SQL Norte for 

Carlos Augusto Oeiras

Introduction - Speaker

- Carlos Oeiras - camoeiras@live.com
  - Data Scientist and Data Professional

Agenda

9:30am Introduction to data analytics with Databricks

10:30am Making predictions based on complex neural network data in Azure

11:45am Q&A

12:00am Interval

Introduction to data analytics with Databricks

Current state of analytics

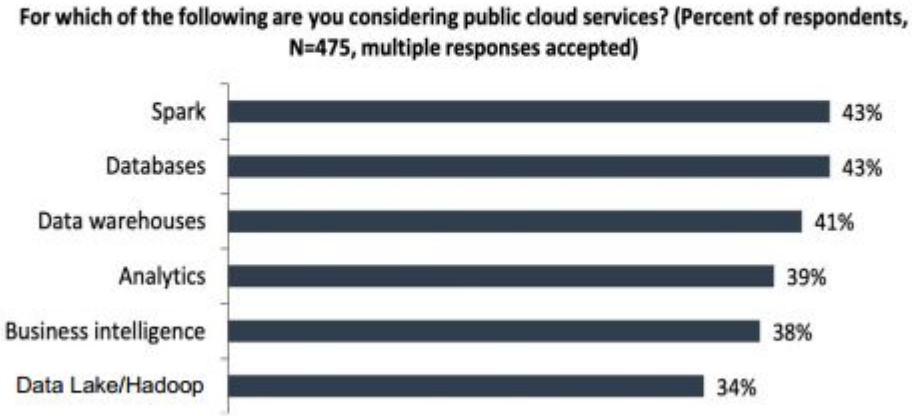
We have more data than ever before, often locked in systems and formats where we can't access it for analytics

On premises systems require excess capacity to process peak needs

Data security is costly and experts are hard to find

Analytic platforms are moving to the cloud

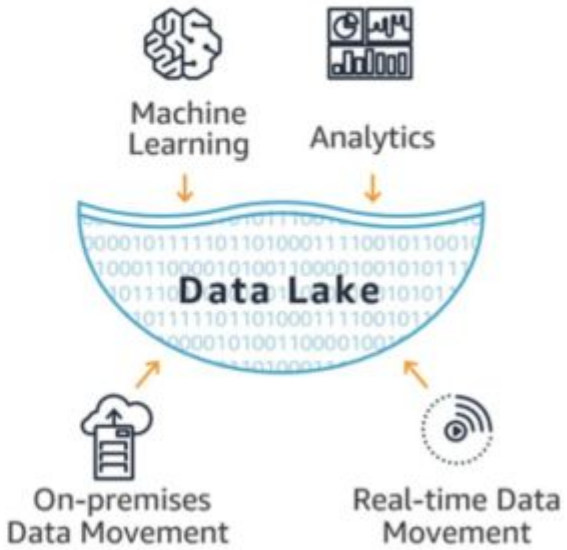
Figure 3. Migration to Cloud for Data Management Services



Source: Enterprise Strategy Group, 2017

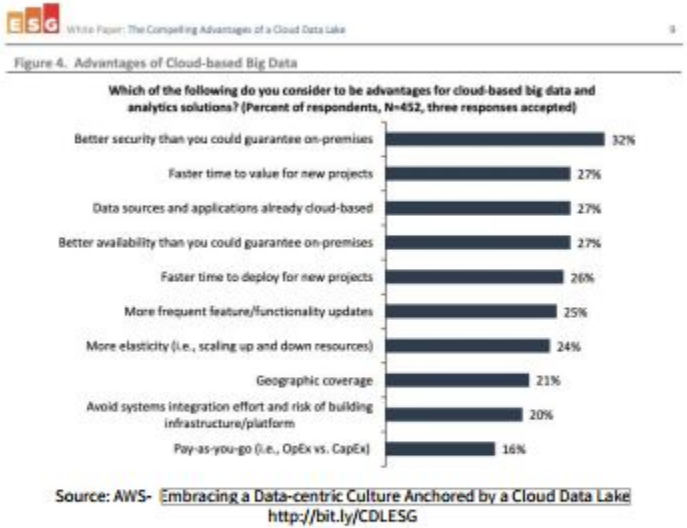
What is a data lake?

A data lake is a centralized repository that allows you to store all your structured and unstructured data at any scale. Organizations see this as their golden repository.



Benefits of cloud data lakes

- Pervasive security features
- Performance and scalability
- Reliability and availability
- Economics
- Integration
- Agility



Source: AWS- Embracing a Data-centric Culture Anchored by a Cloud Data Lake
http://bit.ly/CDLESG

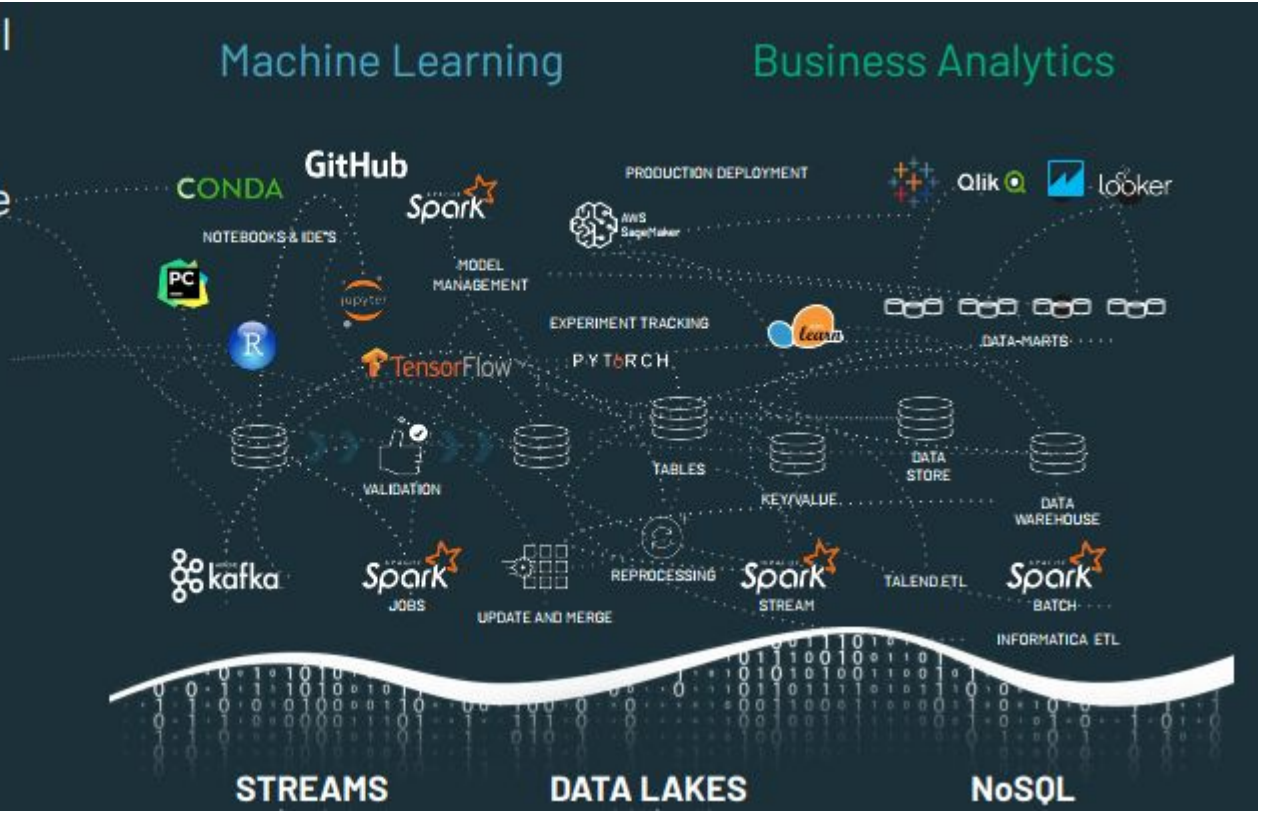
For organizations that are running analytics and machine learning workloads, the cloud makes the most sense, providing the balance of infinite scale and pay by use.

Unified Data Analytics Platform

Accelerating innovation across data science, data engineering, and business analytics

- Global company with 5,000 customers and 450+ partners
- Original creators of popular data and machine learning open source projects
- Empowering data teams to solve the world's toughest problems

Most organizations fail to unlock business value due to data, technology and people silos



Unlocking business value: Four challenges

1

Data is messy, siloed and slow

2

ML is hard, Production is harder

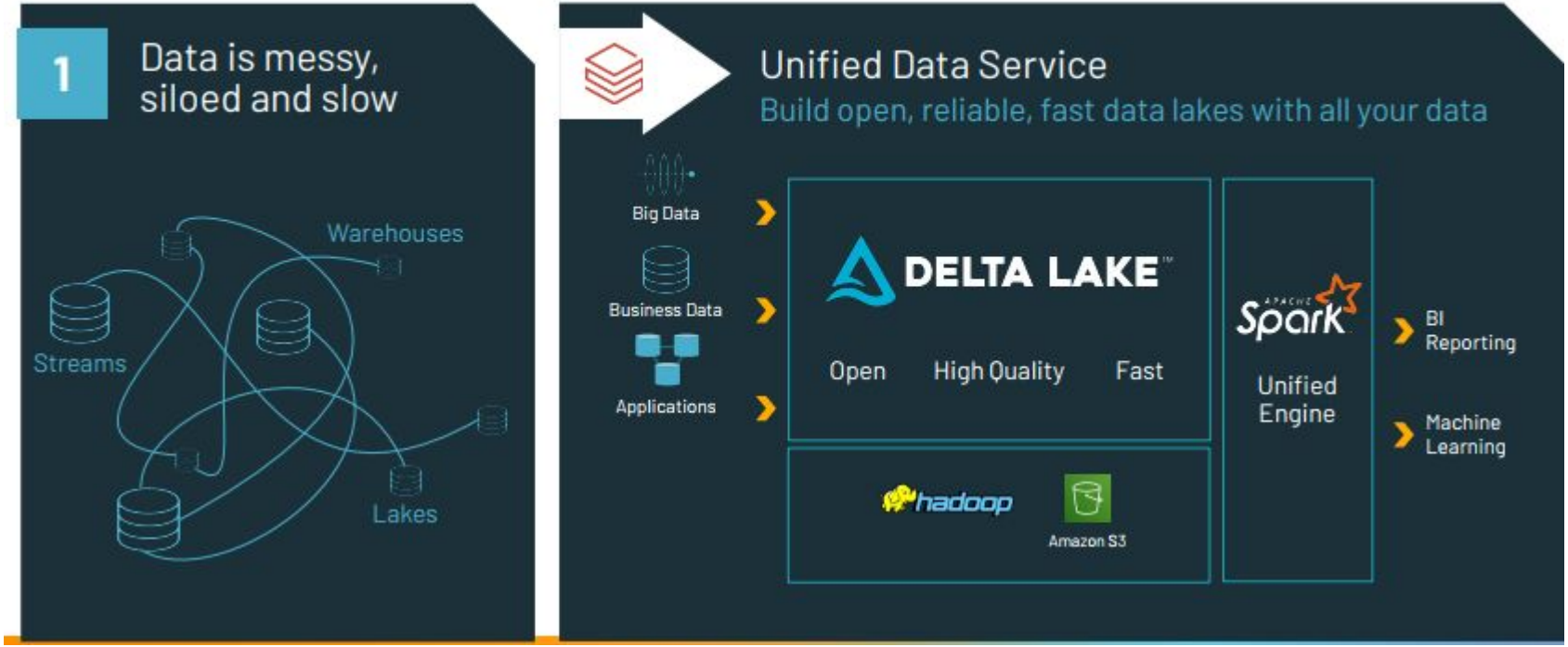
3

BI is limited to a fraction of data

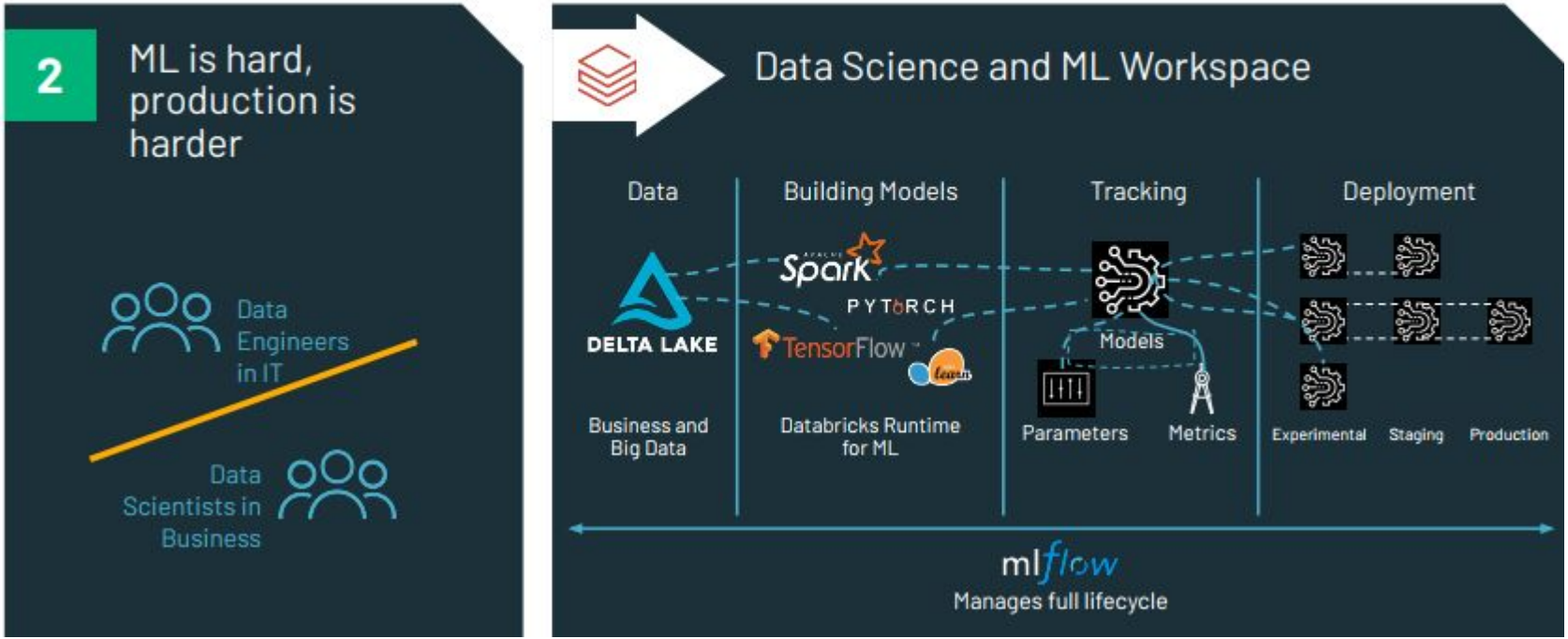
4

Lack of enterprise readiness

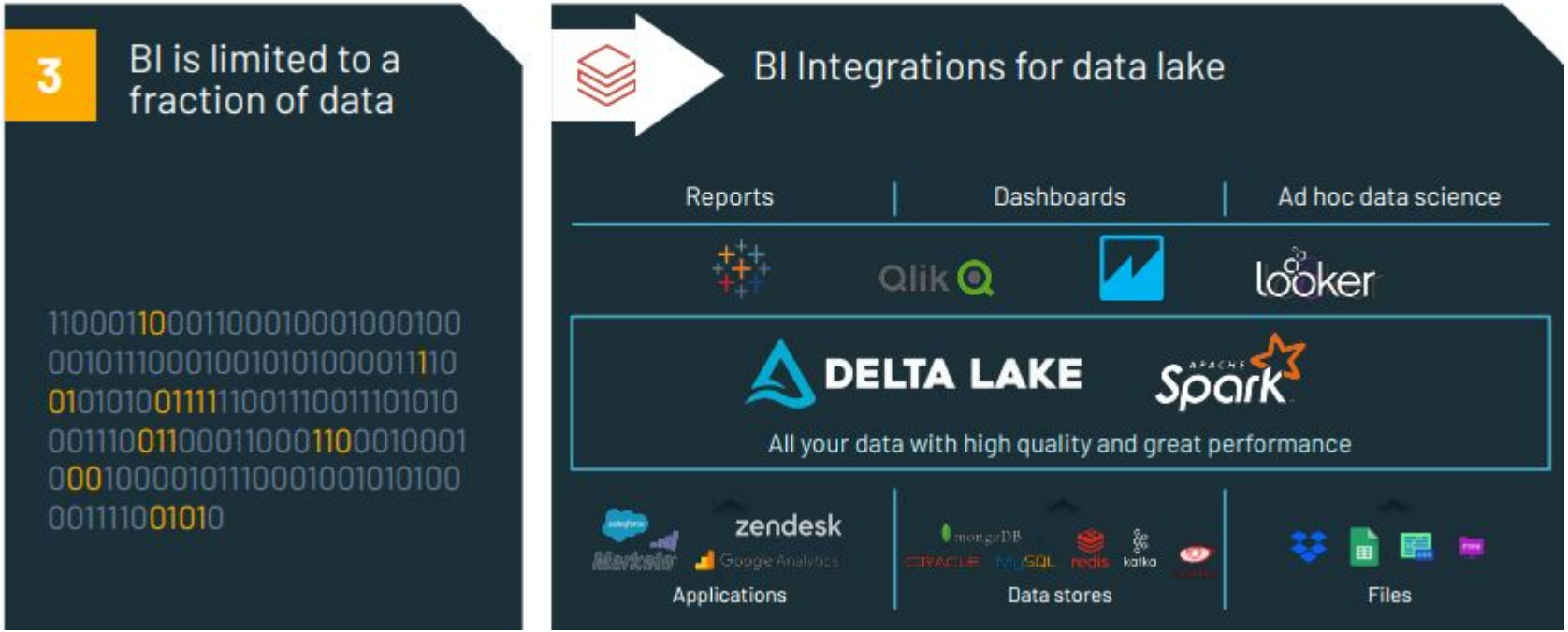
Make all your data ready for BI and ML



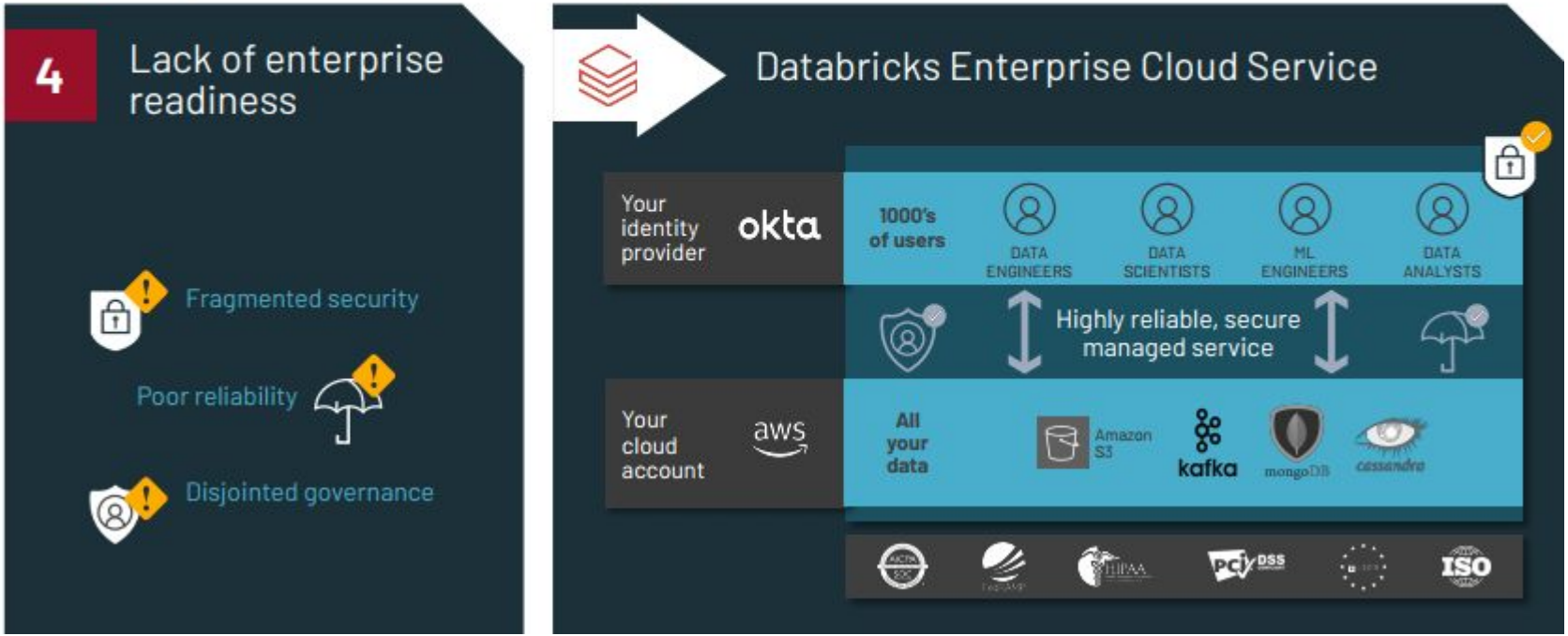
Unify data and ML across the full lifecycle



Enable BI directly on all your source data



Leverage cloud native platform for enterprise grade solution



Databricks Unified Data Analytics Platform

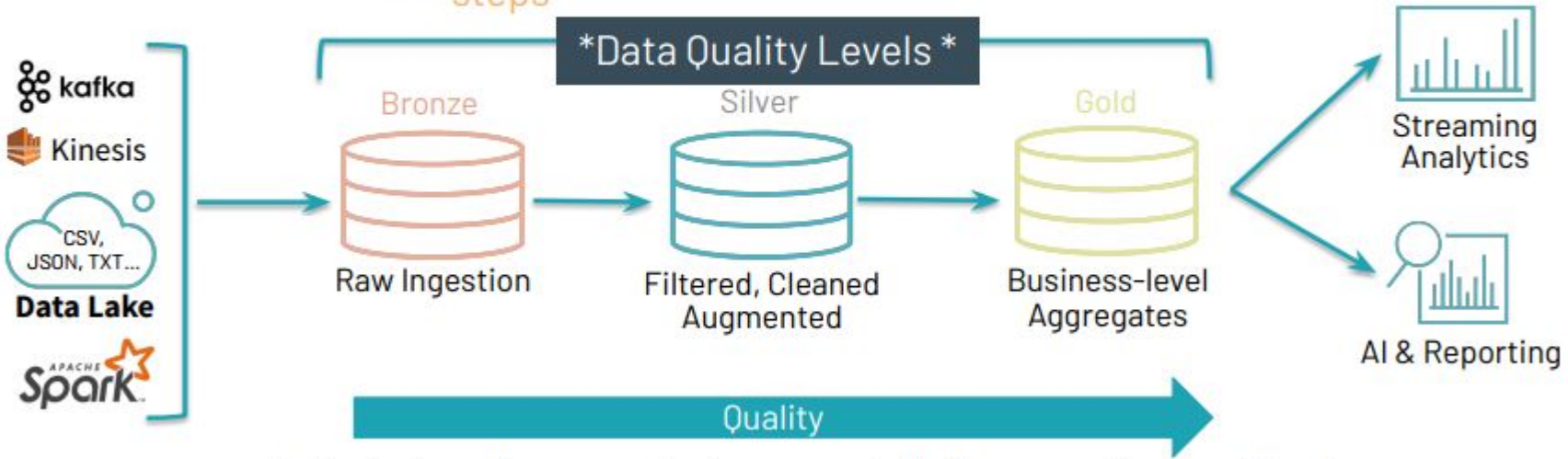
Data Science, ML, and BI on one cloud platform

Access all business and big data in **open data lake**

Securely integrates with your **cloud ecosystem**

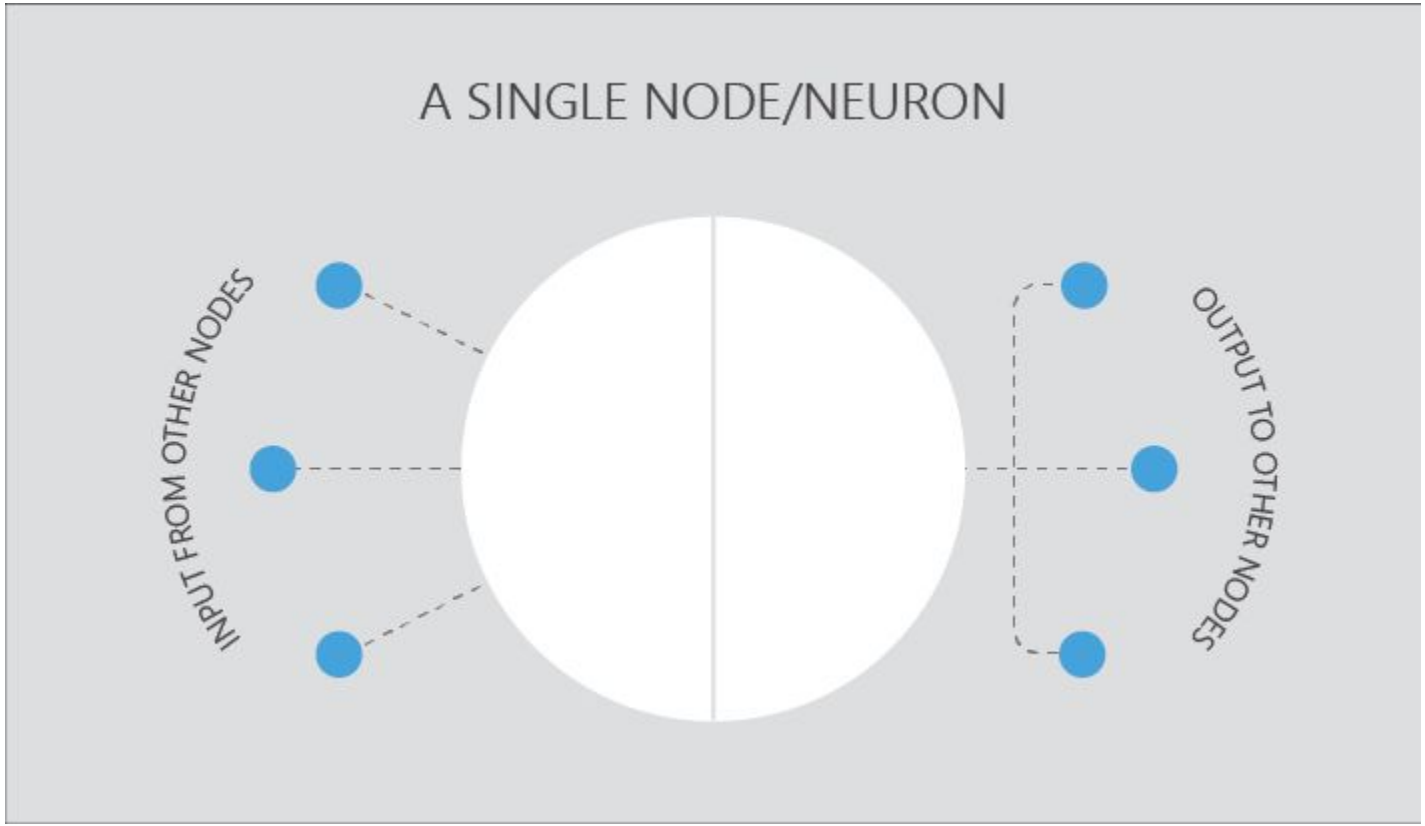
The DELTA LAKE

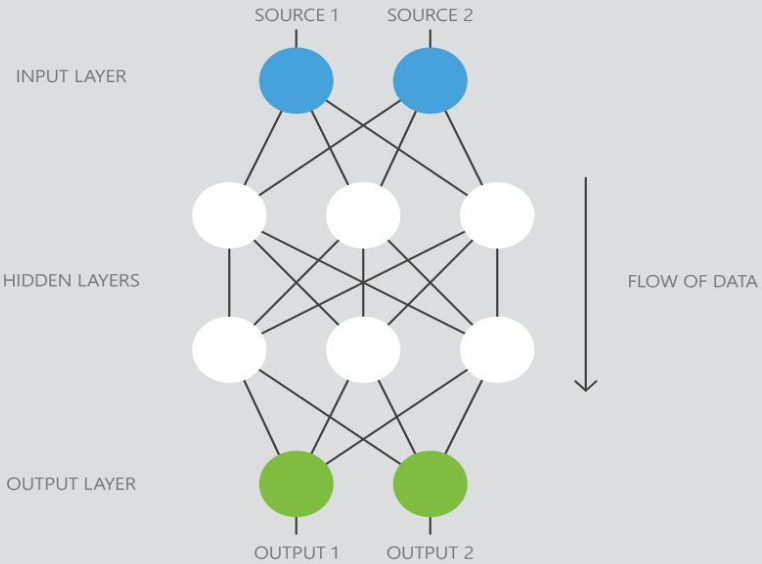
4 Curate your data and refine it in a series of steps



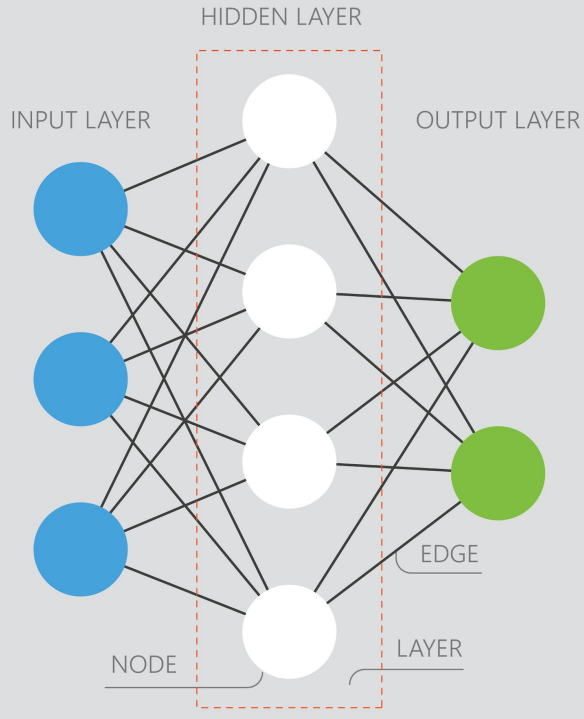
Delta Lake allows you to *incrementally* improve the quality of your data until it is ready for consumption.

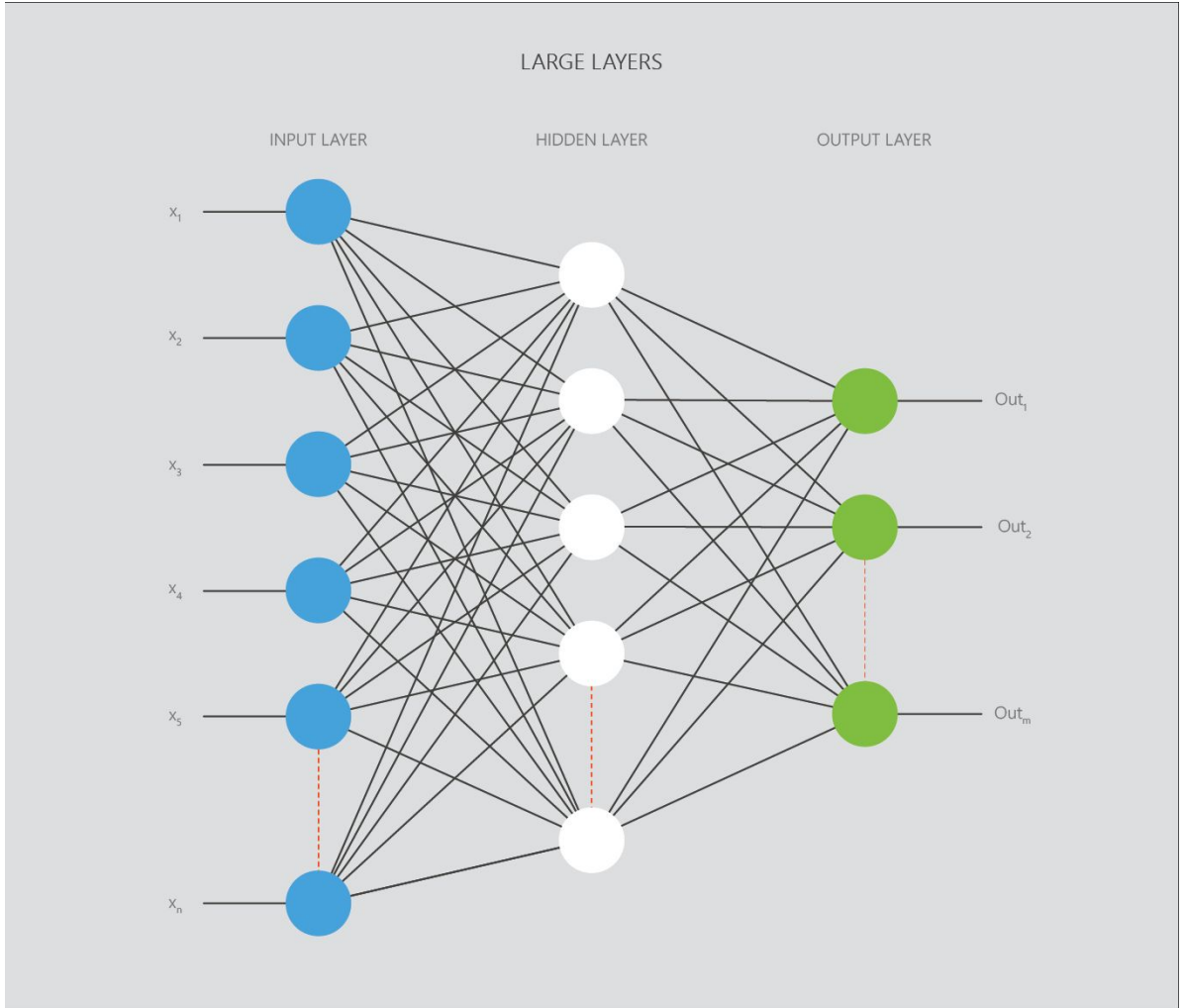
Making predictions based on complex neural network data in Azure

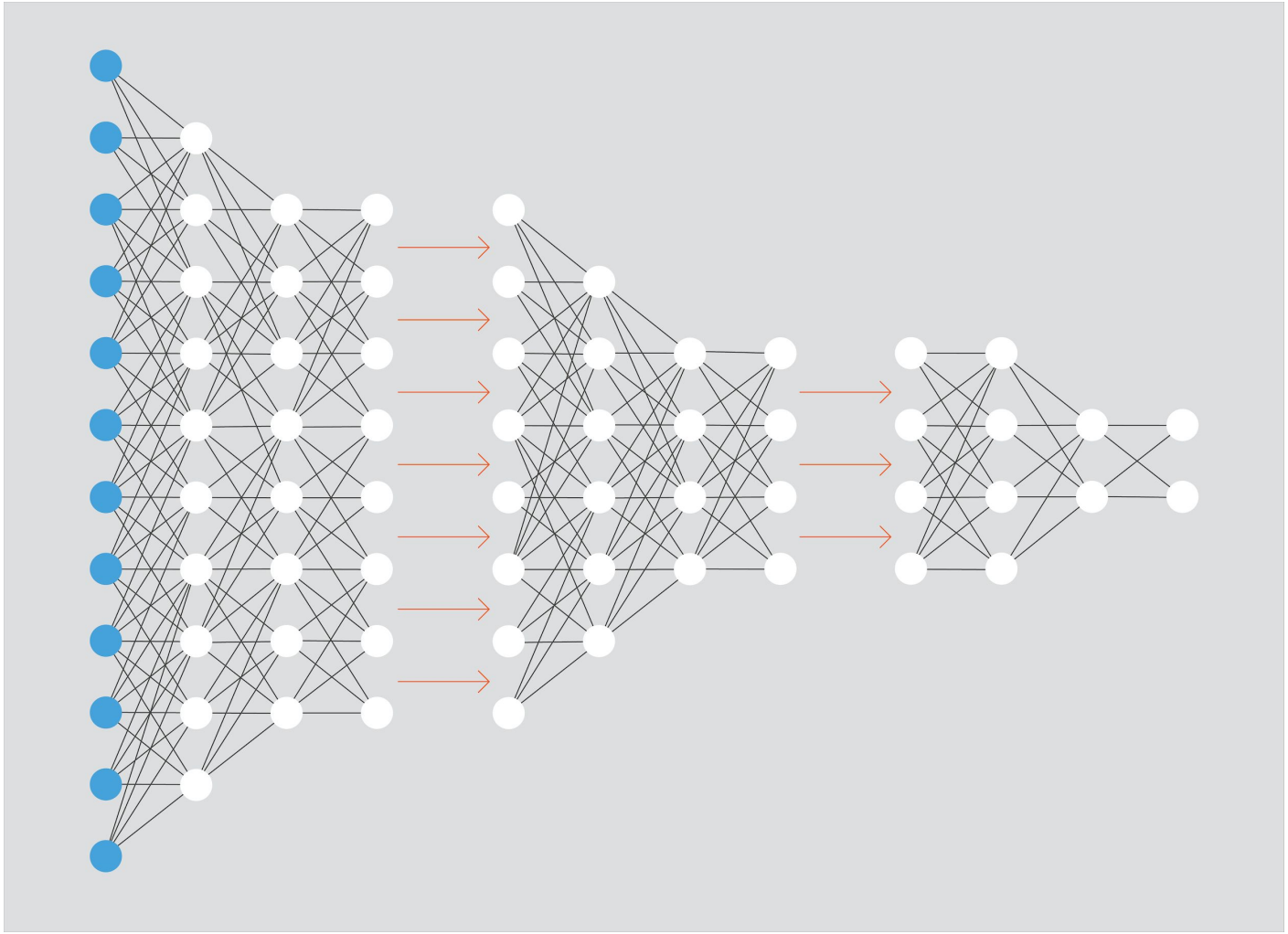

A SINGLE NODE/NEURON
INPUT FROM OTHER NODES
OUTPUT TO OTHER NODES

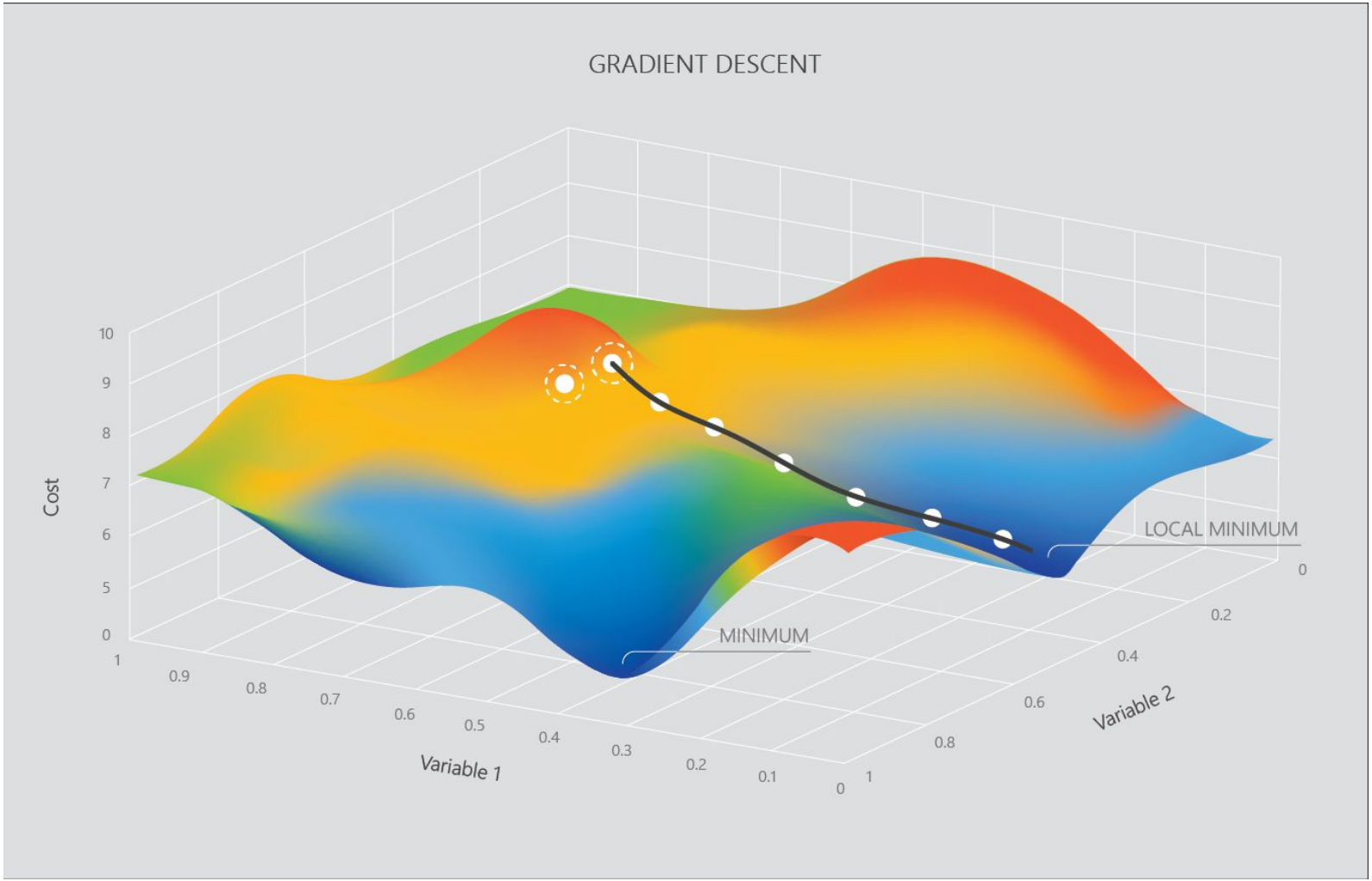

SOURCE 1
SOURCE 2
INPUT LAYER
HIDDEN LAYERS
FLOW OF DATA
OUTPUT LAYER
OUTPUT 1
OUTPUT 2

NETWORK DESCRIPTION

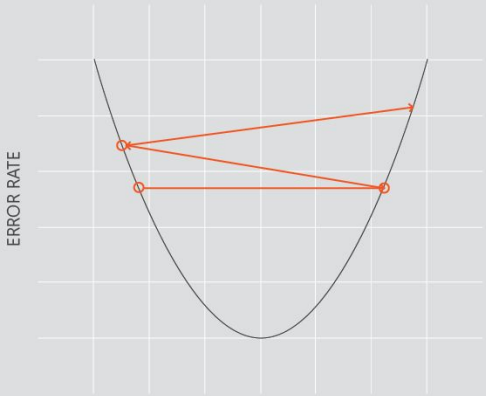



LARGE LAYERS
INPUT LAYER
HIDDEN LAYER
OUTPUT LAYER
x_1
x_2
x_3
x_4
x_5
x_n
Out_1
Out_2
Out_m



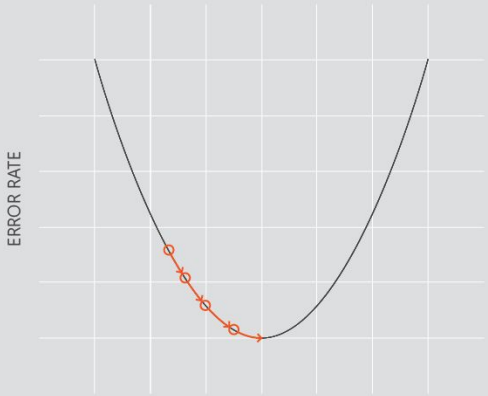

GRADIENT DESCENT
Cost
10
9
8
7
6
5
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Variable 1
1
0.8
0.6
0.4
0.2
0
Variable 2
LOCAL MINIMUM
MINIMUM

TOO HIGH



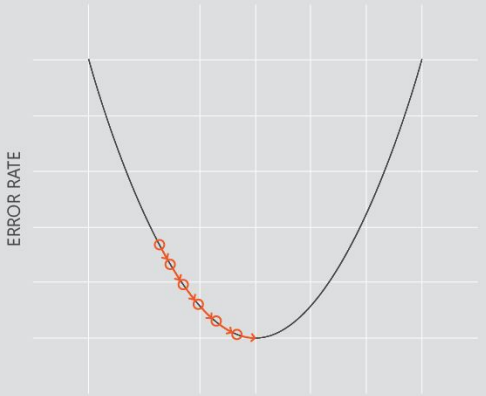
A small learning rate requires many updates before reaching the minimum point.

JUST RIGHT

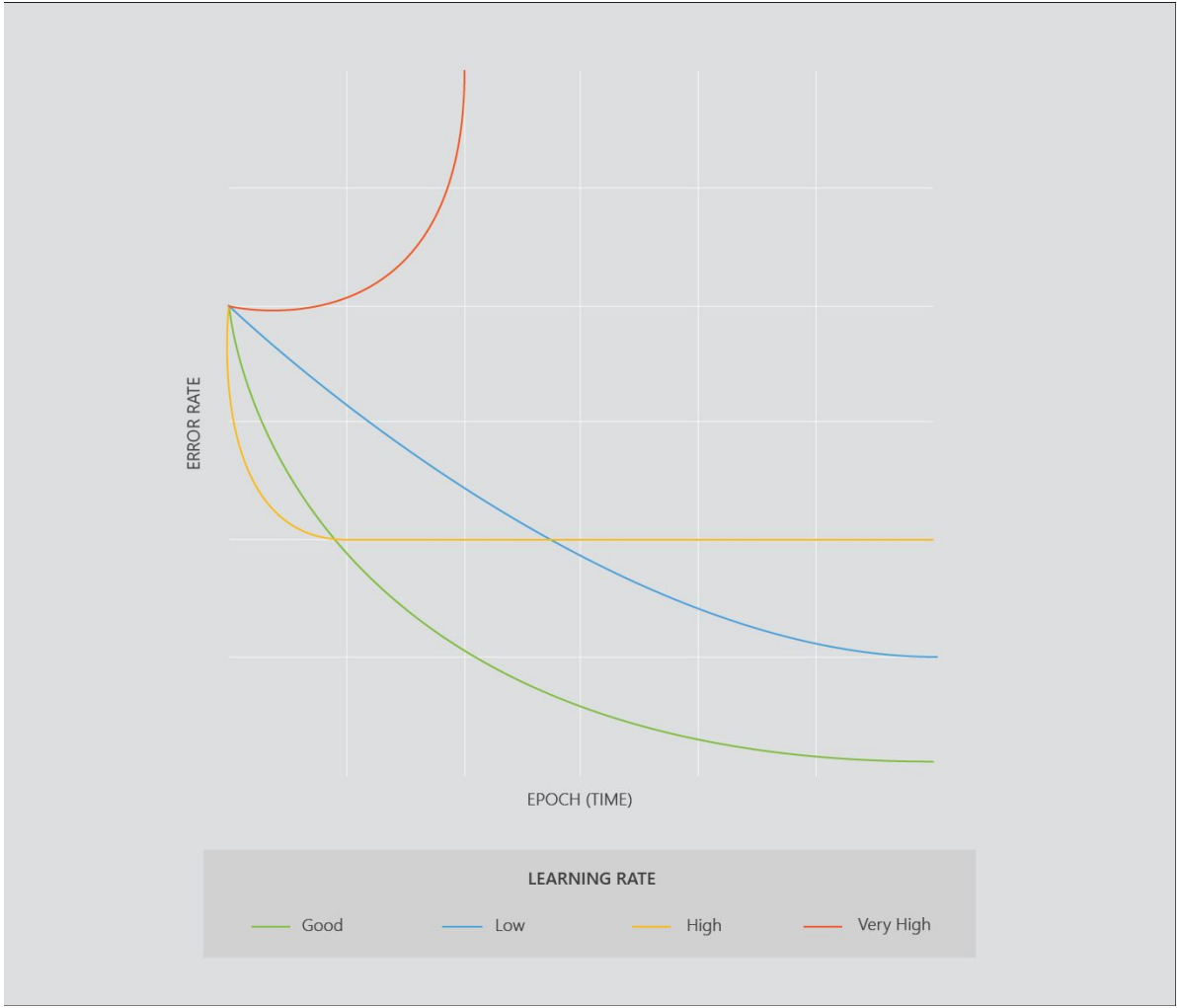


The optimal learning rate swiftly reaches the minimum point.

TOO LOW



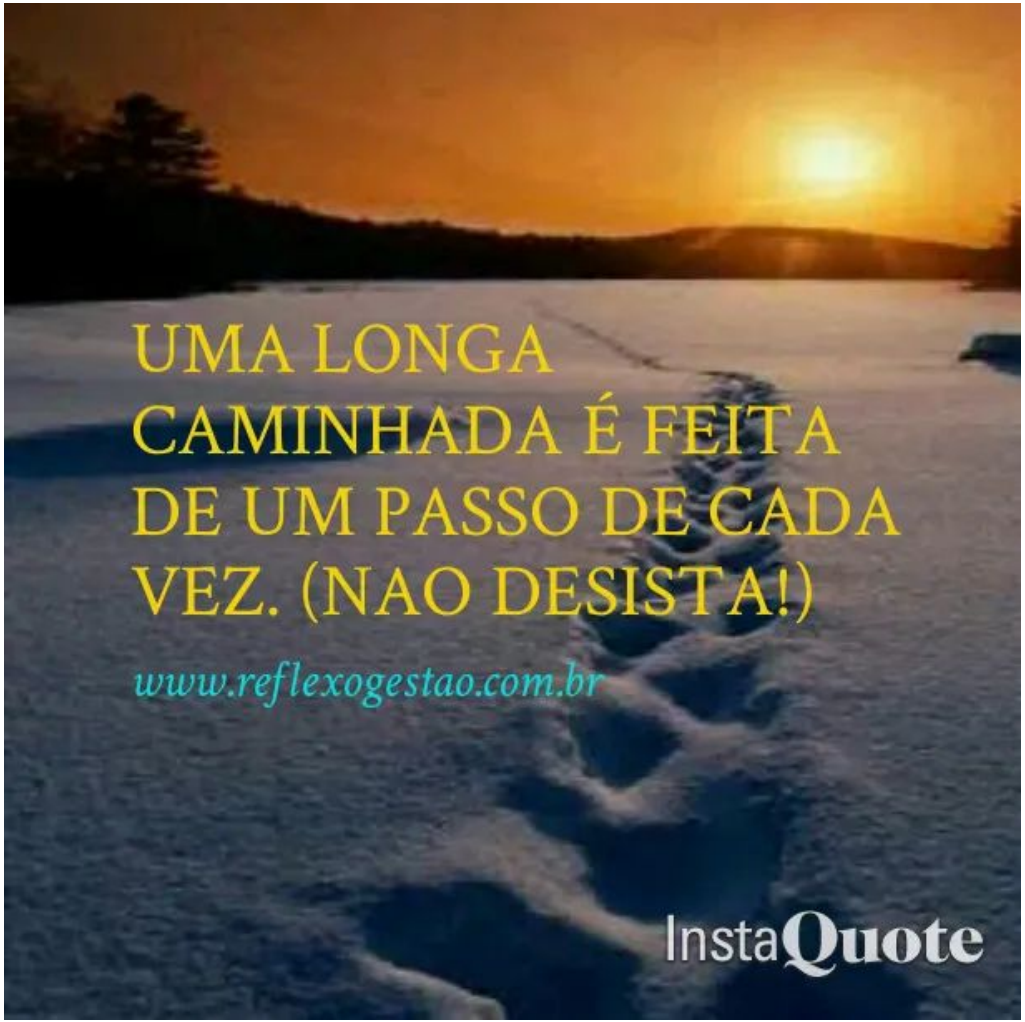
Too large of a learning rate causes drastic updates which lead to divergent behaviours.


ERROR RATE
EPOCH (TIME)
LEARNING RATE
Good
Low
High
Very High

CLASS	VECTORS			
	1	0	0	0
	0	1	0	0
	0	0	1	0
	0	0	0	1

https://docs.microsoft.com/pt-br/learn/ai-school/introduction-to-neural-networks/media/nn-fp-bp.gif

Q&A


UMA LONGA
CAMINHADA É FEITA
DE UM PASSO DE CADA
VEZ. (NAO DESISTA!)
www.reflexogestao.com.br
InstaQuote