

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



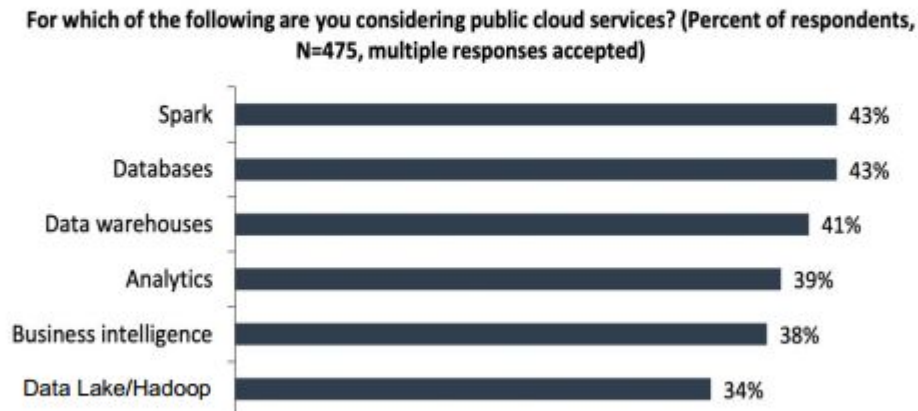
On premises systems require excess capacity to process peak needs



Data security is costly and experts are hard to find

Analytic platforms are moving to the cloud

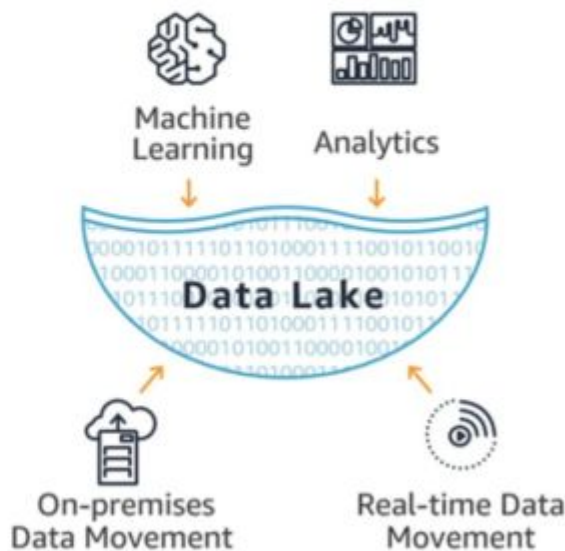
Figure 3. Migration to Cloud for Data Management Services



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

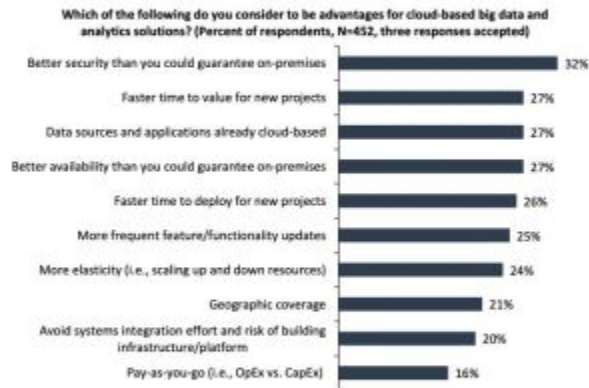
- Pervasive security features
- Performance and scalability
- Reliability and availability
- Economics
- Integration
- Agility



White Paper: The Compelling Advantages of a Cloud Data Lake

9

Figure 4. Advantages of Cloud-based Big Data



Source: AWS- [Embracing a Data-centric Culture Anchored by a Cloud Data Lake](http://bit.ly/CDLESG)
<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

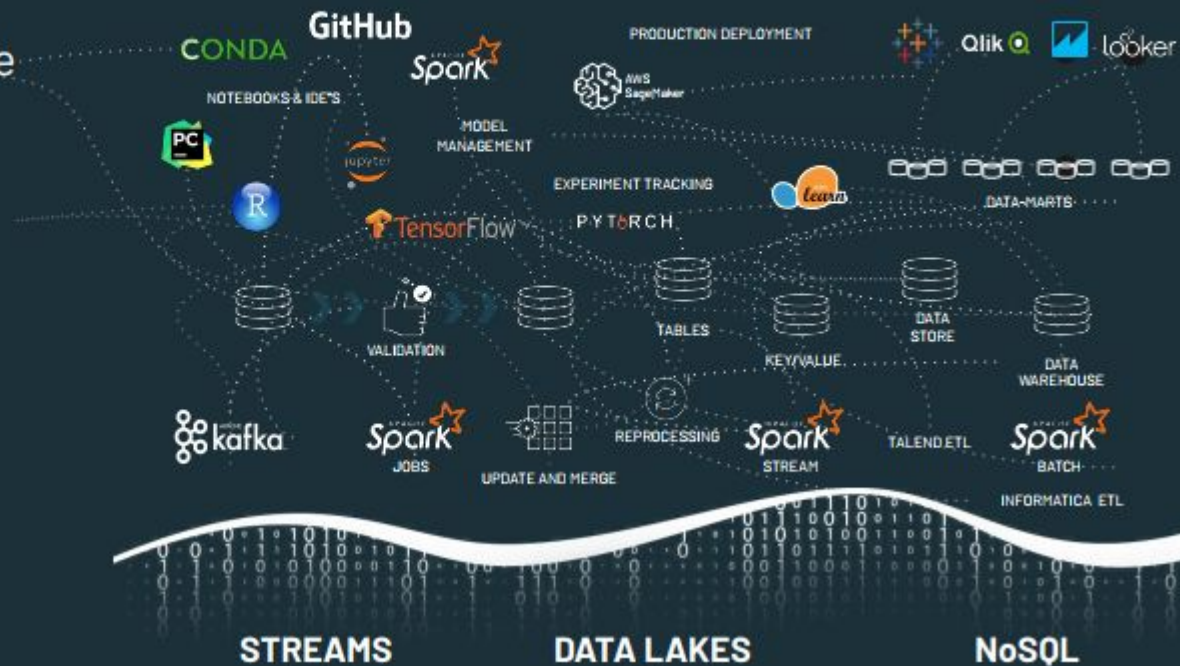
- Global company with 5,000 customers and 450+ partners
- Original creators of popular data and machine learning open source projects
- Empowering data teams to solve the world's toughest problems



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

Machine Learning

Business Analytics



Unlocking business value: Four challenges

1

Data is messy,
siload and slow



2

ML is hard,
Production is
harder



Data
Scientists in
Business



3

BI is limited to a
fraction of data

```
110001100011000100010001
000010111000100101010000
1111001010100111110011001
11010100011100110001100011
00010001000100001011100
01001010100001111001010
```

4

Lack of enterprise
readiness



Fragmented
security

Poor
reliability



Disjointed
governance

Make all your data ready for BI and ML

1

Data is messy,
siloes and slow



Unified Data Service

Build open, reliable, fast data lakes with all your data



Big Data



Business Data



Applications



DELTA LAKE™

Open

High Quality

Fast



Amazon S3



Unified
Engine



BI
Reporting




Machine
Learning

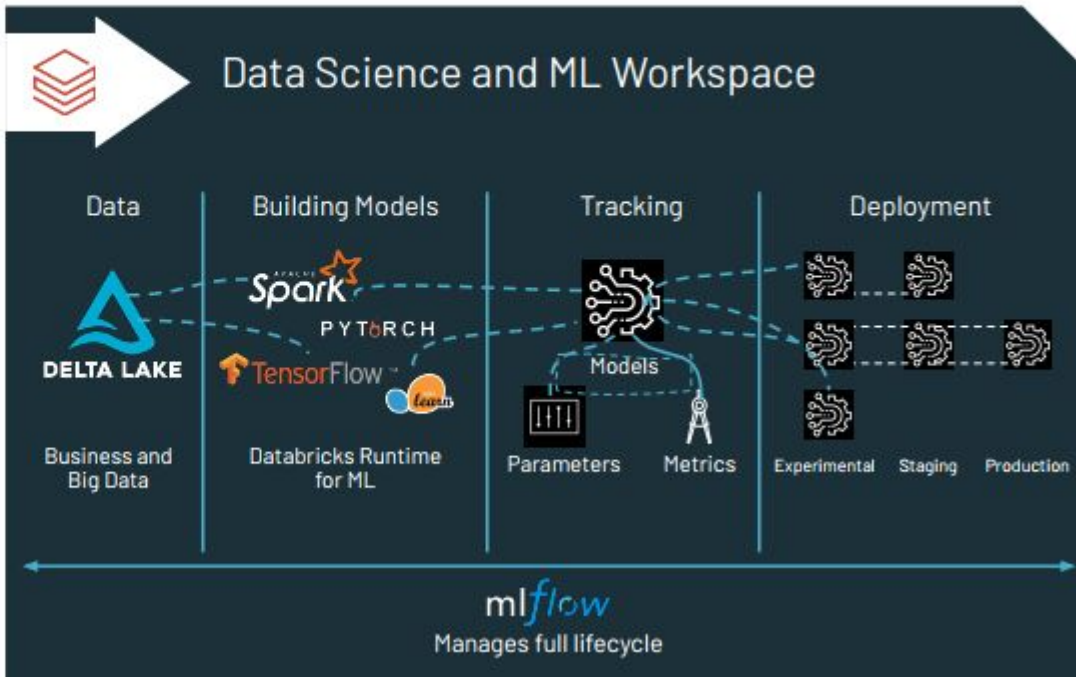
Unify data and ML across the full lifecycle

2

ML is hard,
production is
harder

 Data
Engineers
in IT

 Data
Scientists in
Business



Enable BI directly on all your source data

3

BI is limited to a fraction of data

11000110001100010001000100
00101110001001010100001110
0101010011111001110011101010
001110011000110001100010001
00010000101110001001010100
00111001010



BI Integrations for data lake

Reports



Dashboards



Ad hoc data science



DELTA LAKE



All your data with high quality and great performance



zendesk



Applications



mongodb



oracle

mysql



redis



kafka



kudu



hadoop



google drive



onedrive



dropbox

Files

Leverage cloud native platform for enterprise grade solution

4

Lack of enterprise readiness



Fragmented security

Poor reliability



Disjointed governance



Databricks Enterprise Cloud Service

Your identity provider

okta

1000's of users



DATA ENGINEERS



DATA SCIENTISTS



ML ENGINEERS



DATA ANALYSTS



Highly reliable, secure managed service



Your cloud account

aws

All your data



Amazon S3



kafka



mongoDB



cassandra



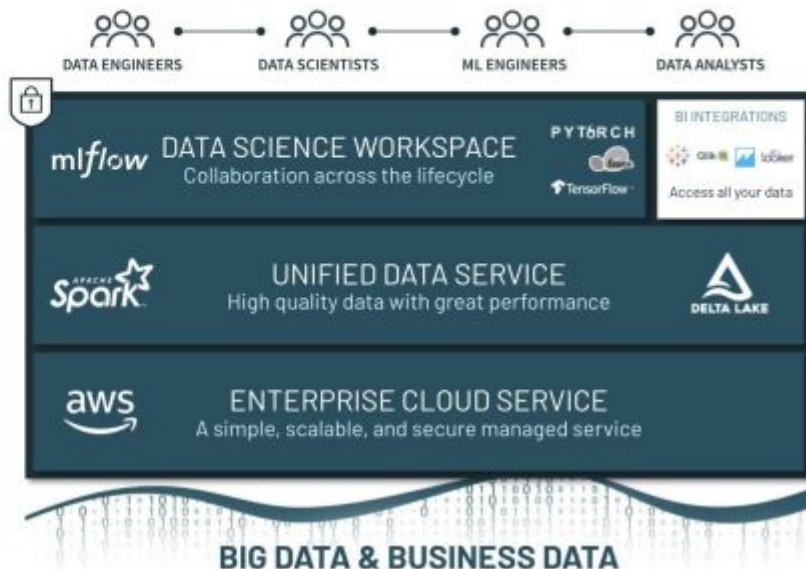
Databricks Unified Data Analytics Platform



Data Science, ML, and BI on
one cloud platform

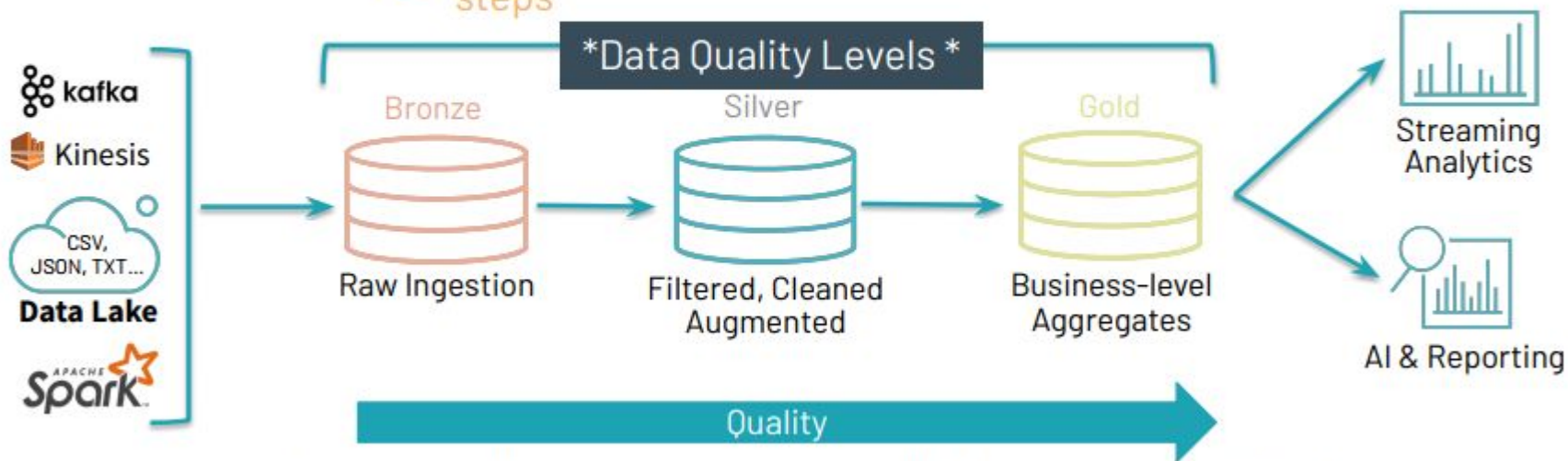
Access all business and big
data in **open data lake**

Securely integrates with your
cloud ecosystem



The DELTA LAKE

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



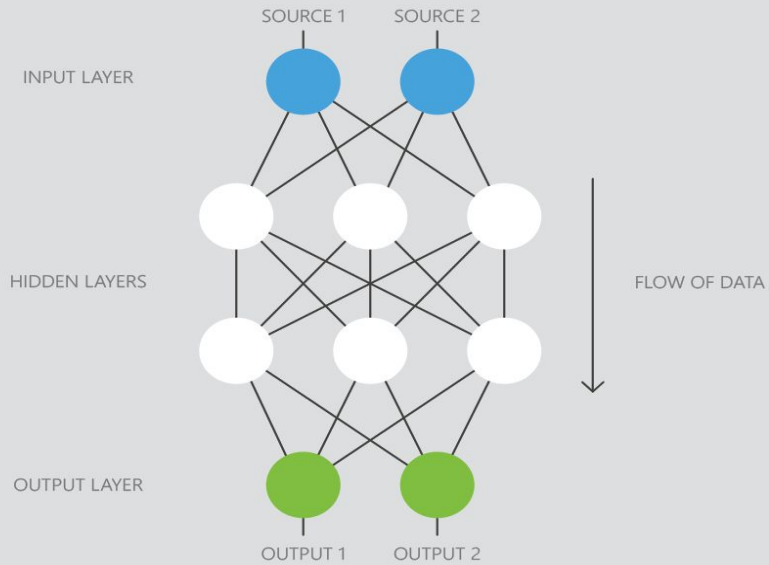
Delta Lake allows you to *incrementally* improve the quality of your data until it is ready for consumption.



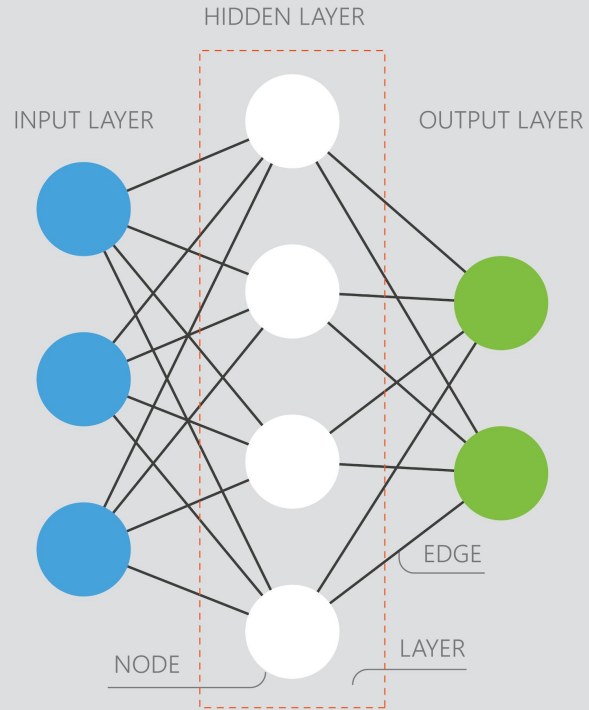
Making predictions based on complex neural network data in Azure

A SINGLE NODE/NEURON

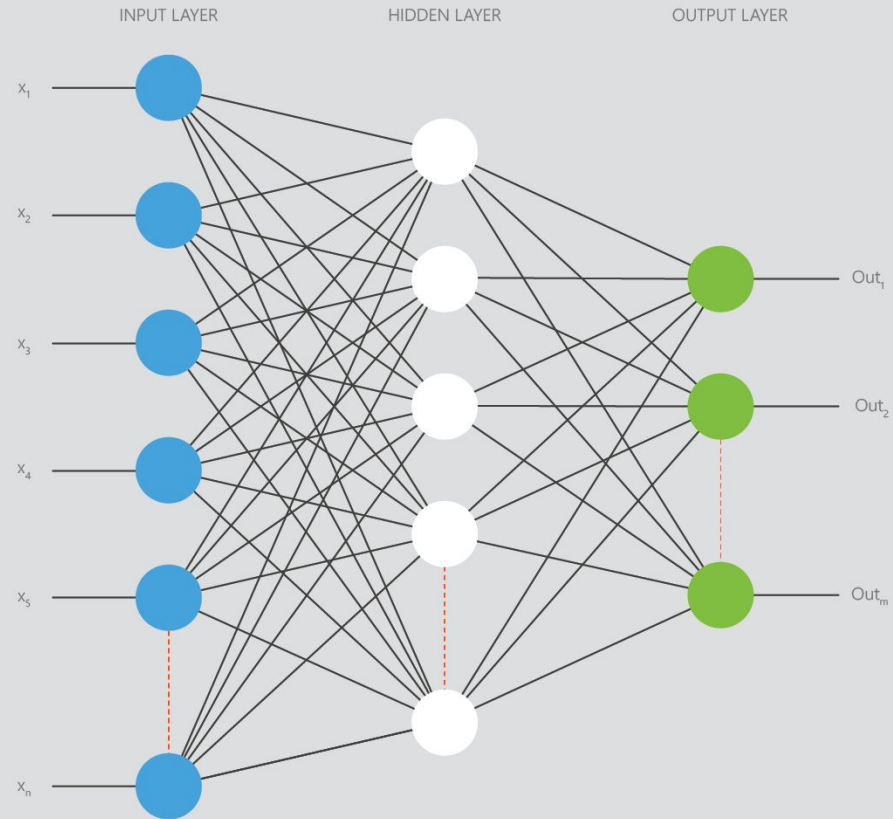


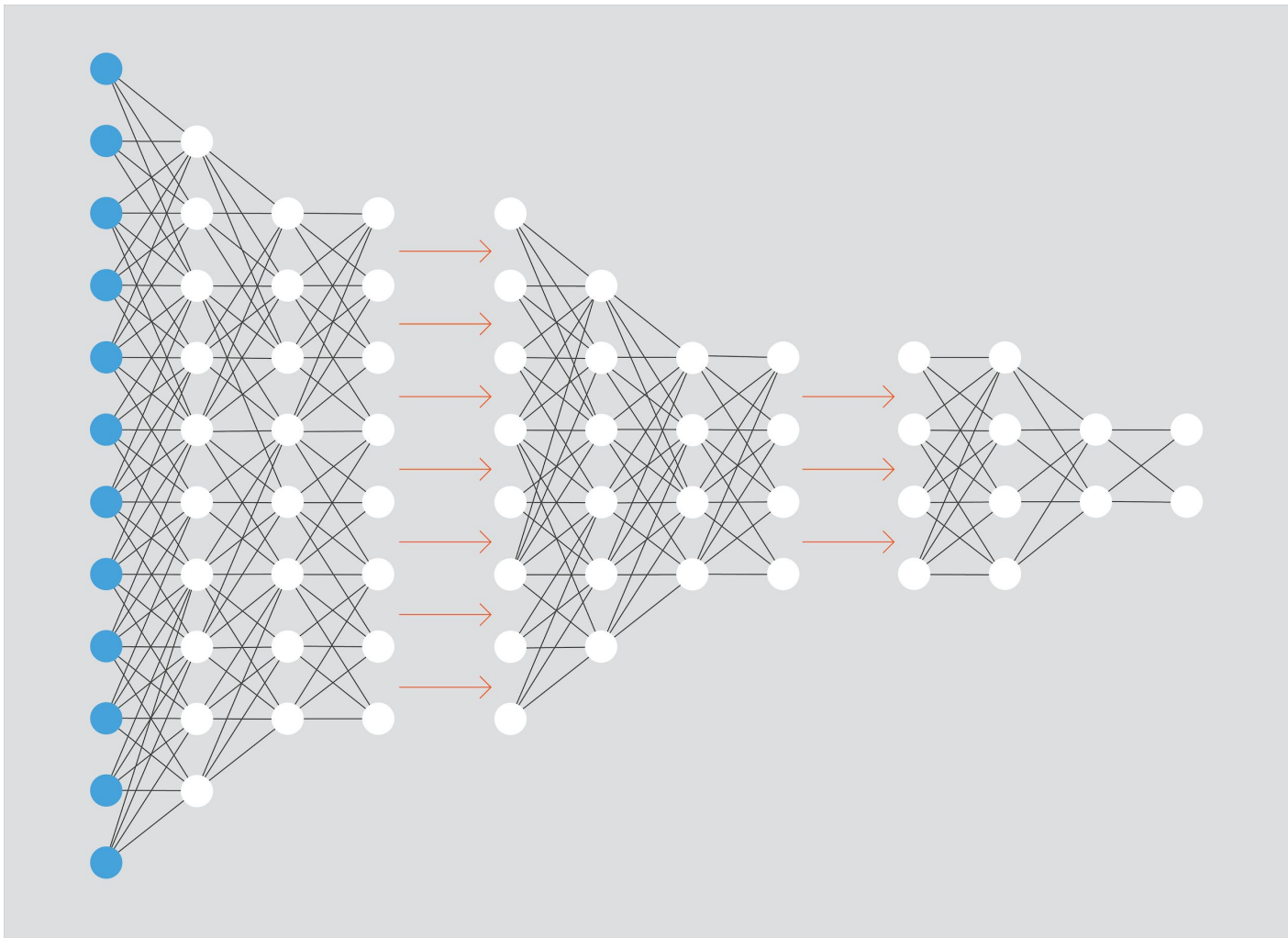


NETWORK DESCRIPTION

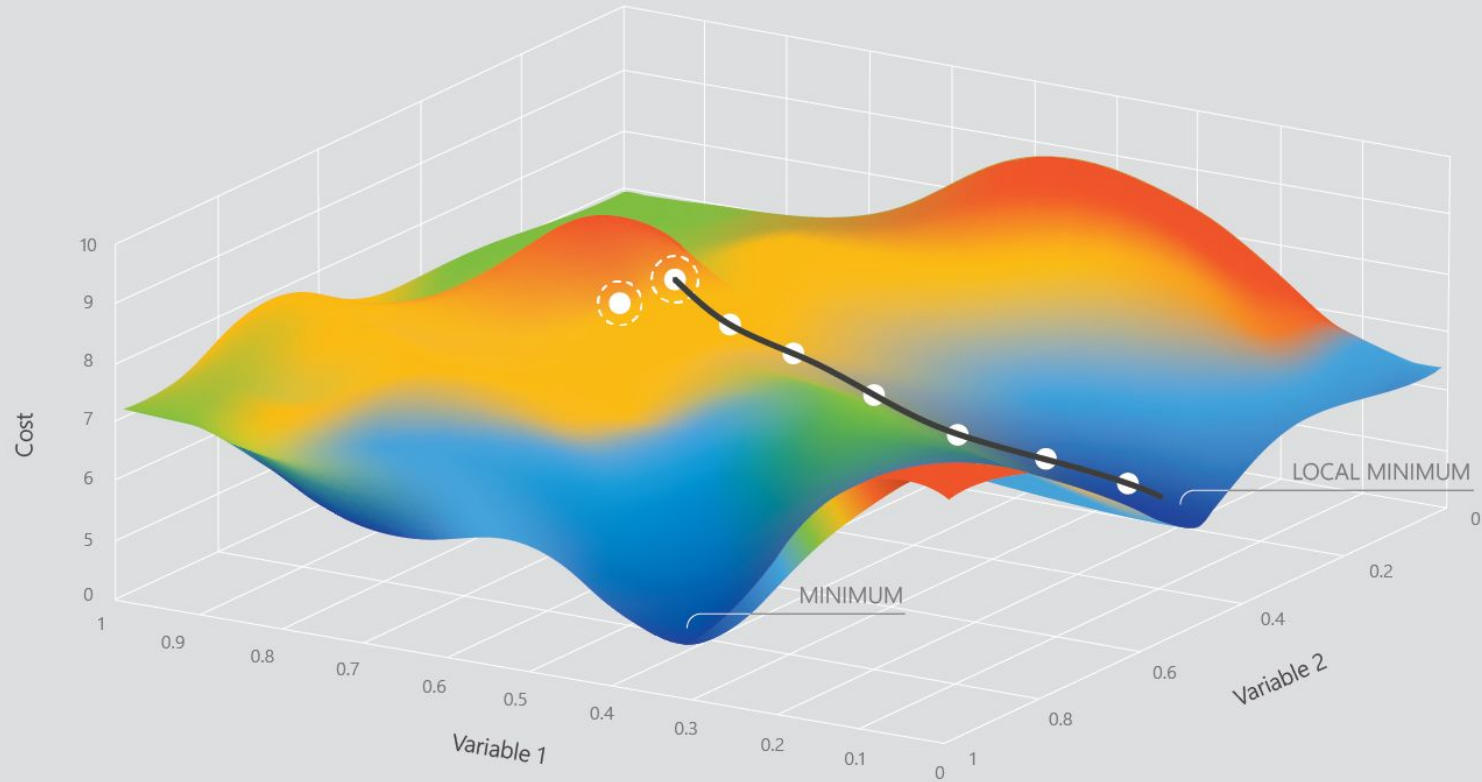


LARGE LAYERS

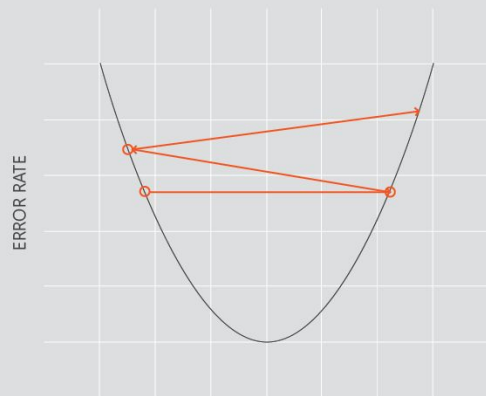




GRADIENT DESCENT

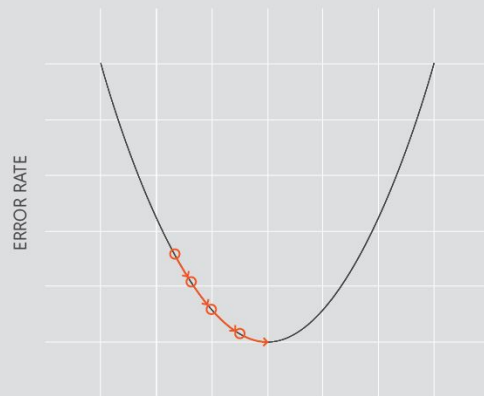


TOO HIGH



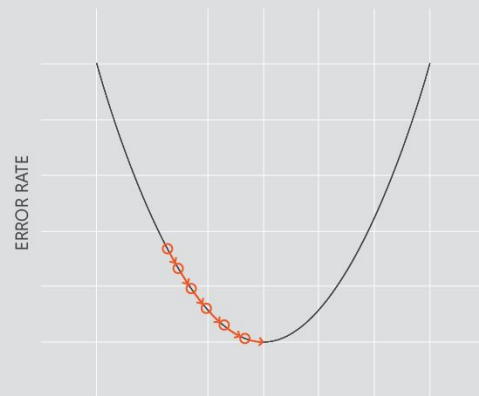
A small learning rate requires many updates before reaching the minimum point.

JUST RIGHT

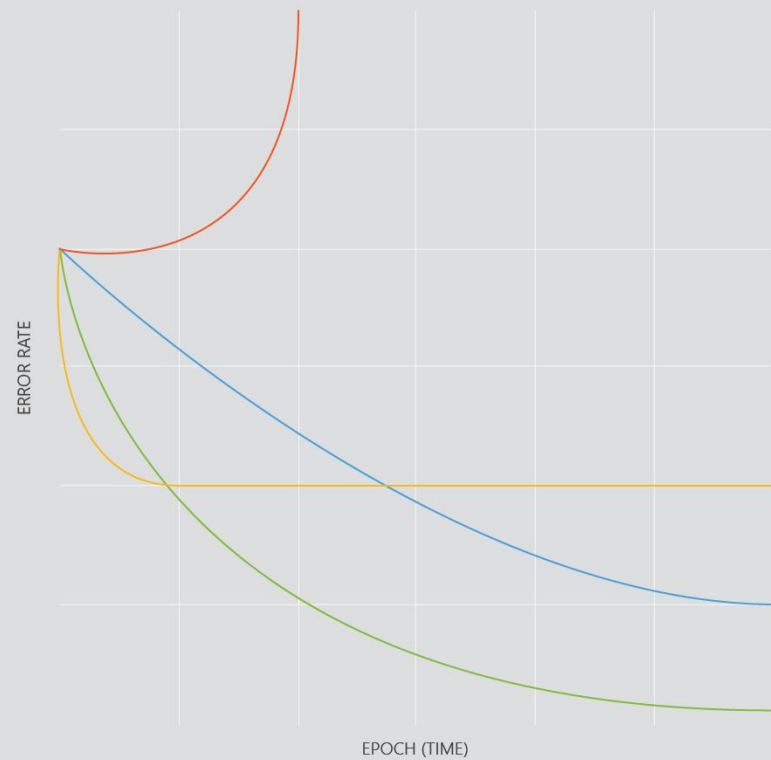


The optimal learning rate swiftly reaches the minimum point.

TOO LOW



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



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

A photograph of a snowy landscape at sunset. The sun is a bright, glowing orb in the upper right, casting a warm orange and yellow light across the sky and snow. In the foreground, a series of footprints are visible in the snow, leading from the bottom center towards the middle ground. The background shows a dark, silhouetted line of trees and hills under the sunset sky.

UMA LONGA
CAMINHADA É FEITA
DE UM PASSO DE CADA
VEZ. (NAO DESISTA!)

www.reflexogestao.com.br

InstaQuote