



# DataOps, MLOps, AIOps...

## Why now?

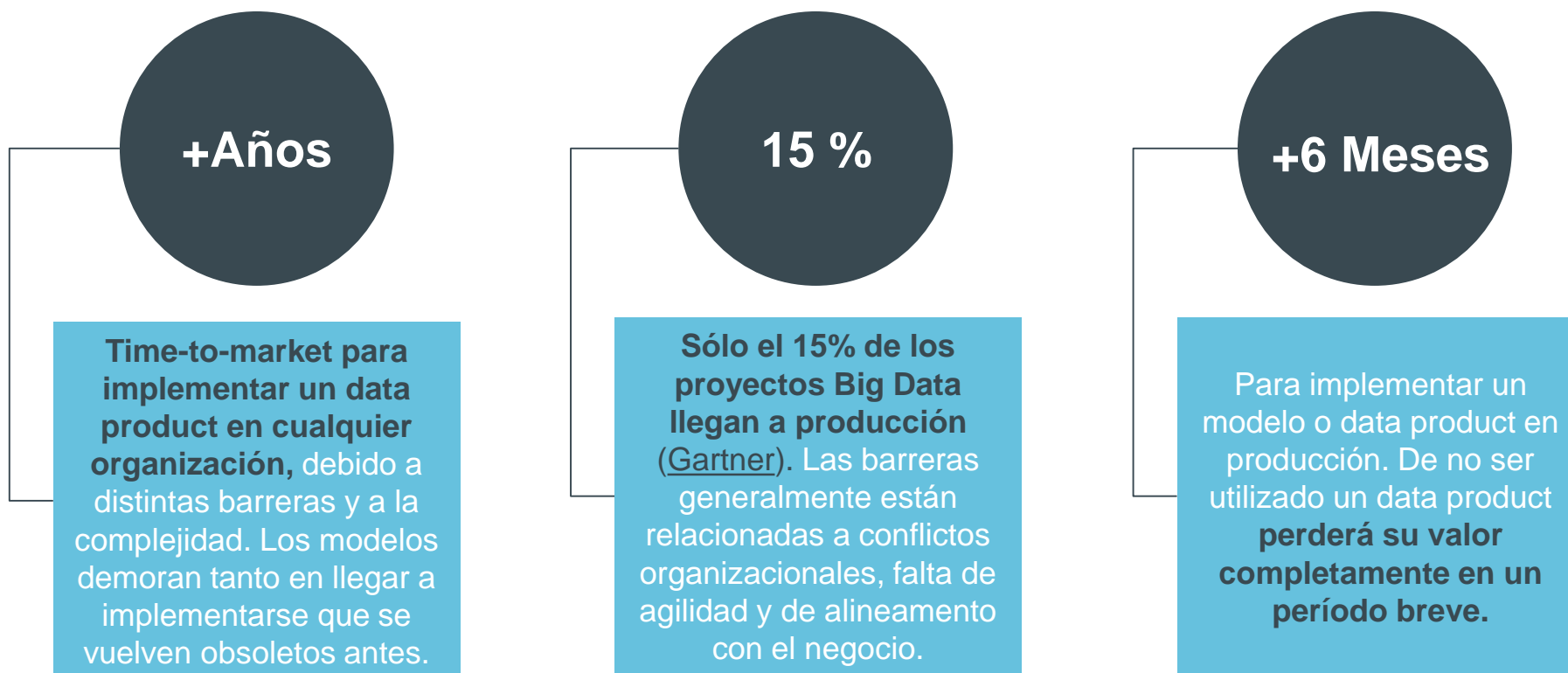
Luis Cajachahua

Senior Data Scientist

Center of Excellence, Teradata Americas

Julio 2020

# Problemas vs Oportunidades



# 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

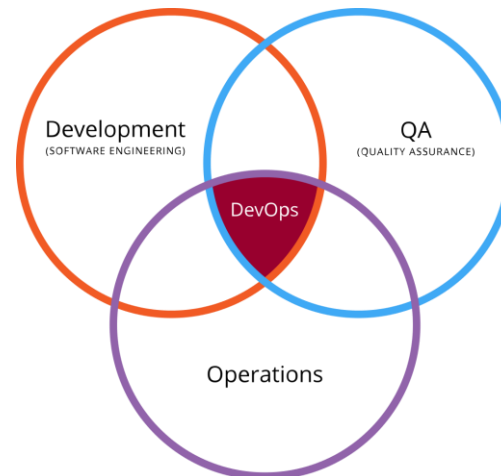
**Dev**elopers & **Op**erations trabajando juntos en equipo

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>



# Herramientas DevOps a través del Proceso

PLAN	SOURCE CODE	BUILD	CI	ANALYZE	REPOSITORY	TEST	RELEASE AND DEPLOY	CONFIG	CONTINUOUS INSPECTION MONITORING
									
									
									
									
									

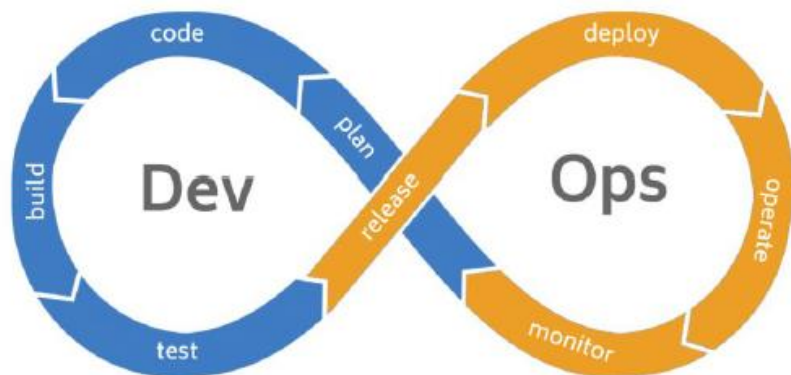
# PERO

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.

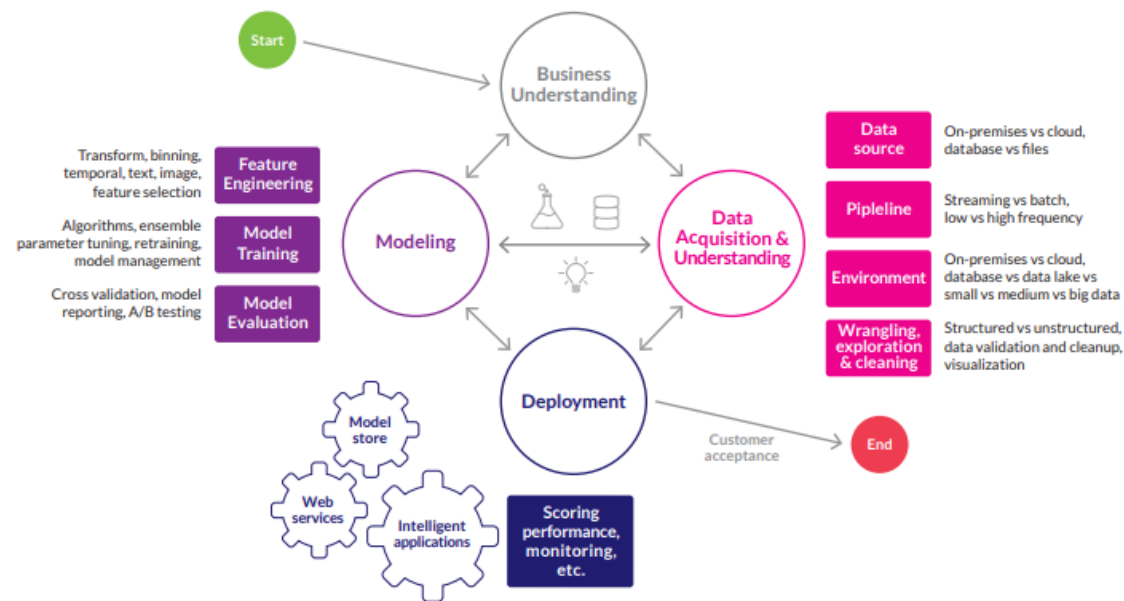
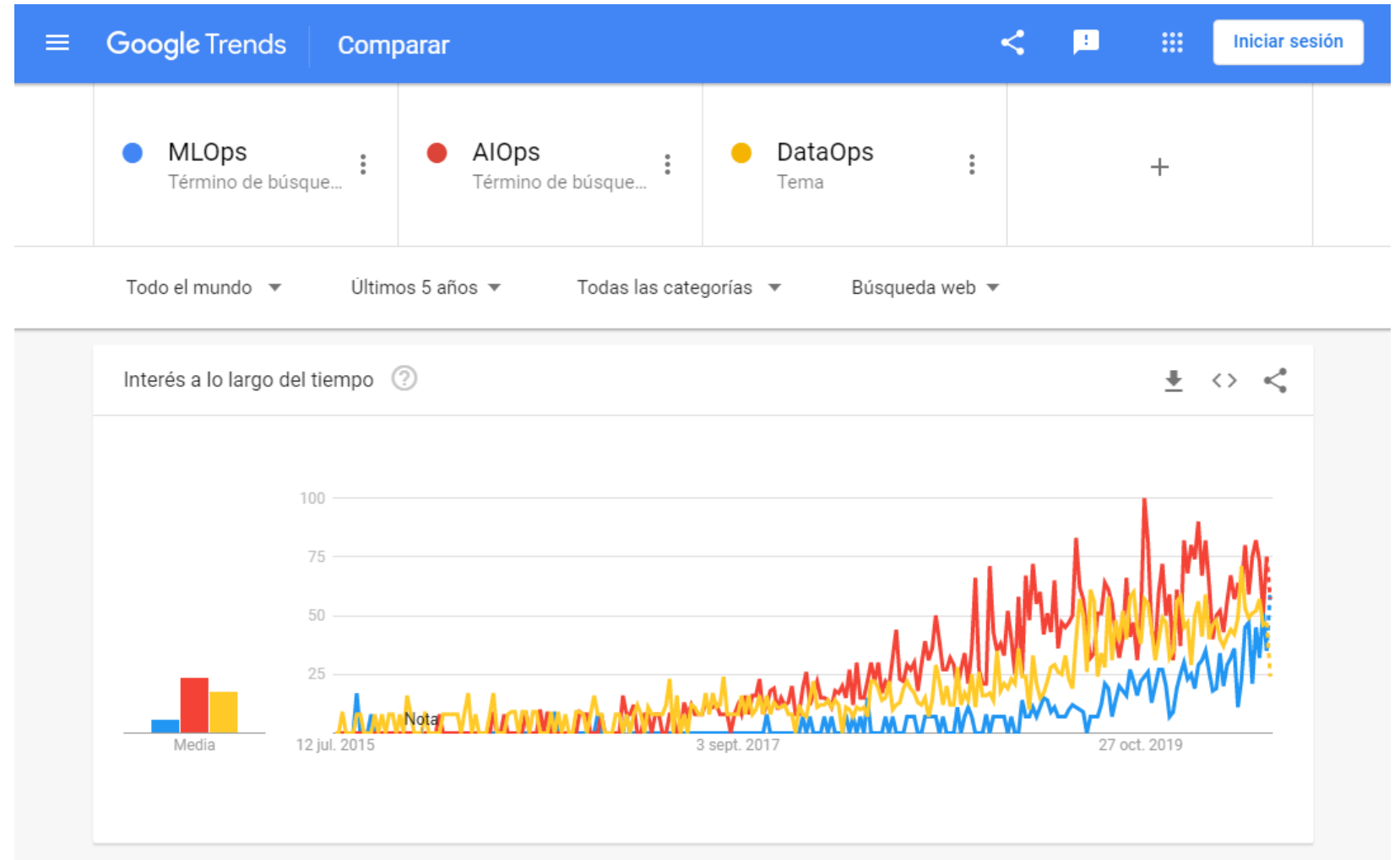


Figure 1. Data science lifecycle

# 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.





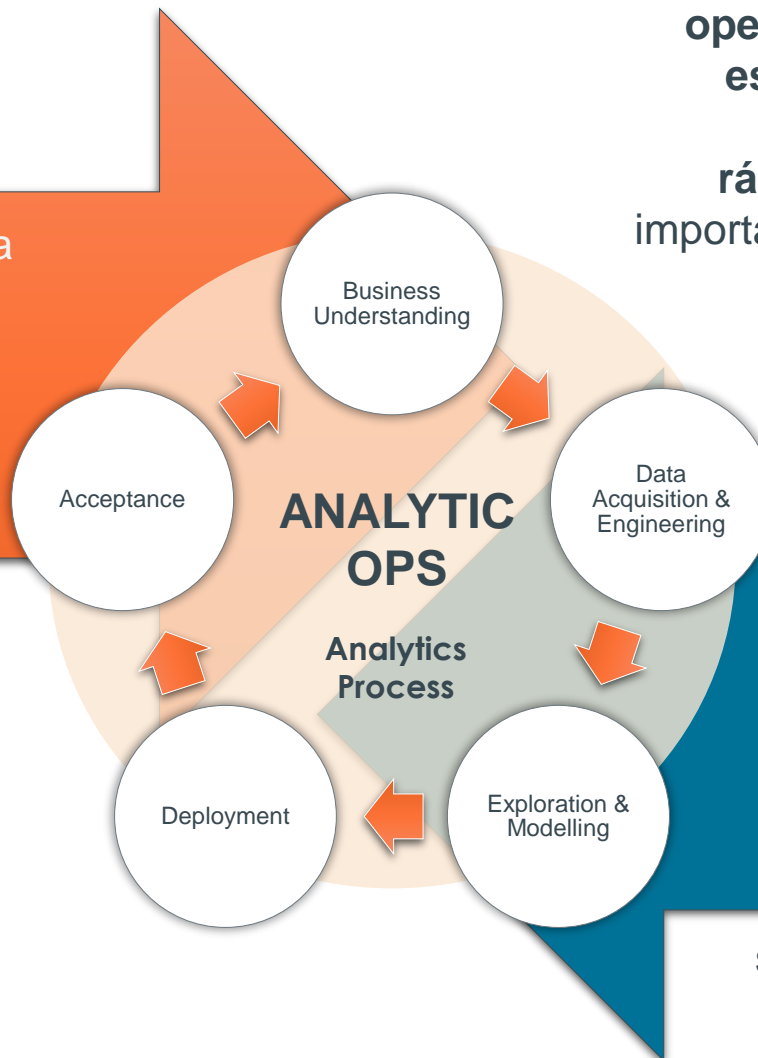
# ¿Y en qué consiste?

## ANALYTICS

- + flexibility
- exploration
- discovery
- modelling
- blue-sky ideation

- external data
- iteration
- data-mining
- statistics
- value-driven

Data Science Practices



**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**

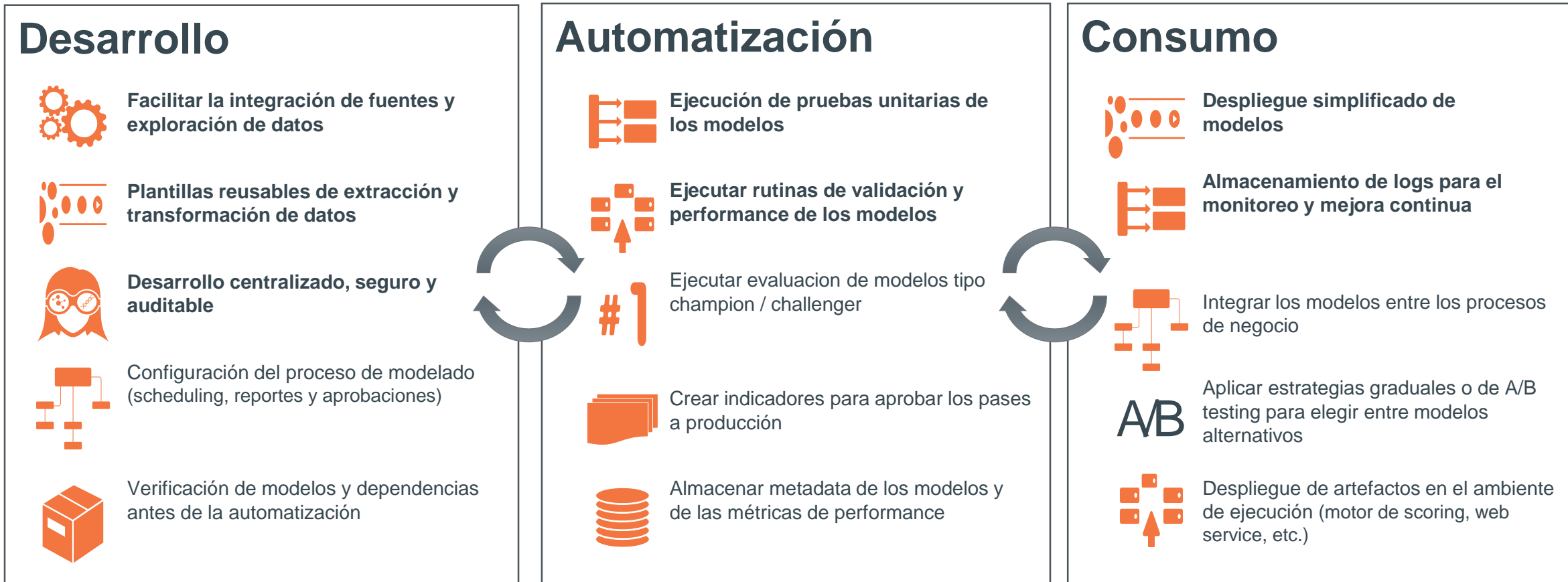
## OPERATIONS

- security
- governance
- compliance
- curation
- deployment

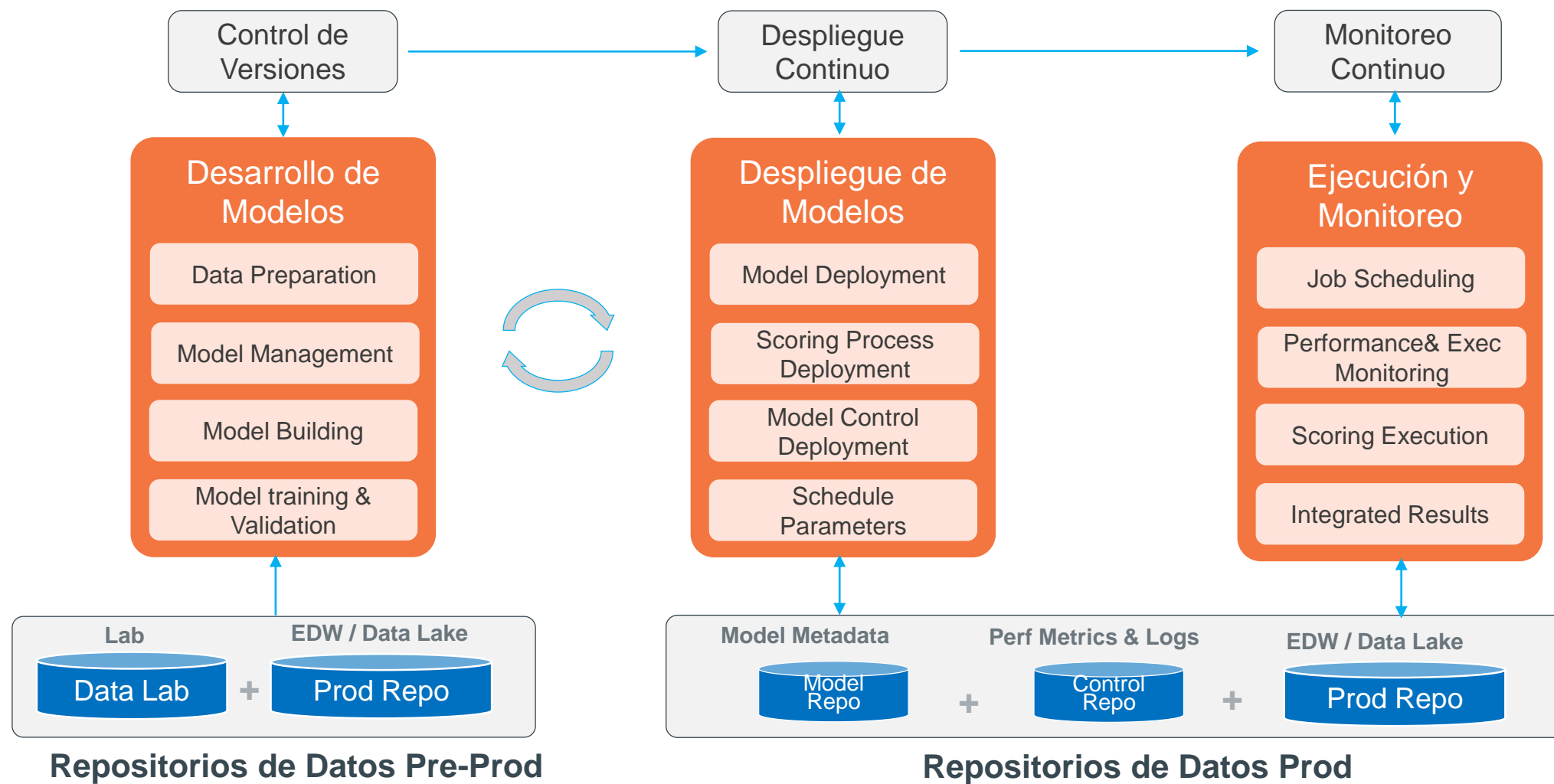
- maintenance
- integration
- testing
- engineering
- process-driven

Software Engineering Practices

# Los tres pilares de AnalyticOps



# Componentes de Arquitectura de AnalyticOps



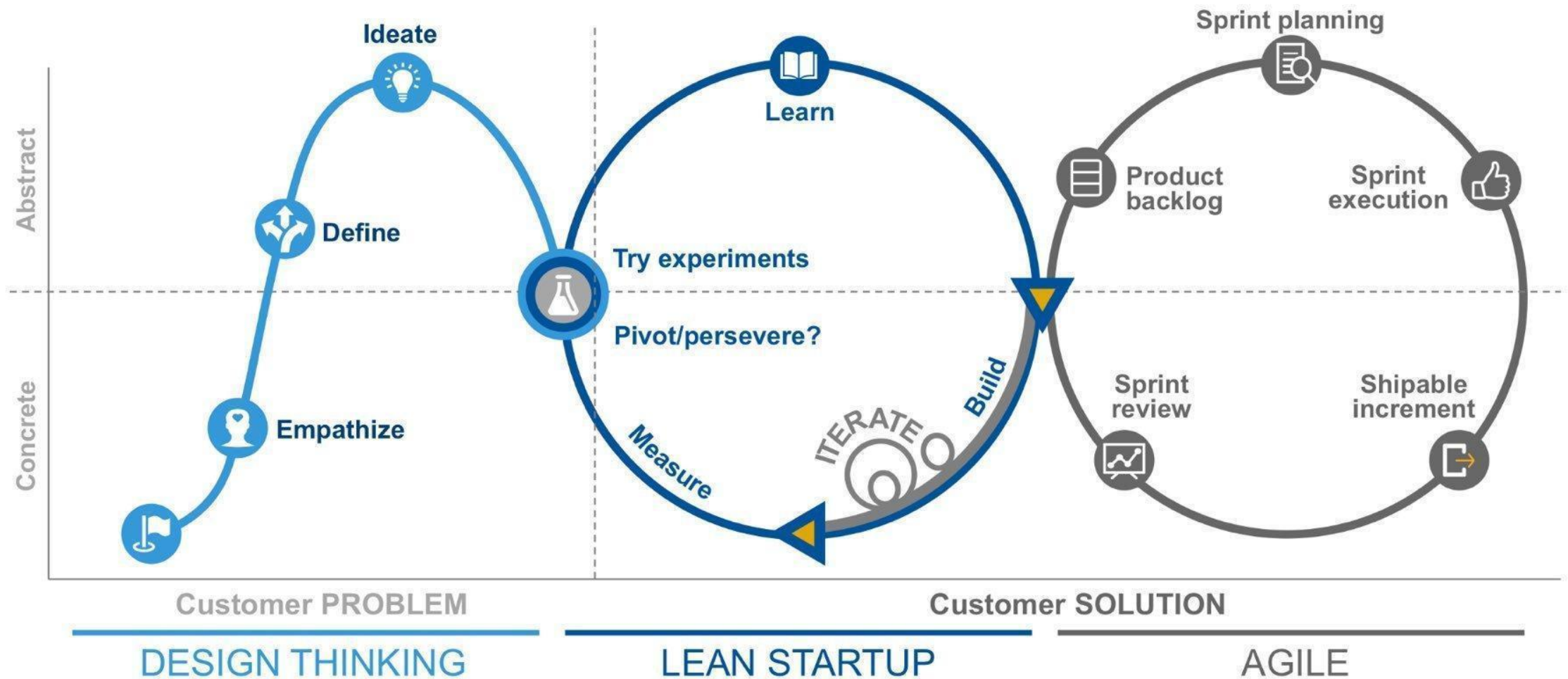
# Y RECUERDA QUE:

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



# ¿Cómo armar un MVP?

## HOW NOT TO BUILD A MINIMUM VIABLE PRODUCT



1



2



3



4

## ALSO HOW NOT TO BUILD A MINIMUM VIABLE PRODUCT



1



2



3



4

## HOW TO BUILD A MINIMUM VIABLE PRODUCT



1



2



3



4

“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.”



# Enterprise Data Catalog

summary diagnosis

Showing Everything

MySQL Analytics / ipps . summ\_top\_drg

Summary Top Diagnosis Related Group

rate per discharge using the Medicare Severity Diagnosis Related Group (MS-DRG) for Fiscal Year (FY) 2011

Discovered new information in the Summary Top Diagnosis Related Group

by DanH

summ\_top\_drg (Summary Top Diagnosis Related Group)

ipps.summ\_top\_drg (Summary Top Diagnosis Related Group)

Account Closure Rules

ipps.summ\_top\_drg (Summary Top Diagnosis Related Group)

SSA Data request

this table for ??? ipps.summ\_top\_drg (Summary Top Diagnosis Related Group) Anand

Dataset on patients and providers

to join the data in ipps.summ\_top\_drg (Summary Top Diagnosis Related Group) and ed\_stats.stu (Students)

Need Data on CS Churn Rates for New Clients

ipps.summ\_top\_drg (Summary Top Diagnosis Related Group)

Data looks like it's almost there but

Hospital Revenue Analysis

Revenue Analysis ipps.summ\_top\_drg (Summary Top Diagnosis Related Group) Hospital Revenue Analysis

Diagnosis Analysis

information on diagnosis analysis. I used this table

ipps.summ\_top\_drg (Summary Top Diagnosis Related Group)

Sportwear API Logging

ipps.summ\_top\_drg (Summary Top Diagnosis Related Group)

I want 2014 data, is that archived?

by Anand

Mazenett owns this ipps.summ\_top\_drg (Summary Top Diagnosis Related Group) is here Phil Mazenett

Data

File Systems

Queries

Articles

Conversations

Reports

MySQL Analytics / ipps . summ\_top\_drg

8 Endorsements

summ\_top\_drg

Summary Top Diagnosis Related Group

Overview

Columns 12

Samples 100

Filters 31

Joins 17

Lineage

Queries 100

Description

The data provided here include hospital-specific charges for the more than 3,000 U.S. hospitals that receive Medicare Inpatient Prospective Payment System (IPPS) payments for the top 100 most frequently billed discharges, paid under Medicare based on a rate per discharge using the Medicare Severity Diagnosis Related Group (MS-DRG) for Fiscal Year (FY) 2011, 2012, and 2013. These DRGs represent more than 7 million discharges or 60 percent of total Medicare IPPS discharges. This data provides insight into doctor payments through medicare.

Hospitals determine what they will charge for items and services provided to patients and these charges are the amount the hospital bills for an item or service. The Total Payment amount includes the MS-DRG amount, bill total per diem, beneficiary primary payer claim payment amount, beneficiary Part A coinsurance amount, beneficiary deductible amount, beneficiary blood deductible amount and DRG outlier amount.

This is usually joined to ipps.lu\_drg (Diagnosis Related Groups Lookup) [ + ] to get the reference descriptions.

This table is normally use by Governance Team.

ipps.ipps\_top\_drg (Slice - IPPS Top 100 Diagnosis Related Groups 2016 [ + ])

Heatmap of Hospital Revenue

bank.customer (Customer) [ + ]

Sample Columns (3 of 12)

Column	Title	Type	Popularity
1 drg	Diagnosis Related Group	text	
2 provider_id	Provider ID	text	
3 provider_name	Provider Name	text	

See all 12 columns...

Sample Content (3 of 10,000+)

provider_state	average_total_payments	average_medicare_payments	hospital_referral_region_descr
Provider State	Average Total Payments	Average Medicare Payments	Hospital Referral Region Description
MD	6936	5928	MD - Takoma Park
NY	9947	8684	NY - East Long Island
FL	5264	4509	FL - Fort Lauderdale

See 100 rows...

Top Users

Jason Ma

Anders Fischer

Gianthomas Volpe

Ben Lumbert

Connor Glazier

Stewards

Steph Yuen

Anand

Anders Fischer (Senior...

Ben Lumbert (Data Gu...

Jason Bondoux (Enter...

Tags

Pii

Banking

Anand's Project

Healthcare

Healthcare Group

Alation-Governed

Diagnosis Related Gro...

User Related

new tag

Raj's Project

Demo

GMI

Properties

Hip issue analysis?

Hey Anders Fischer (Senior Data Analyst)

STARTED BY Bianca Grosso Sep 16 2015 at 9:11 am

I wasn't sure if this table is the right one to analyze the frequency with which medicare patients are discharged for hip issues. Can you confirm it is? If so, can you tell me which column I should be working with?

Thanks!

Like

Gianthomas Volpe Hey Bianca, I think the Healthcare Group has been working on some queries on that topic. Anders Fischer (Senior Data Analyst) can you share any things you have been working on?

Updated Oct 19 2015 at 1:32 pm Like (1) Reply

Anders Fischer Bianca Grosso and Gianthomas Volpe, Yes I have been working on the ipps (Inpatient Prospective Payment System), especially this ipps.summ\_top\_drg (IPPS Top 100 Diagnosis Related Groups 2011) [ + ] table. I wrote up an article to present the insights I found, the queries I used and the overall methodology. You can find the article here: DRG Codes - Costs and Frequency Updated Feb 8 2016 at 12:20 pm Like (2) Reply

Hannah Brown This post has been deleted Updated Oct 18 2015 at 8:37 pm

Sergey Astretsov can you look at this npi.nppes\_provider\_identifier (Nppes Provider Identifier) [ + ] Updated Jan 8 2016 at 6:45 am Like Reply

Mike Lupo This post has been deleted Updated Oct 27 2015 at 1:36 pm

Gianthomas Volpe Should i use table ipps.lu\_drg (Diagnosis Related Groups Lookup) [ + ] or npi (National Provider Index) ? Created Oct 29 2015 at 12:17 pm Like Reply

Mike Lupo This post has been deleted Updated Nov 10 2015 at 8:01 am

Mike Lupo This post has been deleted Updated Nov 12 2015 at 6:43 am

Mike Lupo

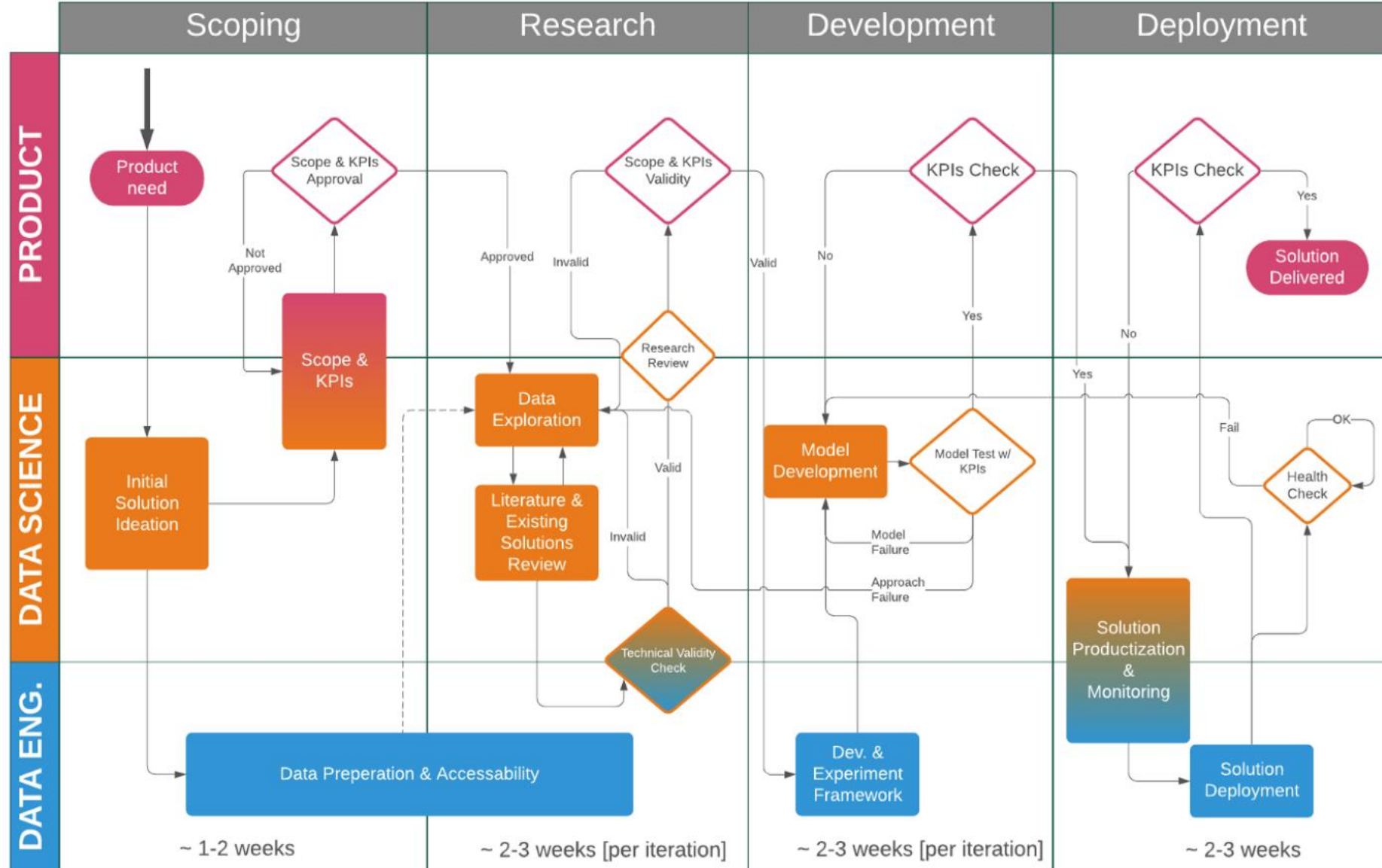
<https://www.youtube.com/watch?v=wrU0k1YNFa8>

Image provided by Alation



# Data Science Project Flow

Shay Palachy | March 1, 2020



# Beneficios Esperados

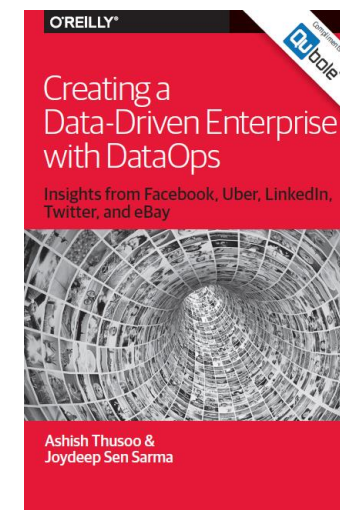
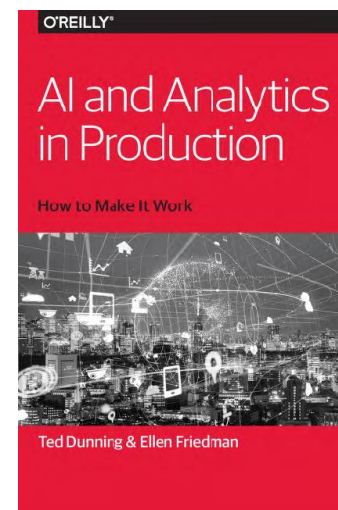
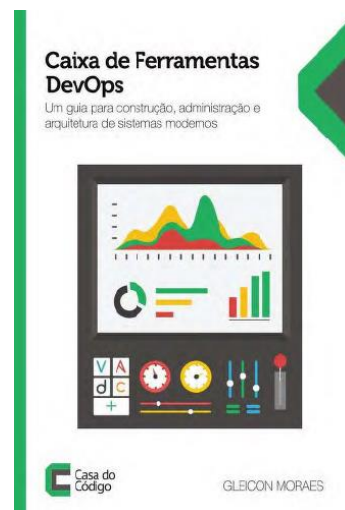
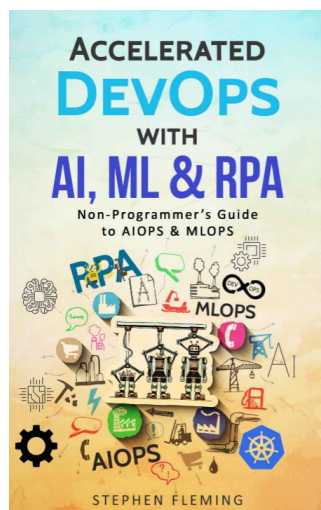
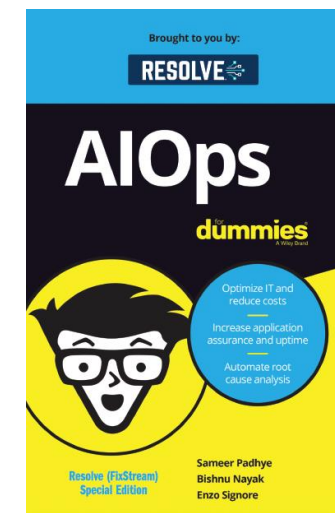
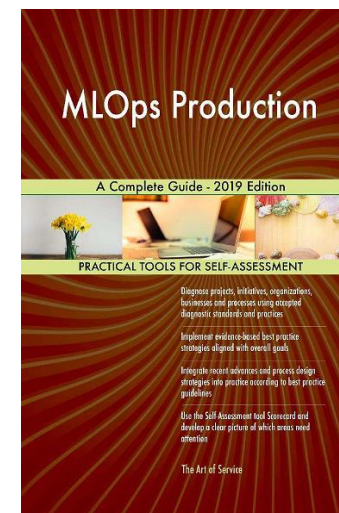
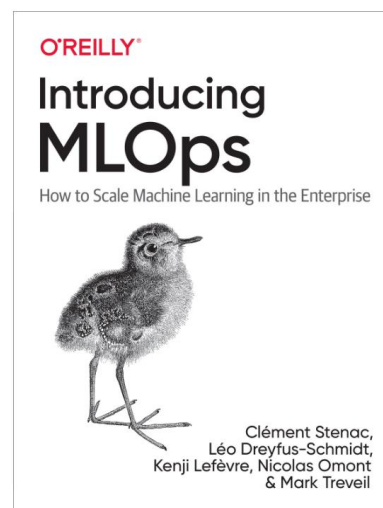
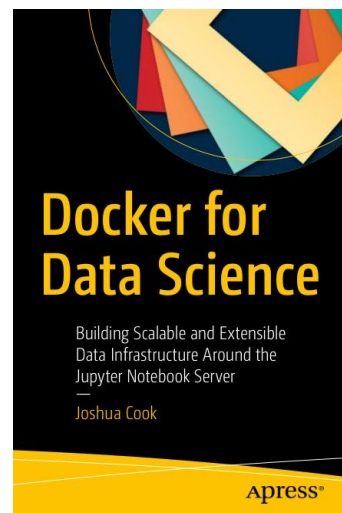
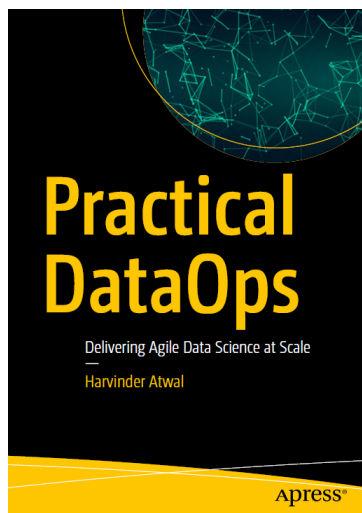


Antes
<b>Proyectos largos</b> – desarrollar sólo un modelo toma muchos meses
<b>Despliegue manual</b> – volátil, sujeto a errores, pérdidas y difícil de auditar
<b>Falta de gobierno</b> – a lo largo de todo el proceso analítico
<b>Silos</b> – Tecnología, Analytics y el Negocio trabajando en silos
<b>Pobre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



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

# Y para más referencias...



# 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.

©2020 Teradata