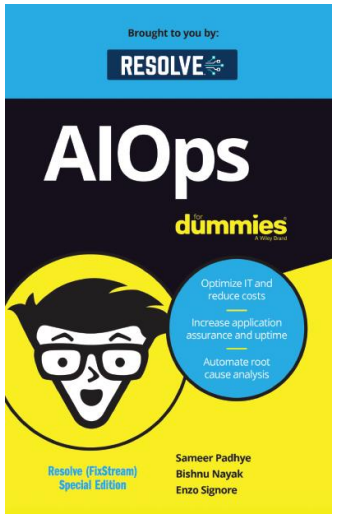
Y para más referencias…

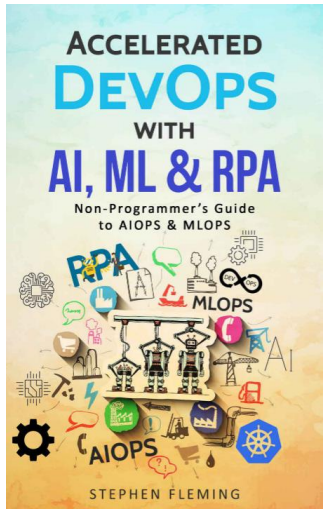


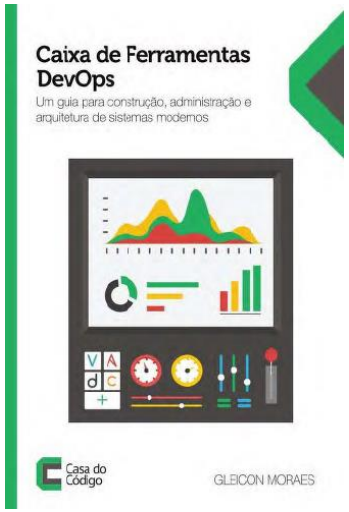


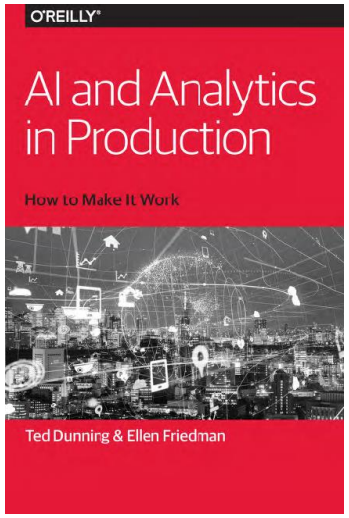


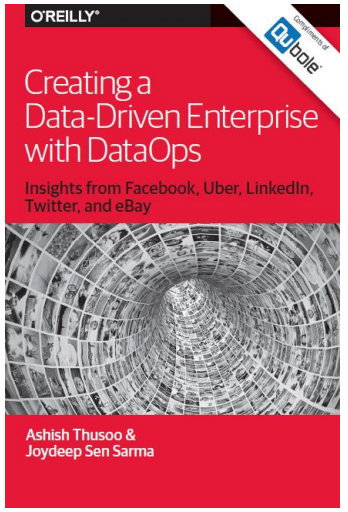












teradata.

Key Takeaways

1

No es suficiente utilizar herramientas DevOps para desarrollar e implementar modelos. Es necesario adoptar una nueva forma de hacer las cosas.

2

Busque experiencia y mejores prácticas. Hay bastantes recursos disponibles para aprender, pero es mejor acudir a quien ya alcanzó cierta madurez.

3

Es un buen momento para actualizar procesos, la agilidad para reaccionar a las innovaciones hará la diferencia de cara al futuro.

Gracias!

teradata.