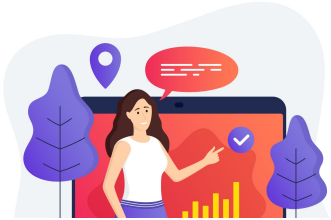
Data Analytics

WHAT IS ANALYTICS



What do they mean by analytics?

A basic definition of analytics

Analytics is **a field of computer science that uses math, statistics, and machine learning to find meaningful patterns in data**. Analytics – or data analytics – involves sifting through massive data sets to discover, interpret, and share new insights and knowledge.


amazon
Uber
The world is changing
NETFLIX

1 How much data is generated every minute?

Source: Domo

41,666,667
messages shared
by WhatsApp users

1,388,889
video / voice calls made
by people worldwide

404,444
hours of video streamed
by Netflix users

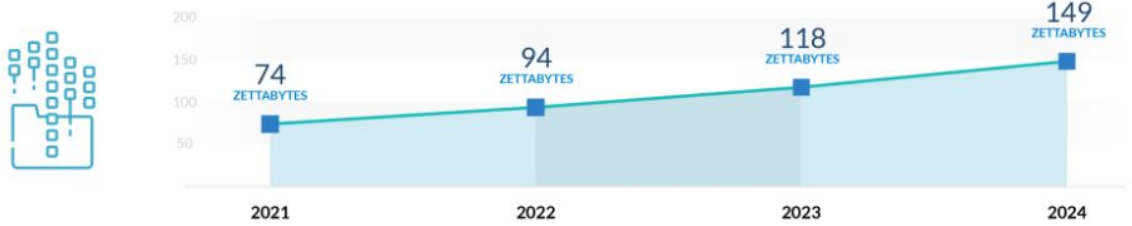
347,222
stories posted by Instagram users

150,000
messages shared by Facebook users

147,000
photos shared by Facebook users

2 Estimated Data Consumption from 2021 to 2024

Source: IDC / Statista



3 Data Growth in 2021

Sources: TechJury, Internet Live Stats, Cisco, PurpleSec

2 TRILLION
searches on Google by the end of 2021

1.134 TRILLION MB
volume of data created every day

3,026,626
emails sent every second, 67% of which are spam

278,108 PETABYTES
global IP data per month by the end of 2021

230,000
new malware versions created every day

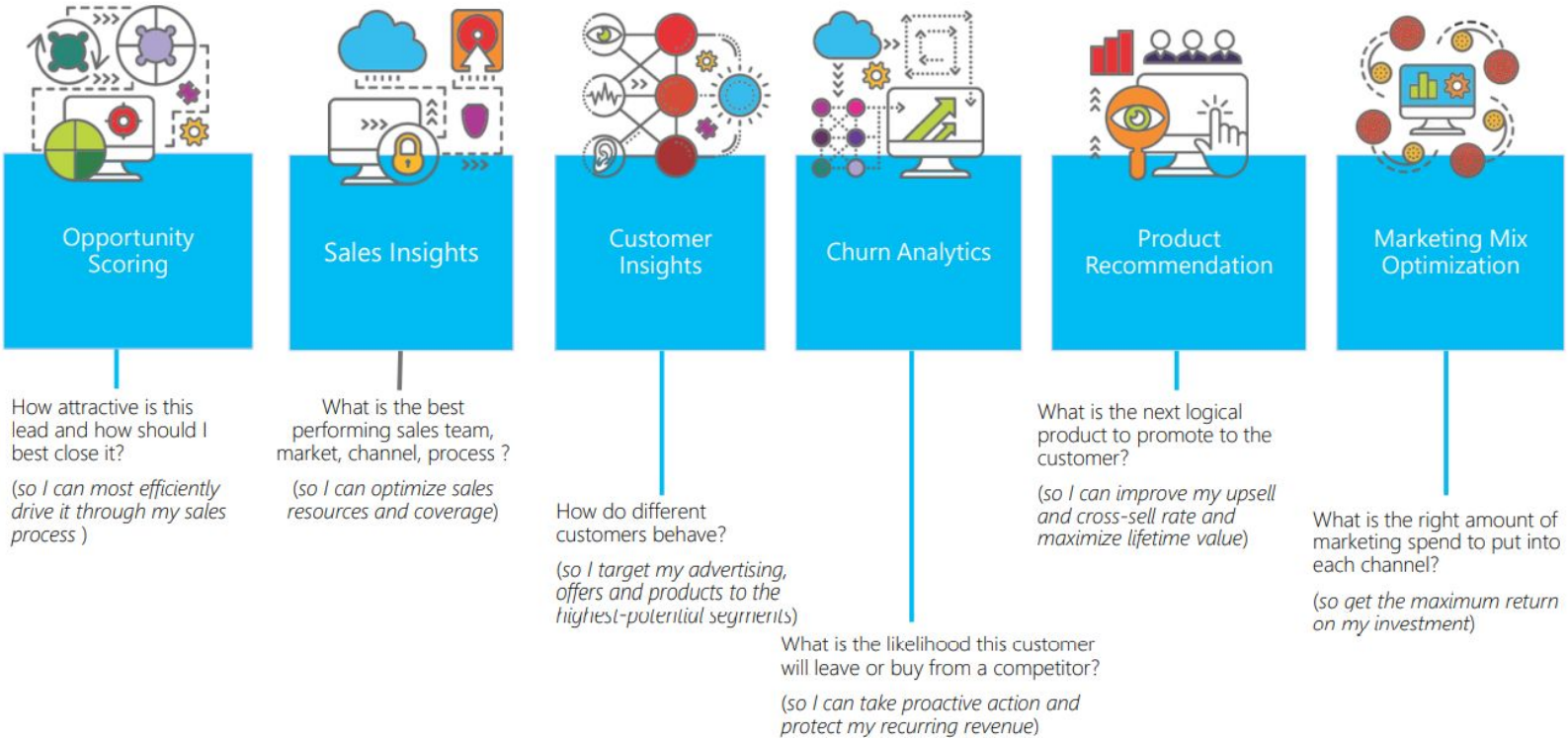
82%
share of video in total global internet
traffic at the end of 2021

Data Analysis

In technical terms **Data analysis** is a process of inspecting, cleansing, transforming and modeling data with the goal of discovering useful information, informing conclusions and supporting decision-making.

Applications that transform Sales

- High-yield use cases:



Applications that transform Finance

High-yield use cases:

Forecasting

What are the range of outcomes I can expect?

(*so I can most efficiently plan and allocate resources*)

Fraud

Which transactions are suspicious?

(*so I can efficiently monitor and prevent it*)

Risk

How risky is this customer or project ?

(*so I can prepare for and mitigate adverse events*)

Auditing

Current State

We are overwhelmed by the amount of data our organization must manage, process, and analyze.

We do not have the IT capabilities to integrate, store and process the data coming from many sources.

Desired Future State

We can manage and process unlimited amounts of data, providing the insight we need to align our strategy and execution.

We can easily integrate data from different sources cutting across sales, finance, payroll, healthcare charge, supply chain, marketing and operations…

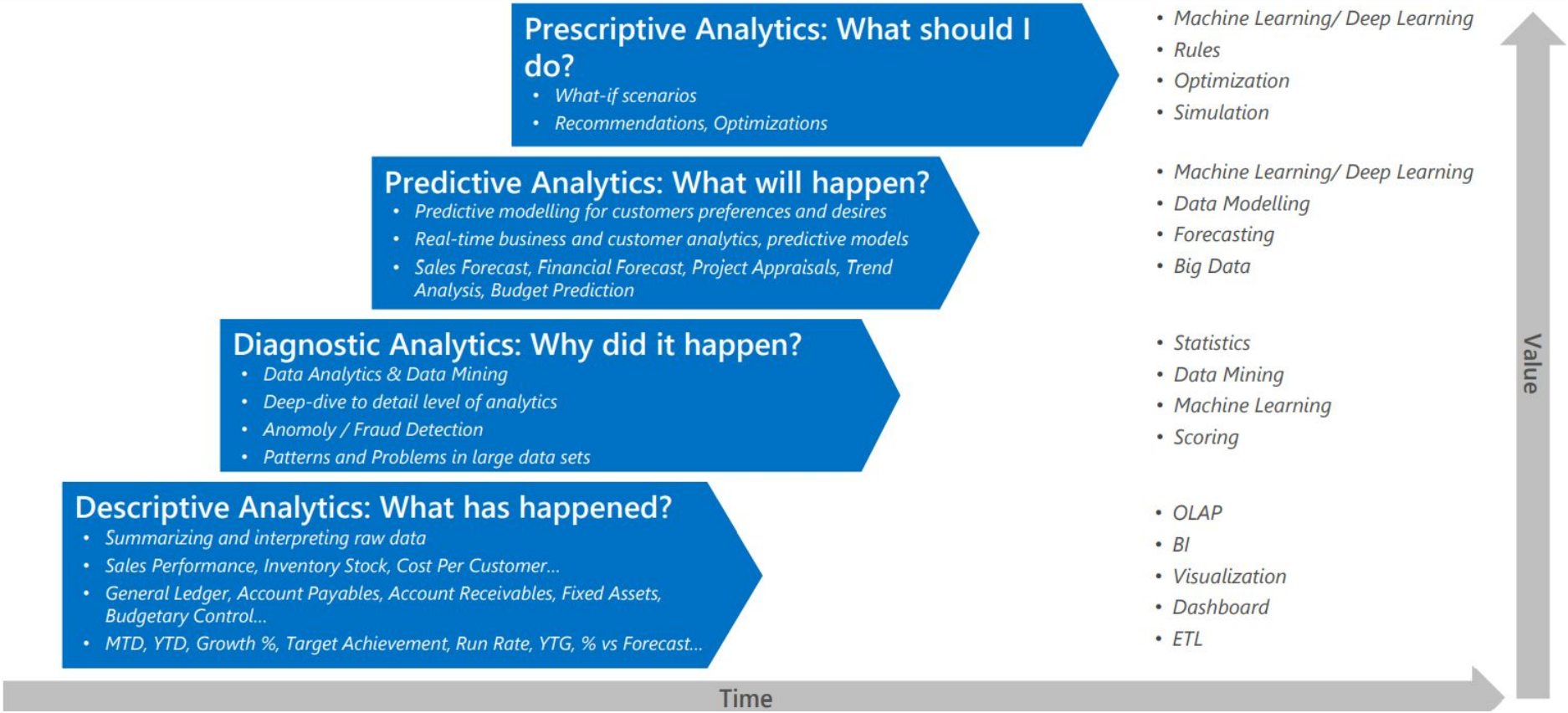
Our focus has shifted from controlling data to being the steward of data and insights.

Auditor

We are also looking to…

- Generate strategic analyses, enterprise risk assessments and business insights
- Develop internal audit plans and suggest quality improvements
- Create predictions and warnings for C levels and stakeholders

Accounting


Prescriptive Analytics: What should I do?
• What-if scenarios
• Recommendations, Optimizations
• Machine Learning/ Deep Learning
• Rules
• Optimization
• Simulation
Predictive Analytics: What will happen?
• Predictive modelling for customers preferences and desires
• Real-time business and customer analytics, predictive models
• Sales Forecast, Financial Forecast, Project Appraisals, Trend Analysis, Budget Prediction
• Machine Learning/ Deep Learning
• Data Modelling
• Forecasting
• Big Data
Diagnostic Analytics: Why did it happen?
• Data Analytics & Data Mining
• Deep-dive to detail level of analytics
• Anomoly / Fraud Detection
• Patterns and Problems in large data sets
• Statistics
• Data Mining
• Machine Learning
• Scoring
Descriptive Analytics: What has happened?
• Summarizing and interpreting raw data
• Sales Performance, Inventory Stock, Cost Per Customer...
• General Ledger, Account Payables, Account Receivables, Fixed Assets, Budgetary Control...
• MTD, YTD, Growth %, Target Achievement, Run Rate, YTG, % vs Forecast...
• OLAP
• BI
• Visualization
• Dashboard
• ETL
Value
Time

MUCH MORE THAN DASHBOARDS

SUMMARY OF TOOLS

DASHBOARDS & REPORTING

"Visible"

PLUMBING

Under the surface

Significant time, resources and expertise requires

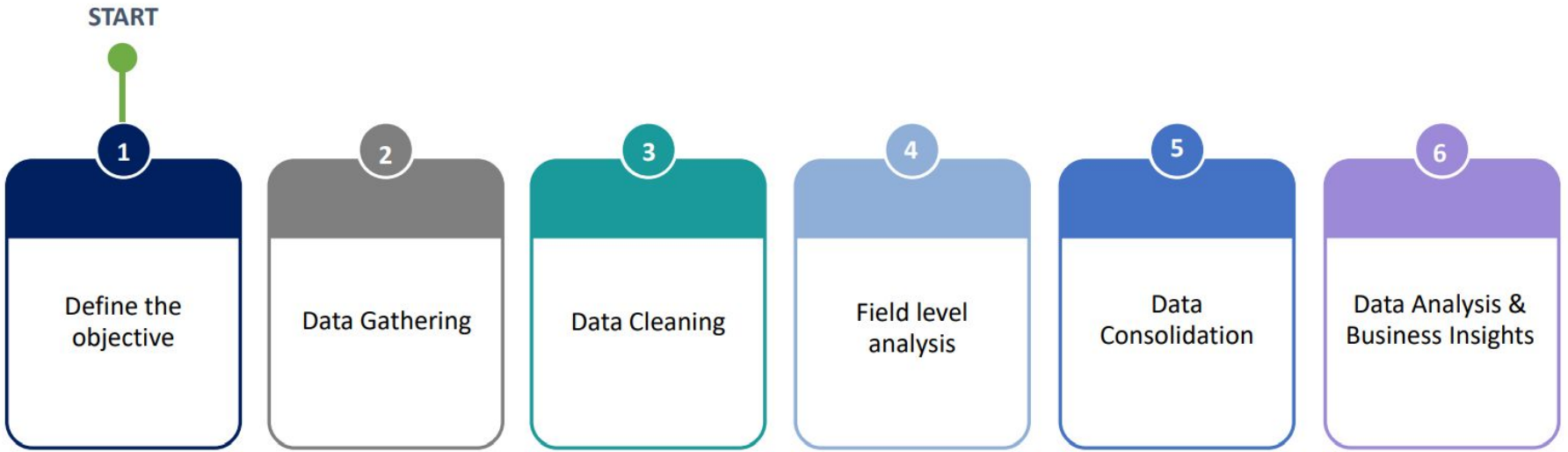
Data Warehouse/ Data Lake Design

Data Mapping & Transformation from Multiple Applications

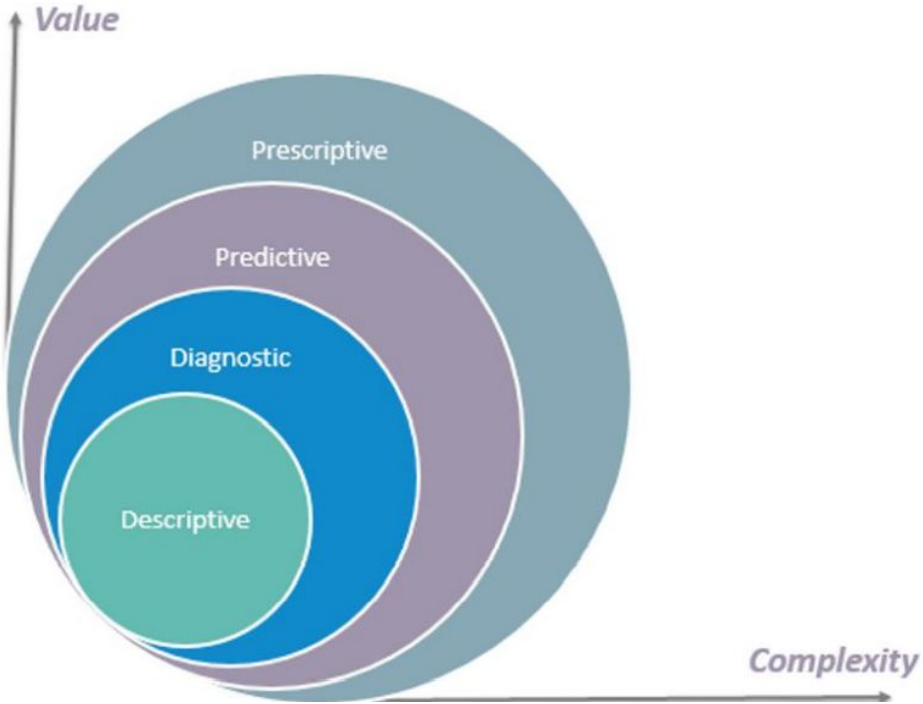
Analytic Models, KPIs, Metrics

Much More…

Data Analytics Process



What is Analytics ?



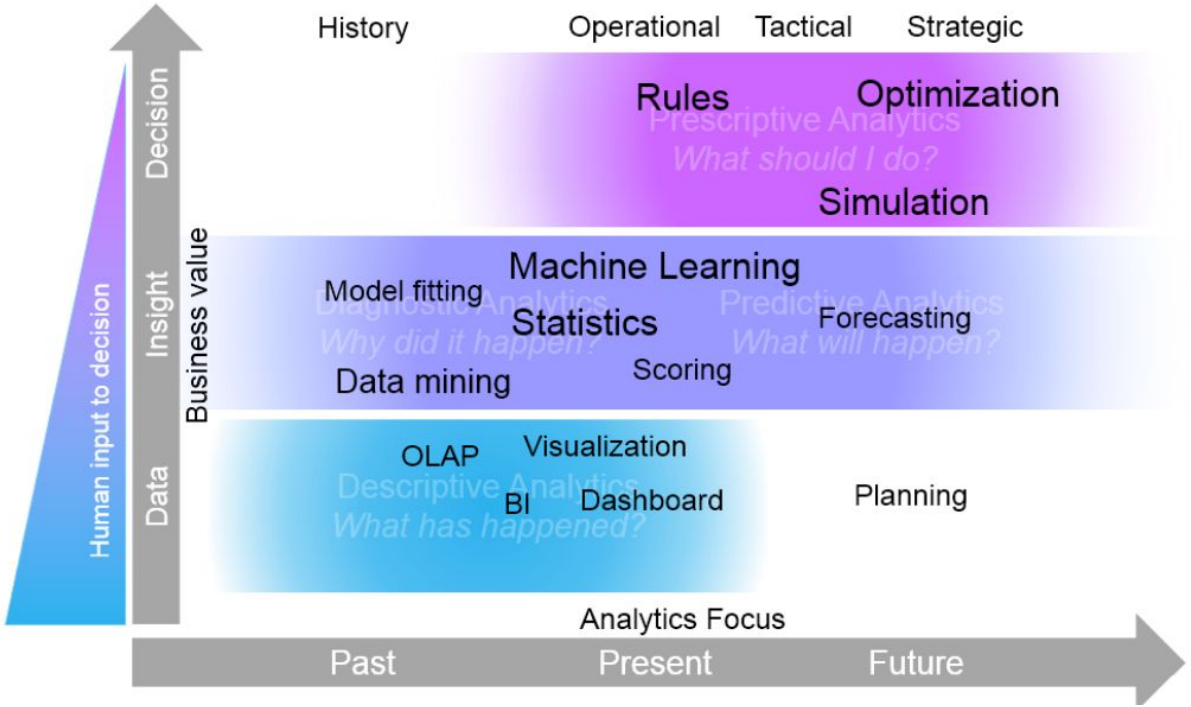
Descriptive: What's happening in my business?
- Comprehensive, accurate and live data
- Effective visualisation

Diagnostic: Why is it happening?
- Ability to drill down to the root-cause
- Ability to isolate all confounding information

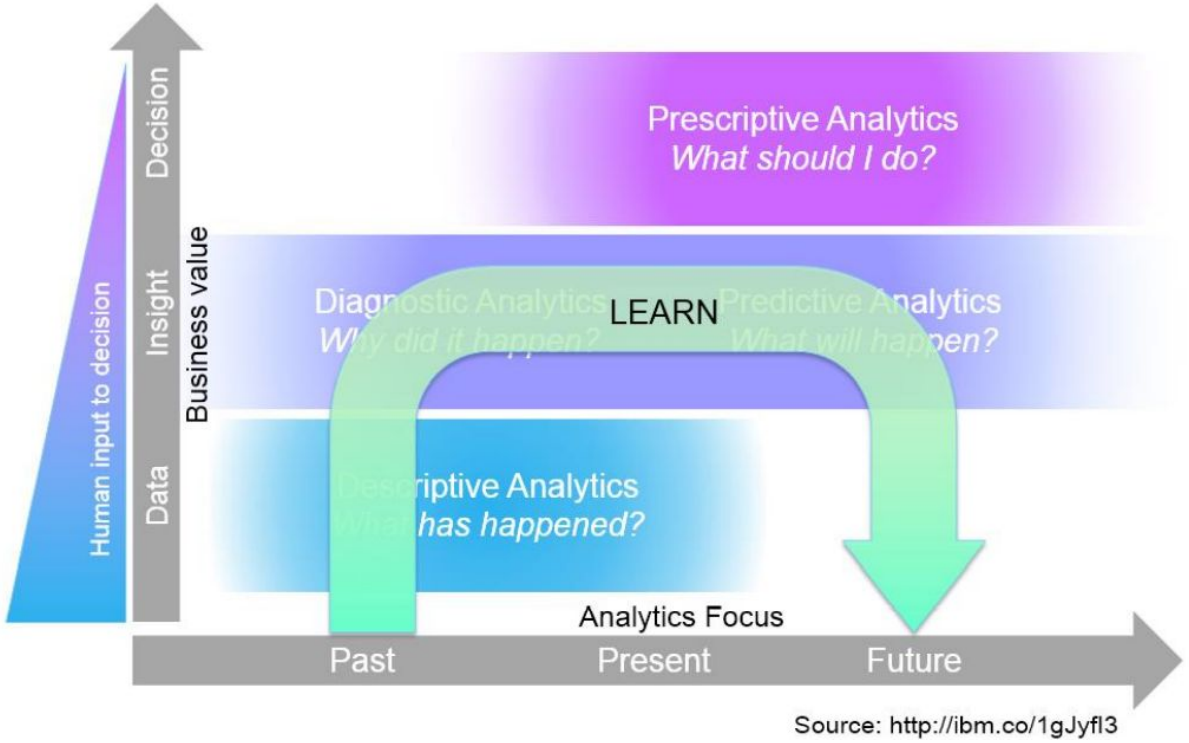
Predictive What's likely to happen?
- Business strategies have remained fairly consistent over time
- Historical patterdns being used to predict specific outcomes using algorithms
- Decisions are automated using algorithms and technology

Prescriptive: What do I need to do?
- Recommended actions and strategies based on champion/ challenger testing strategy outcomes
- Applying advanced analytical techiniques to make specific recommendations

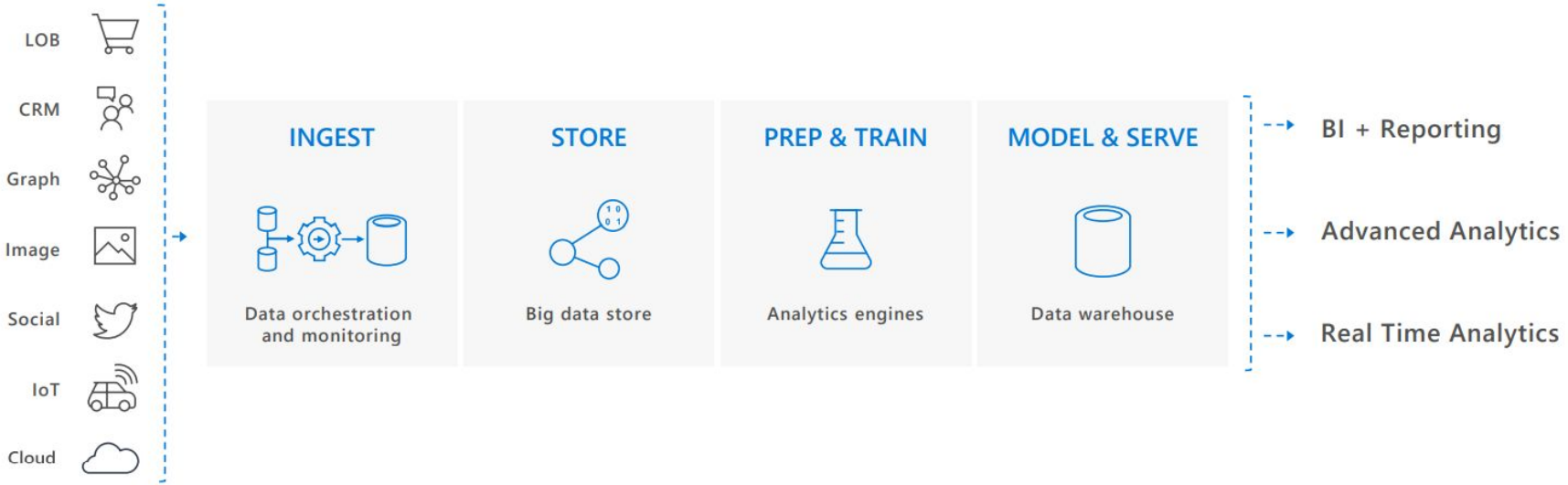
What is Data Analytics ?



What is Data Analytics ?



ANALYTICS SCENARIOS



JUST ANALYTICS

DAILY SELL OUT

SELL OUT	SELL OUT YTD	GROWTH % vs LM	GROWTH % vs LY	TARGET ACHIEVEMENT	ASO
21,442K	18,379K	-54.60%	-38.47%	60.05%	674K

DATE
- 2019
- 2020
  - Jan
  - Feb
  - Mar
  - Apr
  - May
  - Jun
  - Jul
  - Aug
  - Sep
  - Oct
  - Nov
  - Dec

PROVINCE: All

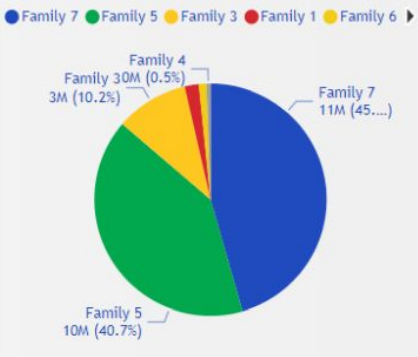
DISTRIBUTOR: All

PRODUCT FAMILY: All

SELLOUT TRENDING

SELLOUT BY CHANNEL

SELLOUT BY PRODUCT



Province	Sell Out Qty	ADS	ASO	ASO Growth % vs LY	Target	vs Target	Remaining vs Target	Remaining Avg/Day	YTD	Growth % vs LM	Growth % vs LY	Sell Out / Sell In
Hồ Chí Minh	4,466,153	259,325.01	61,628	114%	8,910,042.74	70.17%	3,550,659.14	1,775,329.57	5,359,384	-36.53%	-38.46%	77.94%
Phú Yên	3,582,723	115,571.69	74,496	92%	5,102,232.99	50.16%	2,031,327.99	247,937.75	4,094,540	-36.65%	-7.57%	74.69%
Quảng Trị	3,582,679	77,046.85	37,920	114%	3,985,436.31	49.94%	1,995,059.31	599,454.25	1,990,377	-27.16%	-48.72%	85.64%
Hà Nội	1,661,643	64,321.65	101,558	105%	3,315,523.12	90.21%	1,653,880.62	328,447.56	1,993,971	-18.71%	-7.63%	54.32%
Bình Dương	1,020,451	27,431.48	31,115	92%	1,709,586.28	69.64%	178,909.48	429,605.14	1,020,451	-26.88%	-48.83%	49.40%
Yên Bái	987,665	50,976.26	54,327	114%	1,978,487.60	59.90%	200,690.60	297,878.30	1,580,264	-53.90%	-48.54%	140.74%
Đà Nẵng	953,736	39,555.87	25,470	114%	1,366,488.80	49.85%	685,248.80	342,624.40	953,736	-18.84%	-39.03%	72.09%
Huế	816,267	26,331.18	34,650	91%	915,450.45	49.54%	189,880.05	185,636.33	544,178	-17.97%	-8.79%	76.63%
Long An	590,303	11,901.27	44,496	98%	736,527.29	60.11%	293,799.89	73,112.05	516,515	-37.00%	-7.80%	92.58%
Hải Phòng	572,847	24,638.58	34,900	117%	949,819.75	60.31%	90,549.25	236,223.63	477,373	-26.27%	-17.78%	64.01%
Cà Mau	430,997	13,903.11	24,205	113%	482,517.85	49.62%	51,521.35	121,537.67	239,443	-53.80%	-38.47%	97.40%
An Giang	396,232	17,894.33	52,430	117%	789,989.64	70.22%	235,265.54	78,009.62	633,970	-20.11%	-17.54%	83.38%
Total	**24,505,688**	**889,319.32**	**745,625**	**105%**	**30,607,868.91**	**50.04%**	**15,291,813.91**	**4,582,695.96**	**27,568,899**	**-36.45%**	**-48.73%**	**90.40%**

CROSS SELLING DEEP DIVE

23/05/2016 6/06/2018

Product Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Product 1		22%	21%	19%	21%	14%	7%	6%	5%	10%	7%	8%	25%	27%
Product 2	21%		17%	21%	22%	15%	6%	9%	6%	8%	7%	9%	25%	25%
Product 3	23%	20%		22%	23%	14%	7%	8%	6%	6%	8%	8%	26%	27%
Product 4	20%	23%	21%		22%	14%	7%	8%	7%	8%	7%	9%	26%	26%
Product 5	20%	22%	20%	20%		18%	8%	8%	6%	7%	10%	7%	25%	26%
Product 6	19%	21%	17%	18%	26%		7%	9%	6%	7%	8%	10%	24%	28%
Product 7	20%	17%	18%	18%	26%	14%		11%	5%	7%	6%	7%	29%	26%
Product 8	18%	25%	21%	21%	24%	19%	11%		5%	8%	4%	8%	24%	27%
Product 9	20%	24%	22%	23%	24%	16%	6%	7%		9%	9%	9%	36%	16%
Product 10	27%	23%	16%	21%	19%	14%	7%	8%	7%		4%	9%	26%	25%
Product 11	19%	22%	21%	18%	30%	17%	6%	4%	7%	4%		9%	26%	25%
Product 12	20%	22%	17%	22%	18%	18%	6%	7%	6%	8%	8%		24%	30%
Product 13	20%	21%	19%	20%	22%	14%	8%	7%	8%	7%	8%	8%		27%
Product 14	21%	21%	19%	19%	22%	16%	7%	8%	3%	7%	7%	9%	27%	
Product 15	20%	20%	19%	17%	23%	15%	6%	7%	6%	8%	7%	9%	26%	26%
Product 16	23%	24%	20%	19%	21%	17%	5%	6%	7%	9%	9%	8%	26%	27%
Product 17	20%	24%	21%	20%	21%	16%	9%	8%	6%	6%	7%	7%	25%	28%
Product 18	22%	24%	22%	21%	26%	16%	6%	7%	8%	7%	8%	8%	26%	29%
Product 19	19%	25%	20%	20%	21%	17%	9%	7%	6%	7%	7%	7%	28%	27%

Product Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Product 1		164	159	141	160	105	52	48	40	74	53	60	189	201	183	120
Product 2	164		136	166	172	117	45	68	48	61	59	68	197	203	184	121
Product 3	159	136		151	157	94	48	56	45	42	56	53	182	187	170	103
Product 4	141	166	151		159	101	47	57	47	57	48	66	185	185	155	99
Product 5	160	172	157	159		143	67	65	49	52	83	56	202	211	204	110
Product 6	105	117	94	101	143		37	51	33	37	46	54	135	158	135	85
Product 7	52	45	48	47	67	37		29	12	18	16	17	76	67	58	28
Product 8	48	68	56	57	65	51	29		14	22	12	22	64	73	62	32
Product 9	40	48	45	47	49	33	12	14		18	18	19	74	33	51	35
Product 10	74	61	42	57	52	37	18	22	18		12	23	70	67	71	44
Product 11	53	59	56	48	83	46	16	12	18	12		24	72	69	67	46
Product 12	60	68	53	66	56	54	17	22	19	23	24		74	91	79	42
Product 13	189	197	182	185	202	135	76	64	74	70	72	74		256	239	132
Product 14	201	203	187	185	211	158	67	73	33	67	69	91	256		234	136
Product 15	183	184	170	155	204	135	58	62	51	71	67	79	239	234		154
Product 16	120	121	103	99	110	85	28	32	35	44	46	42	132	136	154	
Product 17	101	125	106	104	109	85	44	42	30	30	37	34	128	143	133	66
Product 18	116	129	118	110	135	86	31	39	41	39	41	44	135	151	119	71
Product 19	90	117	96	96	101	79	43	33	28	33	33	35	131	125	116	75

Customer Names	Total Sales	Total Product Purchases
Timothy Barnes	$390,811.00	10
Russell Grant	$365,860.20	8
Gerald Porter	$350,778.50	11
Frank Larson	$341,331.50	9
Johnny Willis	$321,807.70	8
Martin Montgomery	$321,265.00	10
Raymond Cruz	$297,272.30	9
Roger Morrison	$296,743.00	11
Justin Cook	$291,959.20	7
Walter Gonzalez	$285,788.50	9
Total	**$346,326,691.70**	**30**

Expense Management

Department

All

Region

All

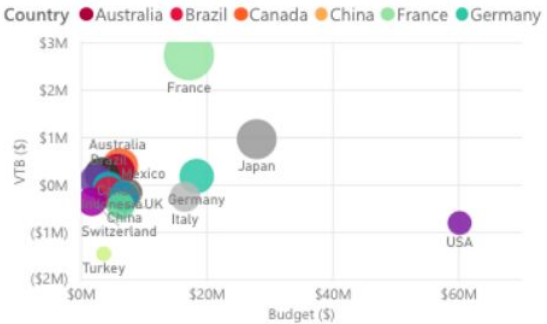
Country

All

Actual ($)	VTB ($)	VTB (%)
$199.5M	$858K	-1.5%

Expense Variance to budget ($) by channel

Expense Variance to budget ($) by country

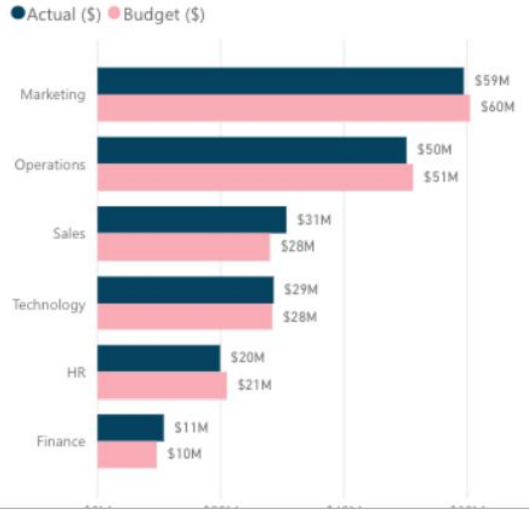


Expense Variance to budget ($) by categ...

Expenses by country

Country	Actual ($)	Budget ($)	YoY ($)	VTB ($)	Average of V
USA	$59,312,555	$60,116,557	$3,763,526.12	($804,001....	-0.9
Japan	$28,853,365	$27,882,355	$1,005,189.70	$971,009.97	-2.7
France	$19,861,429	$17,114,641	$3,709,170.19	$2,746,788...	2.7
Germany	$18,539,070	$18,346,053	$4,490,723.88	$193,017.14	3.7
Italy	$16,257,004	$16,501,948	$2,177,082.95	($244,944....	-8.6
Korea	$7,077,088	$7,237,068	$1,954,752.98	($159,980....	-6.8
China	$6,667,457	$6,896,030	$3,415,836.32	($228,572....	-2.7
Canada	$6,600,430	$6,192,896	$2,626,420.07	$407,533.73	-2.7
India	$6,524,969	$6,757,166	$3,699,399.65	($232,196....	2.5
Australia	$6,050,180	$5,749,313	$4,490,197.59	$300,867.07	5.5
UK	$5,835,903	$6,268,326	$2,353,720.76	($432,423....	2.5
Spain	$4,282,540	$4,325,230	$1,917,568.44	($42,690.14)	-6.9
Brazil	$4,236,666	$4,371,828	$4,031,019.94	($135,162....	-0.5
Mexico	$3,504,258	$3,279,243	$2,788,697.23	$225,015.79	-0.5
Indonesia	$2,446,189	$2,351,443	$2,068,810.56	$94,745.68	-3.4
Total	**$199,510,...**	**$198,651,6...**	**$48,836,7...**	**$858,471.53**	**-1.9**

Actual ($) and Budget ($) by department



Introduction to Data Warehouses

What is Data WareHousing

A **Data Warehousing** (DW) is process for collecting and managing data from varied sources to provide meaningful business insights

Two purposes

- What's the best category?
- How many sales compared to last month?
- What can be improved?

⇒ **Evaluate performance**
⇒ **Decision-making**

Analytical decision making

Operational data keeping

- Receive orders
- React to complaints
- Fill up stock

⇒ **Turn the weel**

OLAP = Online Analytical Processing

OLTP = Online Transactional Processing

Two requirements

- Thousands of records at a time
- Fast query performance
- Historical context
- Usablity

Analytical decision making

Operational data keeping

- One record at a time
- Data input
- No long history

DWH is there to address those analytical data needs!

Two requirements

- Thousands of records at a ti
- Fast
- Hist
- Usa

- One record at a time

Used for reporting and data analysis

DWH is there to address those analytical data needs!

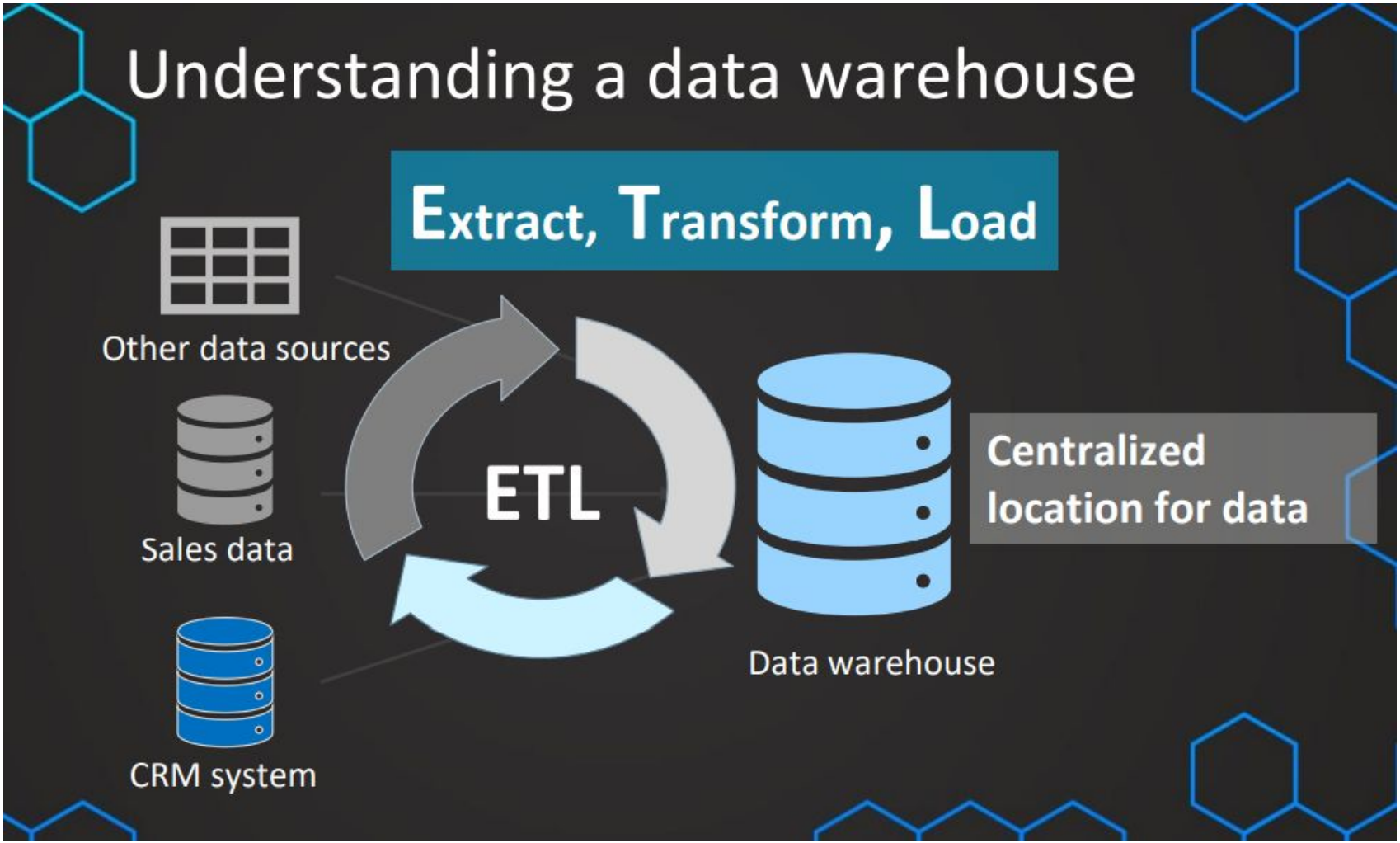
What is a data warehouse?

Data warehouse:

A database used and optimized for analytical purposes.

- ✓ **User friendly**
- ✓ **Fast query performance**
- ✓ **Enabling data analysis**

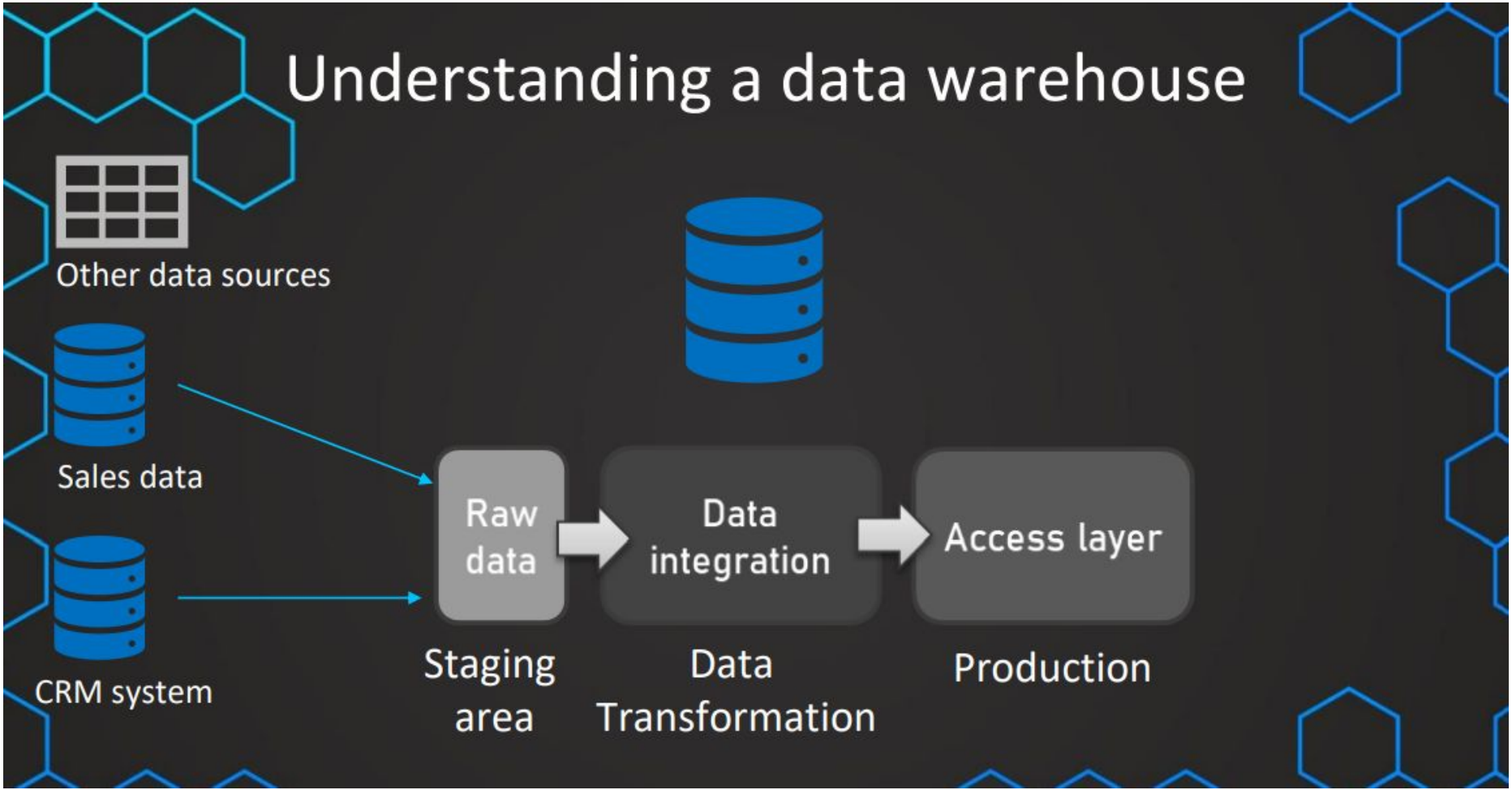
Understanding a data warehouse



Goals of a data warehouse

- ✓ Centralized and consistent location for data
- ✓ Data must be accessible fast (query performance)
- ✓ User-friendly (easy to understand)
- ✓ Must load data consistently and repeatedly (ETL)
- ✓ Reporting and data visualization built on top

Understanding a data warehouse



We create a data warehouse for Business Intelligence...

What is Business Intelligence?

What is Business Intelligence?

Strategies

Technologies

Infrastructures

Data analysis

- Data gathering
- Data storing
- Reporting
- Data visualization
- Data mining
- Predictive analytics

Data warehouse

Raw data

Transform

Meaningful insights

Better decisions

What is Business Intelligence?

BI(Business Intelligence) is a set of processes, architectures, and technologies that convert raw data into meaningful information that drives profitable business actions. It is a suite of software and services to transform data into actionable intelligence and knowledge.

Data Lake or Data Warehouse?

Data Lake or Data Warehouse?

Data lake & data warehouse are

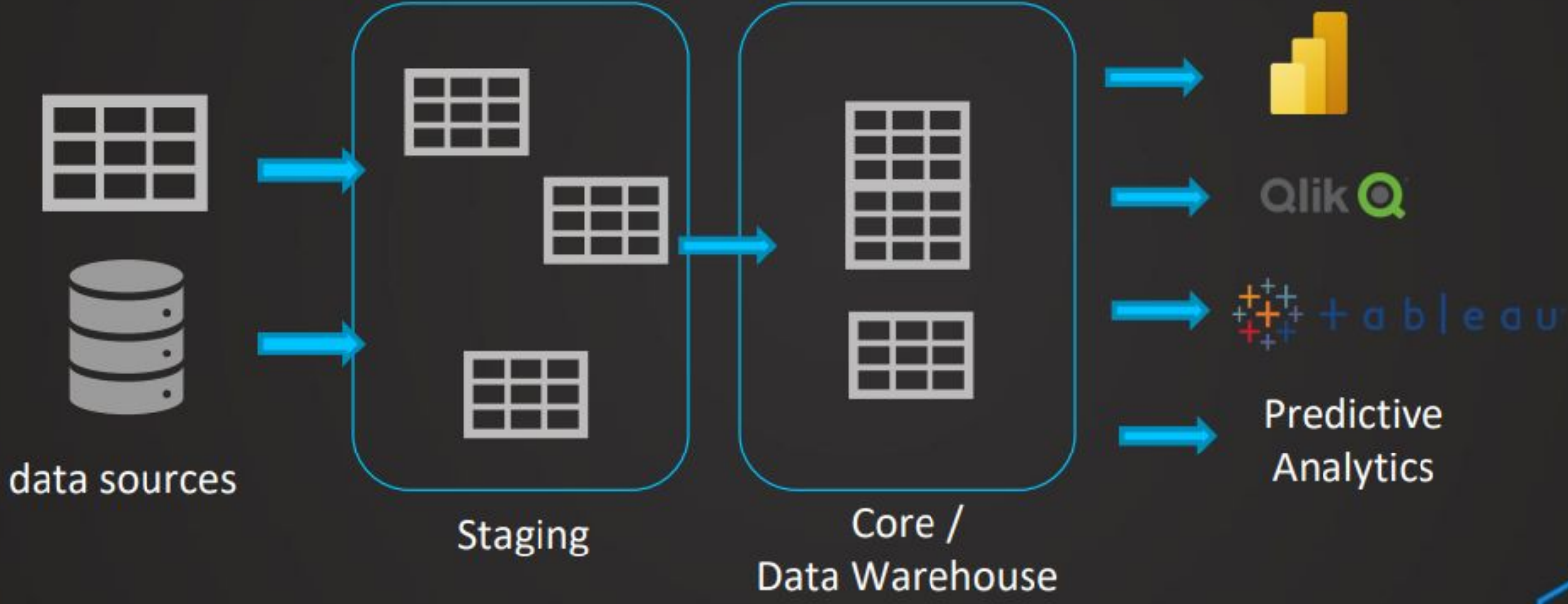
BOTH used as

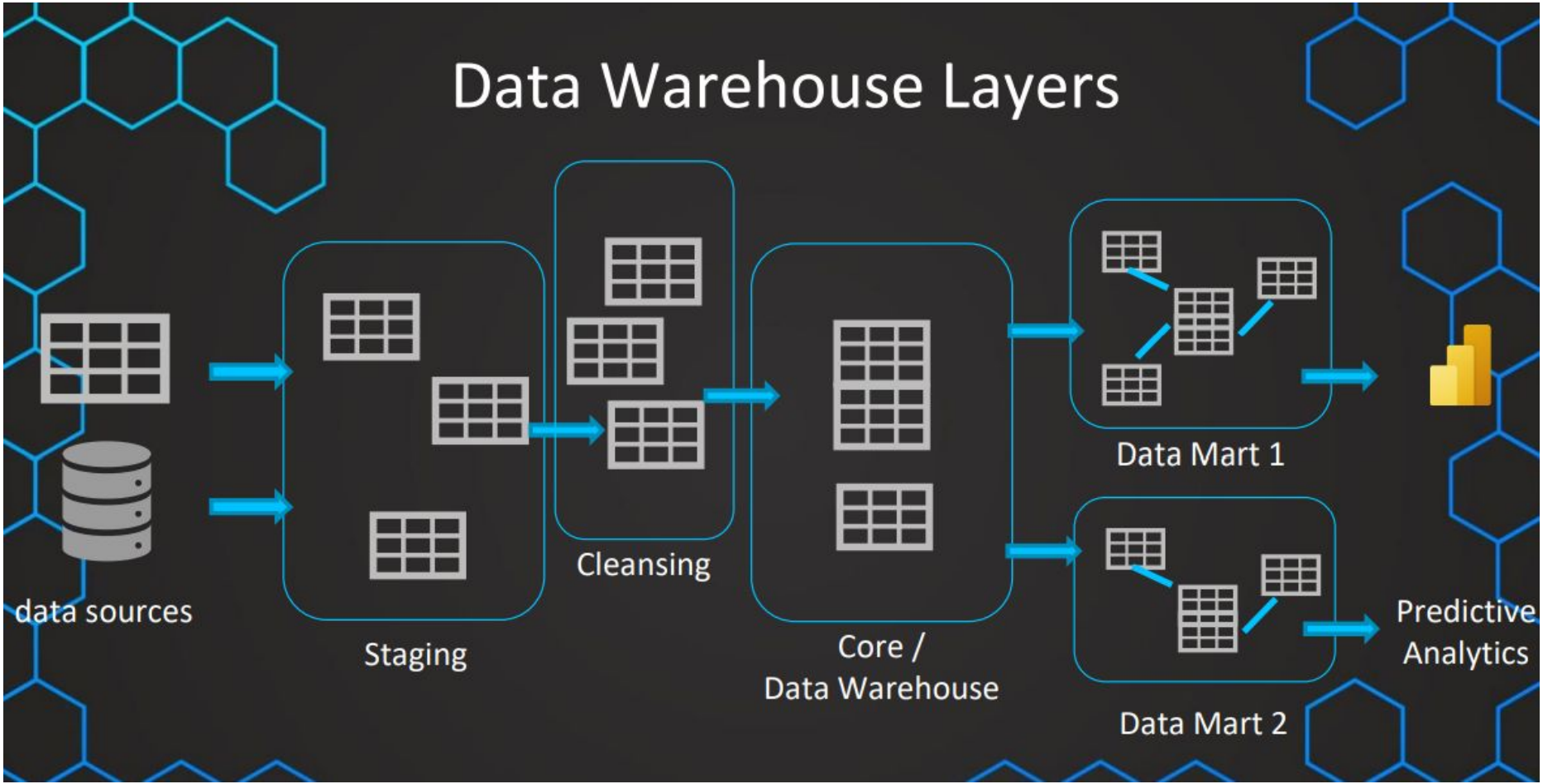
centralized data storage

Data Lake or Data Warehouse?

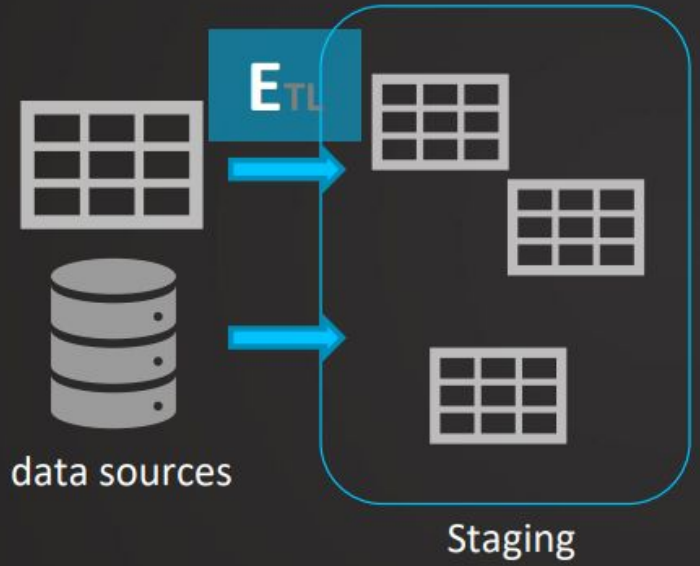
	Data Lake	Data Warehouse
Data	Raw	Processed
Technologies	Big data	Database
Structure	Unstructured	Structured
Usage	Not defined yet	Specific & ready to be used
Users	Data Scientists	Business users & IT

Data Warehouse Layers




Data Warehouse Layers
data sources
Staging
Cleansing
Core /
Data Warehouse
Data Mart 1
Data Mart 2
Predictive
Analytics

Data Warehouse Layers

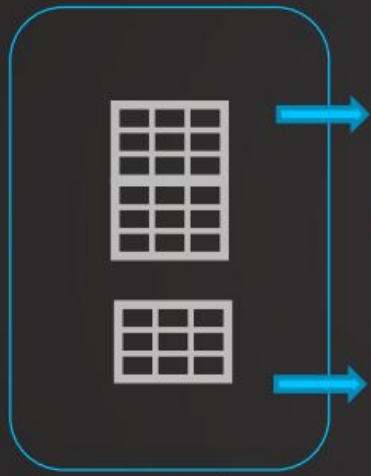


- *"Short time on the source systems"*
- *"Quickly extract"*
- Move the data into relational database
- Start transformations from there

The Staging Layer

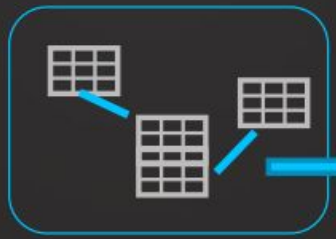
- ✓ Staging Layer is the landing zone extracted data
- ✓ Data in tables and on a separate database
- ✓ As little "touching" as possible
- ✓ We don't charge the source systems
- ✓ Temporary or Persistant Staging Layers

Data Warehouse Layers



Core /
Access Layer /
Data Warehouse

Data Mart 1



Data Mart 2

Predictive
Analytics

- Subset of a DWH
- Dimensional Model
- Can be further aggregated

- *Usability + Acceptance*
- *Performance*

- *Tools*
- *Departments*
- *Regions*
- *Use-cases*

Data Marts

- ✓ Data Mart = Small scale DWH?

⇒ Focus on the business problem

- ✓ Should you use a Data Mart or not?

⇒ Focus on the business problem

The different layers

Staging

- ✓ Landing zone
- ✓ Minimal transformation
- ✓ "Stage" the data in tables

Core

- ✓ Always there
- ✓ Business Logic & Single Point of Truth
- ✓ Can be sometimes the access layer

Mart

- ✓ Access Layer
- ✓ Specific to one use-case
- ✓ Optimized for performance


aws