

Дата-инженер

# Аналитика данных

Ася Гайламазян



# Изменчивые данные, согласованность и визуализация

## Variability — изменчивость данных:

- значение одних и тех же данных может различаться в зависимости от контекста;
- алгоритмы должны быть в состоянии понять контекст и расшифровать точное значение слова в этом контексте.

## Veracity — достоверность данных:

- любой анализ данных бесполезен, если данные недостоверны;
- неточность данных может привести к неправильным решениям.

## Visualization — визуализация данных:

- визуализация делает большие данные доступными для человеческого восприятия;
- визуализация больших объемов сложных данных понятнее для человека, чем электронные таблицы и отчеты.

# Изменчивость данных



VARIABILITY

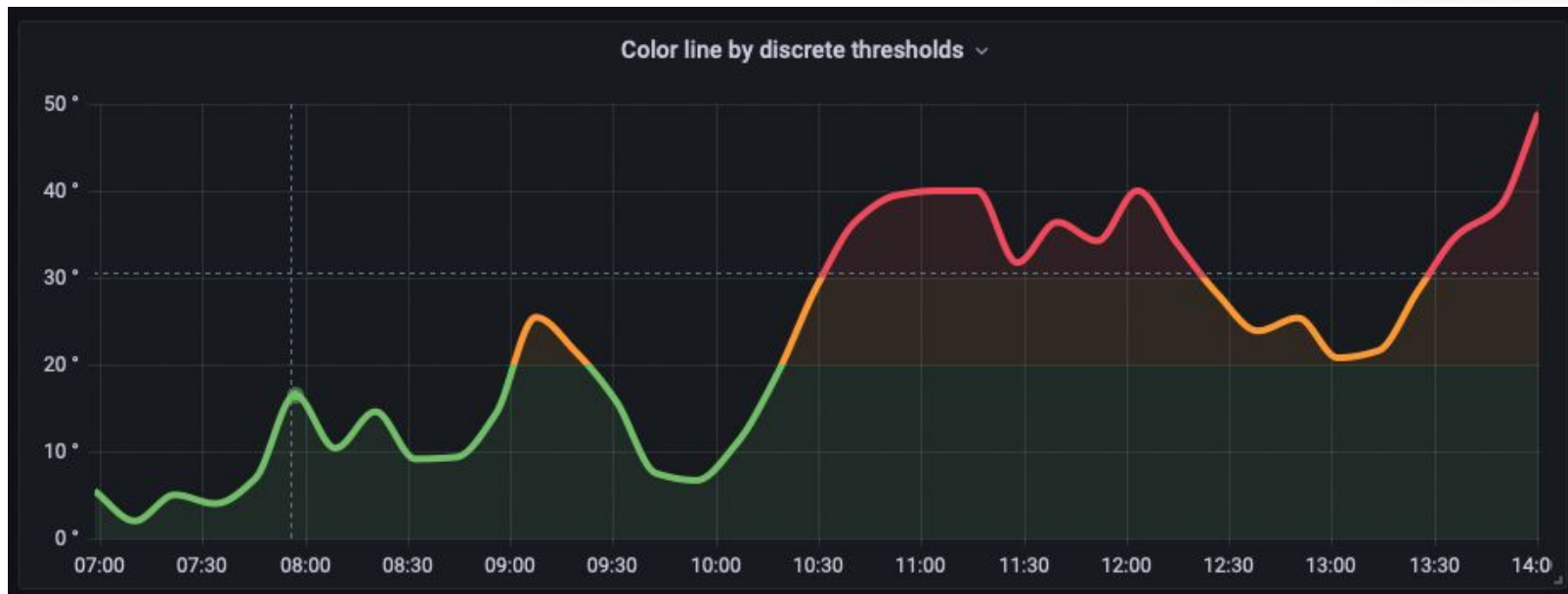
1. Системы потоковой обработки данных.
2. Агрегация и хранение данных.
3. Мониторинг и управление качеством данных.
4. Анализ и прогнозирование

# Эффективная обработка изменчивых данных

1. Непрерывный мониторинг.
2. Установите стандарты качества данных.
3. Используйте статистические методы для анализа и понимания изменчивости данных.
4. Рассмотрите возможность сглаживания или агрегирования данных.
5. Установите надежные процессы обеспечения качества для проверки входящих данных.
6. Используйте машинное обучение.
7. Храните подробную документацию об источниках данных, методах сбора и любой контекстной информации.
8. Рассмотрите возможность использования методов, устойчивых к изменениям или тенденциям в данных.



# Эффективная обработка изменчивых данных











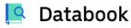


















# Примеры изменчивых данных

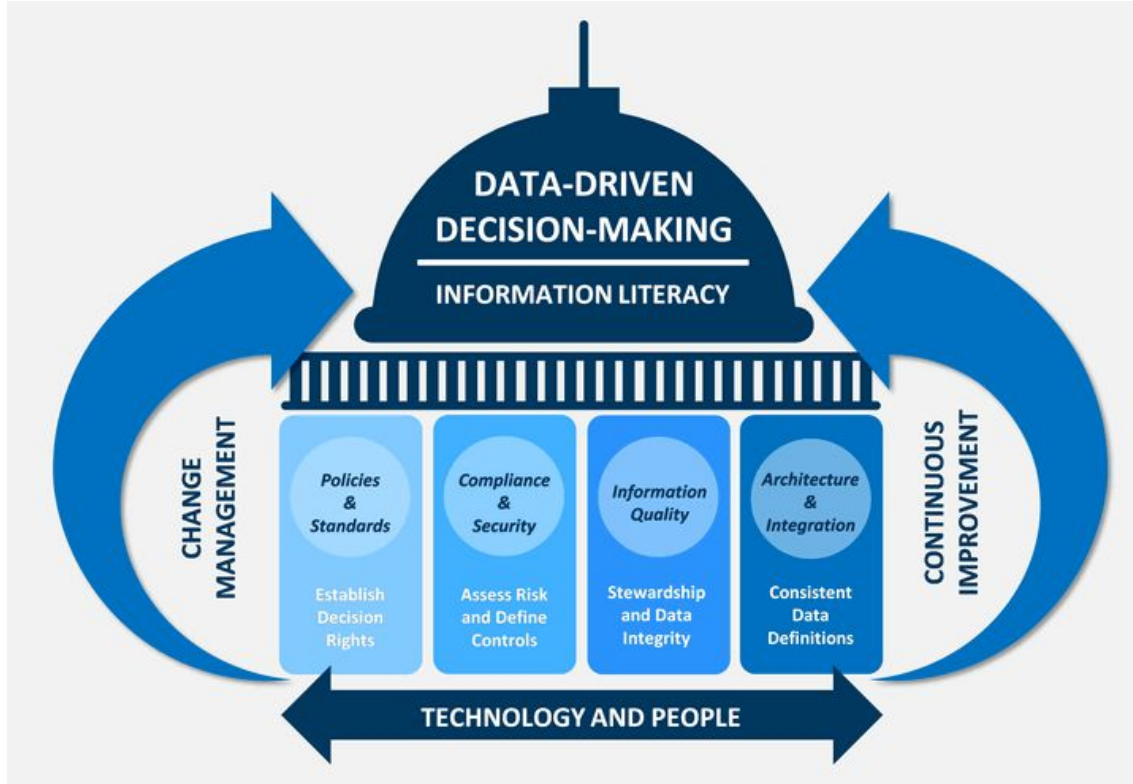


# Каталоги данных

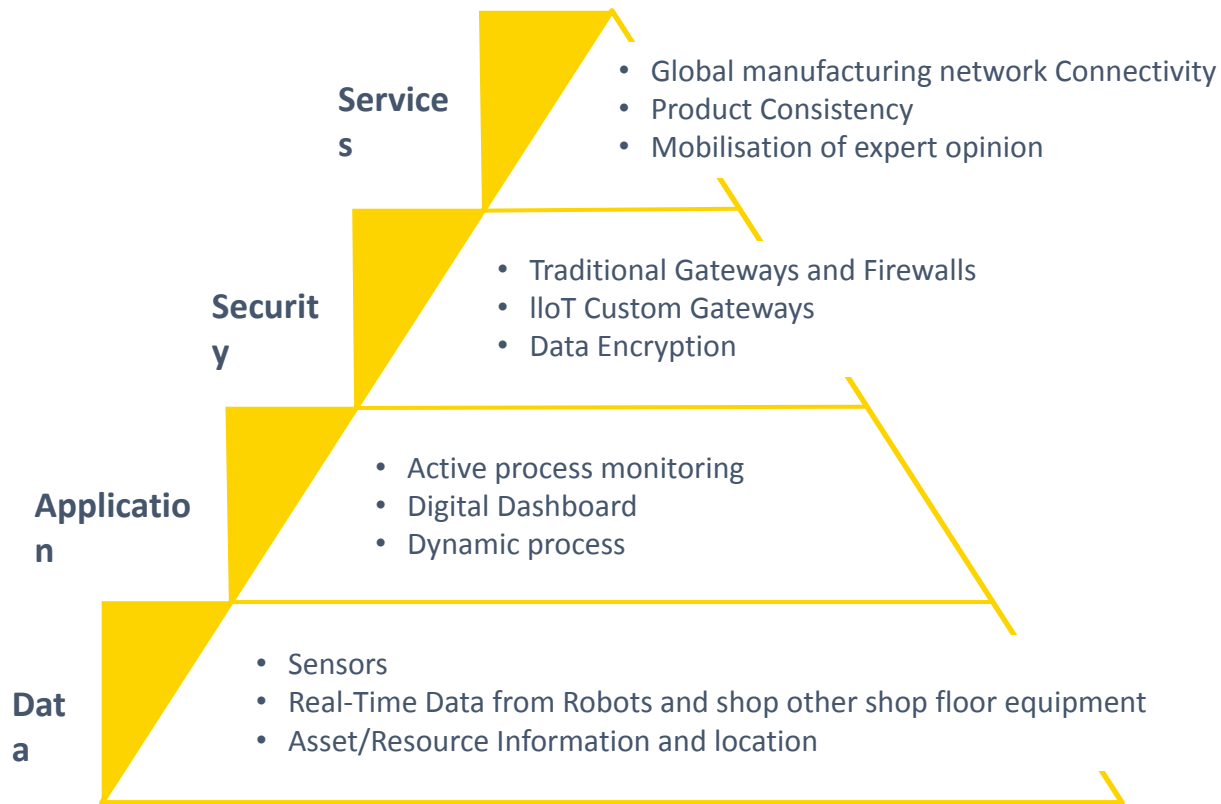
## Catalog Of Data Catalogs(and Data Discoverability Tools)

A	B	C	D	E	G	I	L
   		  	  				
M	O	Q	R	S	T	Z	What did I miss?
 	  			 	 		

# Data governance



# Согласованность данных



# Жизненный цикл данных

