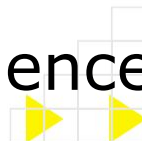




Curso 03: Administração de Dados Complexos em Larga Escala -- Business Intelligence e Big Data --

Prof. Jose Fernando Rodrigues Junior

Objetivo: visão geral da área de Business Intelligence
no contexto de Big Data



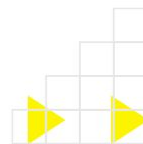
Big Data

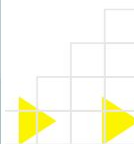
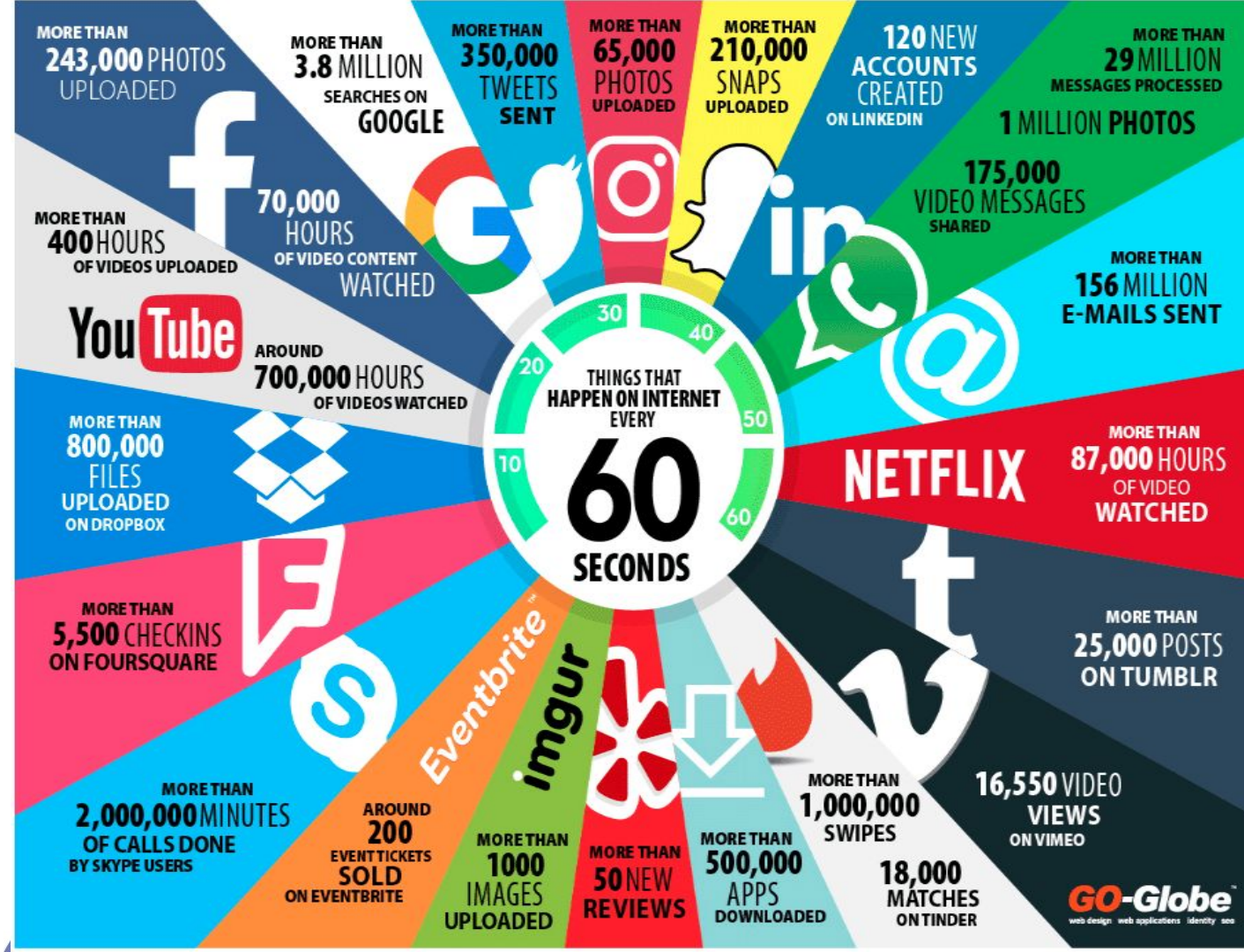


-Definição mais recente: refere-se ao fato de que **tudo o que fazemos está cada vez mais deixando rastros digitais**, os quais podem ser usados para a tomada de decisão;

□ Cartão de crédito, celular, GPS, redes sociais, câmeras fotográficas, sistemas digitais de comunicação, buscadores, mídia digital, *smart devices*,

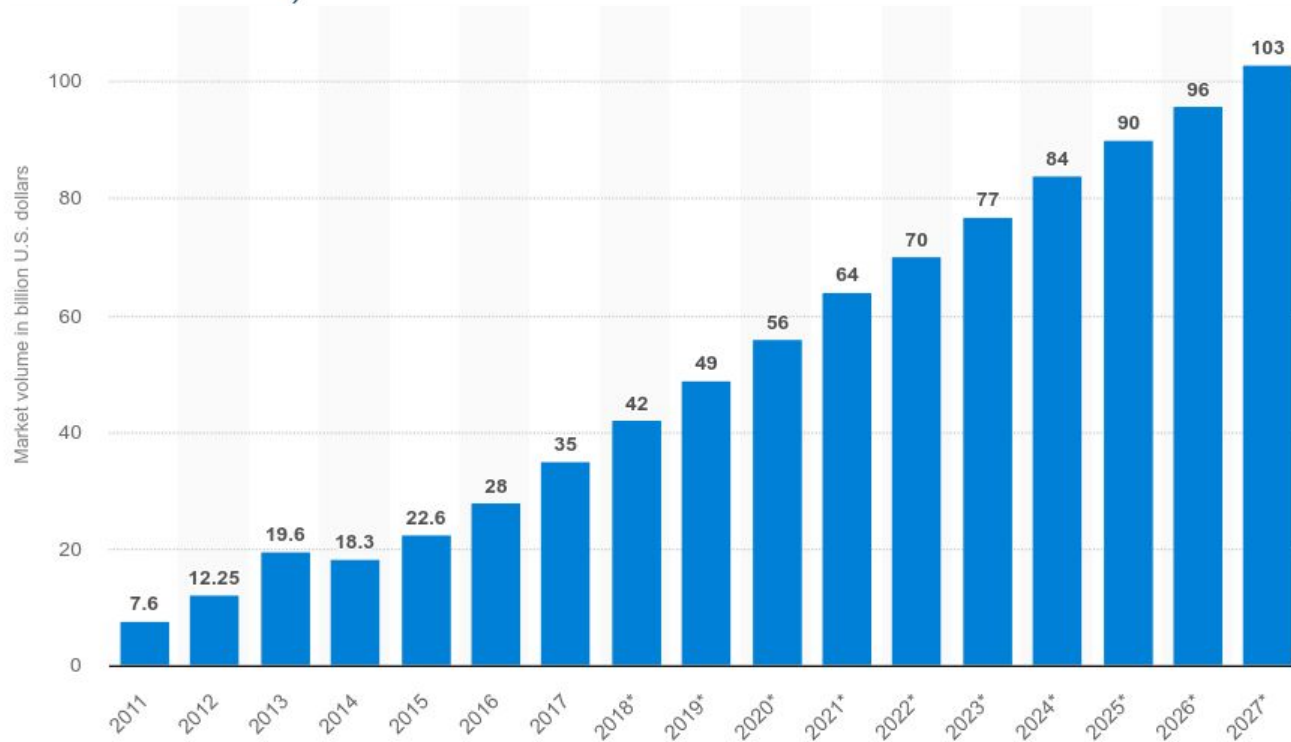
⇒ o uso de Big Data torna-se **mandatório** no momento quando **seus concorrentes** fazem uso dele.





Big data market size revenue forecast worldwide from 2011 to 2027

(in billion U.S. dollars)



[Additional Information](#)

© Statista 2021

[Show source](#)

fonte: <https://www.statista.com/statistics/254266/global-big-data-market-forecast/>



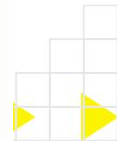
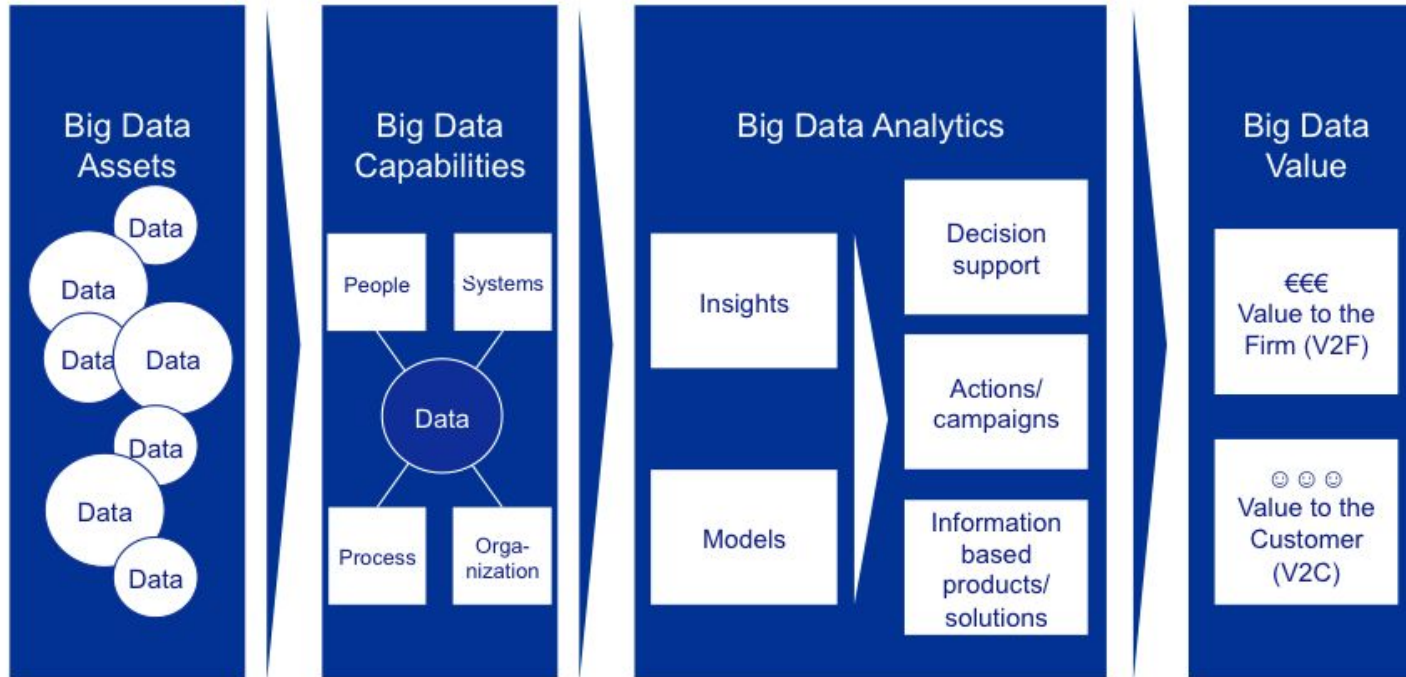


As definições variam entre 3 e 8 vs, um fato que não é muito relevante. O que importa é a compreensão de que se trata de volumes de dados desafiadores.










Big Data Valor

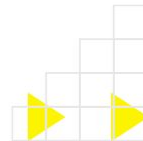
Big Data Value Creation Model





Big Data – Exemplos de valor

-  Empresas de **telecom** podem prever a **taxa de abandono** de clientes (*customer churn*);
-  **Varejistas** podem prever o que será **mais vendido** (e ajustar seus estoques e propaganda);
-  **Seguradoras** podem prever **como seus clientes irão dirigir** (e ajustar os preços de acordo);
-  **Mapas eletrônicos** e dados de tráfego podem **melhorar o trânsito** e a eficiência de cidades.
-  **Distribuidoras** podem **otimizar o tempo de entrega** ao ajustar as rotas dinamicamente;
-  **Redes sociais** podem ser usadas para monitorar **epidemias e política**;
-  **Operadoras de cartão** podem identificar **padrões fraudulentos**.



Applications for Big Data Analytics

Smarter Healthcare



Multi-channel sales



Finance



Log Analysis



Homeland Security



Traffic Control



Telecom



Search Quality



Manufacturing



Trading Analytics



Fraud and Risk



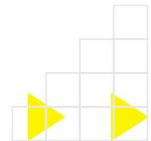
Retail: Churn, NBO



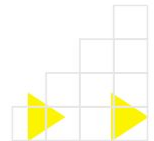
Business Intelligence



- **entendimento da base de dados da empresa**, gerando saídas lógicas e estratégicas;
- **norteia as ações** tomadas por uma organização, com base em dados com potencial estratégico;
- possibilidade de se **acompanhar** o que ocorre no **negócio**, assim como também no **mercado**.



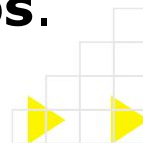
Business Intelligence





Business Intelligence - Tarefas

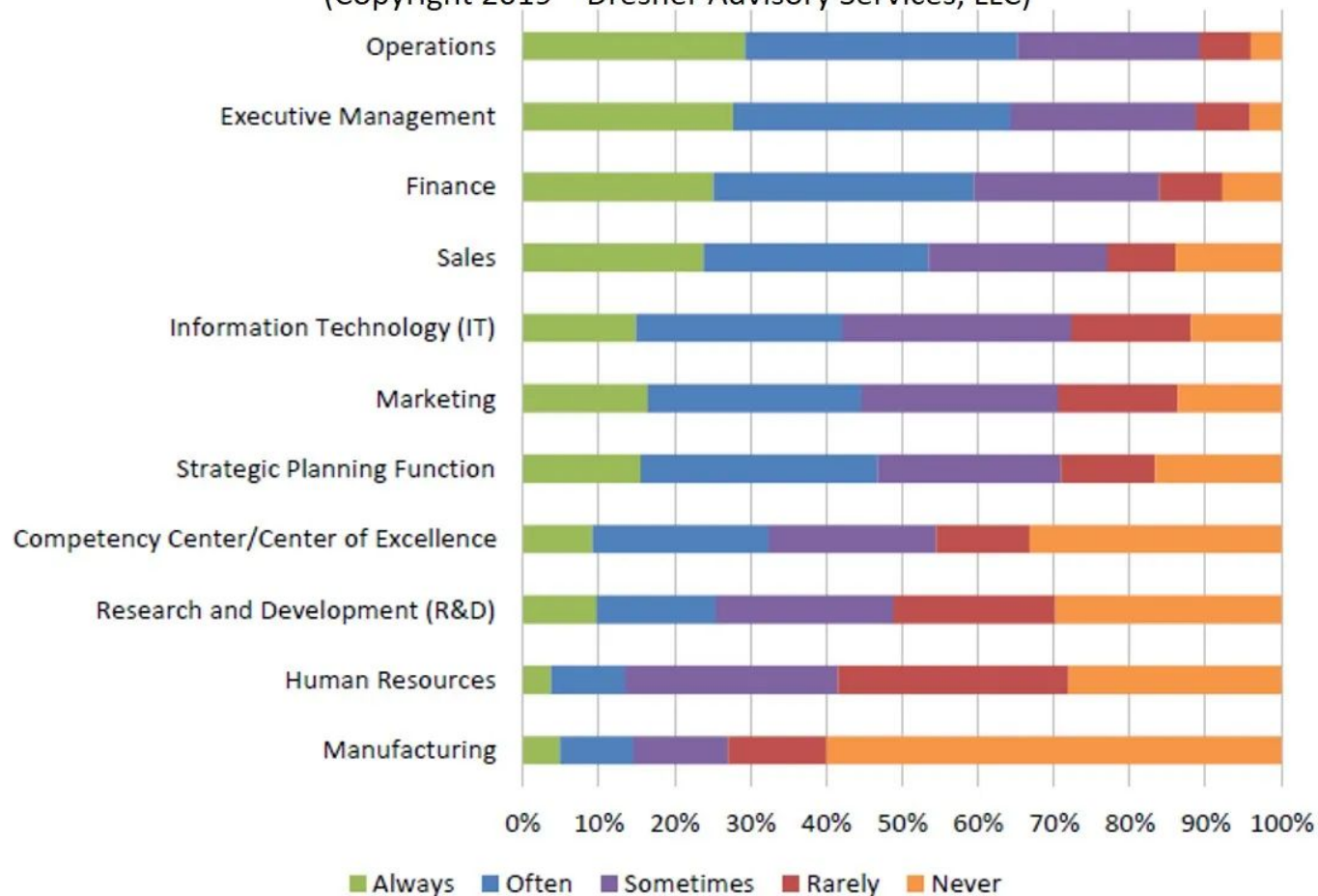
- medição**: criação de medidas que indicam o progresso da empresa em direção às **metas**;
- reporting**: geração de relatórios sobre dados estratégicos e de infraestrutura – **OLAP e visualização**;
- analytics**: uso **exploratório** de dados – mineração de dados, aprendizado de máquina, e predição;
- colaboração**: ferramentas que promovem o trabalho colaborativo **intra e extra** empresa;
- gerência de conhecimento**: promoção do uso das outras quatro tarefas a fim de se alcançar uma **empresa orientada a dados**.

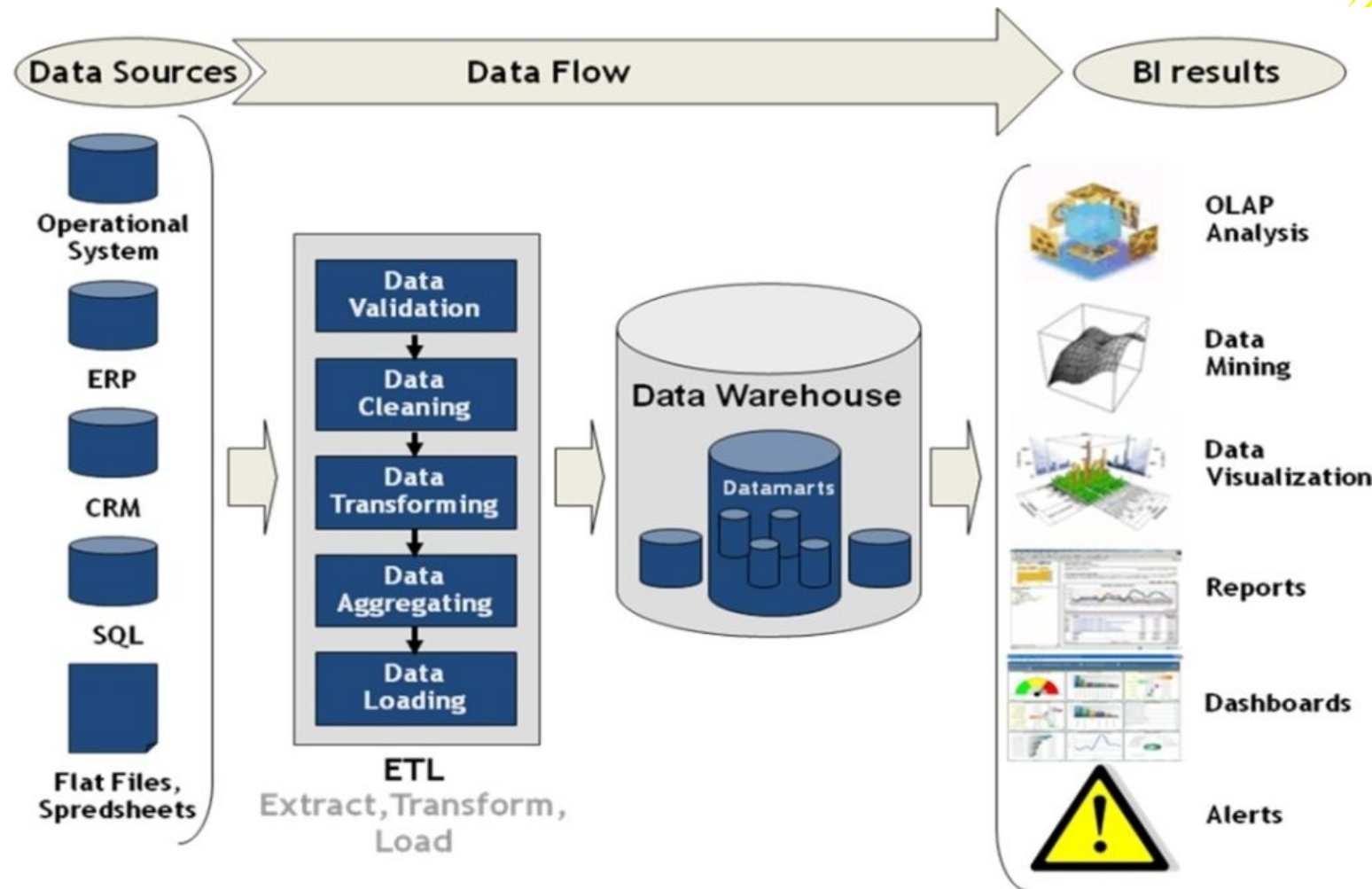




Functions Driving Business Intelligence

(Copyright 2019 – Dresner Advisory Services, LLC)







Business Intelligence - Tools

-DW: fontes de dados, na maioria das vezes heterogêneas organizadas em datamarts ou data warehouses - Apache Hive, Oracle Exadata, IBM Netezza, Microsoft Azure, e outras;

-DW/OLAP: software capaz de sumarizar grandes quantidades de dados de maneira rápida;

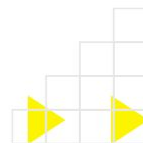
-Visualização: software para exibir os produtos finais do processamento de dados; Tableau, MS PowerBI, IBM Cognos, Metabase, e outros;

-Advanced Analytics:

+Estatística - R, Matlab, Saas;

+Aprendizado de Máquina - Scikit Learn, Saas, Scilab;



































+Inteligência Artificial - em amadurecimento.



INFRASTRUCTURE

The diagram illustrates the landscape of cloud storage and data management services, categorized into six main areas:

- STORAGE:** Includes logos for Amazon S3, Google Cloud Storage, Microsoft Azure Storage, IBM Cloud Object Storage, and Oracle Cloud Infrastructure.
- CLOUDERA:** Includes logos for Cloudera, Hadoop, and HBase.
- DATA LAKES:** Includes logos for AWS Lake Formation, Databricks, and Snowflake.
- DATA WAREHOUSES:** Includes logos for Amazon Redshift, Google BigQuery, Microsoft Azure Synapse Analytics, IBM Cloud Pak for Data, and Oracle Data Warehouse.
- STREAMING / IN-MEMORY:** Includes logos for Amazon Kinesis, Google Cloud Dataflow, Microsoft Azure Stream Analytics, IBM Cloud Streaming, and Oracle Data Streaming.
- CLOUD PLATFORM:** Includes logos for AWS, Google Cloud, Microsoft Azure, IBM Cloud, and Oracle Cloud.

NoSQL DATABASES	NewSQL DATABASES	GRAPH DBs	MPP DBs	SERVER-LESS	CLUSTER SVCS
 Google Cloud  IBM  ORACLE  MarkLogic  Databricks  SCYLLA	 Microsoft Azure  Amazon Aurora  mongoDB  Couchbase  ArangoDB	 SALESFORCE  Pivotal  IBM  ORACLE  Alibaba Cloud  Microsoft Azure	 TERADATA  VERTICA  Cockroach  Snowflake  EcoSQL	 Amazon AWS  Google Cloud  IBM  Microsoft Azure  ORACLE  Alibaba Cloud	 Amazon AWS  Google Cloud  IBM  Microsoft Azure  ORACLE  Alibaba Cloud

ETL / DATA TRANSFORMATION	DATA INTEGRATION	DATA GOVERNANCE	DATA QUALITY
 Alteryx  Pentaho  Trifacta  Informatica Data Integration  DataStage  Paxata  StreamSets  UNIFI  dataform	 Data Services  MuleSoft  Stitch  Fivetran  import.io  Snowflake  Snowpark  Snowflake  Snowflake	 Informatica Data Governance  IBM Data Governance  Collibra  dremio  Alphinity  Watermark  Privasea	 SailPoint  Alation  Buckley Group  MANTA  OKERA  iMANTA  precisely  Monte Carlo

[illegible]

ANALYTICS & MACHINE INTELLIGENCE

The image displays a collection of logos for various data science and analytics platforms, organized into four main categories:

- BI PLATFORMS:** Includes logos for Looker, Tableau, Microsoft Power BI, Alteryx, Qlik, and SAP Business Intelligence.
- VISUALIZATION:** Includes logos for Tableau, Google Data Studio, Periscope Data, Zepel, Looker, and others.
- DATA ANALYST PLATFORMS:** Includes logos for Microsoft, Alteryx, Mode, Databricks, Anodot, Outlier, and others.
- DATA SCIENCE PLATFORMS:** Includes logos for DataRobot, H2O, and others.

DATA SCIENCE

DATA SCIENCE PLATFORMS

MACHINE LEARNING

[illegible]

The image displays a collection of logos for various data science and analytics companies, organized into four distinct categories, each with a red header. The categories and their respective logos are as follows:

- SEARCH**
 - elasticsearch
 - amazon
 - ORACLE INDEX
 - IBM Watson CloudSearch
 - Azure Cognitive Services
 - algolia
 - covio
 - SINEQUA
 - Lucidworks
 - ATTIVO
 - swiftype
 - EXEBL
 - alphasense
 - omni-us
 - MAANA
- LOG ANALYTICS**
 - splunk
 - Google Cloud Logging
 - sumologic
 - scalyr
 - logz
 - timber
 - loggly
 - logz
- SOCIAL ANALYTICS**
 - Hootsuite
 - sprinklr
 - NETBASE
 - synthesio
 - simplereach
 - bitly
 - ETEXE
 - SimilarWeb
- WEB / MOBILE / COMMERCE ANALYTICS**
 - Google Analytics
 - mixpanel
 - SEGMENT
 - Airtable
 - RESCI
 - granify
 - Amplitude

– APPLICATIONS – ENTERPRISE

SALES	MARKETING - B2B	MARKETING - B2C	CUSTOMER EXPERIENCE / SERVICE	HUMAN CAPITAL
          	     	              	                        	      

LEGAL	REGTECH & COMPLIANCE	FINANCE	AUTOMATION & RPA	SECURITY
   	   	   	   	   
   	   	   	   	   

APPLICATIONS – INDUSTRY

ADVERTISING	EDUCATION	REAL ESTATE	GOV'T & INTELLIGENCE	COMMERCE	FINANCE - LENDING	INSURANCE
 Adaptive  MediaMath  critico  IAS  ORACLE AdCloud  Bluebird Interactive  albert  gumgum  Oppler  theTradeDesk  TAFAB	 Lushika  VTS  新东方  Opendee  KNEWTON  Orchard  elearnify  eClerca  SKYLINE  EAS  SPACEMATE  GEOPHY  KORU	 REDFIN  VTS  Opendee  Orchard  elearnify  eClerca  MARK43  ANDRILL  FiscaNet  Quid  PRIMER  SKYLINE  SPACEMATE  GEOPHY	 Palantir  OpenFVG  Dolomint  MARK43  ANDRILL  FiscaNet  Quid  PRIMER  SKYLINE  SPACEMATE  GEOPHY	 FAIR  STITCH FIX  HowGood  STANDARD  AYASDI  KENSHC  ADEPAR  NUMERAL  100Credit  Ateco AI	 affirm  Mondo  ZEST  LALA  UPBANK  BANC  aurea  UStartup  100Credit  Ateco AI	 ROOT  Neuromina  Amenade  ECONOMIA  Shift Technology  CAPE  EVALUATION  TECHTALK  ZILLOS  zesty

HEALTHCARE

batson, ARYUS, METABIOTA, **TEMPOS**, **30Med**, **Phylis**, **AlCur**, **Imagis**, **PAIGE**, **Immunocore**, **lumitox**, **springhealth**, **erilic**, **4111171**, **ezra**, **Cognition Health**, **H.A.O.N**, **zebra**

LIFE SCIENCES

color, **Dharma**, **verily**, **DMNexus**, **gompert**, **brato**, **transgene**, **Genentech**, **SOPHIA**, **insitro**, **Genetech**, **healix**, **Novartis**

TRANSPORTATION

uber, **tesla**, **niro**, **APTIV**, **Qorus**, **Arora**, **nauto**, **ADaptive**, **PIED**, **neo**, **OPTIMUS**, **G7**, **moovit**, **Adapt**, **Adixis**, **ADIR**, **ADIR**

AGRICULTURE

Granular, **Blue River**, **taraneh**, **prospera**, **semios**

INDUSTRIAL

AVAYA, **SIEMENS**, **storm**, **Amper**, **ByteDance**, **YONEX**, **Ecostrux**, **TAUCHUS**, **alicia**, **duetto**, **Logistics**

OTHER

Logistics

- OPEN SOURCE

FRAMEWORKS 	QUERY / DATA FLOW 	DATA ACCESS & DATABASES 	INTEGRATION & PIPELINES 	STREAMING & MESSAGING 	STAT TOOLS & LANGUAGES 	AI OPS & INFRA 	AI / MACHINE LEARNING / DEEP LEARNING 	SEARCH 	LOGGING & MONITORING 	VISUALIZATION 	COLLABORATION 	SECURITY
-----------------------	------------------------------	--	--	--------------------------------------	---------------------------------------	-------------------------------	--	-------------------	-------------------------------------	--------------------------	--------------------------	---------------------

DATA SOURCES & APIs

DATA MARKETPLACES & DISCOVERY

aws Data Exchange, snowflake, DAWEX, EXPLORIM, data.world, narrative

FINANCIAL & ECONOMIC DATA

Bloomberg, THOMSON REUTERS, DOW JONES, Quant, S&P CAPITAL IQ, CB FIELDS, PLAID, SECOND STREET, EVEREST VECTA, predata, MXC, StockTink, xignite, Smartwells, earnest

AIR / SPACE / SEA

Orbital insight, AIRBOTICS, spire, WINDWARD, MarineTraffic, EXOLES, Synspic

PEOPLE / ENTITIES

zoominfo, axciom, Experian, Epsilon, InsideView, Brandwatch, Quantcast, BASIS, SAFEGRAD, Dermyst, melissa

DATA RESOURCES

LOCATION INTELLIGENCE

- Foursquare
- mapbox
- sense360
- primary business
- PlaceIQ
- esri

OTHER

- DATA.GOV
- IMx.GENET
- Lab4Life
- Radax
- Mapillary
- Cuebio
- OpenStreetMap
- ApollonSquare
- CRUX

DATA SERVICES

- QuantumBlack
- Booz | Allen | Hamilton
- ElectrifAI
- fractal
- DataKind
- innorexus
- kaggle
- EXL

INCUBATORS & SCHOOLS


- PivotalPoint
- General Assembly
- DataCamp
- DataLib
- Galvanize
- The Data Incubator

RESEARCH

- OpenAI
- facebook research
- MIRI
- VECTOR INSTITUTE
- AIX

ANALYTICS & MACHINE INTELLIGENCE

DATA LAKES

MGMT / MONITORING	DATA GENERATION & LABELLING	AI OPS	GPU DBS & CLOUD	AI HARDWARE
  	  	  	  	  
  	  	  	  	  
  	  	  	  	  
  	  	  	  	  
  	  	  	  	  
  	  	  	  	  
  	  	  	  	  

VISUALIZATION DATA ANALYSIS

SEARCH	LOG ANALYTICS	SOCIAL ANALYTICS	WEB / MOBILE / COMMERCE ANALYTICS
elasticsearch	splunk	Hootsuite sprinklr	Google Analytics
EMERSON Analytics	Google Cloud Logging	NETBASE	mixpanel SIGHTPT
Amazon AWS Partner Solutions	sumologic	synthesio	Airtable
algolia	scalyr	simplereach	RESCL granity
covoco	timber	bimby	Amplitude
SINEQUA	loggly	EMXEX	
Lucidworks	logpoint		
ATTIVO			
swiftype			
EXPLAIN			
alphasense			
omni-us			
MAANA			

NG - B2C — CUSTOMER E

The collage displays logos for numerous companies across five sectors:

- HEALTHCARE:** flatiron, KYRUUS, METASOFT, bodyfront, 3DMed, Health, TEMPUUS, patientix, AllCur, UtiCare, wintona, Olive, Incept, PAIGE, Immunovacc, lumina, spring health, The Cancer Research Fund, eribic, 4313171, EZGO, Cognition Health, H.A.O.N., zebra.
- LIFE SCIENCES:** Biomimetic, color, DMRVenus, verily, FALZ, 3Dgen, genemay, hato, Innanym, 3Dgen, genemay, hato, insitro, SOPHIA, 3Dgen, Cerevel, helix, vaxxco, moderna.
- TRANSPORTATION:** uber, tesla, NIRO, APTIV, Aeraon, Auroa, nauto, ADAPTIVE, AERO, NEO, OPTIMUS, G7, neo, xosha, xosha, moovit, Daimler, Daimler, Daimler, Daimler.
- AGRICULTURE:** FARMER, Granular, BLUE DRAGON, BLUE RIVER, TARAMIS, prospera, semios.
- INDUSTRIAL:** AVEVA, SIEMENS, LUMINA, KODAK, SCOUT24 TECHNOLOGY, ALICEA, HUNT, ALUMINUM, LUMINA.
- OTHER:** stem, Amper, ByteDance, HUBER, ECOMTECH, ALICEA, LUMINA.

A horizontal collage of logos for various data science and machine learning tools and platforms. The logos are arranged in two rows under category headers.

- TRAINING**: TensorFlow, Caffe, Microsoft Cognitive Toolkit (CNTK), DMLC, Apache Mahout, H2O, DStreams, Spark MLlib, Featurehub, Dataiku, Orange3, Rapidminer, Weka, Rattle, Orange3, Rapidminer, Weka, Rattle.
- SEARCH**: Elasticsearch, Solr, Scaphunk, Sphinx, Elasticsearch, Solr, Scaphunk, Sphinx.
- LOGGING & MONITORING**: Kibana, Logstash, Graylog, Splunk, Prometheus, Grafana, Zabbix, Nagios, PRTG, Solaris, Icinga, Nagios, PRTG, Solaris, Icinga.
- VISUALIZATION**: Superset, Matplotlib, Tableau, PowerBI, QlikView, Looker, Alteryx, SAS Viya, IBM SPSS Modeler, SAP Analytics Cloud, Oracle BI, MicroStrategy, Qlik Sense, SAP Analytics Cloud, Oracle BI, MicroStrategy, Qlik Sense.
- COLLABORATION**: Beaker, JupyterLab, Anaconda, Databricks, AWS SageMaker, Google AI Platform, Azure Machine Learning, IBM Watson Studio, SAP Business Maker, Oracle Data Science, Microsoft Azure Machine Learning, Google AI Platform, AWS SageMaker, IBM Watson Studio, SAP Business Maker, Oracle Data Science.
- SECURITY**: Apache Ranger, Knox, Sentry, SCUMMIO, snyk, SonarQube, Fortify, Checkmarx, Veracode, Synopsys, Black Duck, Snyk, SonarQube, Fortify, Checkmarx, Veracode, Synopsys, Black Duck.

OTHER

- sense360
- DATA.GOV
- IMAGENET
- Leviathan
- Workday Data Science
- APOLLOSCALE
- CRUX

DATA SERVICES

- QUANTUMBLACK
- Booz | Allen | Hamilton
- ElectrifiAI
- DataKind
- fractal
- improplexus
- EXL

INCUBATORS & SCHOOLS

- PAWNLIGHT
- DataCamp
- galvanize
- The Data Incubator

DATA RESOURCES

- GENERAL ASSEMBLY
- DataElixir
- DATA ELIXIR
- THE DATA INCUBATOR

RESEARCH

- OpenAI
- facebook research
- MIRI
- VECTOR INSTITUTE
- AIR2

Data & IA Landscape 2020

INFRASTRUCTURE

STORAGE

HADOOP

DATA LAKES

DATA WAREHOUSES

STREAMING / IN-MEMORY

ANALYTICS & MACHINE INTELLIGENCE

BI PLATFORMS

VISUALIZATION

DATA ANALYST PLATFORMS

APPLICATIONS - ENTERPRISE

SALES

MARKETING - B2B

MARKETING - B2C

CUSTOMER EXPERIENCE / SERVICE

HUMAN CAPITAL

FRAMEWORKS

NoSQL DATABASES

NewSQL DATABASES

ETL / DATA TRANSFORMATION

DATA INTEGRATION

MGMT / MONITORING

DATA GENERATION & LABELING

hadoop HDFS

Spark

hadoop MapReduce

ELI5

Flink

YARN

TEZ

kubernetes

Mesos

docker

CDAP

Red Hat

HighAvailability

HELIX

QUERY / DATA FLOW

Spark SQL

presto

SLAMDATA

Flink

Traffodion

GraphQL

Google Cloud Dataflow

HAWQ

DATA ACCESS & DATABASES

cassandra

mongoDB

redis

CouchDB

Cockroach Labs

riak

Amundsen

druid

Uber Databook

Dataportal

Data Hub

SciDB

Cloud Spanner

OPENTSOB

accumulo

HBASE

EDGE DB

FRAMEWORKS

QUERY / DATA FLOW

DATA ACCESS & DATABASES

ORCHESTRATION & PIPELINES

STREAMING & MESSAGING

STAT TOOLS & LANGUAGES

AI OPS & INFRA

AI / MACHINE LEARNING / DEEP LEARNING

SEARCH

LOGGING & MONITORING

VISUALIZATION

COLLABORATION

SECURITY

DATA SOURCES & APIs

DATA MARKETPLACES & DISCOVERY

FINANCIAL & ECONOMIC DATA

AIR / SPACE / SEA

PEOPLE / ENTITIES

LOCATION INTELLIGENCE

OTHER

DATA RESOURCES

DATA SERVICES

INCUBATORS & SCHOOLS

RESEARCH

ANALYTICS & MACHINE INTELLIGENCE

ANALYTICS & MACHINE INTELLIGENCE

ANALYTICS & MACHINE INTELLIGENCE

SALES	MARKETING - B2B	MARKETING - B2C	CUSTOMER EXPERIENCE / SERVICE	HUMAN CAPITAL
          	       	              	                        	      

[illegible]

⇒ **Leitura - Resilience and Vibrancy: The 2020 Data & AI Landscape**
<https://mattturck.com/data2020/>

2020 Data & AI Landscape

HEALTHCARE	LIFE SCIENCES	TRANSPORTATION	AGRICULTURE	INDUSTRIAL	OTHER
 Flatiron  30Med  AICure  Olive  Health  623Q  Xerus  PethAI  Alcure  Nimble  Lumicore  eribic  HAKR  Teboro  MetabioTA  Tempus  Pheix  Spring Health  Spring Health  Spring Health	 color  Denuvus  Equis  Flanigan  Genetix  Insitro  Genentech  Novartis  Roche  Novartis  Novartis  Novartis  Novartis  Novartis  Novartis  Novartis  Novartis  Novartis	 Uber  Nuro  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv  Aptiv	 Farmers Edge  Granular  John Deere  Blue River  Taranis  Prospero  Sermos  Sermos  Sermos  Sermos  Sermos  Sermos	 Avaya  Siemens  Siemens  Siemens  Siemens  Siemens  Siemens  Siemens  Siemens  Siemens  Siemens  Siemens	 Slalom  Bytedance  Veritas  Veritas  Veritas  Veritas  Veritas  Veritas  Veritas  Veritas  Veritas  Veritas

OPEN SOURCE

[illegible]

DATA SOURCES & APIs

Category	Percentage
DATA MARKETPLACES & DISCOVERY	10%
FINANCIAL & ECONOMIC DATA	15%
DATA SOURCES & APIs	25%
PEOPLE / ENTITIES	15%
LOCATION INTELLIGENCE	10%
OTHER	25%

- DATA RESOURCES

DATA SERVICES

- QUANTUMBLACK
- Booz | Allen | Hamilton
- Electrifi
- fractal
- DATAkind
- innopolis
- TEC
- TECHNICAL

INCUBATORS & SCHOOLS

- FULLSIGHT
- General Assembly
- DataCamp
- Data:live
- NETIS
- INSIGHT
- The Data Incubator

DATA RESOURCES

- OpenAI
- facebook research
- MIRI
- VECTOR INSTITUTE
- CRAIL
- AI2
- ALLEN INSTITUTE FOR ARTIFICIAL INTELLIGENCE