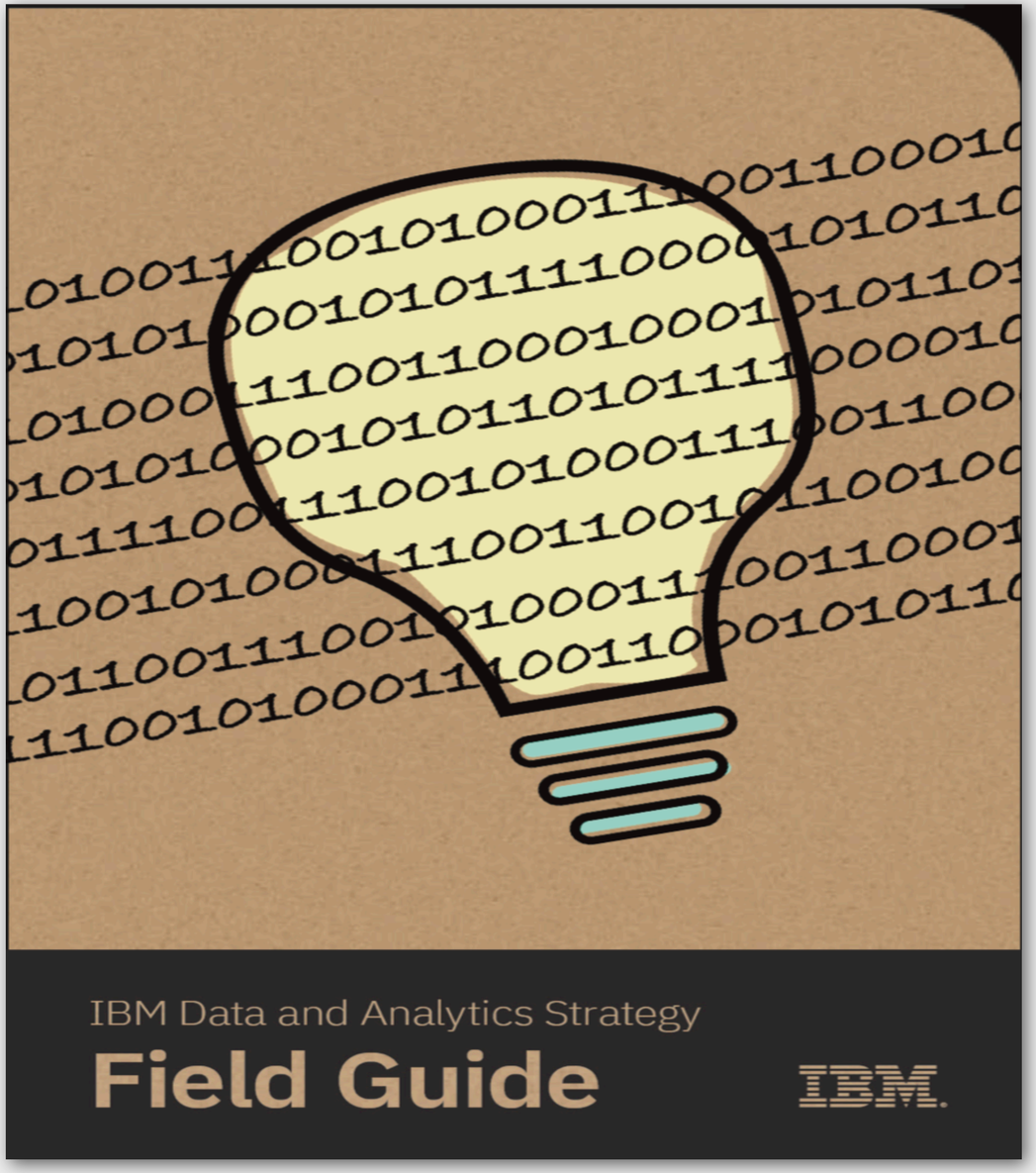
Point of View

Bringing Insights into **Decision Making** with **Data/Analytics** and **AI/Cognitive**

—

Jukka Ruponen
Senior Analytics Architect



ibm.com/cloud/garage/files/data-analytics-field-guide.pdf

Topics addressed:

- Understanding "Business Data" vs "Big Data"
- The Importance of Multi-Cloud Strategy
- About the "I want AI"-phenomenon
- Cornerstones of Business Analytics
- Data Science/AI walkthrough with Watson Studio
- Watson Services - Cognitive building blocks
- Enterprise Insight Platforms with ICP for Data
- Trust & Transparency on AI ("AI Bias")

About

"Business Data" vs "Big Data"

Volume - Scale of the data

The ability to process large amounts of data and what you do with that data.

Variety - Different forms of data

Making sense out of unstructured data by trying to capture all of the data that pertains to our decision-making process.

Velocity - Analysis of streaming data

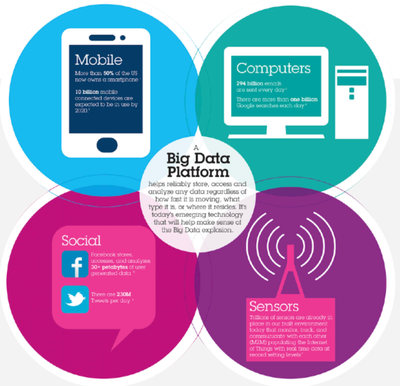
The rate at which data arrives at the enterprise and the time that it takes the enterprise to process and understand that data.

Veracity - Uncertainty of the data

The quality or trustworthiness of the data. The quality or trustworthiness of the data. Tools that help handle big data's veracity discard "noise" and transform the data into trustworthy insights.

Copyright © 2019 IBM Corporation.

“Business Data” vs “Big Data”



"Big data" is used to describe all the data being generated all around us, all the time.
Every device, digital process or social media exchange produce it, networks transmit it, storages persist it and systems process it

What differentiates **big data** from traditional **business data**?

- **Business data** is typically **well structured**, **well known** and **stable**.
  It is created and used in specifically designed business applications and processes (like CRM, SCM and HR systems) and stored and managed in relational databases. This is **facts data.**

- **Big data** is **less structured**, **less known** and **less stable**, because it arrives from disparate new data sources at an alarming **Velocity**, **Volume**, **Variety** and **Veracity**. It is typically created outside of biz-applications and processes, and may not be stored in relational databases. This is **enrichment data.**

To **extract meaningful value out from big data**, you’ll need **processing power**, comprehensive **analytic capabilities** and **skills** in order to **benefit from big data** and **use it to enrich & explain your facts data**.

Copyright (c) IBM Corporation

Example “Insight Pipeline”, describing the way to make sense of data for decision making

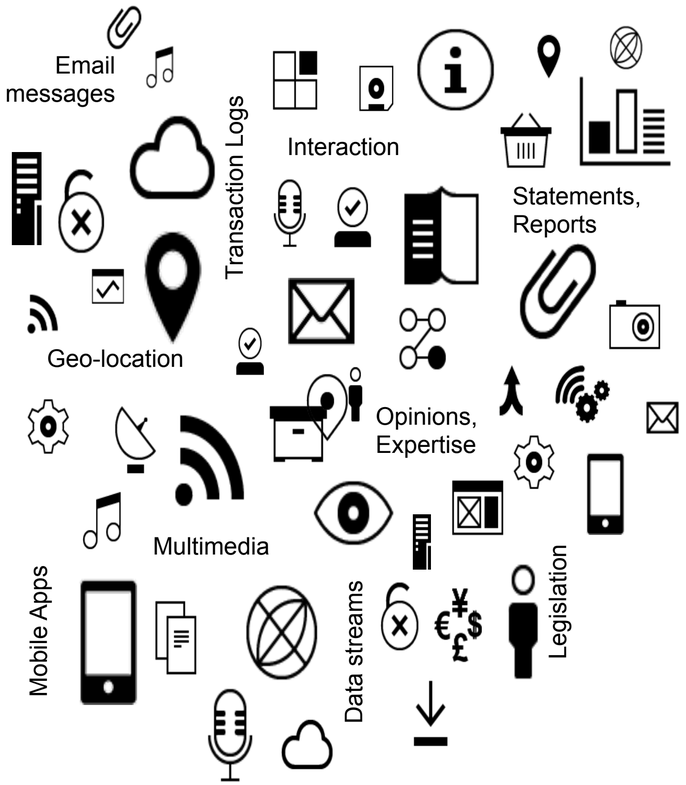
Data/Events Processing & Exploration

Target: Be able to **Discover**, **Explore** and **Process** data at any **Volume**, **Velocity** or **Variety**

Technological capabilities today

Businesses today

Maturity levels



Data, Content, Information

Structured and unstructured data from a variety of sources (persistent or streaming)

Insight Pipeline

> Data discovery
> Data exploration
> Data filtering
> Data cleansing
> Data integration

What data do we have?

What should I be looking for?

What can I ask?

Descriptive Analytics

Business Intelligence

- Reporting
- Forecasting

What has happened?

What I believe should happen?

Predictive Analytics

Advanced Analytics

- Predictions
- Propability

What is likely going to happen, based on history?

Prescriptive Analytics

Automated Decision Management

- Scoring
- Rules
- Optimization

What are my next best actions?

Can I automate these decisions, based on rules?

Cognitive (AI)

Q&A / Recommendations

- Deep Q&A
- Reasoning

Understands natural language and context

Reasoning the answer, based on evidence

Adapts and learns, based on feedback

Assists transparently, but does NOT make the final decision

Robotics (AI)

- Independent
- Decides
- Performs
- Survives
- Human-like

© 2016 IBM Corporation

IBM Big Data & Analytics Hub

ABOUT CONTACT CONTENT BY TYPE SUBSCRIBE

EXPLORE BY TOPIC: USE CASES INDUSTRIES ANALYTICS TECHNOLOGY EVENTS FOR DEVELOPERS BIG DATA & ANALYTICS HEROES

The Latest

GOVERNANCE

BLOG | by Kenneth Duemig

Accelerating time-to-market with fabricated test data

Protecting personal and sensitive data is vital. But, understanding the regulatory environment and available tools is just the first step. There are still challenges when...

ANALYTICS

BLOG | by Roberta Wakerell

IBM Business Intelligence ranks highly... But here's what matters most

The most recent BARC Score report ranks business intelligence (BI) vendors, but more notably states your BI platform must be strategic and able to span...

ANALYTICS

VIDEO | 31:47

Experts on AI Facebook live replay: Fast track your data -- New York City

What will happen to companies who don't embrace data? What do the next five years hold? What's the difference between AI and machine learning? Steve Ardire and Adam...

BIG DATA TECHNOLOGY

PODCAST | 25:49

Making Data Simple: The big data problem

In this first episode of Making Data Simple, we welcome Daniel Hernandez, VP of IBM Analytics Offering Management, who helps us navigate "the big data problem" and shares why he doesn't like the term "big data."

ANALYTICS

BLOG | by Nickolus Plowden

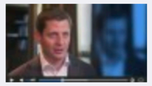
Transforming a city with data and insight

For the first time in human experience, there's the opportunity to transform a city by listening to all of its inhabitants, individually. That's

Most Popular

BLOG
Transforming a city with data and insight

BLOG
Begin your cognitive enterprise journey at DataWorks Summit Sydney



VIDEO
Master your data with data governance

PODCAST
Making Data Simple: The big data problem

VIDEO
Become a data disruptor

BLOG
IBM Business Intelligence ranks highly... But here's what matters most

BLOG
Experts answer your top data science and machine learning questions

Infographics

Show your employees

Presentations

Intelligence in the age of

ibmbigdatahub.com

Real world use cases
Industry insights
Professional's PoVs
Blogs & Videos
Technologies
Techniques
Developer resources
Data science
etc…

IBM

About

Multi-Cloud Strategy

Copyright © 2019 IBM Corporation.

Foundation for Analytics/AI

Solutions

Targeted solutions for enterprise businesses

- Cognitive Maintenance & Quality
- Continuous Engineering
- Factory and Spare Part Logistics
- Customer Experience
- Data Science
- IoT
- Virtual agent
- ...

Cognitive AI

Cognitive building blocks for developers

API Conversation	API Discovery	API Compare and Comply	API DLaaS	API Tone Analyzer	API Personality Insight
API Visual Recognition	API Speech	API Document Conversion	API Nat Language Understanding	TOOL Knowledge Studio	...more

Analytics & ML/DL

Get actionable insights from data

Planning & Scenario Modeling →	Descriptive Analytics →	Data Exploration & Discovery →	Predictive Analytics, Machine Learning Deep Learning →	Prescriptive Analytics
Planning Analytics	**Cognos Analytics**	**Cognos Analytics Watson Explorer**	**Watson ML/DL with Python, R, SPSS etc.**	**Decision Optimization**

Data & Governance

Prepare data for analytics and AI

Metadata	Transform	Quality	Regulations	Publish
Connect, Virtualize →	Events →	SQL →	NoSQL →	Objects

Hybrid Cloud

A highly scalable, security enabled infrastructure

cloud.ibm.com

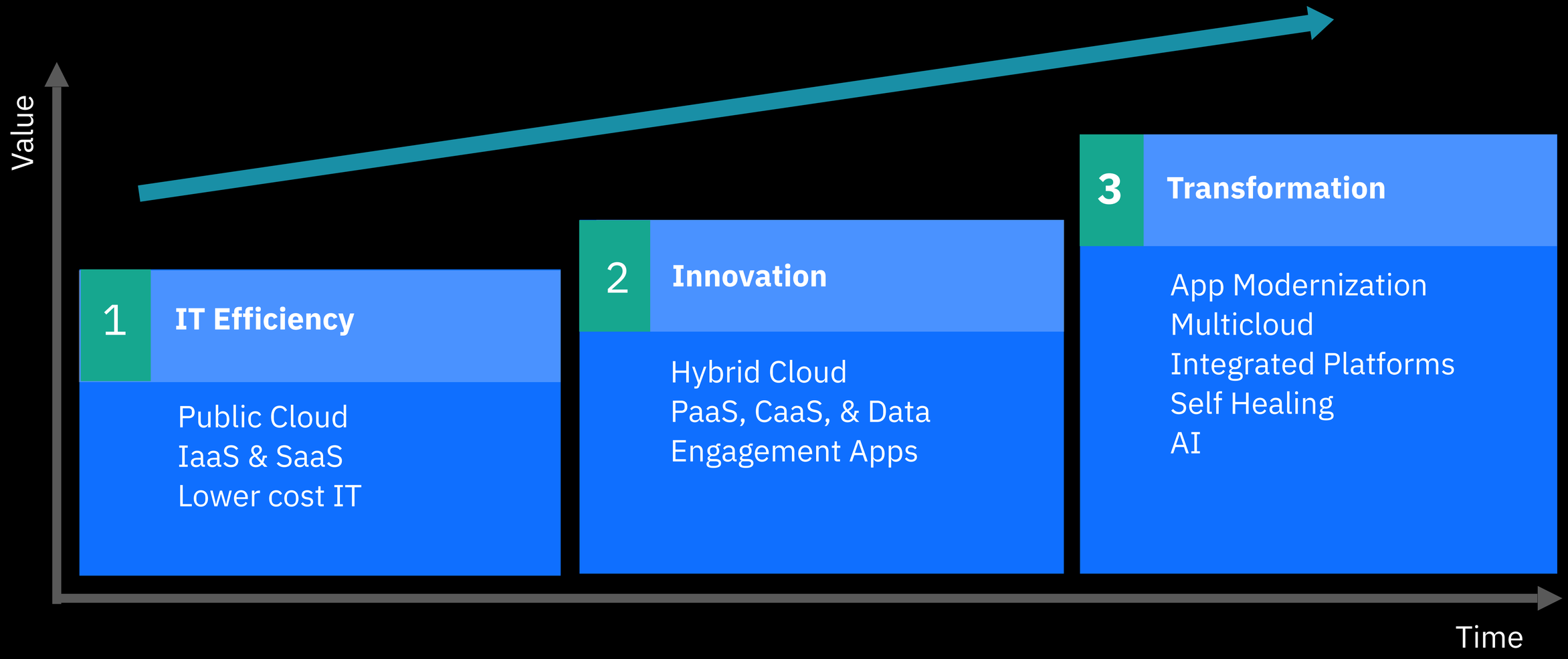
Containerized Microservices on Docker & Kubernetes

Networking	Compute	Security	Dedicated Compute	Virtual Compute	Object Storage	...more

Public | Hybrid | Private

Cloud is *technology* that enables *transformation*

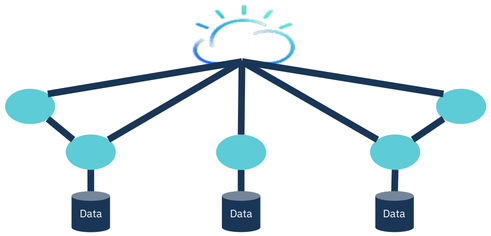
"Cloud" is **NOT** any particular place, location or a service, it is a **WAY** to get forward!



"Cloud Native" means

Agility • **Efficiency** • **Cost Savings**

Microservices

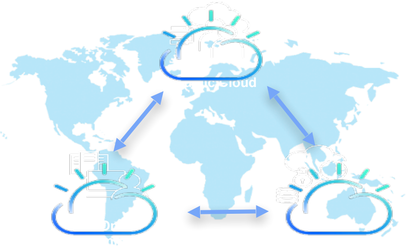


An architecture of an loosely coupled data services, easily refactored to create containerized workloads

Containerized Workloads

Stand-alone workloads composed of micro-services & data that are flexibly deployed, orchestrated and managed

Multi-Cloud Provisioning



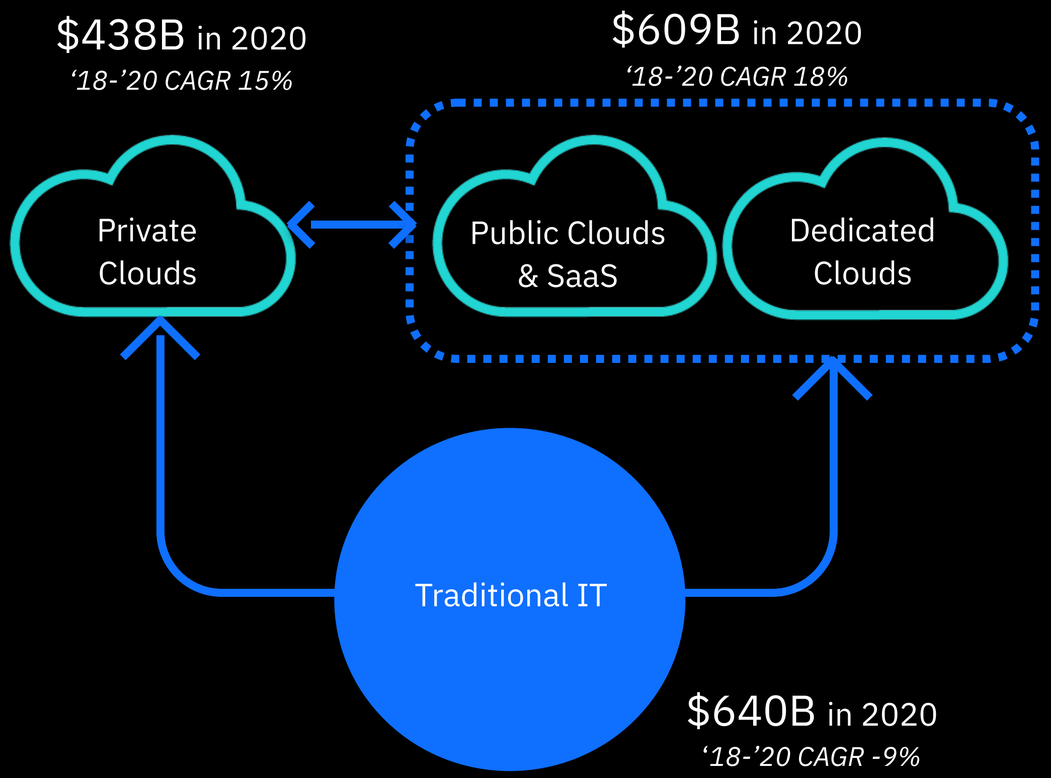
Agile provisioning of containerized workloads in multi-Cloud environments and consumption of Cloud services

CLOUD FOUNDRY · docker · kubernetes · HELM · Terraform

Because of Open Standards, **"Cloud"** can now be deployed & run **ANYWHERE**

©2018 IBM Corporation

Companies are adopting **Multi-Cloud** strategies



A real world look at multicloud

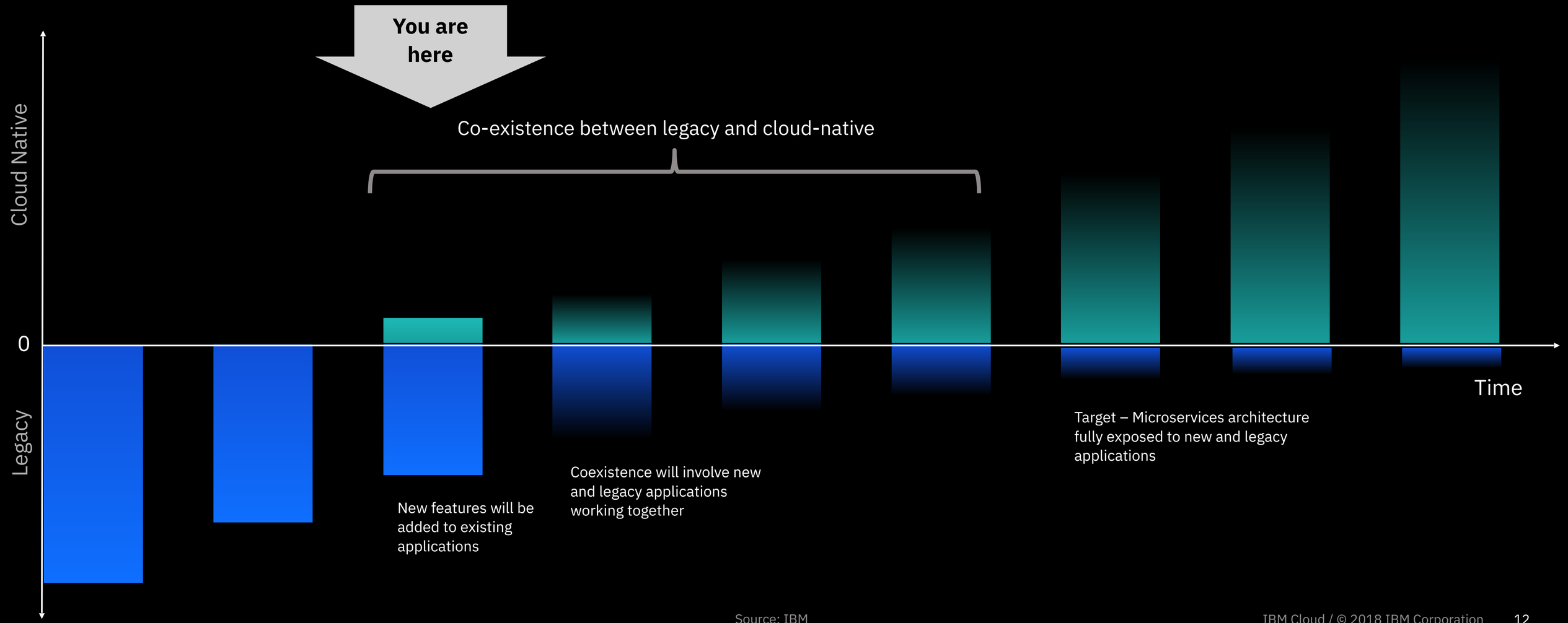
94% of enterprise customers are using multiple cloud environments (public and/or private)

67% of enterprise customers are using more than one public cloud provider (expected to remain constant or increase by 2022)

Movement between clouds	73%	priority concern
Connectivity between clouds	82%	priority concern
Consistency of management	67%	priority concern

Source: IBM MD&I; BCG and McKinsey research

Cloud-native & Legacy-systems will **co-exist** for many years to come



Source: IBM

About

"I Want AI"

Copyright © 2019 IBM Corporation.

“I want AI!“

In a MIT & BCG survey of more than 3,000 executives, managers, and analysts across industries...

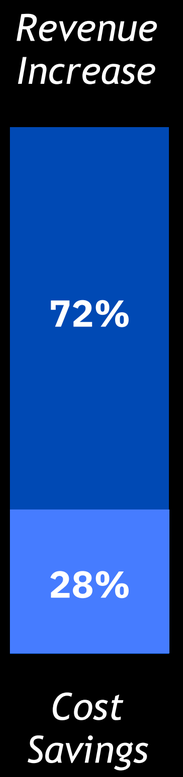
75% believe AI will enable their companies to move into new business

85% believe AI will allow their companies to obtain or sustain a competitive advantage

Reshaping Business With Artificial Intelligence, Sam Ransbotham, David Kiron, Phillip Gerbert, and Martin Reeves, MIT Sloan Management Review, Fall 2017

AI unlocks the value of data in totally new ways

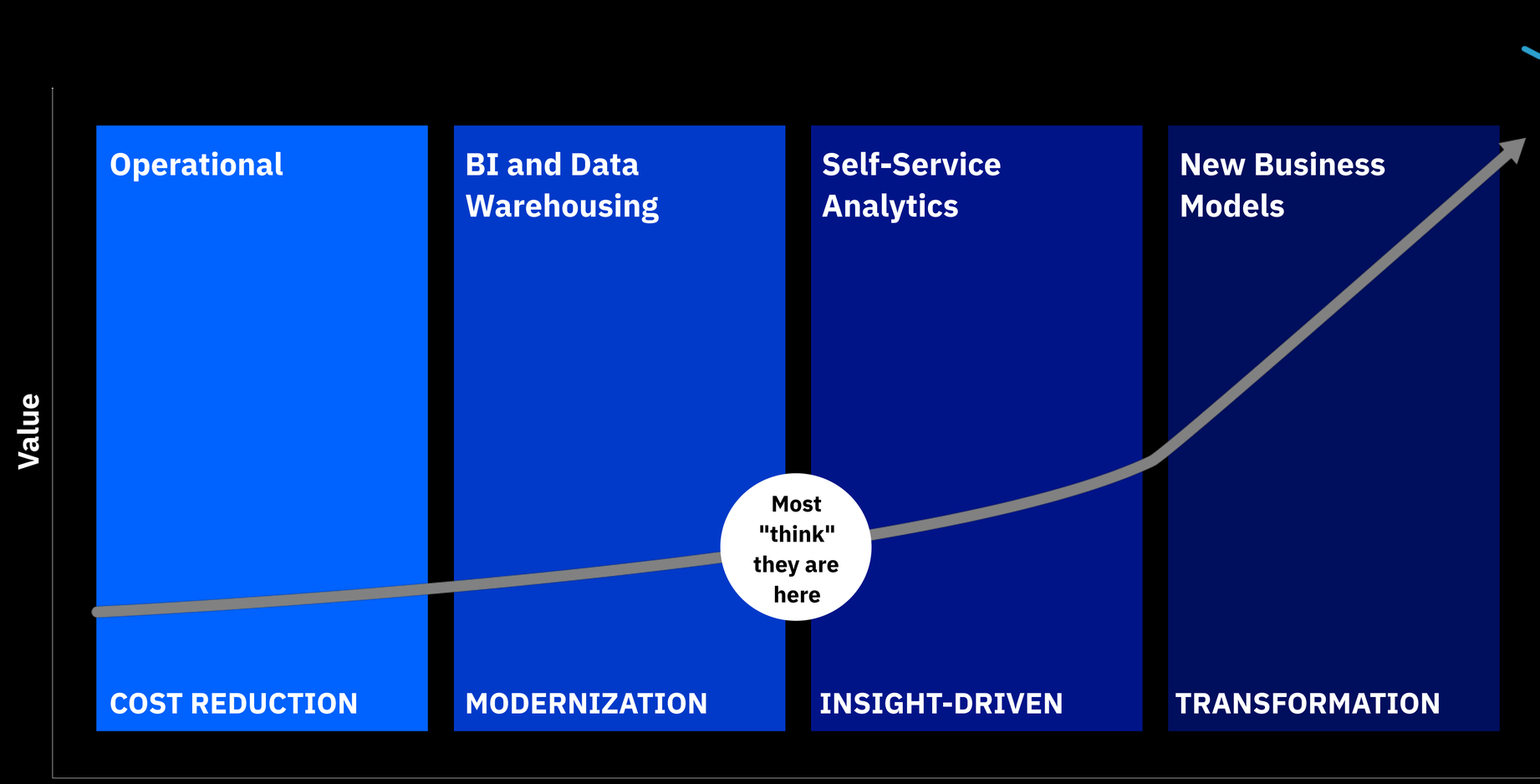
- **Predict and shape future outcomes**
- **Optimize people to do higher value work**
- **Automate decisions, processes, experiences**
- **Reimagine new business models**



How AI pioneers see value

MITSloan

Clients have declared the journey to AI a strategic priority



85%
view AI as a strategic opportunity

MITSloan

In the same global survey of more than 3,000 executives, managers, and analysts across industries...

39%

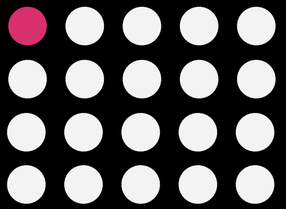
Of all companies have an AI strategy in place

(50% when only counting companies with at least 100,000 employees)

1/5

Has incorporated AI in *some* offerings or products

1/20



Has *extensively* incorporated AI in offerings or process

... why so few?

Reality: Information chaos

Data Lake | Warehouse | Enterprise

Social | Sensor | Dark Data

Open Data | Cloud | Local Files

Structured | Unstructured

IOT | Blockchain | APIs

You can't use data if you can't get it

© 2018 IBM Corporation

“Business professionals spend more than

40%

of their time fixing and validating data before they use it.”

-Forrester

80%

of the time doing analytics isn’t spent doing analytics but doing data preparation

-information-management.com

59%

of business and technology decision makers say it takes months or years to meet new complex requests to turn data into business intelligence insight.

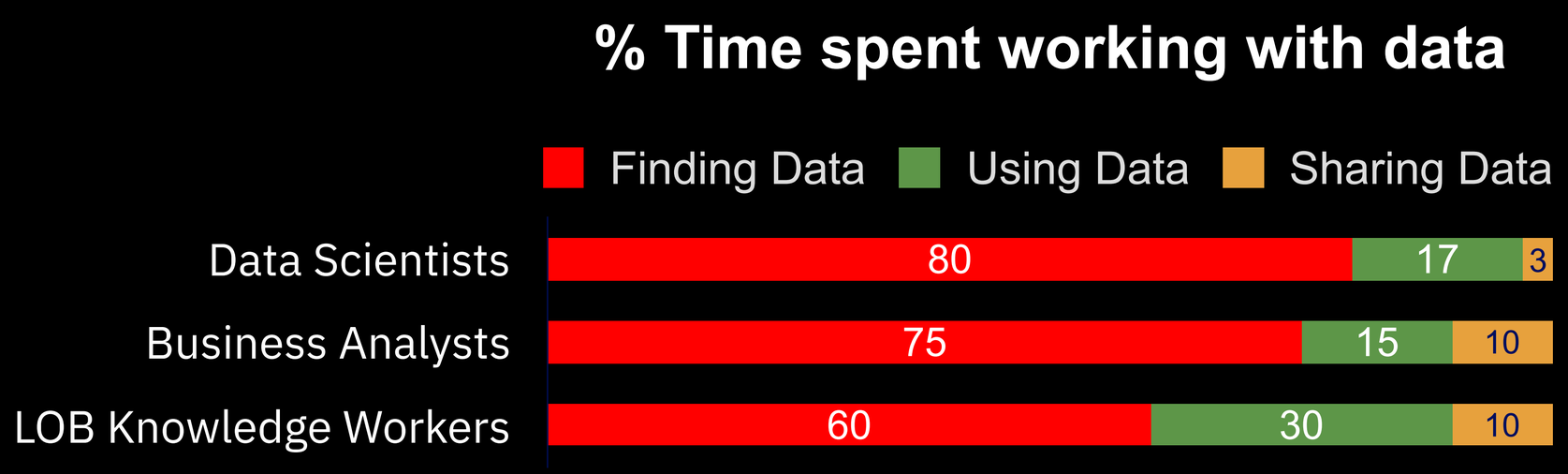
-Forrester

© 2018 IBM Corporation

Knowledge workers may spend as much as

80% of their time **searching for correct data or fixing it**

... imagine if that time was spent actually doing Analytics!



Users spent significantly more time finding the correct data, rather than extracting value from it

© 2018 IBM Corporation

There is no *AI* without *IA*

(information architecture)

81%

do not understand the data required for AI

80%

of data is either inaccessible, untrusted or unanalyzed

> ***No amount of AI algorithmic sophistication will overcome a lack of data [architecture]***
>
> MITSloan

94% are committed to multicloud (hybrid cloud)

74% using AI to modernize existing apps

87% of AI developers use open source

There's hope: What if ...

Data Lake | Warehouse | Enterprise

Social | Sensor | Dark Data

Open Data | Cloud | Local Files

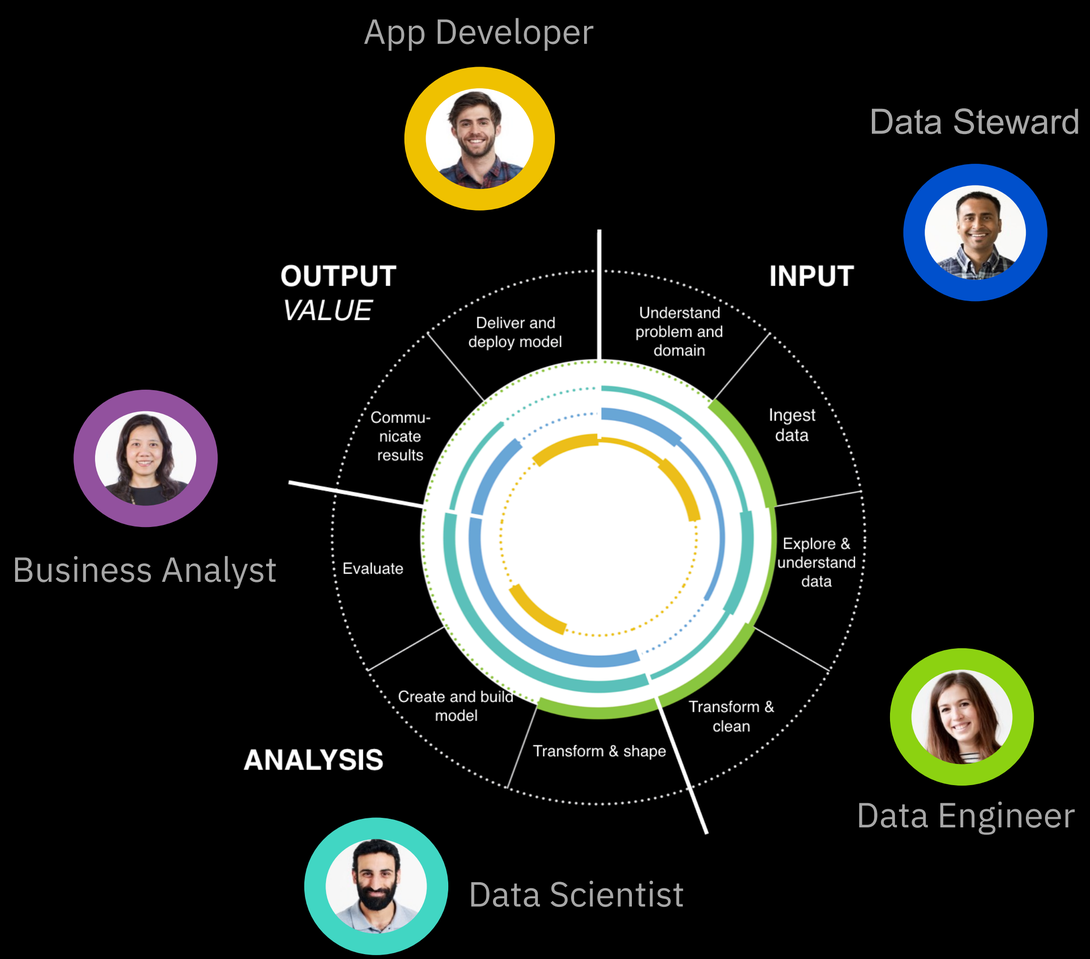
Structured | Unstructured

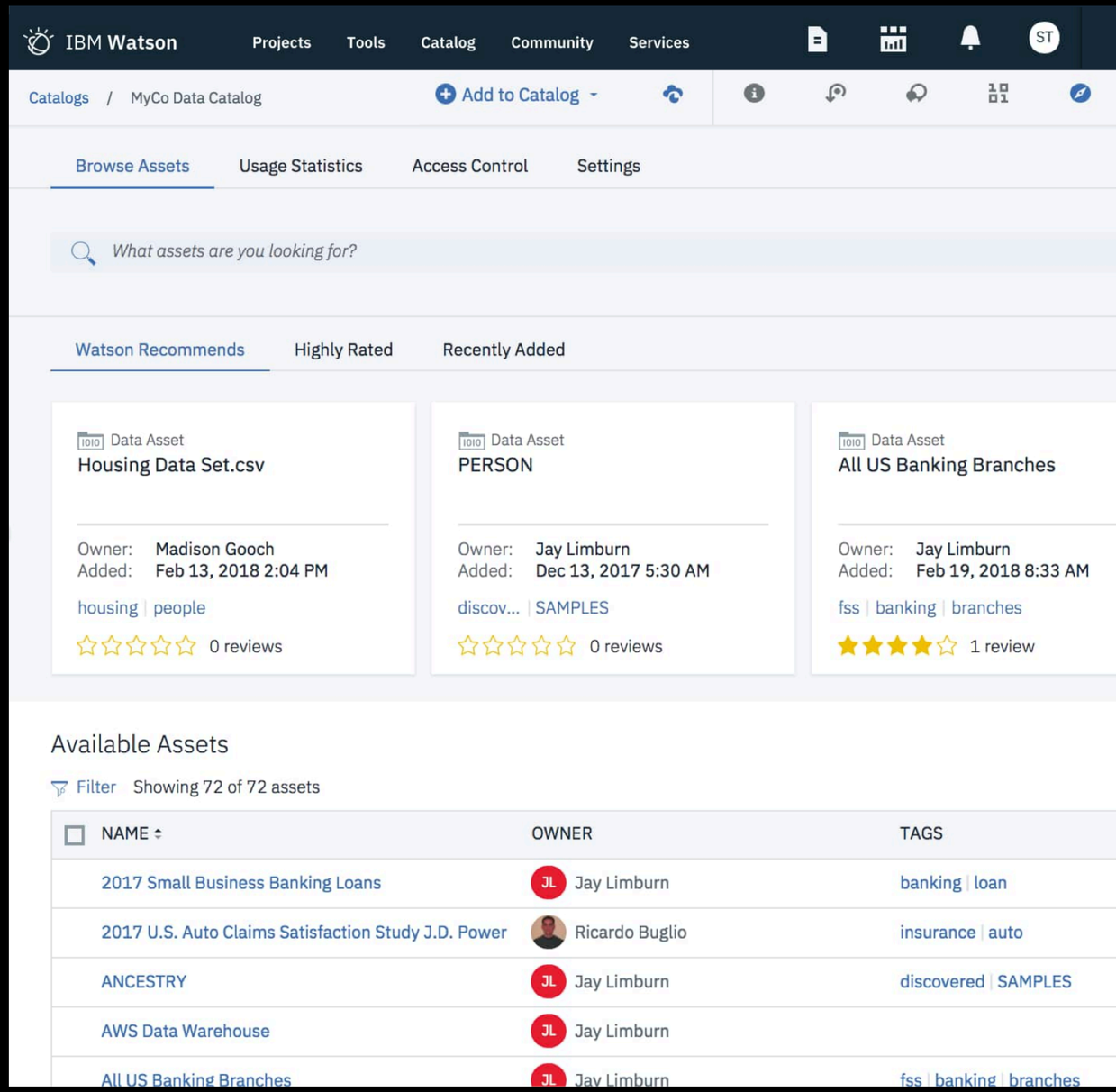
IOT | Blockchain | APIs

*... all that data was **indexed** and **organized**, made **easily available and accessible to all users***

© 2018 IBM Corporation

Analytics/AI is a Team Sport





Example of Watson Knowledge Catalog in Watson Studio

© 2018 IBM Corporation

"Ladders to AI"

A prescriptive, proven approach to **MODERNIZE** your data estate for an **AI** and **multicloud world**

Infuse
– Deploy trusted AI-driven business processes

Analyze
- Scale insights with analytics/AI everywhere (holistic)

Organize
– Create a trusted analytics foundation (with governance)

Collect / Capture
– Make data simple & accessible (and trustworthy)

Strong Foundation – Platform built on “Cloud native architecture”

Foundation for Analytics/AI

Solutions

Targeted solutions for enterprise businesses

- Cognitive Maintenance & Quality
- Continuous Engineering
- Factory and Spare Part Logistics
- Customer Experience
- Data Science
- IoT
- Virtual agent
- ...

Cognitive AI

Cognitive building blocks for developers

API Conversation	API Discovery	API Compare and Comply	API DLaaS	API Tone Analyzer	API Personality Insight
API Visual Recognition	API Speech	API Document Conversion	API Nat Language Understanding	TOOL Knowledge Studio	...more

Analytics & ML/DL

Get actionable insights from data

Planning & Scenario Modeling →	Descriptive Analytics →	Data Exploration & Discovery →	Predictive Analytics, Machine Learning Deep Learning →	Prescriptive Analytics
Planning Analytics	**Cognos Analytics**	**Cognos Analytics Watson Explorer**	**Watson ML/DL with Python, R, SPSS etc.**	**Decision Optimization**

Data & Governance

Prepare data for analytics and AI

Metadata	Transform	Quality	Regulations	Publish
Connect, Virtualize →	Events →	SQL →	NoSQL →	Objects

Hybrid Cloud

A highly scalable, security enabled infrastructure

cloud.ibm.com

Containerized Microservices on Docker & Kubernetes

Networking	Compute	Security	Dedicated Compute	Virtual Compute	Object Storage	...more

Public | Hybrid | Private

About the

Cornerstones of Business Analytics

Descriptive - Diagnostic - Predictive - Prescriptive

Foundation for Analytics/AI

Solutions

Targeted solutions for enterprise businesses

- Cognitive Maintenance & Quality
- Continuous Engineering
- Factory and Spare Part Logistics
- Customer Experience
- Data Science
- IoT
- Virtual agent
- ...

Cognitive AI

Cognitive building blocks for developers

API Conversation	API Discovery	API Compare and Comply	API DLaaS	API Tone Analyzer	API Personality Insight
API Visual Recognition	API Speech	API Document Conversion	API Nat Language Understanding	TOOL Knowledge Studio	...more

Analytics & ML/DL

Get actionable insights from data

Planning & Scenario Modeling →	Descriptive Analytics →	Data Exploration & Discovery →	Predictive Analytics, Machine Learning Deep Learning →	Prescriptive Analytics
Planning Analytics	**Cognos Analytics**	**Cognos Analytics Watson Explorer**	**Watson ML/DL with Python, R, SPSS etc.**	**Decision Optimization**

Data & Governance

Prepare data for analytics and AI

Metadata	Transform	Quality	Regulations	Publish
Connect, Virtualize →	Events →	SQL →	NoSQL →	Objects

Hybrid Cloud

A highly scalable, security enabled infrastructure

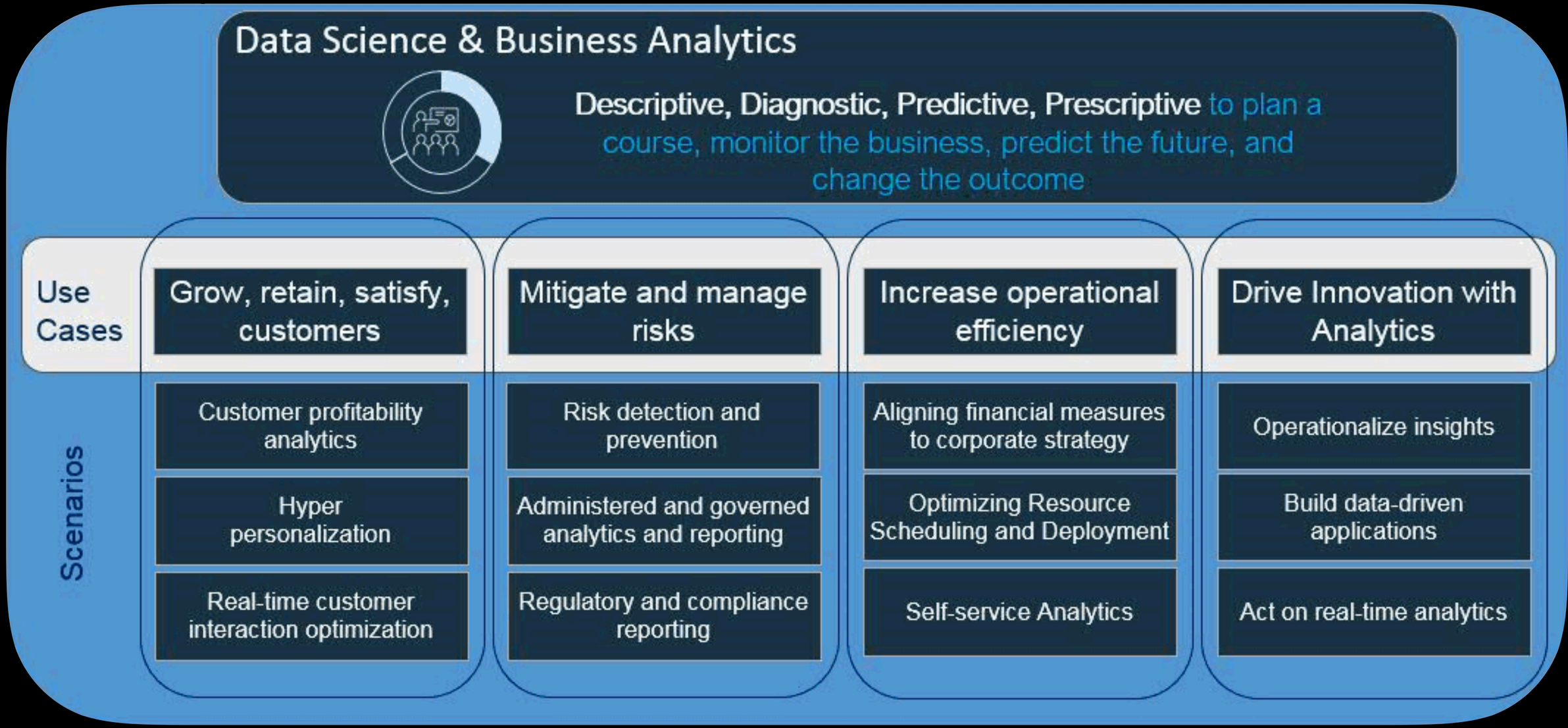
cloud.ibm.com

Containerized Microservices on Docker & Kubernetes

Networking	Compute	Security	Dedicated Compute	Virtual Compute	Object Storage	...more

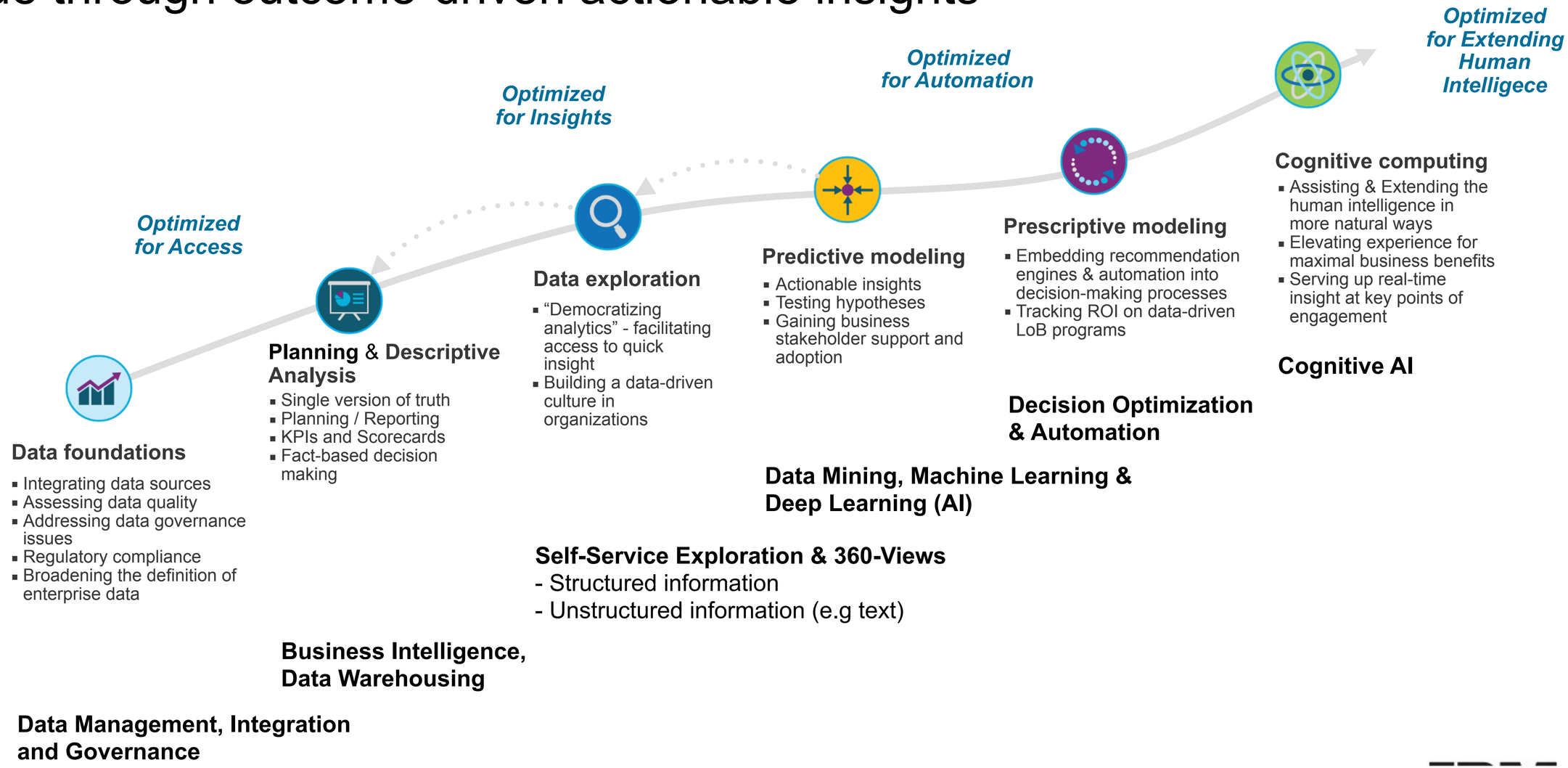
Public | Hybrid | Private

Popular Analytics & Data Science use cases across business/industries



©2018 IBM Corporation

As the analytics/ai maturity level of increases, it helps organization to progress on their transformation journey, to capture and create increasing value through outcome-driven actionable insights



© 2016 IBM Corporation

IBM Business Analytics solutions

What is our plan?

IBM Planning Analytics

Planning

Planning, budgeting, and forecasting

What happened?

IBM Cognos Analytics (& WS)

Descriptive

Get in touch with reality, a single source of the truth, visibility

Why did it happen?

IBM Cognos Analytics (& WS)

Understanding why something happened is a critical step

IBM Watson Explorer (& WS)

Data Augmentation and Contextualization

Diagnostic

Structured Data

Unstructured Data

What will happen next?

IBM Watson Studio, containing
IBM SPSS, Open Source ML/DL and more

Predictive

Understand the most likely future scenario and its business implications

What should we do?

IBM Watson Studio, containing
IBM Decision Optimization

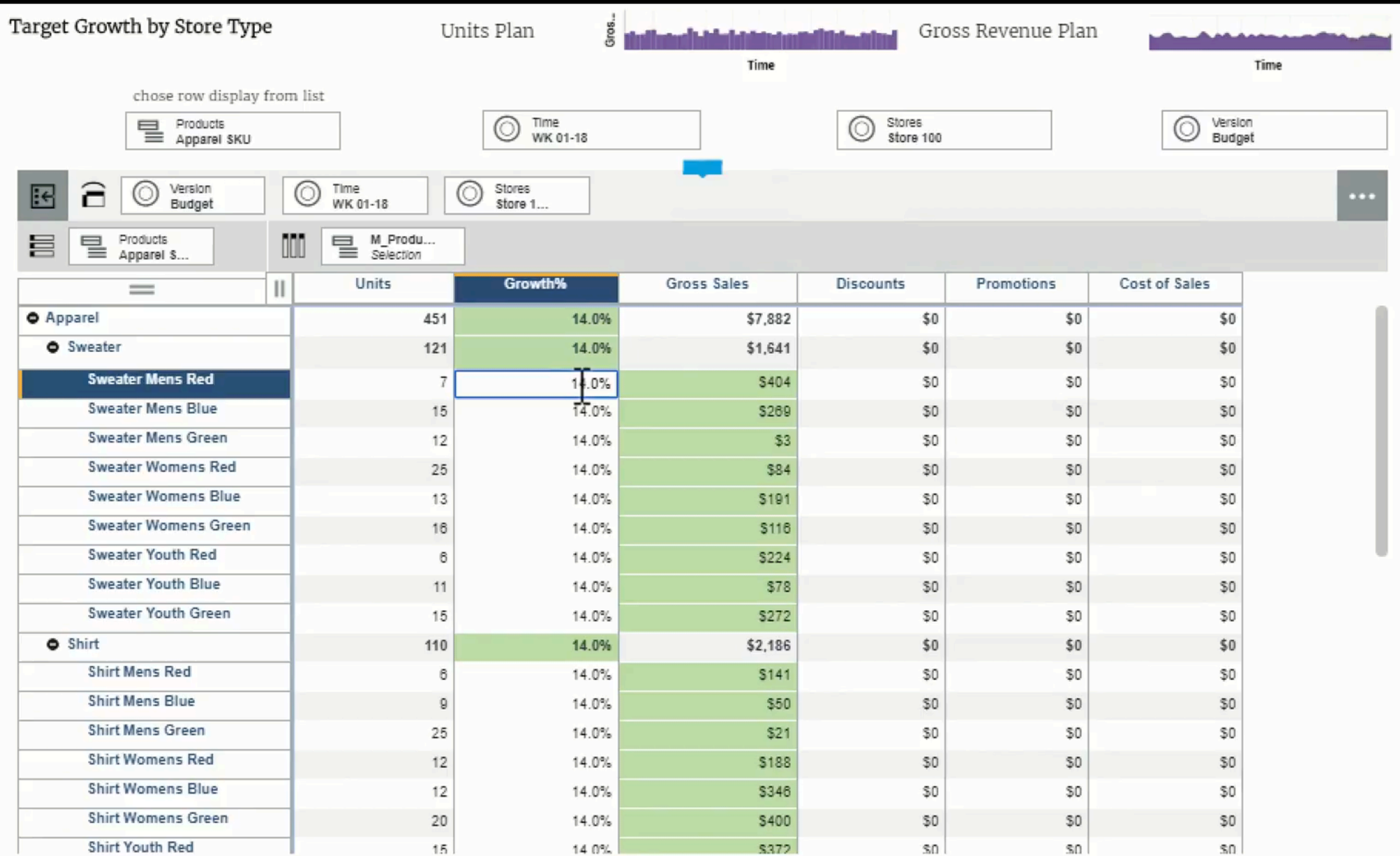
Prescriptive

Collaborate for maximum business value, informed by advanced analytics

Governed & Trusted Data

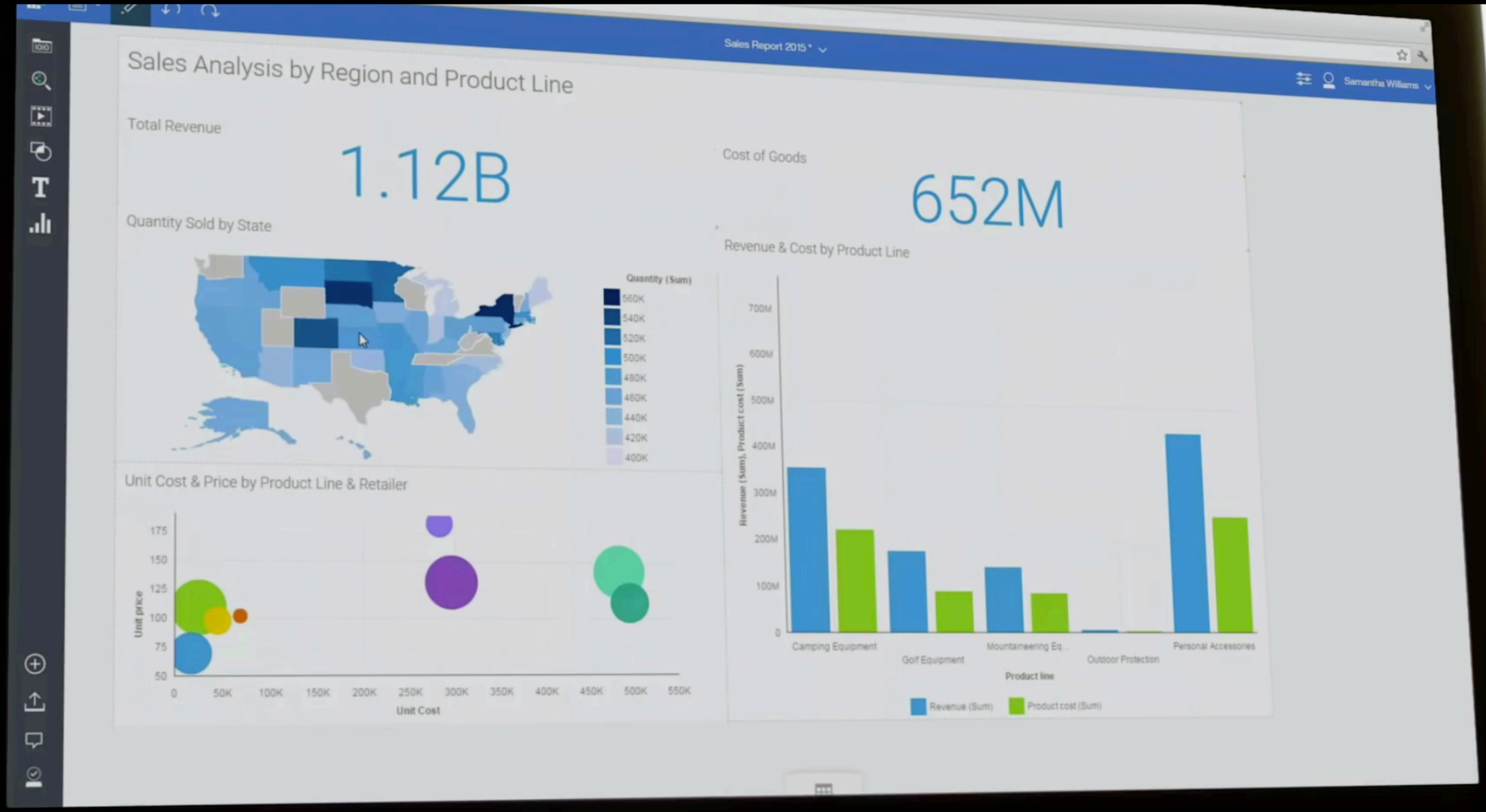
ibm.com/analytics/business-analytics

Planning Analytics in short

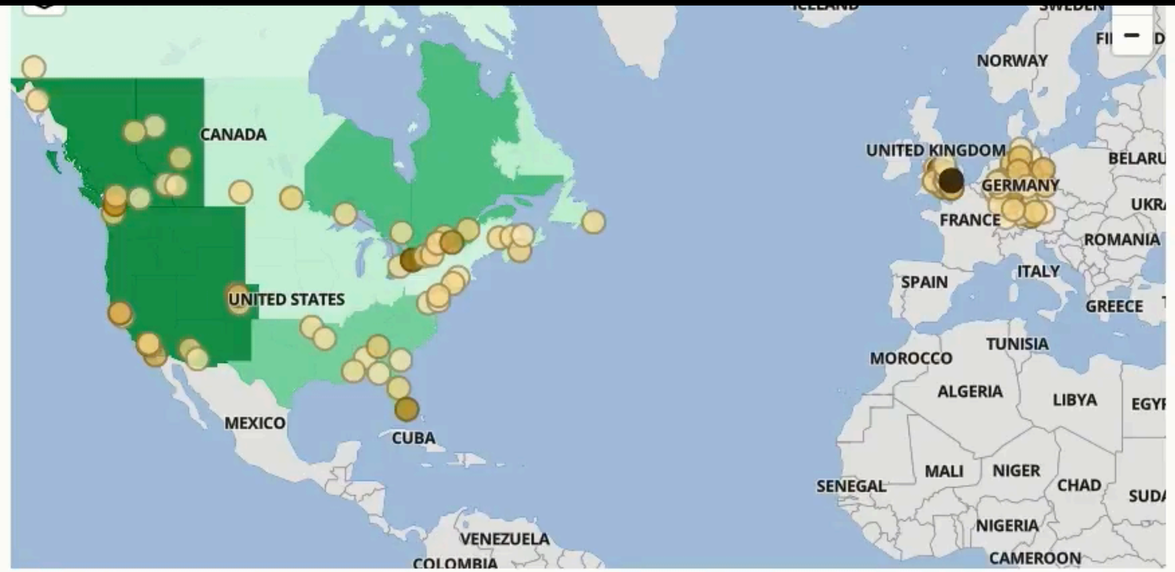


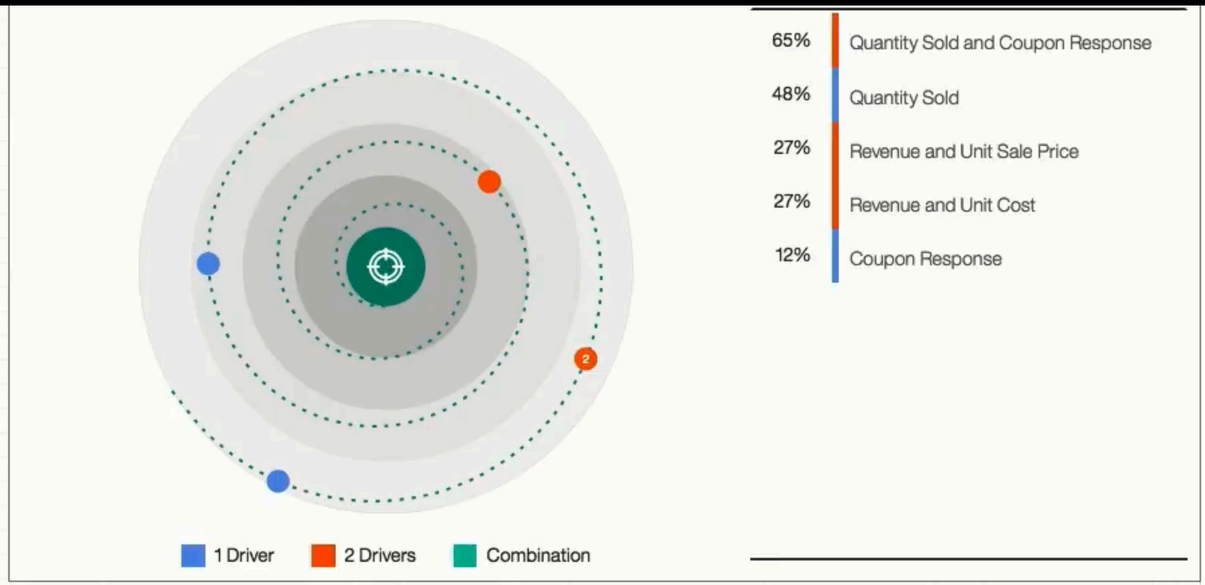
	Units	Growth%	Gross Sales	Discounts	Promotions	Cost of Sales
Apparel	451	14.0%	$7,882	$0	$0	$0
Sweater	121	14.0%	$1,641	$0	$0	$0
Sweater Mens Red	7	14.0%	$404	$0	$0	$0
Sweater Mens Blue	15	14.0%	$269	$0	$0	$0
Sweater Mens Green	12	14.0%	$3	$0	$0	$0
Sweater Womens Red	25	14.0%	$84	$0	$0	$0
Sweater Womens Blue	13	14.0%	$191	$0	$0	$0
Sweater Womens Green	16	14.0%	$116	$0	$0	$0
Sweater Youth Red	6	14.0%	$224	$0	$0	$0
Sweater Youth Blue	11	14.0%	$78	$0	$0	$0
Sweater Youth Green	15	14.0%	$272	$0	$0	$0
Shirt	110	14.0%	$2,186	$0	$0	$0
Shirt Mens Red	6	14.0%	$141	$0	$0	$0
Shirt Mens Blue	9	14.0%	$50	$0	$0	$0
Shirt Mens Green	25	14.0%	$21	$0	$0	$0
Shirt Womens Red	12	14.0%	$188	$0	$0	$0
Shirt Womens Blue	12	14.0%	$346	$0	$0	$0
Shirt Womens Green	20	14.0%	$400	$0	$0	$0
Shirt Youth Red	15	14.0%	$372	$0	$0	$0

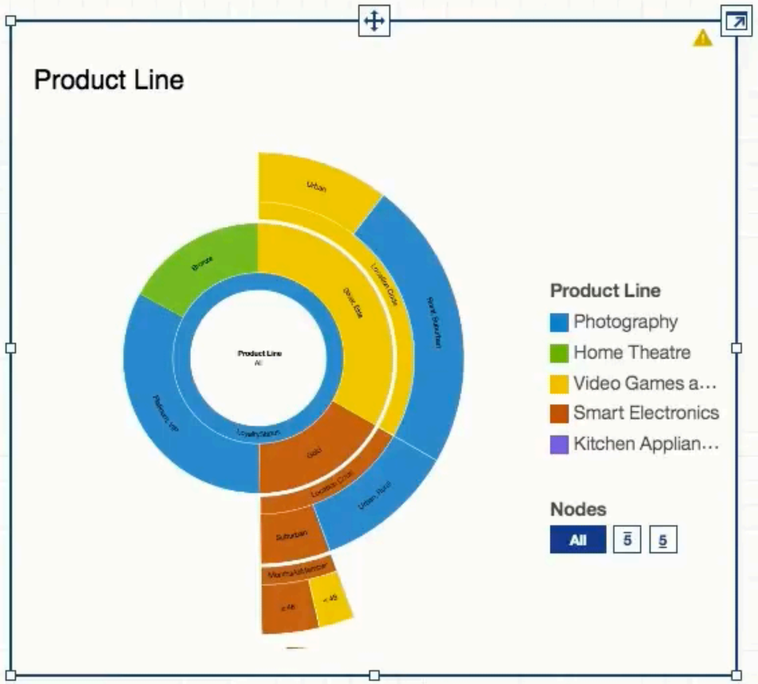
Cognos Analytics BI in short

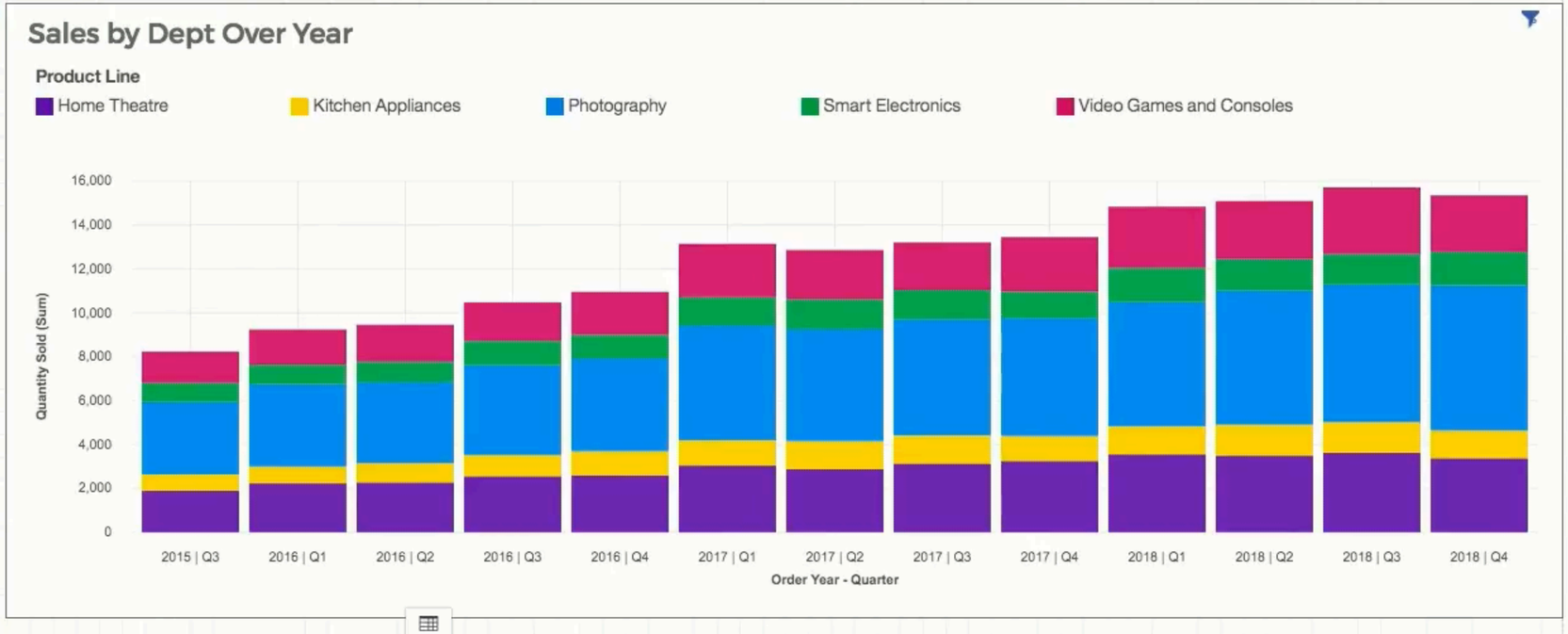


Cognos Analytics Exploration in short









IBM Business Analytics

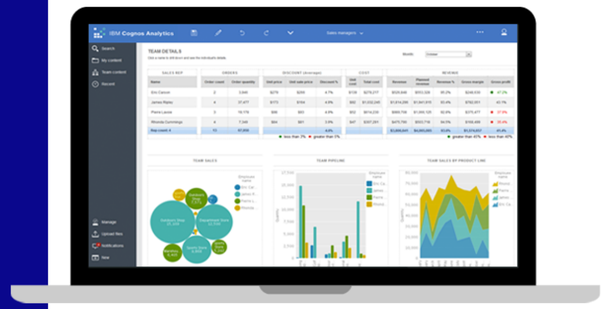
ibm.biz/BdjPh6

Featured solutions

Planning Analytics
Descriptive Analytics (BI)
Diagnostic/Explorative Analytics
Predictive Analytics
Prescriptive Analytics
Cognitive

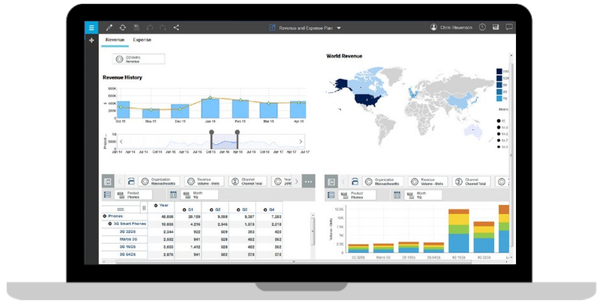
IBM Cognos Analytics

Become more analytics-driven with an integrated solution for all your managed and self-service needs. Empower people to work with data and find their own answers within a governed framework.



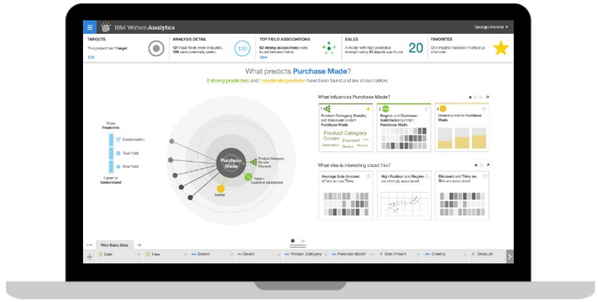
IBM Planning Analytics

Planning, budgeting, forecasting and multidimensional analysis--powered by IBM TM1.



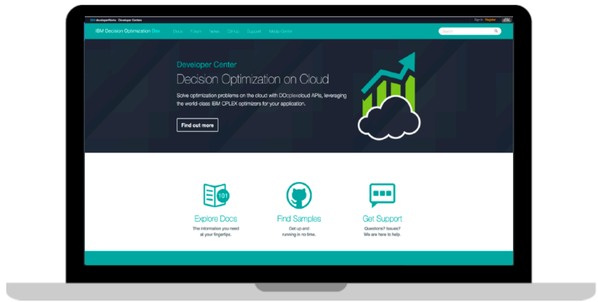
IBM Cognos Analytics Diagnostics/Exploration

Gain a deeper understanding of your data with easy-to-use analytics. Interact conversationally to explore and visualize business insights.



IBM Decision Optimization Center

Uses powerful analytics to solve tough planning and scheduling challenges to drive better outcomes.



Each image contains a weblink to demos just shown previously

© IBM Corporation

About

Data Science / AI with **Watson Studio**

Foundation for Analytics/AI

Solutions

Targeted solutions for enterprise businesses

- Cognitive Maintenance & Quality
- Continuous Engineering
- Factory and Spare Part Logistics
- Customer Experience
- Data Science
- IoT
- Virtual agent
- ...

Cognitive AI

Cognitive building blocks for developers

API Conversation	API Discovery	API Compare and Comply	API DLaaS	API Tone Analyzer	API Personality Insight
API Visual Recognition	API Speech	API Document Conversion	API Nat Language Understanding	TOOL Knowledge Studio	...more

Analytics & ML/DL

Get actionable insights from data

Planning & Scenario Modeling →	Descriptive Analytics →	Data Exploration & Discovery →	Predictive Analytics, Machine Learning Deep Learning →	Prescriptive Analytics
Planning Analytics	**Cognos Analytics**	**Watson Analytics Watson Explorer**	**Watson ML/DL with Python, R, SPSS etc.**	**Decision Optimization**

Data & Governance

Prepare data for analytics and AI

Metadata	Transform	Quality	Regulations	Publish
Connect, Virtualize →	Events →	SQL →	NoSQL →	Objects

Hybrid Cloud

A highly scalable, security enabled infrastructure

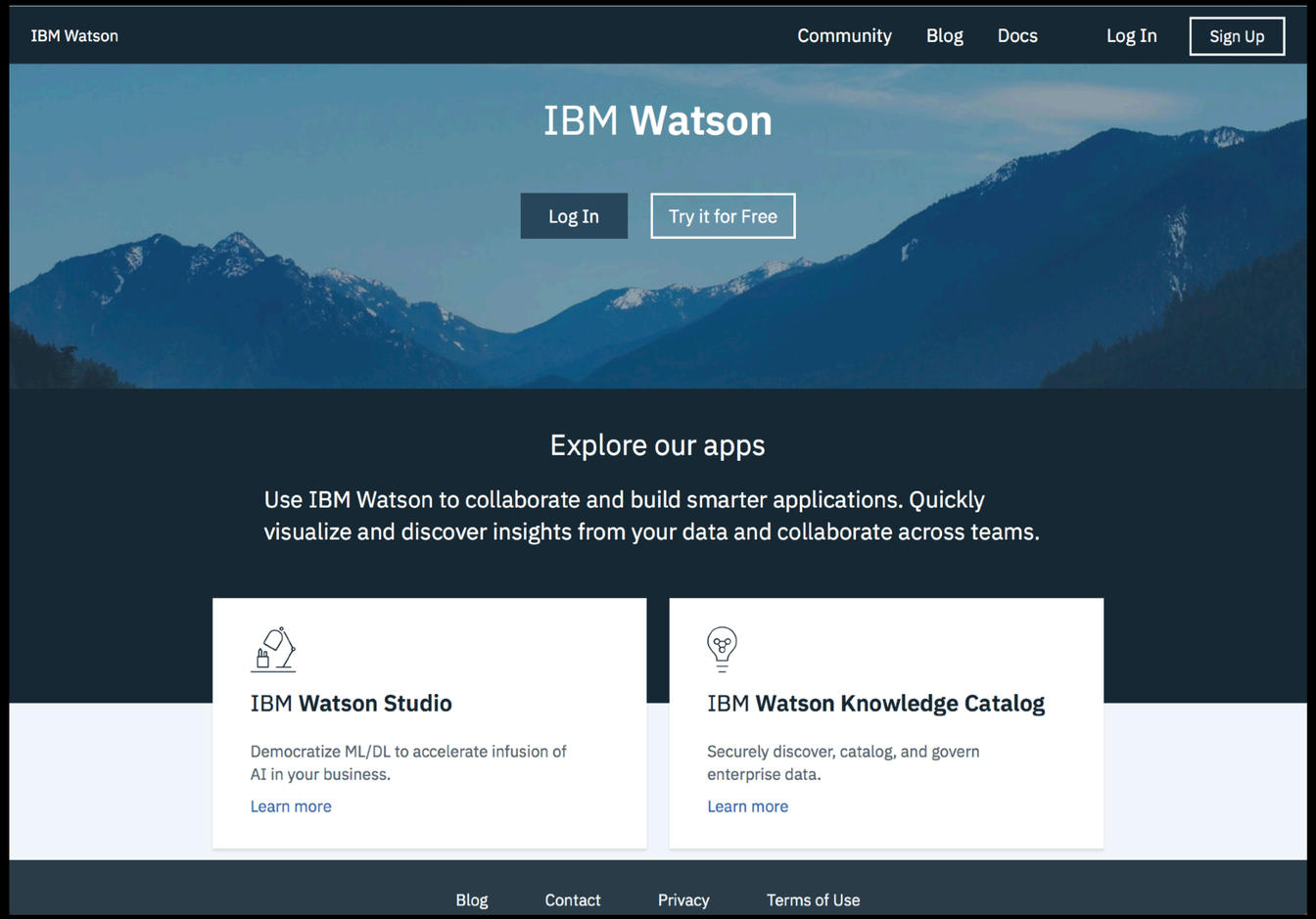
cloud.ibm.com

Containerized Microservices on Docker & Kubernetes

Networking	Compute	Security	Dedicated Compute	Virtual Compute	Object Storage	...more

Public | Hybrid | Private

Watson Studio



Watson Studio Introduction video: youtube.com/watch?v=mg25cDnuT84
Watson Studio Overview video: youtube.com/watch?v=TiS-LGfNoSo
Watson Studio Resources: ibm.com/cloud/watson-studio/resources
Related Articles & Tutorials: medium.com/ibm-watson

Once within the Watson Studio, look at the **Community** section for much more!

Watson Studio is the best predictive & multimodal Machine Learning solution on market today!

©2018 IBM Corporation

Watson Studio Desktop is free for students!

IBM

Search

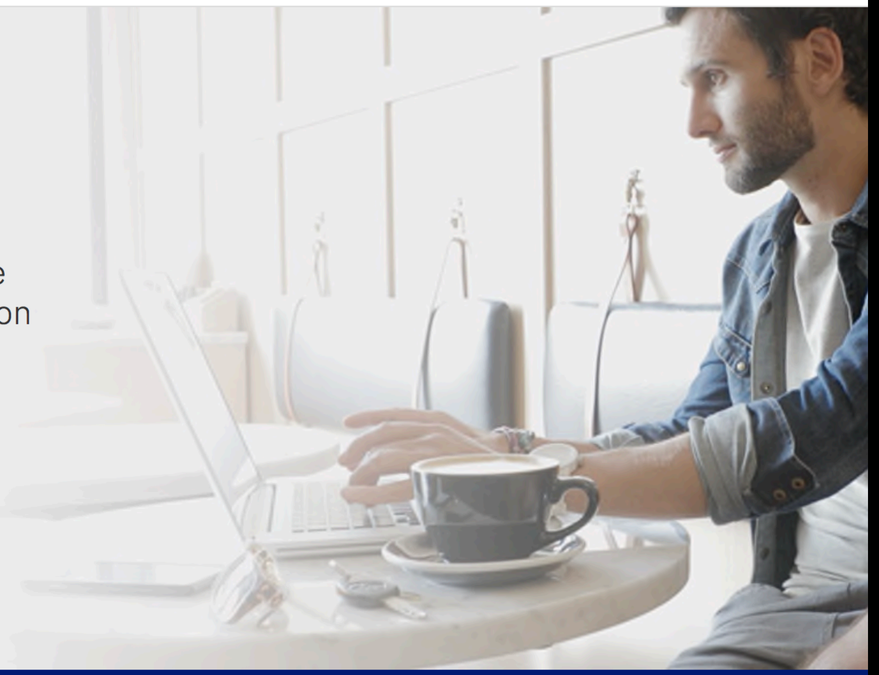
Watson Studio Desktop Details Pricing Resources Try for free

IBM Watson Studio Desktop

Empower your data science and AI teams to refine data, visually build models and deploy using data on the desktop for anytime, anywhere access

Try for free | View pricing and buy

→ Get free student edition



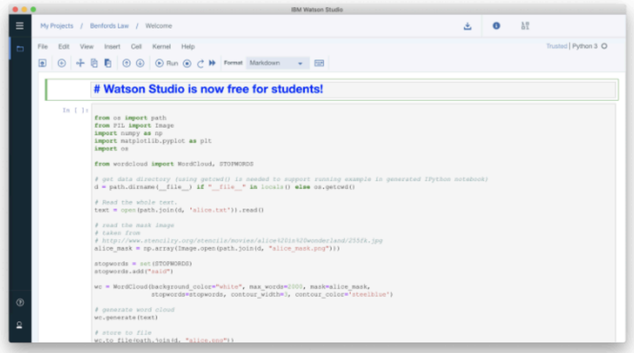
Need a cloud or on-premises solution?
→ Learn more about other IBM Watson Studio deployment options

Get started with AI on your desktop

Let's talk

Watson Studio Desktop is now free for academia

Doug Stauber Follow
Feb 12 · 2 min read



Machine Learning, Data Science, and Predictive Analytics techniques are in strong demand. That's why since its launch, IBM Watson Studio has proven to be very popular with academia. Thousands of students and faculty have been drawn to Watson Studio for its powerful open source and code-free data analysis tools. Now, this all-in-one platform for data science is free to students and faculty with unlimited use with Watson Studio Desktop.

Access it: ibm.com/products/watson-studio-desktop
Read more about it: ibm.biz/Bd2QMa

©2018 IBM Corporation

ibm.com/cloud/watson-studio
ibm.com/products/watson-studio-desktop

youtube.com/watch?v=mg25cDnuT84

New connection

Connect to Data Anywhere

Db2-WH-Merja
Db2 Warehouse

IBM services

- BigInsights HDFS
- Cloud Object Storage
- Cloud Object Storage (infrastructure)
- Cloudant
- Compose for MySQL
- Compose for PostgreSQL
- Db2
- Db2 Big SQL
- Db2 for i
- Db2 for z/OS
- Db2 Hosted
- Db2 on Cloud
- Db2 Warehouse
- Informix
- Object Storage OpenStack Swift
- Object Storage OpenStack Swift (infrastructure)
- PureData for Analytics
- Watson Analytics

Third-party services

- Amazon Redshift
- Amazon S3
- Apache Hive
- Cloudera Impala
- Dropbox
- FTP
- Google BigQuery
- Google Cloud Storage
- Hortonworks HDFS
- Looker
- Microsoft Azure Data Lake Store
- Microsoft Azure SQL Database
- Microsoft SQL Server
- MySQL
- Oracle
- Pivotal Greenplum
- PostgreSQL
- Salesforce.com
- Sybase
- Sybase IQ
- Tableau
- Teradata

Select a data source to begin

Cancel

Data & AI Asset Catalog

IBM **Watson Studio** | 1867965 - IBM Forum ...

Catalogs / Forum Catalog

Add to Catalog
- Local files
- Connected asset
- Connection
- Import assets

Browse Assets | Access Control | Settings

What assets are you looking for?

Watson Recommends | Highly Rated | Recently Added

Collapse

Data Asset
customers.csv
Owner: Jukka Ruponen
Added: Mar 23, 2019 12:13 PM
custo... | churn | telco | + 2
0 reviews

Data Asset
calls.csv
Owner: Jukka Ruponen
Added: Mar 23, 2019 1:27 AM
telco | sample | calls | + 1
0 reviews

Filter

Asset types
- Data Asset (2)

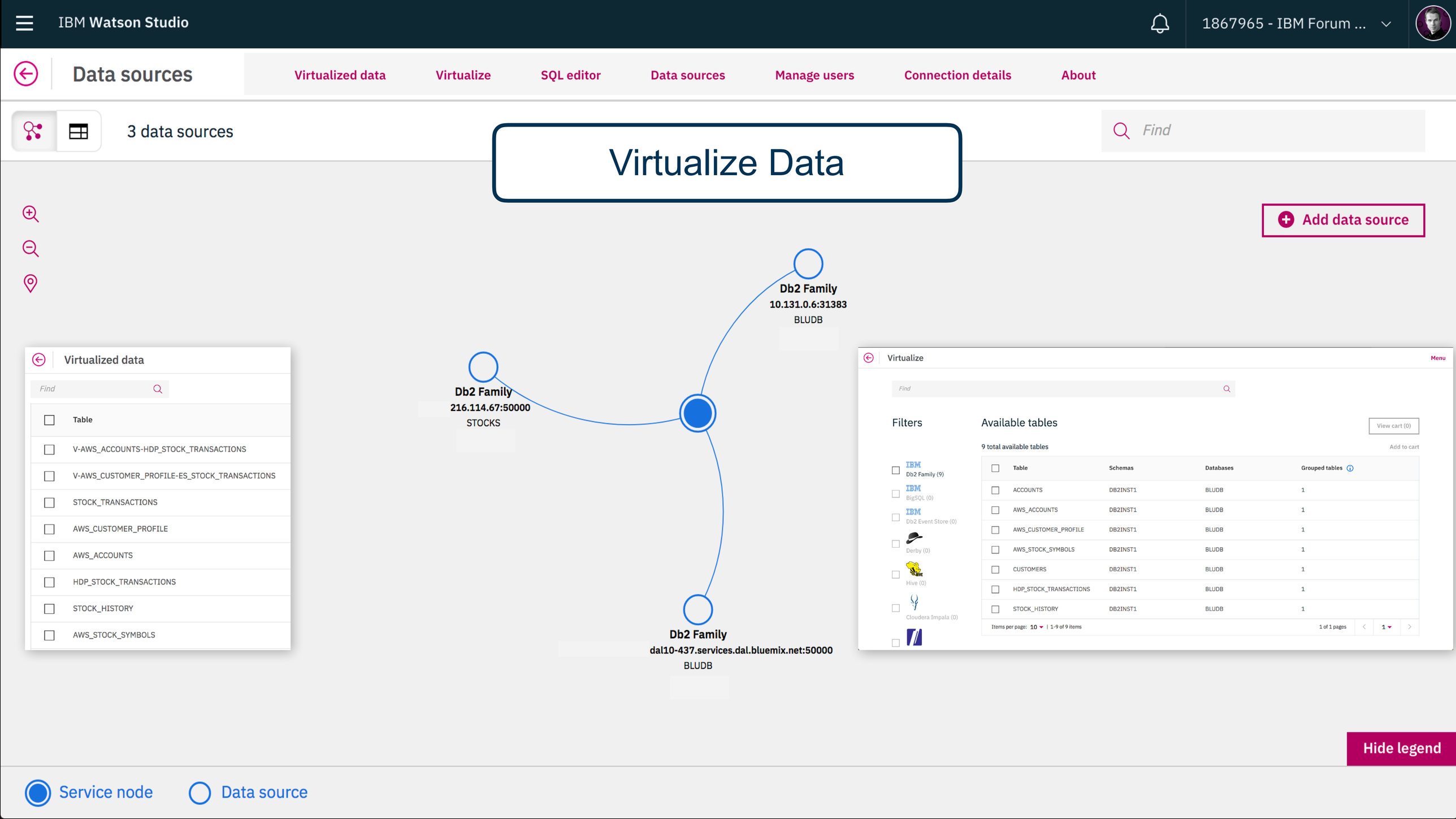
Tags
- churn (2)

Available Assets

Showing 2 of 2 assets

NAME	OWNER	TAGS	TYPE	DATE ADDED
calls.csv	Jukka Ruponen	telco | sample | calls | + 1	Data Asset	Mar 23, 2019
customers.csv	Jukka Ruponen	customer | churn | telco | + 2	Data Asset	Mar 23, 2019

Cookie Preferences


IBM Watson Studio
1867965 - IBM Forum ...
Data sources
Virtualized data
Virtualize
SQL editor
Data sources
Manage users
Connection details
About
3 data sources
Find
Virtualize Data
Add data source
Db2 Family
10.131.0.6:31383
BLUDB
Db2 Family
216.114.67:50000
STOCKS
Db2 Family
dal10-437.services.dal.bluemix.net:50000
BLUDB
Virtualized data
Find
Table
V-AWS_ACCOUNTS-HDP_STOCK_TRANSACTIONS
V-AWS_CUSTOMER_PROFILE-ES_STOCK_TRANSACTIONS
STOCK_TRANSACTIONS
AWS_CUSTOMER_PROFILE
AWS_ACCOUNTS
HDP_STOCK_TRANSACTIONS
STOCK_HISTORY
AWS_STOCK_SYMBOLS
Virtualize
Menu
Find
Filters
Available tables
View cart (0)
9 total available tables
Add to cart
IBM Db2 Family (9)
IBM BigSQL (0)
IBM Db2 Event Store (0)
Derby (0)
Hive (0)
Cloudera Impala (0)
Table
Schemas
Databases
Grouped tables
ACCOUNTS DB2INST1 BLUDB 1
AWS_ACCOUNTS DB2INST1 BLUDB 1
AWS_CUSTOMER_PROFILE DB2INST1 BLUDB 1
AWS_STOCK_SYMBOLS DB2INST1 BLUDB 1
CUSTOMERS DB2INST1 BLUDB 1
HDP_STOCK_TRANSACTIONS DB2INST1 BLUDB 1
STOCK_HISTORY DB2INST1 BLUDB 1
Items per page: 10 | 1-9 of 9 items
1 of 1 pages
Hide legend
Service node
Data source

Business Glossary/Terms

Customer

Term

Industry Accelerators

★★★★☆
1 Ratings

Description
A person or organization buying goods or services from the organization.

Select your rating:
★★★★☆

New Comment:

Write a comment

0 (Maximum 2000 characters) .

Submit

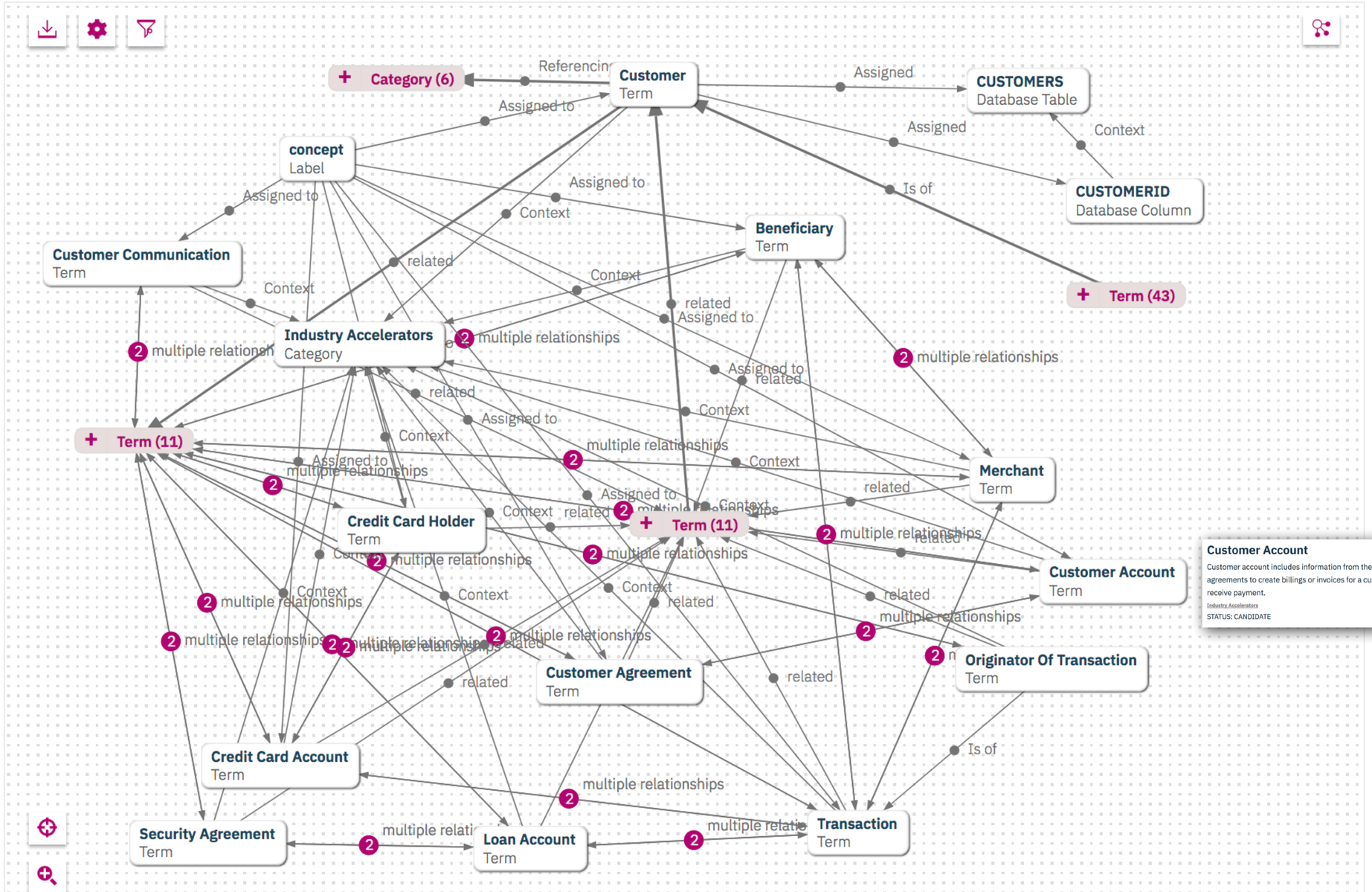
All Comments (1)

admin, admin - 3/28/2019

Explains relationships to term 'Customer'

Delete

Relationships



Understanding & Profiling Data

IBM **Watson Studio** | 1867965 - IBM Forum ...

My Projects / SLUSH 18 Standard / churn_prep_data_v

Preview | **Profile** | **Lineage**

Current profile: 164 classifiers

Last profile: 28 Mar 2019 - 2:19 am View Log

Columns: 26

Rows: 309

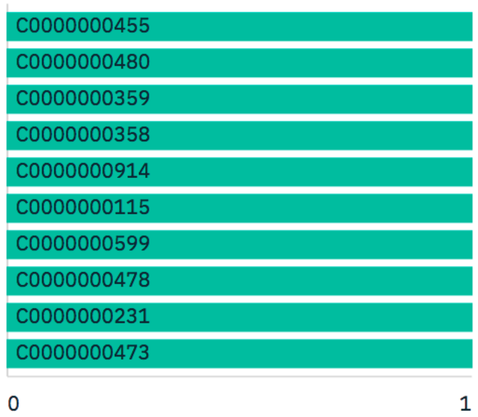
Delete | Update Profile

customer_id
Type: Varchar

• *Customer Number*

100% 0% 0%

FREQUENCY



Showing 10 of 100 | View All

STATISTICS

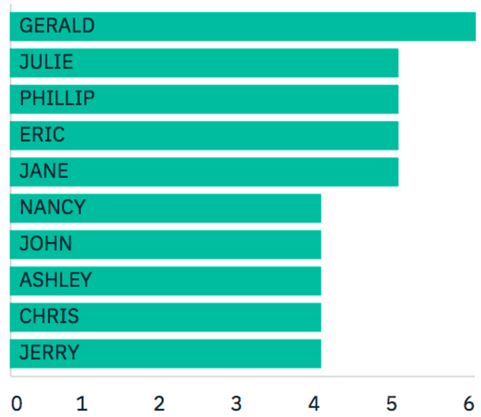
Unique	309
Minimum Length	11
Maximum Length	11

first_name
Type: Varchar

• *First Name*

100% 0% 0%

FREQUENCY



Showing 10 of 100 | View All

STATISTICS

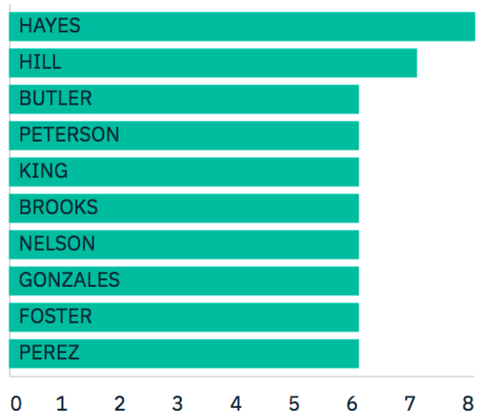
Unique	70
Minimum Length	3
Maximum Length	10

last_name
Type: Varchar

Last Name

100% 0% 0%

FREQUENCY



Showing 10 of 94 | View All

STATISTICS

Unique	16
Minimum Length	3
Maximum Length	10

twitter_handle
Type: Varchar

• *Not classified*

31% 0% 69%

FREQUENCY

Showing 10 of 96 | View All

STATISTICS

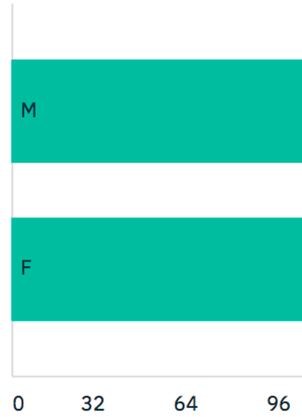
Unique	95
Minimum Length	9
Maximum Length	18

gender
Type: Varchar

• *Gender*

100% 0%

FREQUENCY



Showing 2 of 2

STATISTICS

Unique	
Minimum Length	
Maximum Length	

Refine Data for Analytics/AI

× Operation

Code an operation to cleanse and shape your data

Steps

Help

Search operations

FREQUENTLY USED

- Calculate
- Convert column type
- Filter
- Math
- Remove
- Rename
- Sort ascending
- Sort descending
- Substitute
- Text

CLEANSE

- Convert column value to missing
- Remove duplicates
- Remove empty rows
- Replace missing values
- Replace substring

CUSTOMERID String	FIRSTNAME String	LASTNAME String	BIRTHDATE String	STREET String	CITY String	STATE String
100784	Randy	Downs	1995-10-31	School Street	Abie	NE
100785	Janice	Durham	1969-03-29	High Street	Achilles	VA
100786	Timothy	Combs	1982-09-16	Durham Road	Lubbock	TX
100787	Shirley	Hopkins	1959-08-29	13th Street	Alden	IL
100788	Willie	Cotton	1980-05-05	Grant Street	Amityville	NY
100789	Katherine	Hill	1986-08-18	Surrey Lane	Assonet	MA
100790	Charles	Stephenson	1991-08-02	3rd Avenue	Atlanta	GA
100791	Denise	Mcintosh	1994-01-19	Lake Street	Allendale	NJ
100792	Johnny	Franks	1966-08-11	Elm Street	Akron	IA
100793	Marilyn	Kelley	1963-06-13	Pine Street	Aliquippa	PA
100794	Scott	Hanson	1991-05-14	Academy Street	Glendale	CA
100795	Samantha	Wallace	1967-11-22	Walnut Street	Boulder	CO
100796	Gerald	Warner	1994-10-03	Laurel Lane	Portsmouth	VA
100797	Amber	Glass	1981-07-20	7th Street	Brooklyn	NY
100798	Gregory	Schneider	1985-10-20	11th Street	Sikeston	MO
100799	Megan	Robinson	1995-05-17	Maple Street	Charleston	SC
100800	Bryan	Russell	1962-03-09	Franklin Avenue	Brooklyn	NY
100801	Maria	Sawyer	1988-05-07	Grant Street	Milwaukee	WI

OPERATION MENU

Refine data by using the operations menu

Use the Operation menu to search for an operation or browse the operation categories, including frequently-used operations.

Data assets containing personal or restricted data

Confidential
0
View all

Personally Identifiable Information
1
View all

Sensitive Personal Information
1
View all

These data sets contain sensitive personal data (SPI) which is defined as personal data consisting of information relating to an individual with regard to racial or ethnic origin;political opinions; religious beliefs or other beliefs of a similar nature; trade union membership; physical or mental health or condition; sexual life; or any criminal or alleged criminal history of a person.

Operational policies

1
Data policies

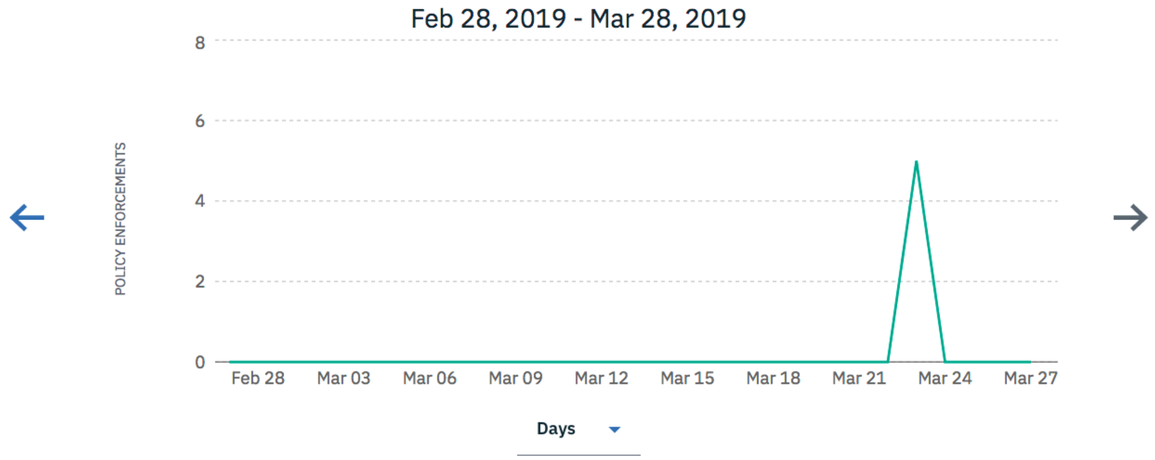
1
Data policy rules

Automatic enforcement

5
enforcements in last 30 days

0%
from last month

Policy enforcements over time



Ensure
Regulatory Compliance

Top ten most enforced policies

Name	Enforced	Most common outcome
PII policy Mask Personally Indentifiable Data	5	Access Transform

Cookie Preferences

Add Analytics/AI Projects

IBM **Watson Studio** | 1867965 - IBM Forum ...

My Projects / SLUSH 18 Standard

Add to project

Overview | Assets | Environments | Bookmarks | Deployments | Access Control | Settings

SLUSH 18 Standard

Last Updated: Jan 10 2019

Readme

24	0	2
Assets	Bookmarks	Collaborators

Date created
Dec 03 2018

Description
No description available

Storage
Cloud Object Storage
53.18 MB used

Collaborators View all (2)

NS Nitesh Sood
Viewer

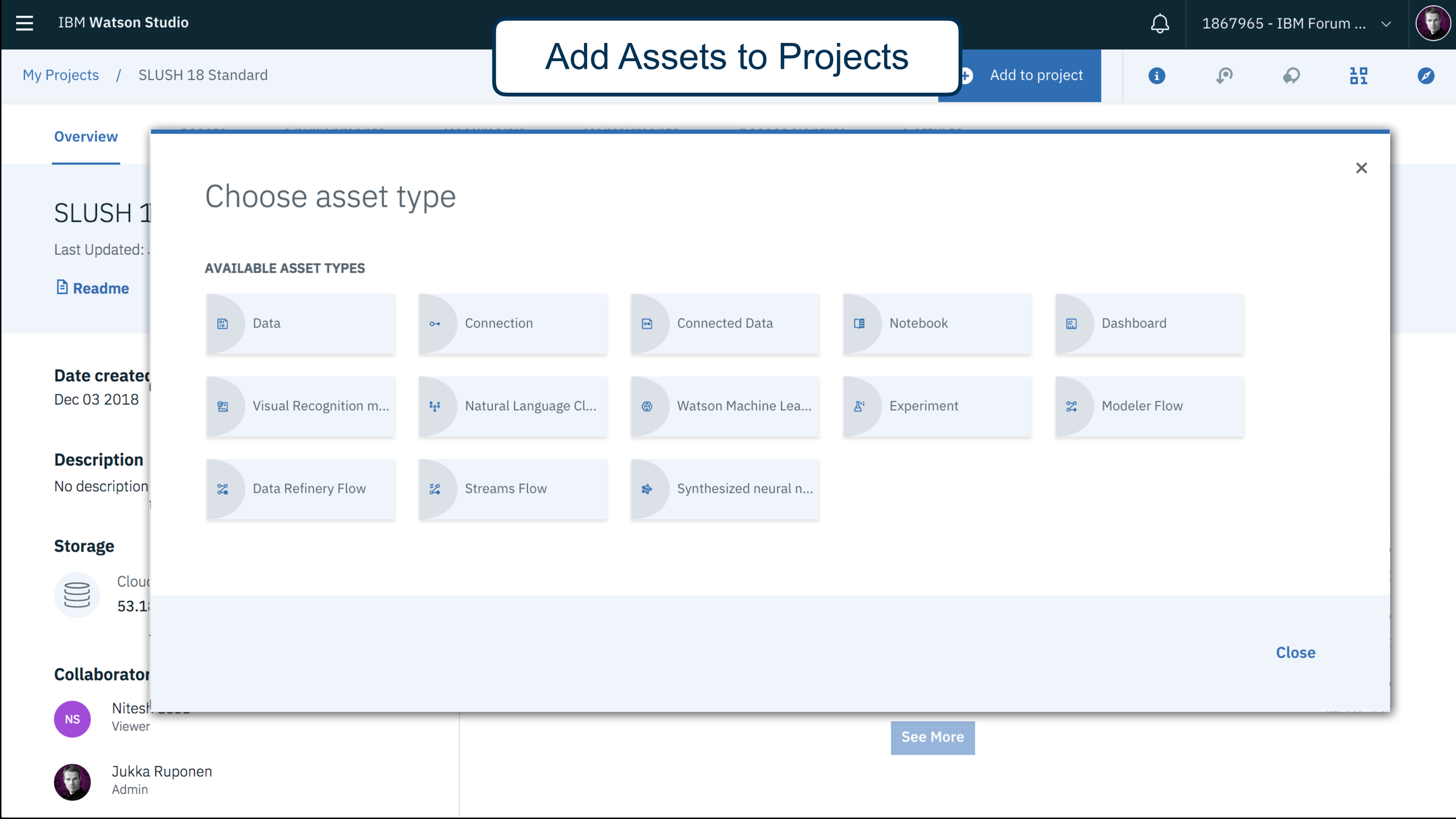
Jukka Ruponen
Admin

Recent activity

Search activities

Activity	Date
WD Discovery process has **completed** for connection **Db2 WH dsx** to project **SLUSH 18 Standard**	2019/03/16 @ 11:14p
WD Discovery process has **started** for connection **Db2 WH dsx** to project **SLUSH 18 Standard**	2019/03/16 @ 11:13p
Jukka Ruponen added Nitesh Sood to **SLUSH 18 Standard**	2019/01/10 @ 10:53a

See More


IBM Watson Studio
1867965 - IBM Forum ...
Add Assets to Projects
My Projects / SLUSH 18 Standard
Add to project
Overview
SLUSH 1
Last Updated:
Readme
Date created
Dec 03 2018
Description
No description
Storage
Collaborator
NS
Viewer
Jukka Ruponen
Admin
See More
Choose asset type
AVAILABLE ASSET TYPES
Data
Connection
Connected Data
Notebook
Dashboard
Visual Recognition m...
Natural Language Cl...
Watson Machine Lea...
Experiment
Modeler Flow
Data Refinery Flow
Streams Flow
Synthesized neural n...
Close

Add Assets to Projects

IBM **Watson Studio** | 1867965 - IBM Forum ...

My Projects / SLUSH 18 Standard

Add to project

Overview | Assets | Environments | Bookmarks | Deployments | Access Control | Settings

What assets are you looking for?

Data assets View all (15)

Models

Notebooks

New notebook

NAME	SHARED	SCHEDULED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
Predict Churn (Use Model)				Python 3.5	Jukka Ruponen	17 Mar 2019	
JSON data conversion				Python 3.5	Jukka Ruponen	16 Mar 2019	
German Credit Lab				Python 3.5	Jukka Ruponen	17 Mar 2019	
German Credit Lab (published)				Python 3.5	Jukka Ruponen	26 Mar 2019	

Experiments

New experiment

NAME	CREATED BY	LAST MODIFIED	ACTIONS
You don't have any Experiments yet.			

Load | Files | Catalog

ADD FROM CATALOG

Select source catalog

Find in Catalog

Recently created assets:

- calls.csv
- customers.csv
- flow_eb7aef99c59443b68e01c07d352127
- profiling_47cd9c42-c504-465f-915a-0748
- profiling_47cd9c42-c504-465f-915a-0748
- profiling_87ae97ba-e6cf-4aa6-9fee-9a2ae
- profiling_87ae97ba-e6cf-4aa6-9fee-9a2ae
- profiling_93bfc7b9-b94a-4bbf-84e9-0a88c

PROGRESS

IBM **Watson Studio**

1867965 - IBM Forum ...

Many Assets in Projects

Home > Projects > JR FS19 WSL

JR FS19 WSL

Created by **admin**
on **3 Feb 2019, 11:13 PM**

Assets 73 | Data Sources 1 | Jobs 0 | Environments 9 | Collaborators 7

- Recent
- Data sets 40
- Notebooks 15
- Scripts 2
- Models 2
- Model groups 0
- Analytics dashboards 0
- Data Refinery flows 0
- RStudio 8
- Modeler flows 6
- Watson Explorer collections 0

CustomerChurn_el SPSS Modeler flow • 28 Mar 20...	CUSTOMERS Data set • 25 Mar 2019, 9:33 P...	Sample Streams (basic) Jupyter notebook • 14 Mar 20...
History_Transactions_... Data set • 3 Feb 2019, 11:13 P...	Current_Transactions_... Data set • 3 Feb 2019, 11:13 P...	brakeEventModel.rds RStudio • 3 Feb 2019, 11:13 PM
ATM_CleanData.csv Data set • 3 Feb 2019, 11:13 P...	ATM_Data.csv Data set • 3 Feb 2019, 11:13 P...	Online_Transactions.csv Data set • 3 Feb 2019, 11:13 P...
historical_brake_even... RStudio • 3 Feb 2019, 11:13 PM	Customer_Segmentati... SPSS Modeler flow • 3 Feb 201...	TelcoChurnZeppelin Zeppelin notebook • 3 Feb 201...

Many Assets in Projects

JR FS19 WSL

Assets 73 | Data Sources 1 | Jobs 0 | Environments 9 | Collaborators 7

- Add data set
- Add script
- Add notebook
- Add model
- Add Model Group
- Add Modeler Flow
- Open RStudio
- Add Watson Explorer Collection
- Add data source
- Add job
- Add streaming analytics notebook

- Recent
- Data sets 40
- Notebooks 15
  - Jupyter 14
  - Zeppelin 1
- Scripts 2
- Models 2
- Model groups 0
- Analytics dashboards 0
- Data Refinery flows 0
- RStudio 8
- Modeler flows 6
- Watson Explorer collections 0

Search by notebook name

Notebooks 15

Name	Type	Environment	
Sample Streams (basic)	Jupyter	Jupyter with Python 3.6, Spark 2.3.2	
TelcoChurnZeppelin	Zeppelin	Zeppelin with Anaconda2, Python 2.7	
DriverClassification	Jupyter	Jupyter with Python 2.7, Scala 2.11, R 3.4.3, Spark 2.0	
TelcoChurn_Scoring	Jupyter	Jupyter with Python 2.7, Scala 2.11, R 3.4.3, Spark 2.0.2	3 Feb 2019, 11:13 PM
TelcoChurn_SparkML_35	Jupyter	Jupyter with Python 3.5, Scala 2.11, R 3.4.3, Spark 2.2.1	3 Feb 2019, 11:13 PM
MortgageDefault-XGBoost-BuildSave	Jupyter	Jupyter with Python 2.7, Scala 2.11, R 3.4.3, Spark 2.0.2	3 Feb 2019, 11:13 PM

SPSS Modeler Assets

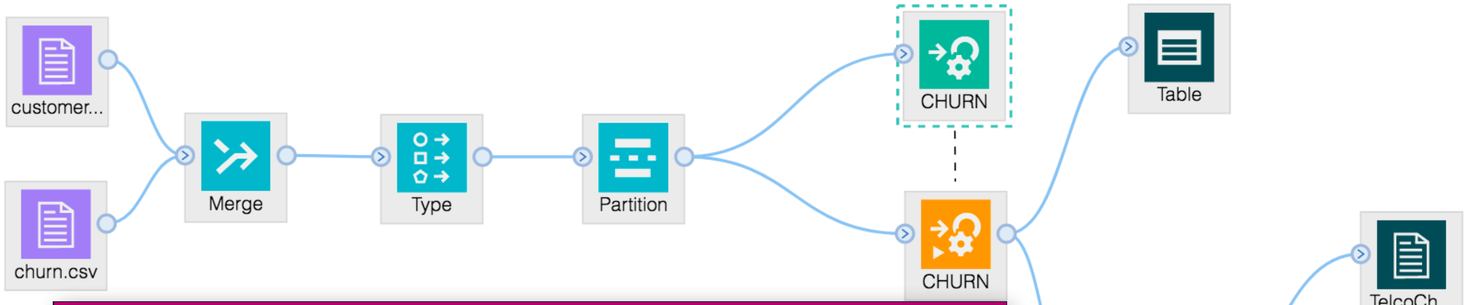
IBM Watson Studio

1867965 - IBM Forum ...

Projects > JR FS19 WSL > CustomerChurn_el

Search Nodes

- Import
- Record Operations
- Field Operations
- Graphs
- Modeling
  - Auto Classifier
  - Auto Numeric
  - Bayes Net
  - C5.0
  - C&R Tree
  - CHAID
  - Quest
  - Tree-AS
  - Random Trees
  - Random Forest
  - Decision List
  - Time Series
  - GenLin



NGDISTANCEBILLT	USAGE	RATEPLAN	CHURN	PARTITION	$XF-CHURN	$XFC-CHURN
nl_discount	229.640	3.000	T	1_Training	T	0.984
andard	75.290	2.000	F	1_Training	F	0.819
andard	47.250	3.000	F	1_Training	F	0.782
andard	59.010	1.000	F	2_Testing	F	0.788
nl_discount	28.140	1.000	F	1_Training	F	0.974
andard	58.870	1.000	F	1_Training	F	0.983
nl_discount	58.720	1.000	F	1_Training	F	0.833
andard	34.170	3.000	F	1_Training	F	0.927
andard	48.350	2.000	F	1_Training	F	0.829
nl_discount	15.980	4.000	F	1_Training	F	0.987

Close

CHURN

- FIELDS
- BUILD OPTIONS
  - **Model Name**
  - (•) Auto
  - ( ) Custom
  - 128
  - [x] Use partitioned data
  - [x] Build model for each split
  - **Rank models by**
  - Lift (flag targets)
  - **Rank models using**
  - ( ) Training partition
  - (•) Test partition
  - **Number of models to use**
  - 3
  - [x] Calculate predictor importance
- CRITERIA FOR FLAG TARGETS
- EXPERT
- DISCARD
- ENSEMBLE
- ANNOTATIONS

Cancel Save

Notebook Assets

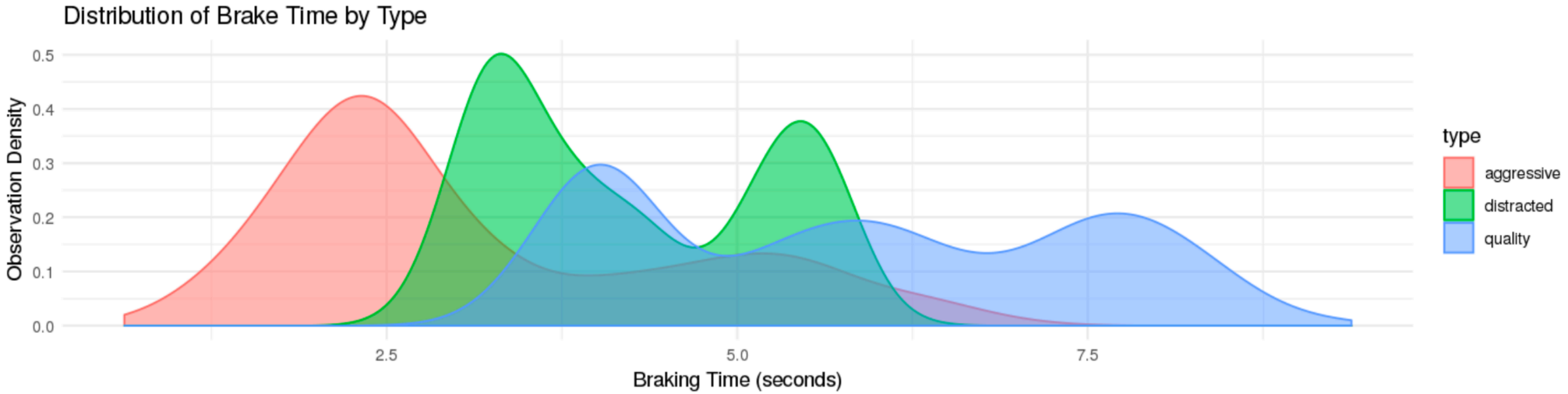
IBM **Watson Studio** | 1867965 - IBM Forum ...

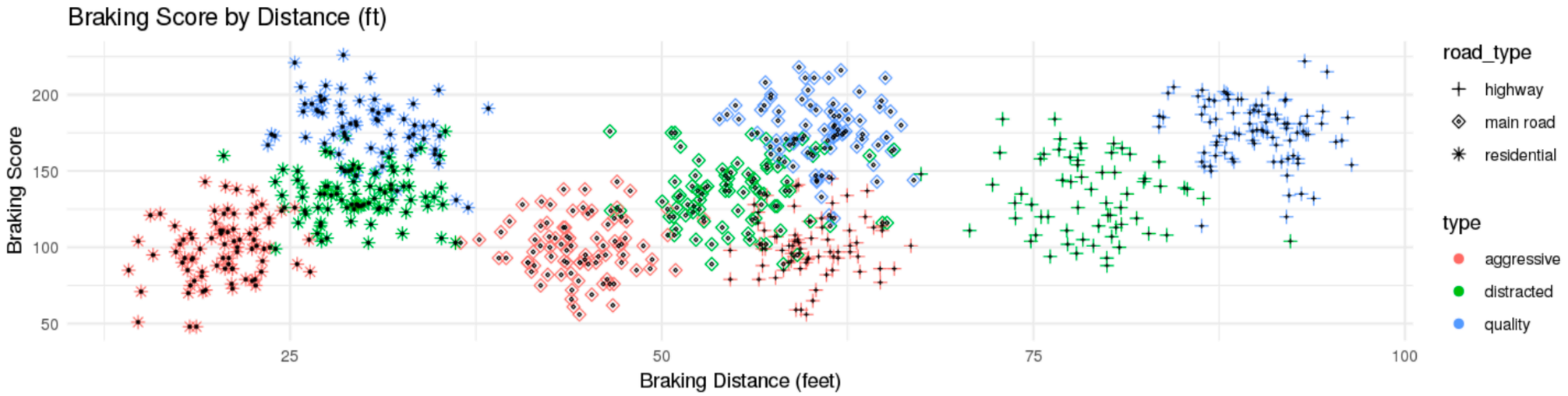
JR FS19 WSL > Notebooks > DriverClassification

File Edit View Insert Cell Kernel Widgets Help | R 3.4.3 with Watson Studio Spark 2.0.2

Markdown | CellToolbar

```
ggplot(aggDF, aes(x = road_type, y = abs_events)) +
    geom_bar(aes(fill = type), stat = 'identity') +
    coord_flip() +
    labs(x = "# of ABS Events", y = "Road Type", title = "ABS Events by Road Type and Event Type") +
    theme_minimal()
```





RStudio Assets

IBM **Watson Studio** | 1867965 - IBM Forum ...

Projects > JR FS19 WSL > RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

ui.R

Run App

```
#
# This is the user-interface definition of a Shiny web application. You can
# run the application by clicking 'Run App' above.
#
# Find out more about building applications with Shiny here:
#
#    http://shiny.rstudio.com/
#
newEventsDF <- read.csv("./new_brake_events.csv")
oldEventsDF <- read.csv("./historical_brake_events.csv")

library(shiny)

# Define UI for application that draws a histogram
shinyUI(fluidPage(

  # Application title
  titlePanel("Analyze Recent Brake Events"),

  # Sidebar with a slider input for number of bins
  sidebarLayout(
    sidebarPanel(
      selectInput("xvar",
                  label = h3("Choose X-axis Variable"),
                  choices = colnames(newEventsDF[, -1])
```

1:1 (Top Level) R Script

Console Terminal

```
    dataTableOutput, renderDataTable

data.table 1.11.4  Latest news: http://r-datatable.com
randomForest 4.6-14
Type rfNews() to see new features/changes/bug fixes.

Attaching package: 'randomForest'

The following object is masked from 'package:ggplot2':

    margin

test

>
```

Environment History Connections

Import Dataset | List

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder | Upload | Delete | Rename | More

Home

Name	Size	Modified
demoBrakeEvents		
ibm-sparkaas-demos		

Visualization Assets

IBM Watson Studio

1867965 - IBM Forum ...

All tabs

Time: 2013 1

Region

Products

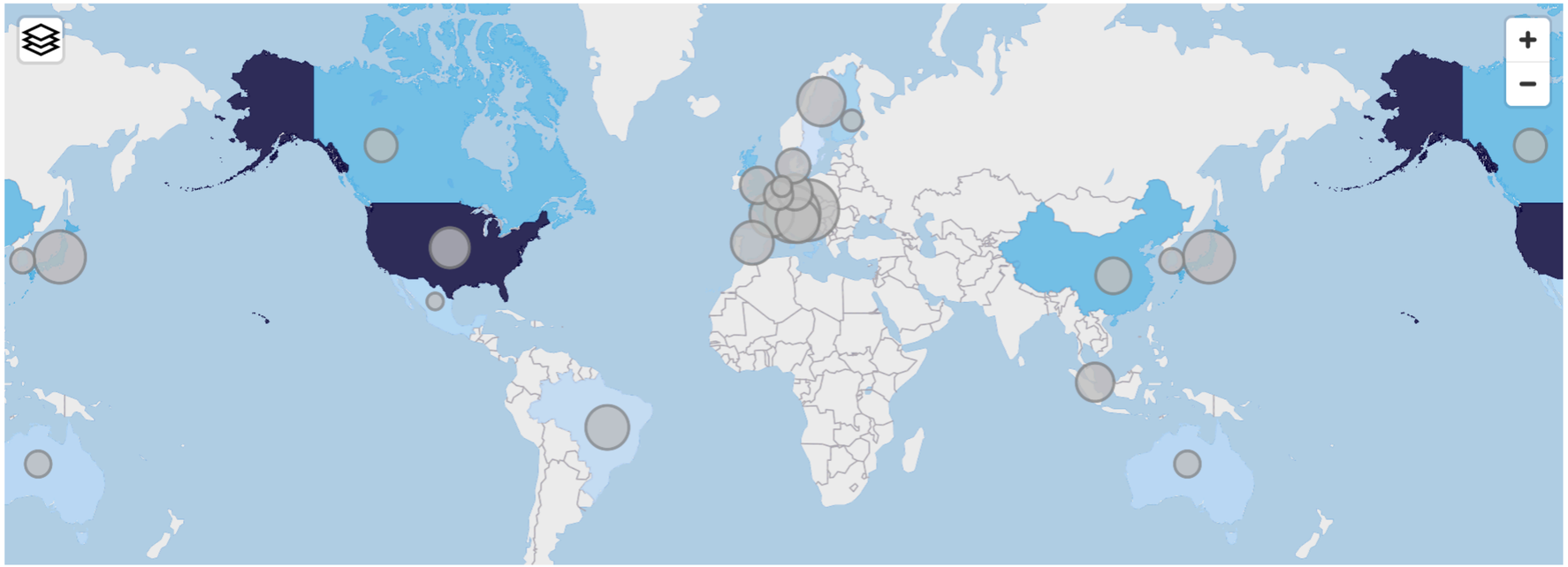
This tab

Drag and drop data here to filter this tab.

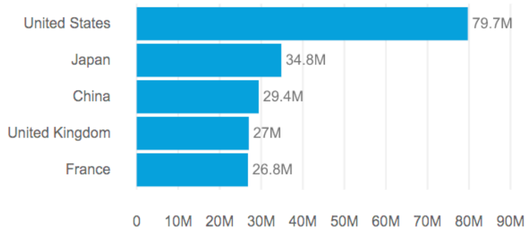
World Sales | Sales Trends | Sales Breakdown | Sales Numbers

REVENUE	GROSS PROFIT	PROFIT MARGIN %	QUANTITY SOLD	AVG DISCOUNT %
€1.12B	€465M	41.6%	19.6M	2%

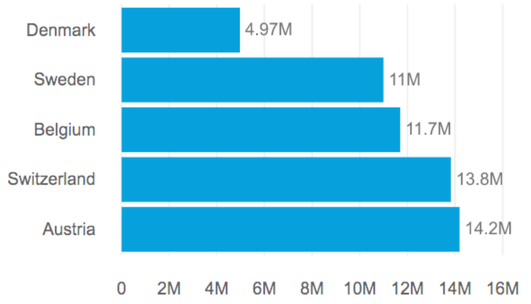
REVENUE AND GROSS MARGIN by COUNTRY



TOP 5 COUNTRIES by GROSS PROFIT



BOTTOM 5 COUNTRIES by GROSS PROFIT



Powered by IBM Cognos Analytics

Analyze Text/Language

Dashboard

Show Documents

Expression / claim_product

Column: claim_product

Row: Expression

Expression	apple juice	pine juice	chocolate	lemon tea	minerals	muffin	N/A	milk chocolate	chocolate muffin	cookie	apple juice (bottle)	peach juice	chocolate cookie	apple jelly	black coffee	orange juice (bottle)	strawberry ice cream
allergic	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	4 0.72	0 0.00	0 0.00	0 0.00	2 0.28	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
dirty	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	4 7.25	0 0.00	0 0.00	0 0.00	0 0.00
tear	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	4 3.38	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
have strange-odor	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
love	0 0.00	2 0.57	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	3 3.15	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
TOO tight	0 0.00	0 0.00	0 0.00	0 0.00	3 2.35	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1 0.19	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
full	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
have rancid-odor	1 0.05	0 0.00	0 0.00	0 0.00	0 0.00	1 0.06	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00

Facet Analysis

Pairs

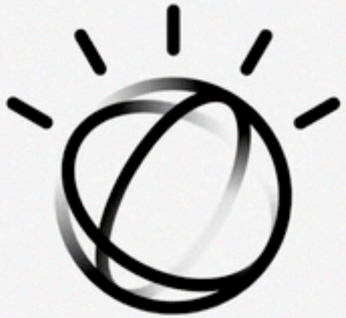
25

Target Facets

- label
- Part of Speech
- Phrase Constituent
- Sentiment Expression
  - Phrase
  - Expression
  - Target
- date
- client_age
- client_location
- client_segment
- claim_id
- claim_product
- client_sex
- claim_product_line

Show Always
Show Subfacet
Append Mode

Analyze



Watson Explorer

youtube.com/watch?v=feLLE-MrLhE

Leverage Communities

IBM **Watson Studio** | 1867965 - IBM Forum ... | Create new

All filters | What are you looking for?

Popular filters: Spark | Deep Learning | Brunel

Featured

Sort by: **Featured**

ARTICLE
Bringing the Human-in-the-loop needed for AI...
AUTHOR: IBM Watson
DATE: Oct 10, 2018
TOPIC: Watson
FORMAT: Web page
♥ 4

TUTORIAL
Detecting Whisky brands with Core ML and IBM...
AUTHOR: Martin Mitrevski
DATE: Apr 25, 2018
LEVEL: Beginner
TOPIC: Watson
♥ 22

TUTORIAL
Build Deep Learning Architectures With...
AUTHOR: developerWorks TV
DATE: Apr 02, 2018
LEVEL: Beginner
TOPIC: Deep Learning +2
♥ 26

ARTICLE
Introducing IBM Watson Studio
AUTHOR: Armand Ruiz
DATE: Mar 20, 2018
TOPIC: Watson
FORMAT: Web page
♥ 21

All content

ARTICLE
Automating and Accelerating Hyperparameter...
AUTHOR: IBM Watson
DATE: Mar 27, 2019
TOPIC: Deep Learning +1
FORMAT: Web page
♥ 0

TUTORIAL
Classify spectrograms using a visual...
AUTHOR: IBM
DATE: Mar 26, 2019
LEVEL: Beginner
TOPIC: Watson
♥ 0

ARTICLE
A Beginner's Guide to Machine Learning
AUTHOR: Noteworthy - Shaurya Bhandari
DATE: Mar 25, 2019
TOPIC: Machine Learning
FORMAT: Web page
♥ 0

NOTEBOOK
Data model with Streaming Analytics and...
AUTHOR: IBM
DATE: Mar 22, 2019
TOPIC: Science & Technology
♥ 11

Discovery
Add a cognitive search and content analytics engine to applications.

Add

Language Translator
Translate text, documents, and websites from one language to another. Create industry or region-specific t

Add

Machine Learning
IBM Watson Machine Learning - make smarter decisions, solve tough problems, and improve user outcomes

Add

Natural Language Classifier
Natural Language Classifier uses advanced natural language processing and machine learning techniques to creat

Add

Natural Language Understanding
Analyze text to extract meta-data from content such as concepts, entities, emotion, relations, sentime

Add

Personality Insights
The Watson Personality Insights derives insights from transactional and social media data to identify psych

Add

Speech to Text
Low-latency, streaming transcription

Add

Text to Speech
Synthesizes natural-sounding speech from text.

Add

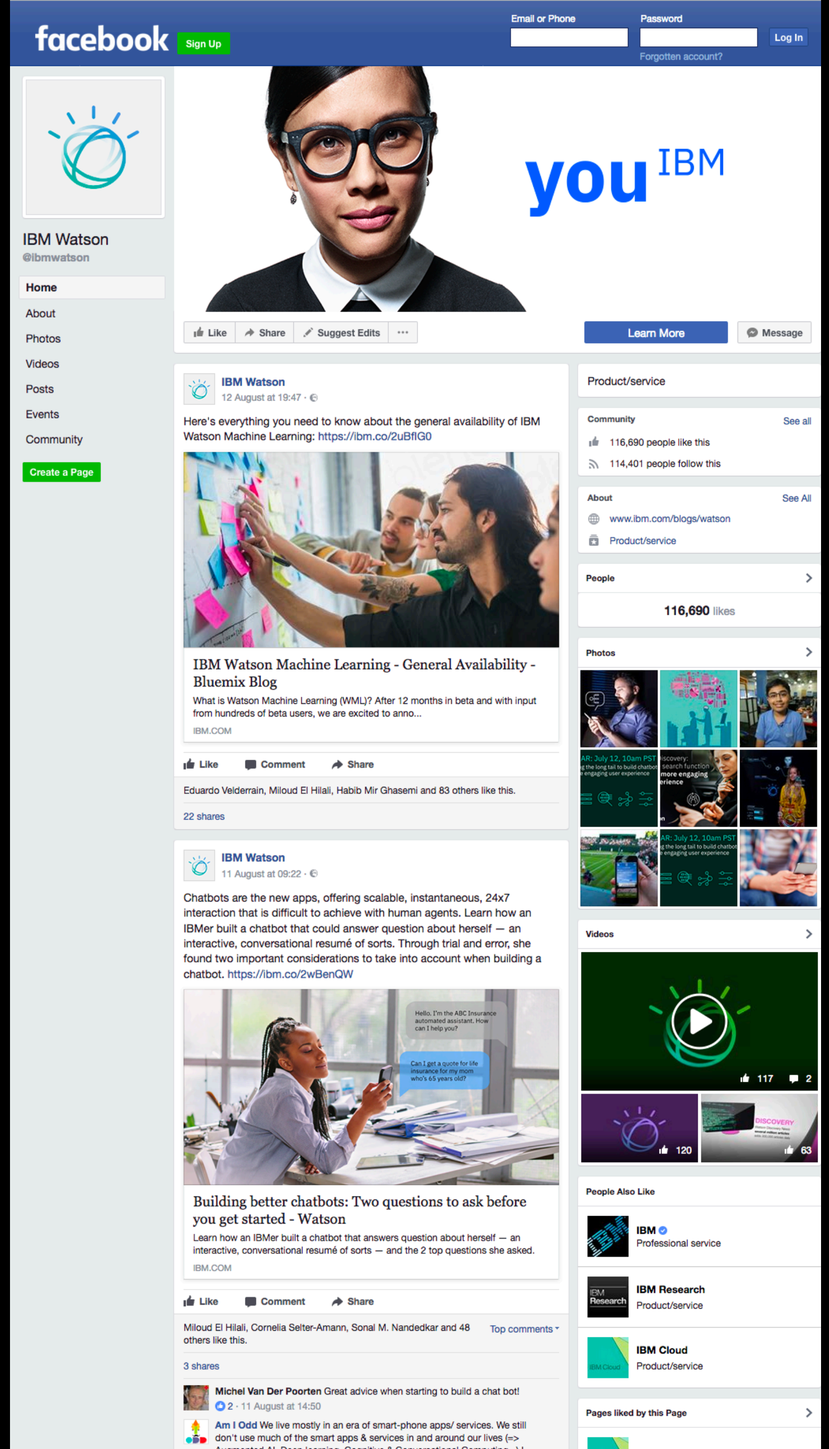
Tone Analyzer
Tone Analyzer uses linguistic analysis to detect three types of tones from communications: emotion, s

Add

Visual Recognition
Find meaning in visual content! Analyze images for scenes, objects, faces, and other content. Choose a def

Add

Add Watson Cognitive Services



One nice example of using **Watson Studio** along with **Watson Services** to analyze social media for ...

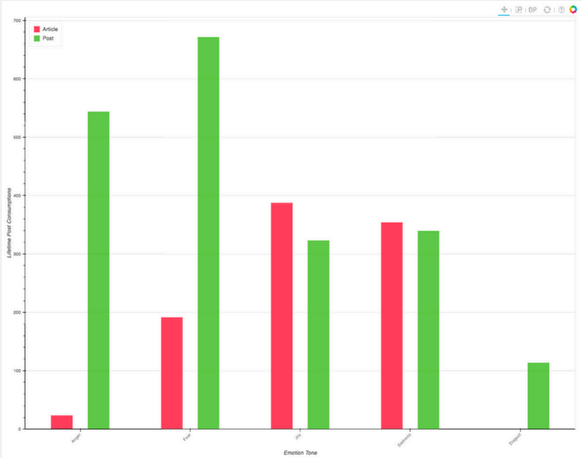
- **content**, **concepts** and **meaning** (text, images, ratings....)
- understanding **sentiment**, **likes/dislikes**
- capturing **emotional tones**
- pointing to **related articles** and **content**
- recognizing **key influencers**
- understanding their **personalities**
- etc…

... to get prepare for targeted actions

© IBM Corporation

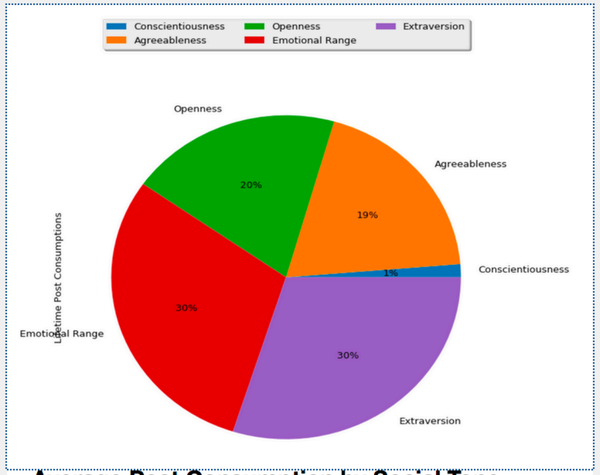
SOCIAL MEDIA ANALYSIS

Filter: **ALL** | NLU | Tone Analyzer | Sentence Tone | Article Text | Post Text | Thumbnail Text



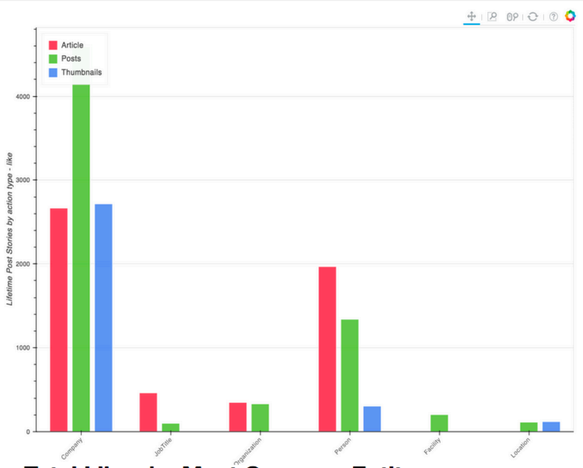
Average Post Consumption by Emotion Tone

When Post Consumption(any interaction with the post) is highest, the most common Emotion Tones in the post text are Fear and Anger



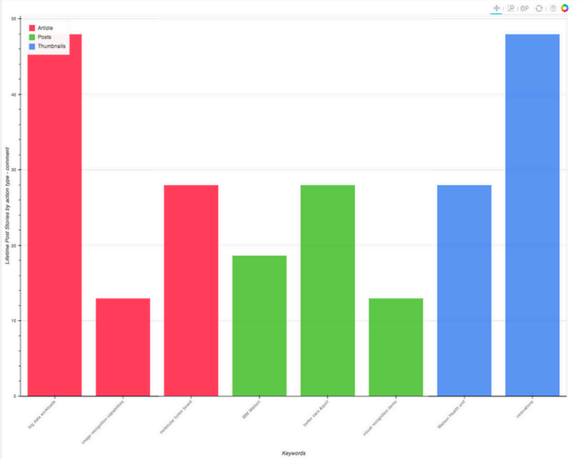
Average Post Consumption by Social Tone

When Post Consumption is highest, the Social Tone is typically Emotional Range/Neuroticism, or Extraversion

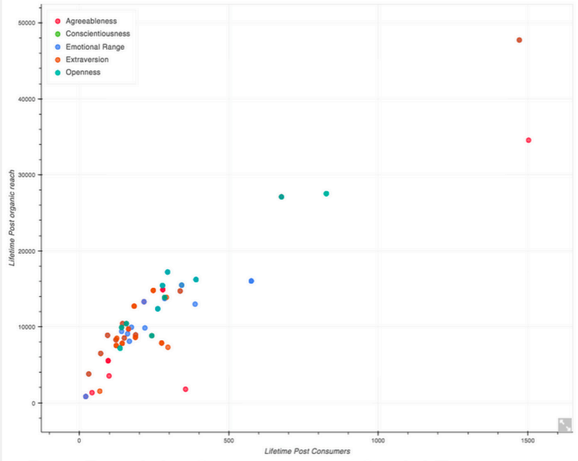


Total Likes by Most Common Entity

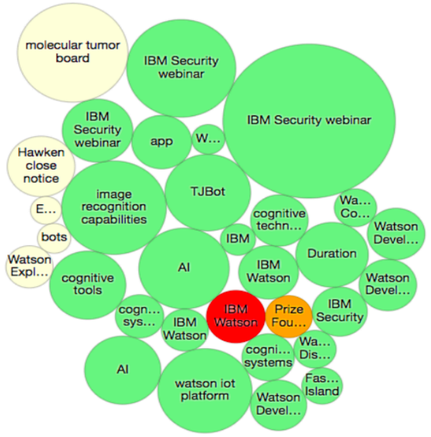
When a post is well liked the post, article or thumbnail text typically contains the 'Company' entity



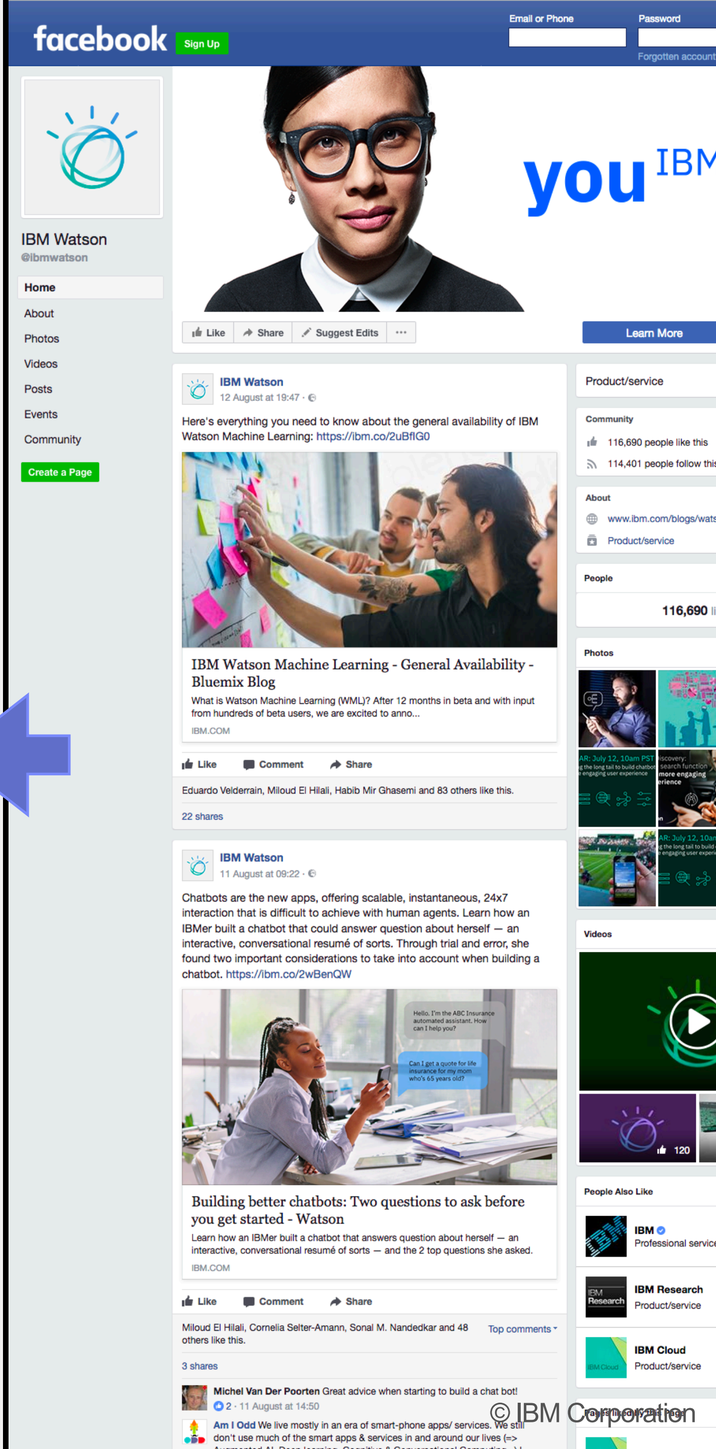
Avg. Number of Comments by Keywords



Post Reach by Consumers w/ Social Tone



Shares by Most Common Article Keywords



© IBM Corporation

About

Watson Services (Cognitive)

Copyright © 2019 IBM Corporation.

Foundation for Analytics/AI

Solutions

Targeted solutions for enterprise businesses

- Cognitive Maintenance & Quality
- Continuous Engineering
- Factory and Spare Part Logistics
- Customer Experience
- Data Science
- IoT
- Virtual agent
- ...

Cognitive AI

Cognitive building blocks for developers

API Conversation	API Discovery	API Compare and Comply	API DLaaS	API Tone Analyzer	API Personality Insight
API Visual Recognition	API Speech	API Document Conversion	API Nat Language Understanding	TOOL Knowledge Studio	...more

Analytics & ML/DL

Get actionable insights from data

Planning & Scenario Modeling →	Descriptive Analytics →	Data Exploration & Discovery →	Predictive Analytics, Machine Learning Deep Learning →	Prescriptive Analytics
Planning Analytics	**Cognos Analytics**	**Watson Analytics Watson Explorer**	**Watson ML/DL with Python, R, SPSS etc.**	**Decision Optimization**

Data & Governance

Prepare data for analytics and AI

Metadata	Transform	Quality	Regulations	Publish
Connect, Virtualize →	Events →	SQL →	NoSQL →	Objects

Hybrid Cloud

A highly scalable, security enabled infrastructure

cloud.ibm.com

Containerized Microservices on Docker & Kubernetes

Networking	Compute	Security	Dedicated Compute	Virtual Compute	Object Storage	...more

Public | Hybrid | Private

Watson services (APIs) are cognitive "building blocks" that you can teach for your use!

How Watson sees, hears, conversates, feels, interprets, finds...

Vision

- Visual Recognition

Speech

- Text to Speech
- Speech to Text

Discovery

- Discovery News
- Knowledge Studio
- Discovery Service
- Natural Language Understanding
- Retrieve and Rank

Empathy

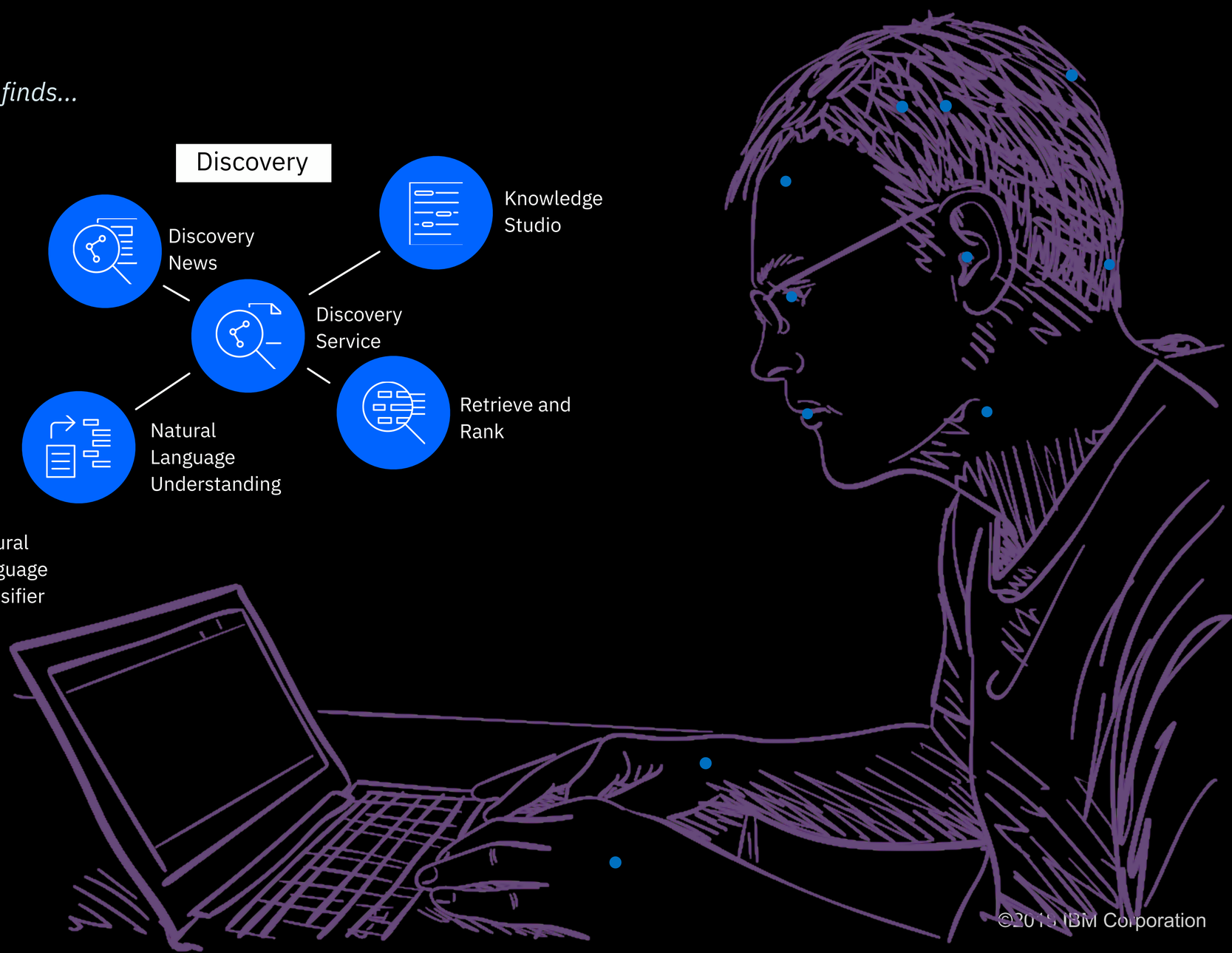
- Tone Analyzer
- Personality Insights

Language

- Natural Language Classifier
- Language Translator

Conversation

- Conversation Service
- Virtual Agent



©2015 IBM Corporation

MACHINE COGNITION

HUMAN COGNITION

SENSE, INTERPRET, PERCEIVE

EVALUATE, REASON, UNDERSTAND

OBSERVE, COMMUNICATE, EXECUTE

(AI)

Speech to Text

Visual Recognition

(I)

Back brain
senses, vision, hearing, touch...

Right brain
visual memory, perceiving dimensions, distances, coordination...

(AI)

Natural Language Classifier

Natural Language Understanding

Knowledge Studio (NLP)

Personality Insights

Tone Analyzer

Discovery News

Discovery Service

Virtual Agent

Conversation

Computation (ML/DL)

Language Translator

(I)

Front brain
feelings, evaluate meaning, reasoning, planning actions, choises, targets, decisions

Left brain
language, producing and understanding speech, writing, calculating...

(AI)

Text to Speech

(I)

Hearing
capture sounds, verbal communication

Vision
capture sight, visual communication

Vocal cords
produce sounds/speech, verbal communication

(Logic)

Process Workflow

Application Logic

UI & Visualization

(Logic)

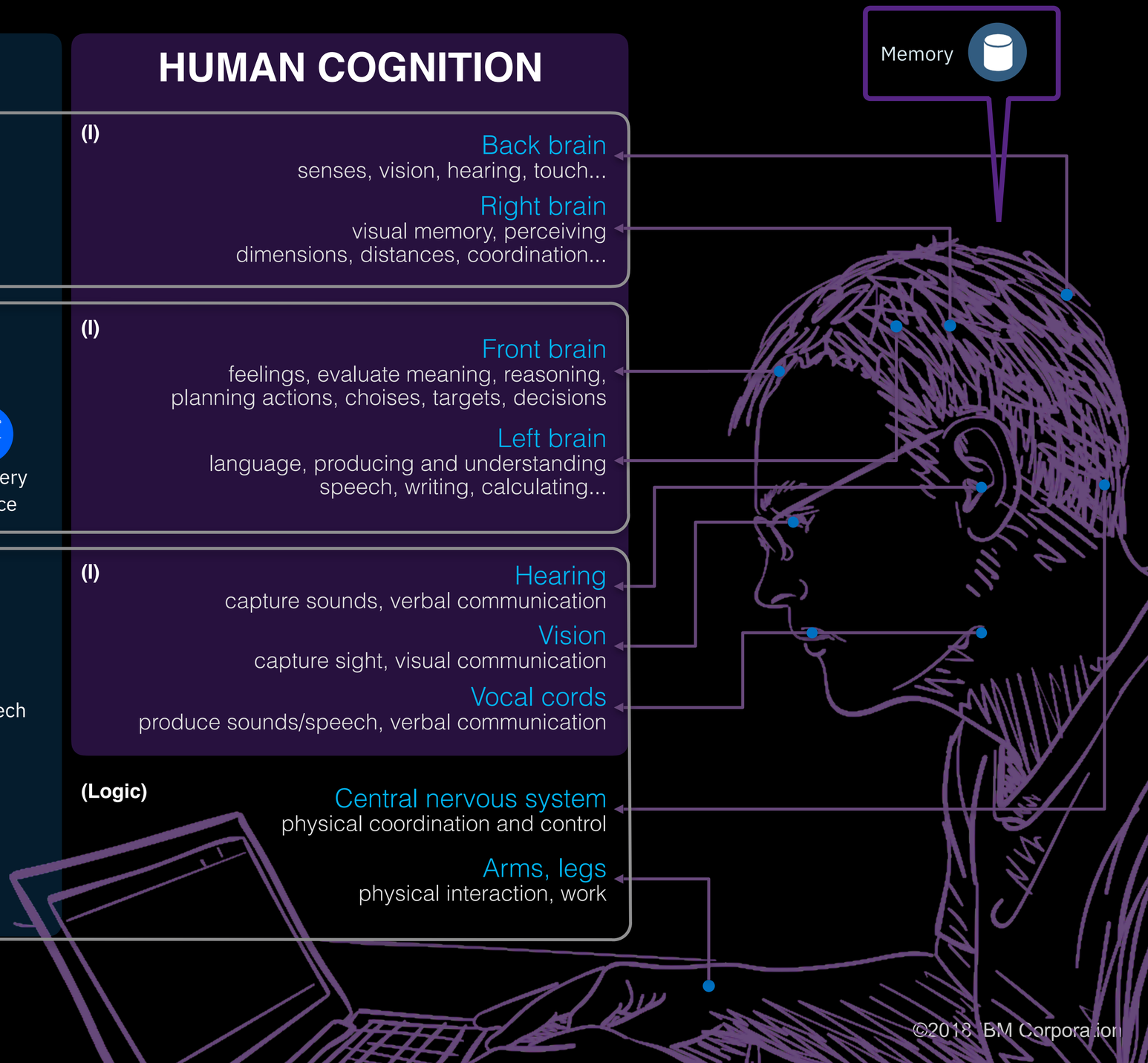
Central nervous system
physical coordination and control

Arms, legs
physical interaction, work

Events

Data

Memory



©2018 IBM Corporation

Aerialtronics, a Netherlands-based manufacturer of drones used to monitor enterprise assets, such as inspecting wind turbines, oil rigs, cell towers, and monitoring traffic. Aerialtronics will use the IBM Watson IoT Platform and visual recognition APIs to analyze images and identify specific areas of concern such as loose or frayed cabling and damaged equipment that could impact the quality of telecommunications service to consumers.

Watson IoT

Visual Recognition

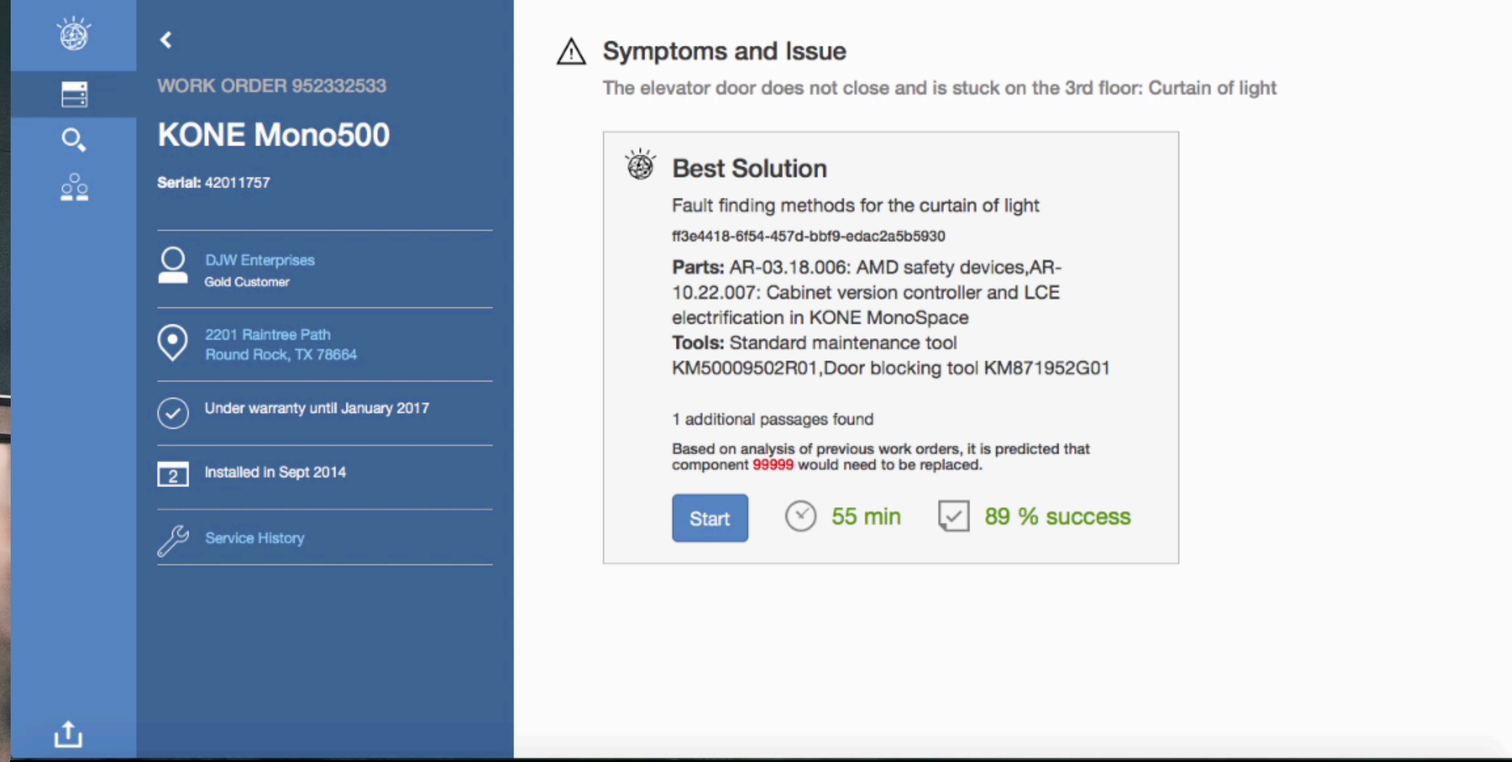
AERIALTRONICS
REMOTELY PILOTED AIRCRAFT SYSTEMS

With Watson you can move millions of people smarter

KONE moves 1 billion passengers throughout the world every day, creating an enormous maintenance challenge

KONE equipped their elevators and escalators with IoT sensors, connected to the IBM Cloud and Watson, to analyze for wear and tear, bumpy ride, and misalignments

Solution makes preventative maintenance possible — saving downtime for KONE and their riders



Note: This is an illustrative showcase demo only - Not an actual solution!

Watson IoT

Retrieve and Rank

Discovery Service

©2018 IBM Corporation

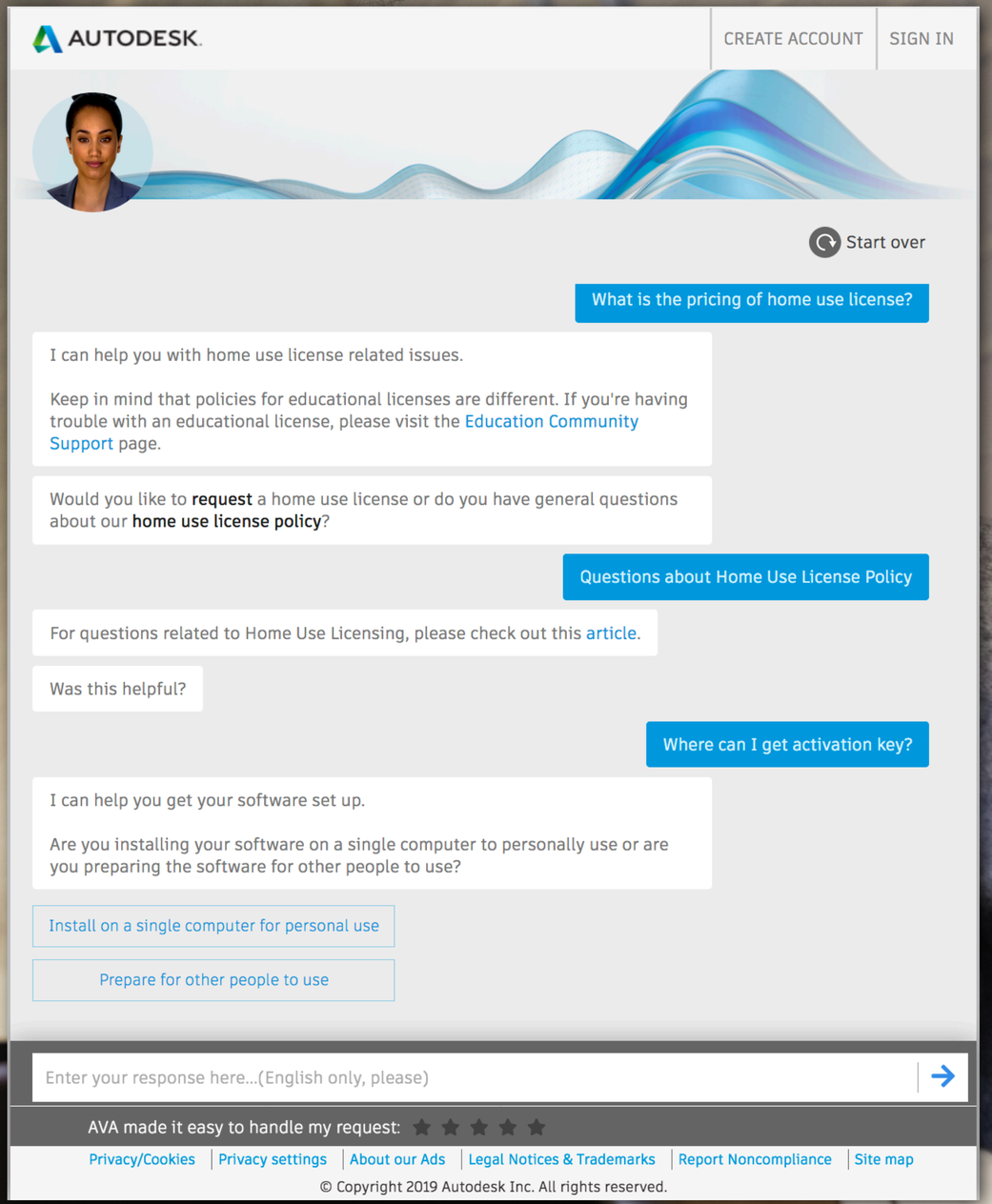
With Watson AI you can talk to customers on their terms

https://ava.autodesk.com

Autodesk was faced with the challenge of scaling real-time customer service and support after shifting to a subscription-based business model

With Watson, Autodesk created a solution that supports 20,000 conversations per month and recognizes 40 distinct use cases, freeing agents to focus on customers with complex issues

The solution cut resolution time from 1.5 days to just 5.4 minutes for most inquiries



Assistant Service

Natural Language Understanding

AUTODESK.

TED

What is the relationship between income inequality and crime?

Powered by
IBM Watson™

How Not To Be Ignorant About The World

Hans Rosling
Global health expert; data visionary

Even the most worldly and well-traveled among us will have their perspectives shifted by Hans Rosling. A professor of global health at Sweden's Karolinska Institute, his current work focuses on **dispelling common myths about the so-called developing world**, which (he points out) is no longer worlds away from the West. In fact, most of the Third World is on the same trajectory toward health and prosperity, and many countries are moving twice as fast as the west did.

What sets Rosling apart isn't just his apt observations of broad social and economic trends, but the stunning way he presents them. Guaranteed: **You've never seen data presented like this.** By any logic, a presentation that tracks global health and poverty trends should be, in a word: boring. But in Rosling's hands, data sings. Trends come to life. And the big picture — usually hazy at best — snaps into sharp focus.

Rosling's presentations are grounded in solid statistics (often drawn from United Nations data), illustrated by the visualization software he developed. The animations transform development statistics into **moving bubbles and flowing curves that make global trends clear, intuitive and even playful**. During his legendary presentations, Rosling takes this one step farther, narrating the animations with a sportscaster's flair.

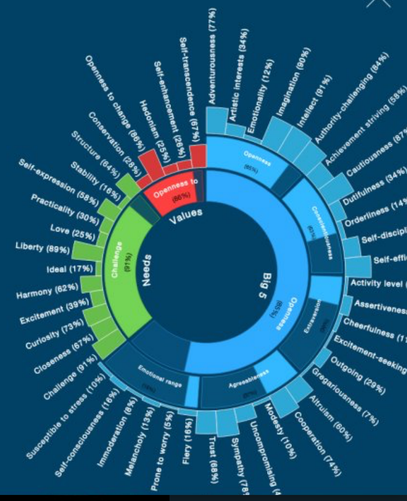
Presence Insights Summary

heartfelt and tranquil.

self-assured: you tend to feel calm and self-assured. calm-seeking: you prefer activities that are quiet, calm, and safe. independent: you have a strong desire to have time to yourself.

Choices are driven by a desire for prestige.

Considers independence to guide a large part of what is done: you like to set your own goals to decide how to best achieve them. Relatively unconcerned with taking pleasure in life: you prefer activities with a purpose greater than just personal enjoyment.



watson.ted.com

© IBM Corporation

TED

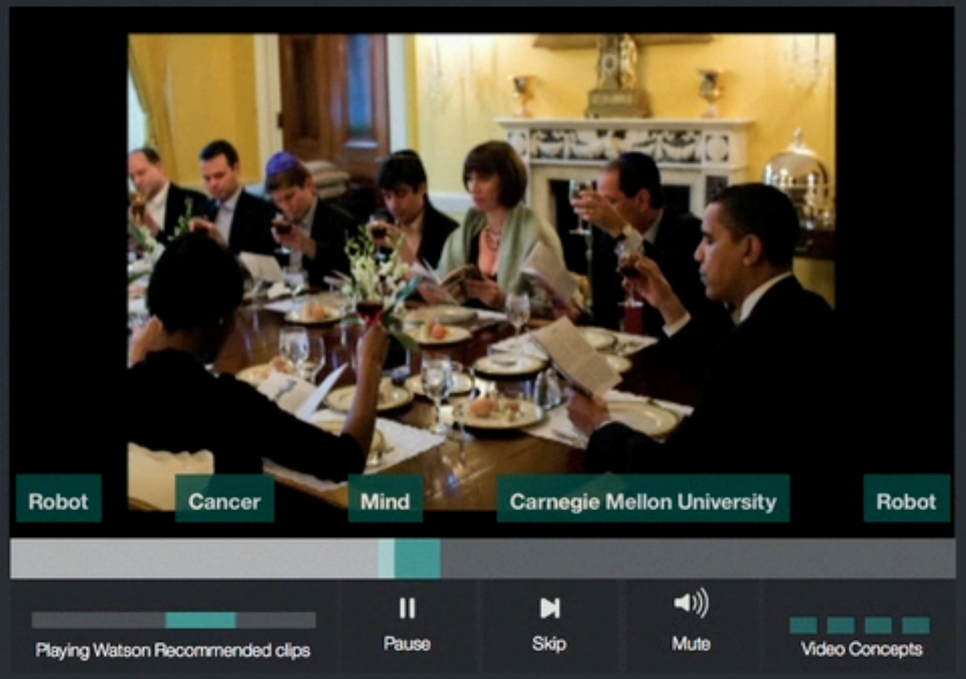
5 Videos with possible answers to your question. Change your question

SHARE

will robots take over human

We found some interesting concepts in this playlist you might like to explore:

Recommended by IBM Watson™ | Videos concepts | Democracy | Robotics | Sociology | Mathematics | Walking



Guy Hoffman
Robots with "soul"
Psychology | Music | Passover

Ken Goldberg
4 lessons from robots about being human
Robotics | Surgery | World Wide Web

Auke Ijspeert
A robot that runs and swims like a salamander
Mathematics | Walking | Simulation

Daniel Suarez
The kill decision shouldn't belong to a robot
Democracy | Sociology | RQ-170 Sentinel

P.W. Singer
Military robots and the future of war
Basketball | Iraq War | Al-Qaeda

How does Watson do it?

Watson uses its understanding of human language to comprehend all of the vi[illegible] sections that may address your question or talks that contain concepts that you may find interesting based on your

TED --- --- ---

Copyright © 2018 IBM Corporation

watson.ted.com

About

"Enterprise Insight Platforms" with IBM Cloud Private for Data

Copyright © 2019 IBM Corporation.

Foundation for Analytics/AI - Context of an Enterprise Insight Platform

Solutions

Targeted solutions for enterprise businesses

- Cognitive Maintenance & Quality
- Continuous Engineering
- Factory and Spare Part Logistics
- Customer Experience
- Data Science
- IoT
- Virtual agent
- ...

Cognitive AI

Cognitive building blocks for developers

API Conversation	API Discovery	API Compare and Comply	API DLaaS	API Tone Analyzer	API Personality Insight
API Visual Recognition	API Speech	API Document Conversion	API Nat Language Understanding	TOOL Knowledge Studio	...more

Analytics & ML/DL

Get actionable insights from data

Planning & Scenario Modeling →	Descriptive Analytics →	Data Exploration & Discovery →	Predictive Analytics, Machine Learning Deep Learning →	Prescriptive Analytics
Planning Analytics	**Cognos Analytics**	**Watson Analytics Watson Explorer**	**Watson ML/DL with Python, R, SPSS etc.**	**Decision Optimization**

Data & Governance

Prepare data for analytics and AI

Metadata	Transform	Quality	Regulations	Publish
Connect, Virtualize →	Events →	SQL →	NoSQL →	Objects

Hybrid Cloud

A highly scalable, security enabled infrastructure

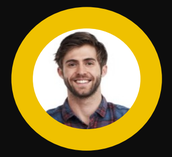
cloud.ibm.com

Containerized Microservices on Docker & Kubernetes

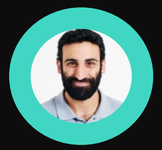
Networking	Compute	Security	Dedicated Compute	Virtual Compute	Object Storage	...more

Public | Hybrid | Private

Accelerating **Enterprise Insight**



Developers



Researchers,
Data Scientists

RESEARCH & DEVELOPMENT

Test

Development

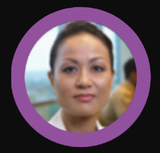
Model building

Refining Data

NEW INSIGHTS

Continuous Delivery of New Applications

Continuous Delivery of New Insights



Business Users

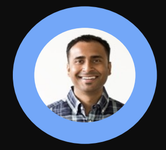
App Deployment

New Requirements & Engagement

NEW DATA

Monitoring Retraining

Acquiring Data/ Self Service



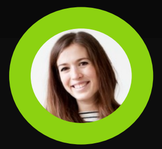
Data Stewards

New Data

Search for Data

BUSINESS PROCESSES, EVENTS & DATA

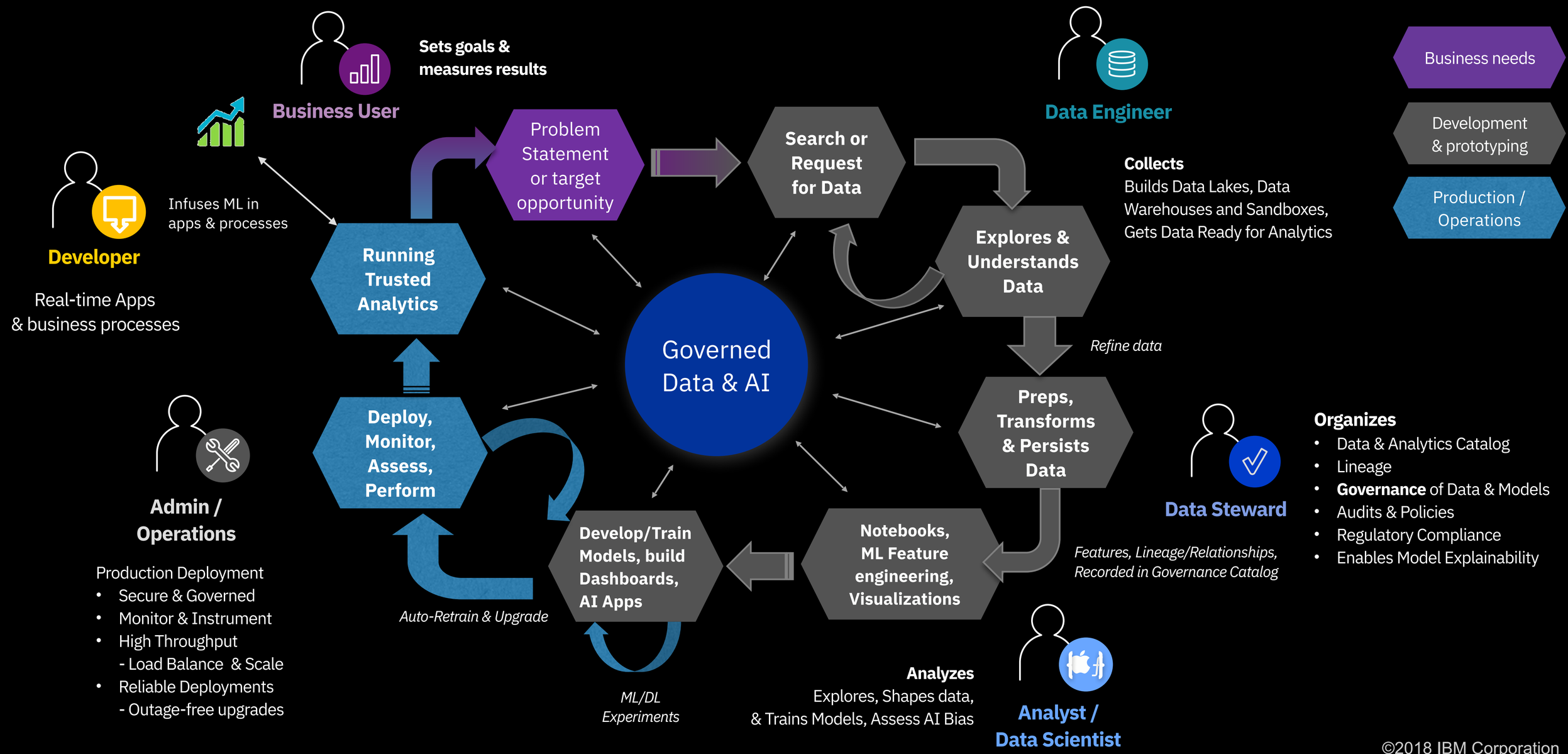
Analysts



Data Engineers

Copyright © 2019 IBM Corporation

End-to-End Scenario: **Accelerating Insights with Governance**



©2018 IBM Corporation

The hard way

(read as “most information architectures”)

- **Hard & Slow**
- **Siloed Data & Workflows**
- **Multiple Disjunct Stacks**

Dream

(read as "ideal IA with AI" :-)

- **Easy & Fast**
- **Pre-Integrated & Governed**
- **Multi-Cloud enabled**

What makes an "Enterprise Insight Platform"?

As defined by Forrester Wave™ when revewing different vendor's solutions on the market:

★ ***Data management tools or services***

The vendors offer well-rounded sets of integrated and complementary data management tools or services geared for analytics that include some, but not necessarily all, of the following: ***data persistence***, ***data catalogs and governance tools***, *and* ***data integration and operations services***.

★ ***Multiple analytical tools or services***

The vendors offer well-rounded sets of analytics tools and frameworks, such as ***business intelligence tooling***, ***predictive analytics***, ***machine learning (ML) and artificial intelligence (AI) services***, ***streaming analytics capabilities***, *and* ***text and entity analytics services***.

★ ***Insight application development tools or services***

The vendors offer some insight application development support services that deliver a ***unified insight team experience***, ***insight application management***, *and* ***insight activation capabilities***. *Vendors should also provide some* ***insight solution accelerators***, *ranging from fully packaged systems to templates and best practices.*

★ ***Unified platform management tools or services***

The vendors offer unified platform management capabilities that include tools for ***automated platform provisioning***, ***monitoring, and scaling*** *as well as* ***services for maintaining availability and reliability***.

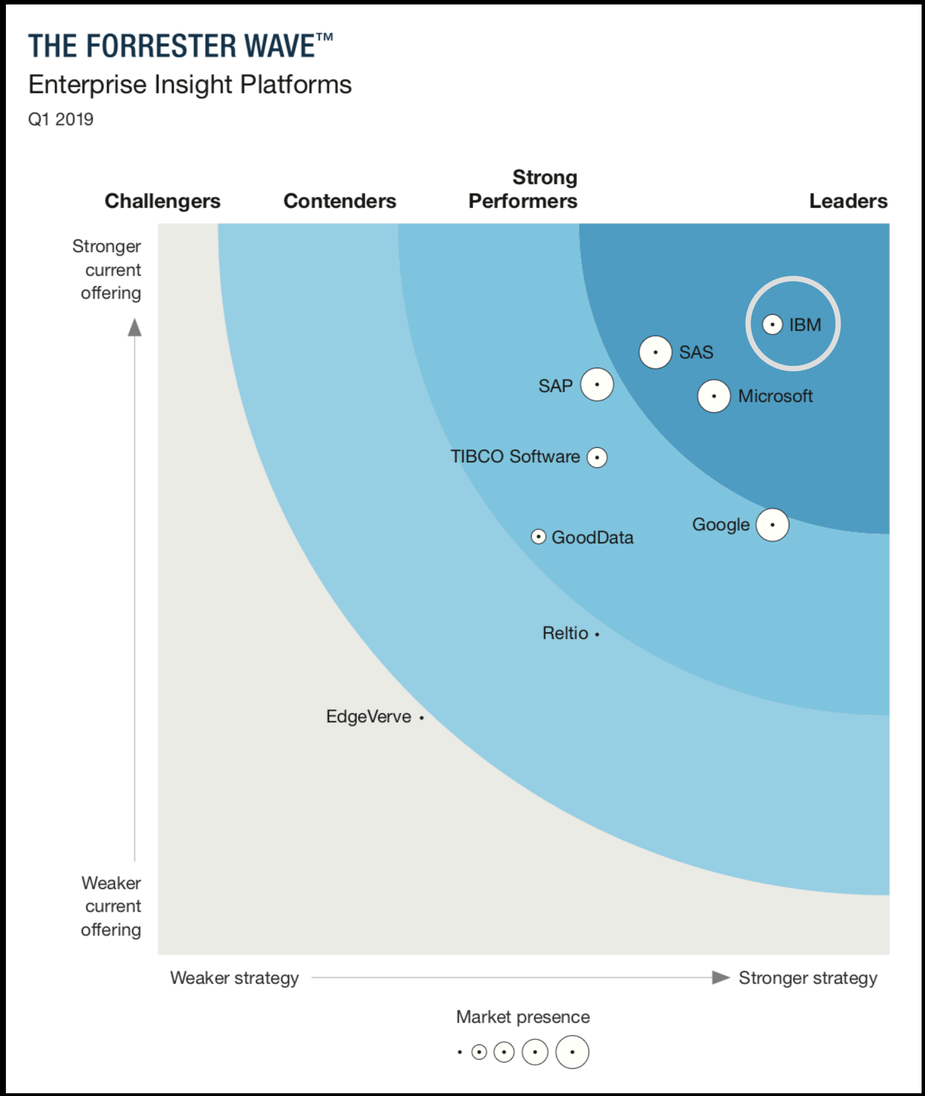
★ ***A strategic focus on insight platform enablement***

Integrating data management and analytics is a conventional vendor approach to increasing the value of their individual tools; in fact, this is so common that it would be impractical as a differentiator. Therefore, we selected only those vendors that told us that ***building an integrated insight platform was important***. *This led us to exclude some vendors that CIOs might otherwise have expected to see in our analysis.*

Source & Full Report: https://www.forrester.com/report/The+Forrester+Wave+Enterprise+Insight+Platforms+Q1+2019/-/E-RES141393
Summary: https://www.ibm.com/downloads/cas/JXNMQN6O

Leaders in Enterprise Insight Platforms

by Forrester Wave™ Q1 2019



★ *IBM has **ML-assisted data cataloging and governance tools** at the core of its offering.*

★ *IBM has an **impressive portfolio of individual data management and analytics capabilities** that have consistently scored well on individual component Forrester Waves.*

★ *With **IBM Cloud Private for Data**, IBM has pre-integrated capabilities that **allow clients to be productive in a week or less**.*

★ *We were also impressed with its **ML-assisted data cataloging and governance tools**.*

★ *IBM's platform **uses Kubernetes to deploy on-premises or into the public cloud**.*

★ *Lastly, IBM's **support for different insight team personas through tailored but unified experiences is commendable**. Firms looking to unify the work of insight teams will do well on this platform.*

Full report: https://www.forrester.com/report/The+Forrester+Wave+Enterprise+Insight+Platforms+Q1+2019/-/E-RES141393
Summary: https://www.ibm.com/downloads/cas/JXNMQN6O

IBM Cloud Private for Data with Watson Studio

Foundational “out of the box” multi-cloud & multi-modal data & AI services

The Ladder to AI

Collect

- Data virtualization
- Data warehousing
- Databases on-demand
- Data source ingestion
- Distributed processing

***Powered by**: IBM & Open Source data storage technologies*

Organize

- Discovery & search
- Data transformation
- Data cataloging
- Business glossary
- Policies, rules & privacy

***Powered by**: IBM Data Governance & Catalogging technologies*

Analyze

- Data visualization
- Machine learning learning
- Model build & deploy
- Model management
- Dashboards

***Powered by**: Open source ML/DL, Watson Studio and Cognos Anaytics*

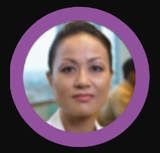
Multicloud Services

- Logging
- Monitoring
- Metering
- Persistent Storage
- Kubernetes
- Security
- Identity Access Mgmt.
- Docker Registry / Helm

IBM Cloud Private for Data

IBM Cloud · RED HAT OPENSHIFT · openstack · amazon web services · Azure

BUSINESS PROCESSES, EVENTS & DATA



Business Users

Businesses

Services

Products

Assets

Utilities

Facilities

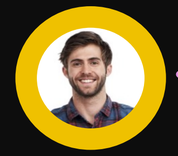
Infrastructures

Analysts

Remember:

"I need insight, where's my AI?"

Developers



RESEARCH & DEVELOPMENT

Business R&D

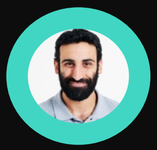
Services R&D

Products R&D

Industry R&D

Academic R&D

Gov/EU Programs



Researchers, Data Scientists

Copyright © 2019 IBM Corporation

Self-Service IA with AI in Action !

Analysts

Developers

BUSINESS PROCESSES, EVENTS & DATA

Business Users

Businesses

Services

Products

Assets

Utilities

Facilities

Infrastructures

Insights

Custom AI

Feedback

Data

Terms

RESEARCH & DEVELOPMENT

Business R&D

Services R&D

Products R&D

Industry R&D

Academic R&D

Gov/EU Programs

Deploy

Develop

Researchers, Data Scientists

Governed Data

New Insights & Opportunities via Operationalized Analytics & AI

Watson Studio
IBM Decision Optimization
Prescriptive
Collaborate for maximum business value, informed by advanced analytics

Watson Studio, SPSS, IBM Data Science Experience
Predictive
Understand the most likely future scenario and its business implications

Diagnostic
IBM Cognos Analytics
Understanding why something happened is a critical step
IBM Watson Explorer
Data Augmentation and Contextualization

IBM Cognos Analytics
Descriptive
Get in touch with reality, a single source of the truth, visibility

RUN — Watson Machine Learning — Analytics & AI Deployment — Watson AI Services — **MONITOR**

Validate AI Trust & Transparency

ML Runtimes & Frameworks: Spark, learn, Spark MLlib, ANACONDA, dmlc XGBoost

ML/DL Model Deployment

DL Runtimes & Frameworks: TensorFlow, Caffe, torch, Keras

TRAIN — Watson Studio — Analytics & AI Projects — Accelerators & Add-ons — **MODEL**

ML/DL Authoring Tools: python, Scala, jupyter, SPSS, R Studio, H_2O.ai, Dashboards & Visuals, Decision Optimization

ML/DL DevOps

ML/DL Asset Cataloging

Add-on Catalog

SHARE / PUBLISH — Data/AI Catalog - Governance Catalog w. Biz Glossary — **REQUEST / SEARCH**

DATA INGESTION & INTEGRATION

Data Engineers

GOVERNANCE, POLICIES & REGULATIONS

Data Stewards

DISCOVER	UNDERSTAND	PREPARE	CONSUME
Data Virtualization	Metadata Management	Consent Management	Regulatory Compliance	Industry Models	Data Preparation	Data Asset Cataloging	Search & Understand
Data Federation	Auto Discovery	Data Profiling	Data Quality	Business Glossary	Data Masking	Policy Activation	Integration for Productive Use
VIRTUAL DATA LAKE				Policy Management			

DATA LAKE: ETL | STREAMS | EVENT STORE | RDBMS | NOSQL | HADOOP | OBJECT STORAGE

PROVISION — IBM Cloud Private for Data — Terraform, HELM, docker, kubernetes

HYBRID-CLOUD INFRASTRUCTURE

vmware, intel, Power Systems, IBM Cloud, RED HAT OPENSHIFT, openstack, amazon web services, Azure, Google Cloud

RUN - MANAGE - SCALE

Copyright © 2019 IBM Corporation

IBM Cloud Private for Data with Watson Studio
Top Use-Cases by Industry Vertical

Over 24 Data Science / AI use cases across 15 industry verticals

(*applicability varies by organization/customer*)

IBM differentiation :

- ✓ **Operationalize ML/AI in a matter of minutes** (Deploy, Scale & Manage models with minimum effort)
- ✓ **Comprehensive Data & ML/AI Governance** (Regulations Compliance, Lineage and Provenance – Who created, when, what data was used, comments, ratings etc.)

Use Case(s)	Aerospace & Defense	Automotive	Banking	Chemicals & Petrooleum	Consumer	Education	Electronics	Energy, Environmental & Utilities	Financial Markets	Government	Healthcare & Life Services	Insurance	Industrial Products	Telco, Media & Entertainment	Travel & Transportation
Predictive Maintenance	X							X							X
Real time analytics (IOT)	X	X		X										X	
Customer Churn / Retention		X	X											X	
Anomaly Detection	X	X					X						X		
Regulatory Compliance			X						X			X			
Anti-Money Laundering (AML)			X												
Cross Sell / Up-Sell			X		X										
Demand Forecasting				X			X	X							
Inventory Optimization					X										
Retention & Time to Degree						X									
Application modernization						X									
Student Safety						X									
Predictive Customer Insights			X					X							
Counter Fraud & Payments									X		X				
Clounter Party Credit-risk									X						
Client Insights for Wealth Management									X						
Threat Prediction & Prevention										X					
Patient Diagnosis											X				
Data Privacy											X				
Client Risk Scoring			X									X			
Targeted Ads														X	
Intrusion Detection	X													X	
Route Optimization															X

Building

Trust & Transparency on AI

Copyright © 2019 IBM Corporation.

ICP for Data with Watson Studio in Action

Analysts

Developers

New Insights & Opportunities via Operationalized Analytics & AI

BUSINESS PROCESSES, EVENTS & DATA

Business Users

Businesses

Services

Products

Assets

Utilities

Facilities

Infrastructures

Insights

Custom AI

Feedback

Data

Terms

RESEARCH & DEVELOPMENT

Business R&D

Services R&D

Products R&D

Industry R&D

Academic R&D

Gov/EU Programs

Validate

Develop

Researchers, Data Scientists

Governed Data

Watson Studio
IBM Decision Optimization
Prescriptive
Collaborate for maximum value, informed by adv

Watson Studio, SPSS, IBM Data Science Experience
Predictive

IBM Cognos Analytics
Understanding why something happened is a critical step

IBM Cognos Analytics
Descriptive
Get in touch with reality, a single source of the truth, visibility

RUN

Validate AI Trust & Transparency

ML Runtimes & Frameworks

TRAIN

ML/DL Authoring Tools

Watson AI Services

MONITOR

torch

Keras

DL Runtimes & Frameworks

Accelerators & Add-ons

MODEL

ML/DL DevOps

ML/DL Asset Cataloging

Add-on Catalog

AI Trust & Transparency

SHARE / PUBLISH

Data/AI Catalog - Governance Catalog w. Biz Glossary

REQUEST / SEARCH

DISCOVER

Data Virtualization

Data Federation

VIRTUAL DATA LAKE

Metadata Management

Consent Management

Regulatory Compliance

Auto Discovery

Data Profiling

Data Quality

UNDERSTAND

Industry Models

Business Glossary

Policy Management

PREPARE

Data Preparation

Data Masking

CONSUME

Data Asset Cataloging

Search & Understand

Policy Activation

Integration for Productive Use

DATA LAKE

ETL

STREAMS

EVENT STORE

RDBMS

NOSQL

HADOOP

OBJECT STORAGE

DATA INGESTION & INTEGRATION

Data Engineers

GOVERNANCE, POLICIES & REGULATIONS

Data Stewards

PROVISION

IBM Cloud Private for Data

Terraform

HELM

docker

kubernetes

HYBRID-CLOUD INFRASTRUCTURE

vmware

intel

Power Systems

IBM Cloud

RED HAT OPENSHIFT

openstack

amazon web services

Azure

Google Cloud

RUN - MANAGE - SCALE

Copyright © 2019 IBM Corporation

Main issues/concerns in AI

Short-term:

- Bias
- Explainability
- Transparency
- Accountability
- Data responsibility
- Value alignment
- Ethics/morality
- Impact on jobs and society

Long-term?

- Singularity & Superintelligence
- Switch-off problem

Copyright © 2019 IBM Corporation.

Bias in AI

Bias: Prejudice for or against something

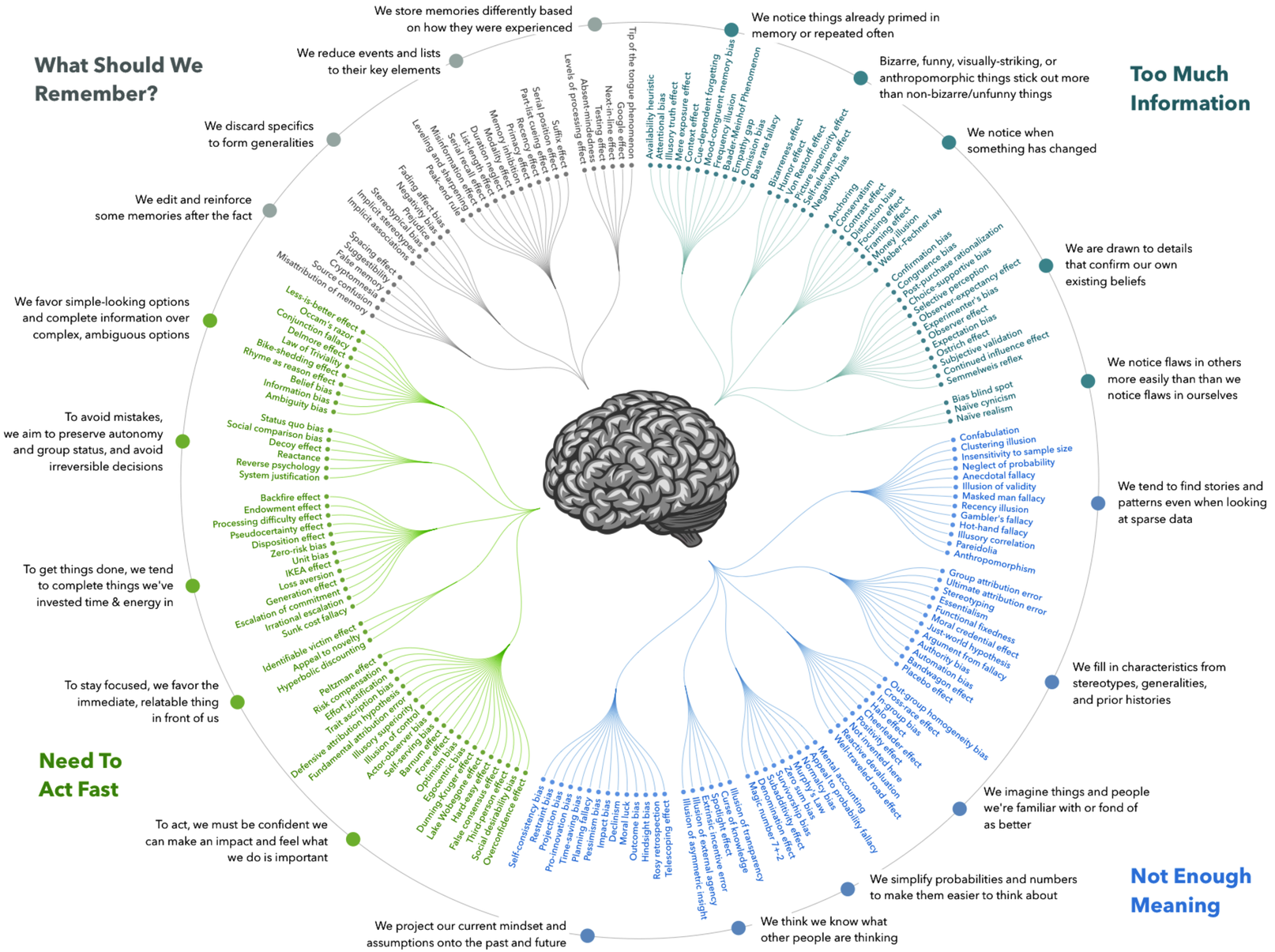
- As a consequence of bias, one could behave unfairly to certain groups, compared to others

What makes AI biased?

- Trained on data provided by people
  - people are always biased
- Learnt from examples
  - generalizing to situations never seen before

Copyright © 2019 IBM Corporation.

COGNITIVE BIAS CODEX

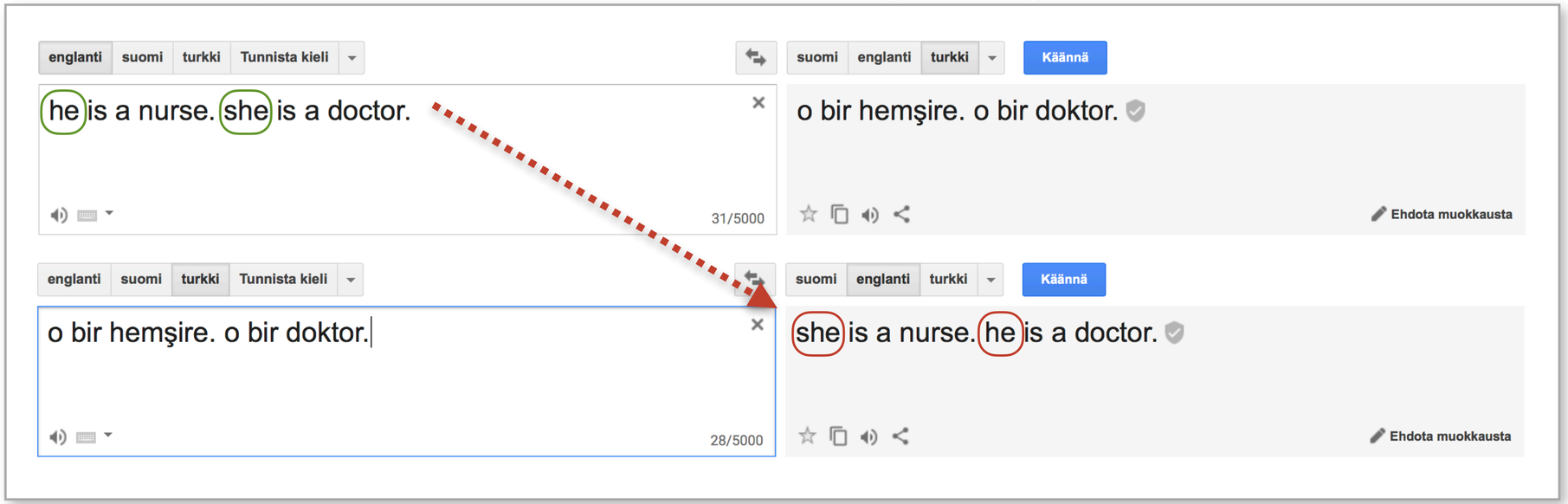


creative commons attribution · share-alike

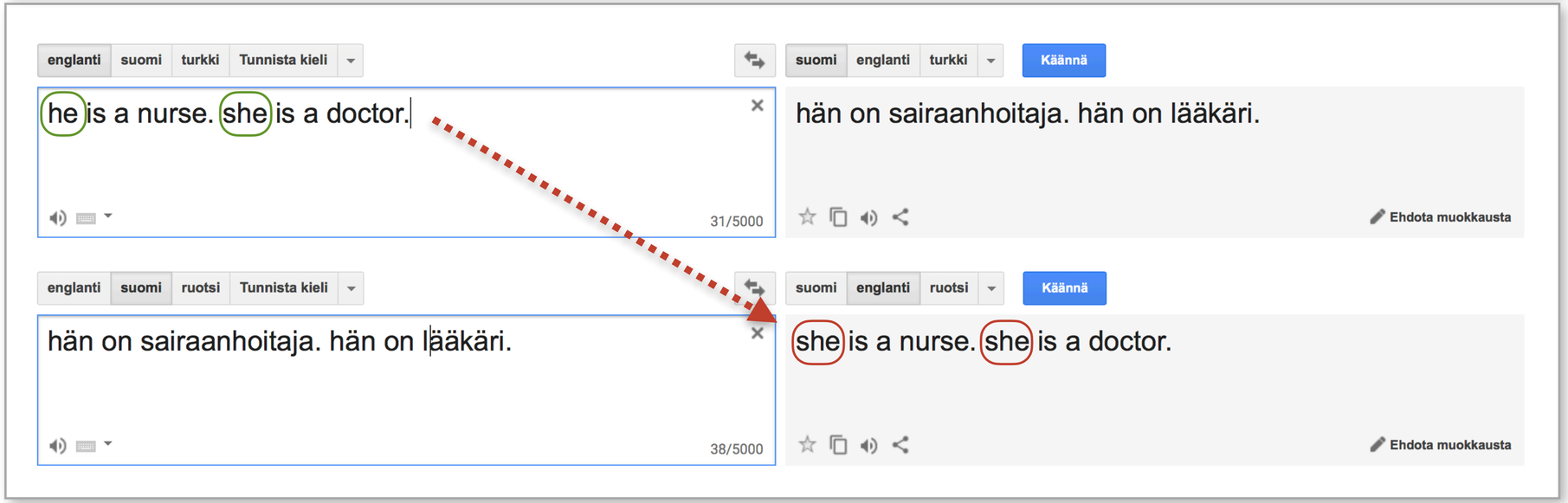
© 2019 IBM Corporation.

Source: https://commons.wikimedia.org/wiki/File:The_Cognitive_Bias_Codex_-_180%2B_biases,_designed_by_John_Manoogian_III_(jm3).png

Example: AI-based language translation



Is it trained with an assumption that most nurses are women and most doctors are men?



Is it trained with an assumption that most healthcare professionals are women?

Ref: https://en.wikipedia.org/wiki/Google_Neural_Machine_Translation

Copyright © 2019 IBM Corporation.

Example: Gender classification from pictures

Commerical visual recognition services were reported to have trouble in classifying gender of dark skinned female faces.

"Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,"

Classifier	Metric	DF	DM	LF	LM
MSFT	PPV(%)	76.2	**100**	**100**	**100**
	Error Rate(%)	**23.8**	0.0	0.0	0.0
	TPR(%)	**100**	84.2	**100**	**100**
	FPR(%)	**15.8**	0.0	0.0	0.0
Face++	PPV(%)	64.0	99.5	**100**	**100**
	Error Rate(%)	**36.0**	0.5	0.0	0.0
	TPR(%)	99.0	77.8	**100**	96.9
	FPR(%)	**22.2**	1.03	3.08	0.0
IBM	PPV(%)	66.9	94.3	**100**	98.4
	Error Rate(%)	**33.1**	5.7	0.0	1.6
	TPR(%)	90.4	78.0	96.4	**100**
	FPR(%)	**22.0**	9.7	0.0	3.6

Source: http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Mitigating Bias in AI Models

Country	Total	Male	Female	Lighter Male	Darker Male	Lighter Female	Darker Female
Finland	194	113	81	113	0	81	0
Iceland	63	39	24	39	0	24	0
Rwanda	26	16	10	0	16	0	10
Senegal	161	95	66	0	95	0	66
South Africa	424	246	178	63	183	27	151
Sweden	349	187	162	180	7	158	4
All	1217	696	521	395	301	290	231
Errors @ score threshold = 0.99	15	7	8	1	6	0	8
Error as %	1.23%	1.005%	1.535%	0.253%	1.99%	0	3.46%

Source: https://www.ibm.com/blogs/research/2018/02/mitigating-bias-ai-models/

Copyright © 2019 IBM Corporation.

Example: Chatbot turning to racist, search algorithms promoting child abuse, conspiracy and extreme content...

Facebook and YouTube should have learned from Microsoft's racist chatbot

- Facebook and YouTube have recently come under fire for offensive search suggestions.
- Microsoft made a Twitter chatbot in 2016 that was trained to say outrageous things by users, but Facebook and YouTube don't seem to have learned from the mistake.
- Psychological studies have shown that people are drawn to negative and offensive content, so engagement maximization drives the popularity of this content.

Ingrid Angulo | @Ingrid__Angulo

Published 4:03 PM ET Sat, 17 March 2018

CNBC

Microsoft showed us in 2016 that it only takes hours for internet users to turn an innocent chatbot into a racist. Two years later, Facebook and YouTube haven't learned from that mistake.

Facebook came under fire on Thursday night after users noticed search suggestions alluding to child abuse and other vulgar and upsetting results when people started typing "video of..." Facebook promptly apologized and removed the predictions.

YouTube has also been the subject of investigations regarding how it highlights extreme content. On Monday, Youtube users highlighted the prevalence of conspiracy theories and extreme content in the website's autocomplete search box.

Source: https://www.cnbc.com/2018/03/17/facebook-and-youtube-should-learn-from-microsoft-tay-racist-chatbot.html

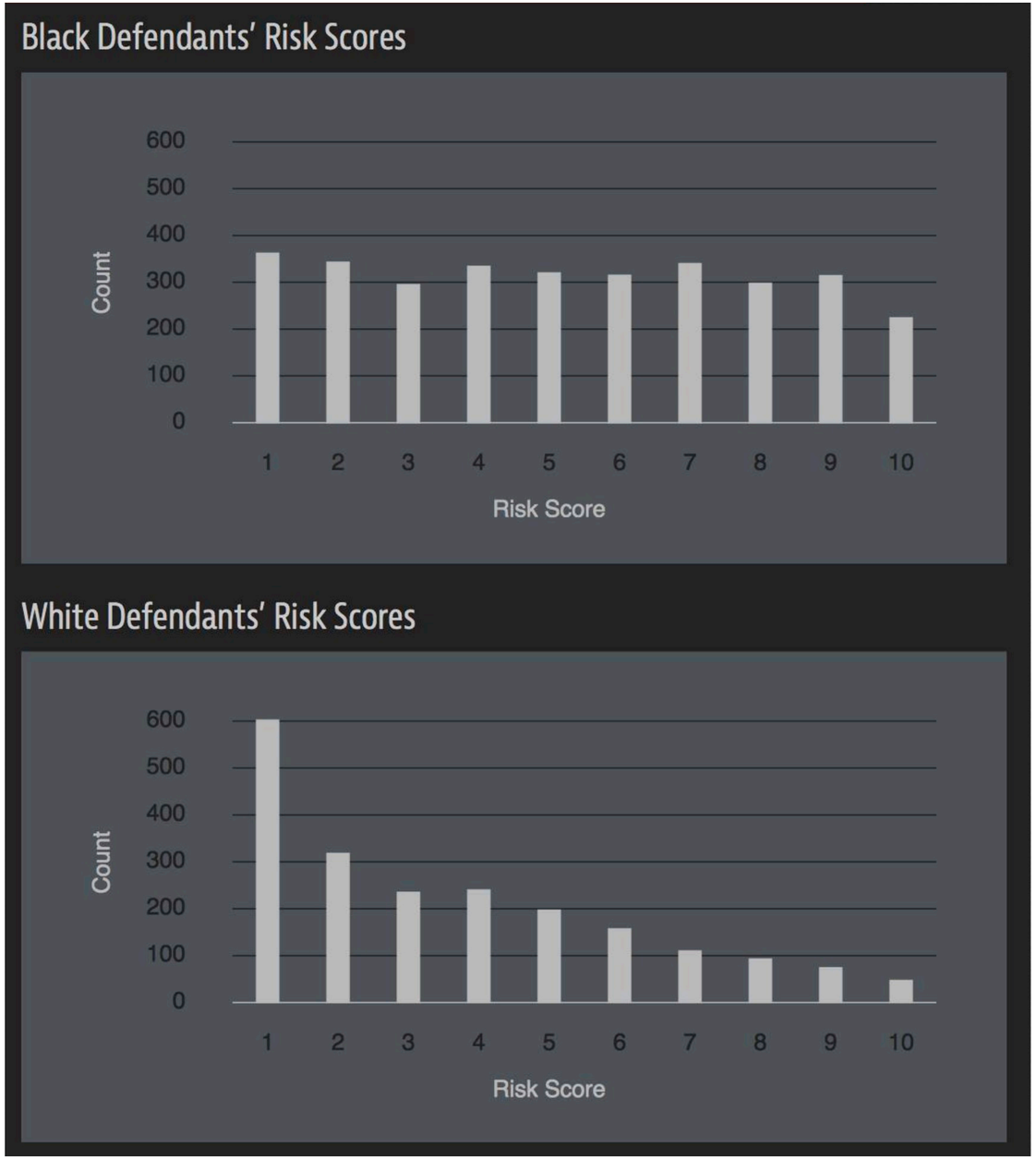
Copyright © 2019 IBM Corporation.

Example: Judicial system

The tool correctly predicts recividism 61% of the time.

African Americans are almost twice as likely as White Americans to be labeled a higher risk, but not actually re-offend.

Opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes.



Source: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Copyright © 2019 IBM Corporation.

Towards Composable Bias Rating of AI Services

Towards Composable Bias Rating of AI Services

Biplav Srivastava and Francesca Rossi
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA 10598

Abstract

A new wave of decision-support systems are being built today using AI services that draw insights from data (like text and video) and incorporate them in human-in-the-loop assistance. However, just as we expect humans to be ethical, the same expectation needs to be met by automated systems that increasingly get delegated to act on their behalf. A very important aspect of an ethical behavior is to avoid (intended, perceived, or accidental) bias. Bias occurs when the data distribution is not representative enough of the natural phenomenon one wants to model and reason about. The possibly biased behavior of a service is hard to detect and handle if the AI service is merely being used and not developed from scratch, since the training data set is not available. In this situation, we envisage a 3rd party rating agency that is independent of the API producer or consumer and has its own set of biased and unbiased data, with customizable distributions. We propose a 2-step rating approach that generates bias ratings signifying whether the AI service is unbiased compensating, data-sensitive biased, or biased. The approach also works on composite services. We implement it in the context of text translation and report interesting results.

Introduction

The popular approach for building software applications today is by reusing any existing capability from others exposed as Application Programming Interfaces (APIs), and developing new code for the rest, as well as glue code to connect them (Vukovic *et al.* 2016). Service catalogs facilitate API discovery by enabling search by an API's functional (e.g., description) and non-functional capabilities (e.g., cost, availability). Most API catalogs, whether public, like ProgrammableWeb (Mulesoft 2017), or private by cloud vendors, list services based on metadata and cost. As adoption of such AI services increases that draw insights from data and get incorporated into the human-in-the-loop decision-making, the expectation of ethical decisions from humans gets extended to automated systems that increasingly get delegated to act on their behalf, or that recommend decisions to humans.

There are many aspects of an ethical behavior that we expect from a decision making entity. Prominent among them are alignment to common norms, transparency, fairness, diversity, and interpretability. In particular, fairness refers to the behavior that treats all elements of a certain class in the same way. A more precise term for fairness is bias. In an ethical system, it is important to avoid behaving in a way that presents intended, perceived or accidental bias.

More precisely, bias occurs with respect to an attribute (such as gender or race) when the data distribution is not representative enough of the natural phenomenon (that is, the distribution of the attribute's values) that one wants to model and reason about. For example, if we search for images of engineers in ImageNet, we will get very few womens, in percentage which is much lower than the actual percentage of women engineers in real life. If such dataset would be used to train a system that is intended to make decisions (or help humans make decisions) about engineers, the system would possibly not treat women engineers in a fair way. Bias has been shown in many existing AI systems that are currently used, for example in the algorithm used by the US judicial system to predict which criminals have a high probability of reoffending, which has been shown to be biased against African Americans (Angwin *et al.* 2016).

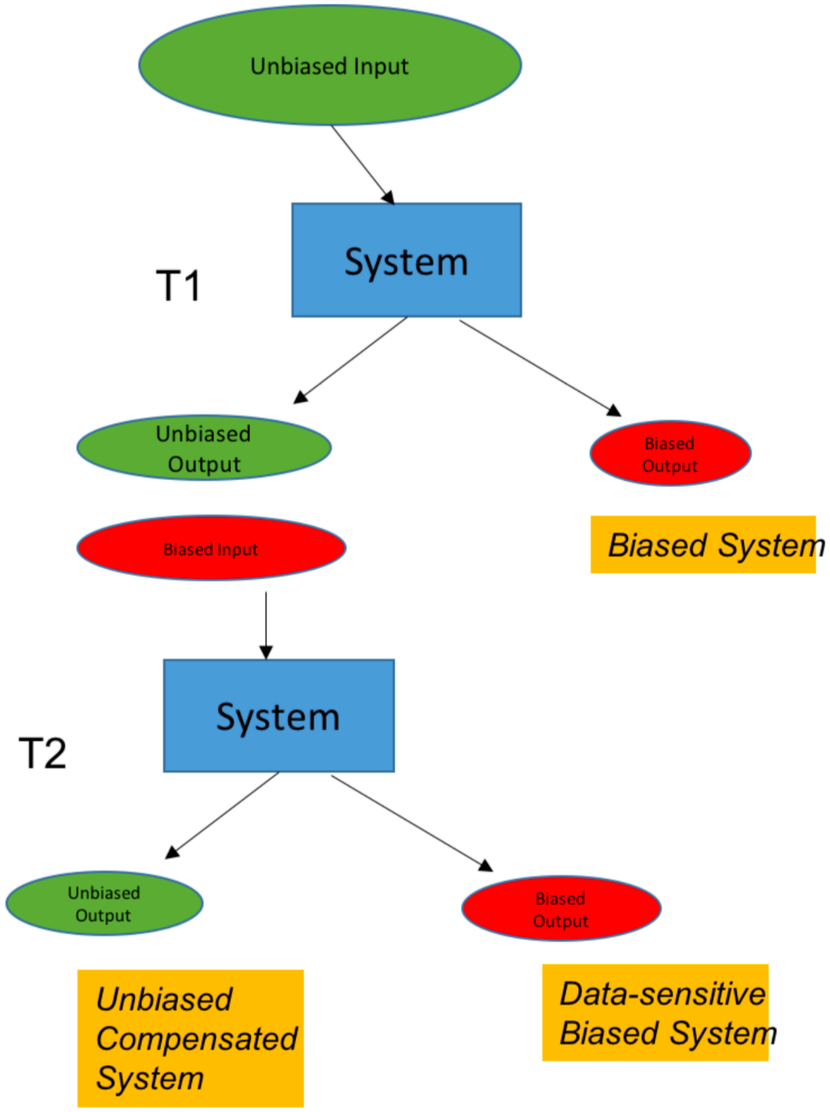
If the dataset used for training the system is available, it is easy to check if it is biased, and there are technical solutions that allow to partially remove the bias. However, if the AI service is merely being used by a consumer and not developed from scratch, so the training data set is not available, the possibly biased behavior of the service is hard to detect and handle.

In this paper we consider this scenario and show how to detect bias through a two-step test approach, and to rate the AI service according to the kind of bias that has been recognized. We consider bias as any abnormal distribution of values of an attribute from one or more baseline distributions that are considered unbiased (or normal). For example, the attribute *Gender* may have values *He*, *She* and *Other*, and attribute *Place of Worship* may have attribute values *Church, Mosque, Temple, Synagogue, Other*. We will focus on gender for the rest of the paper but the discussion applies to any attribute of interest.

Illustration and Running Example

Let us consider a simple hypothetical AI system called *UniversalSocialRepeater* (USR) that takes an English input text

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.



Source: http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_65.pdf

Copyright © 2019 IBM Corporation.

CODE

Models
Code Patterns
Open Projects

CONTENT

Announcements
Articles
Courses
Series
Tutorials
Videos

COMMUNITY

Blogs
Events

RELATED AI TOPICS

Conversation
Data science
Deep learning
Machine learning

OPEN PROJECT

AI Fairness 360

The AI Fairness 360 toolkit (AIF360) is an open source software toolkit that can help detect and remove bias in machine learning models

Get the Code

Published November 14, 2018

Analytics Artificial intelligence Machine Learning

The AI Fairness 360 toolkit (AIF360) is an open source software toolkit that can help detect and remove bias in machine learning models. It enables developers to use state-of-the-art algorithms to regularly check for unwanted biases from entering their machine learning pipeline and to mitigate any biases that are discovered.

SOCIAL

CONTENTS

Why the AI Fairness 360 toolkit?
Why should I contribute?

RESOURCES

IBM / AIF360

Watch 52 Star 465 Fork 86

Code Issues 2 Pull requests 3 Projects 0 Wiki Insights

Join GitHub today

GitHub is home to over 31 million developers working together to host and review code, manage projects, and build software together.

Sign up

Dismiss

https://github.com/IBM/AIF360

A comprehensive set of fairness metrics for datasets and machine learning models, explanations for these metrics, and algorithms to mitigate bias in datasets and models. http://aif360.mybluemix.net/

ai fairness-ai fairness fairness-testing fairness-awareness-model bias-detection bias bias-correction bias-reduction bias-finder artificial-intelligence discrimination ibm-research-ai ibm-research machine-learning deep-learning

119 commits 17 branches 3 releases 13 contributors Apache-2.0

Branch: master New pull request Find File Clone or download

michaelhind Add files via upload Latest commit 7fbe0cd on 25 Jan

aif360	Merge pull request #65 from IBM/style_fixes	2 months ago
docs	Add files via upload	2 months ago
examples	Merge pull request #62 from IBM/credit_tutor_reweigh	2 months ago

AI Fairness 360 - Demo

Data Check Mitigate Compare

aif360.mybluemix.net

Back Next

2. Check bias metrics

Dataset: German credit scoring

Mitigation: none

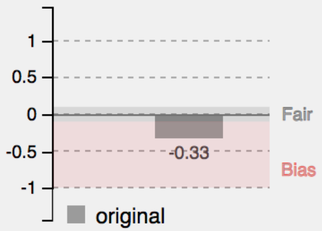
Protected Attribute: Age

Privileged Group: *Old*, Unprivileged Group: *Young*

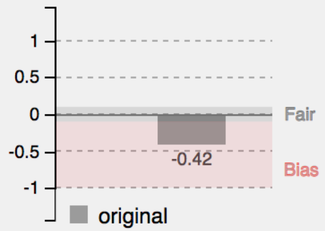
Accuracy with no mitigation applied is 76%

With default thresholds, bias against unprivileged group detected in 4 out of 5 metrics

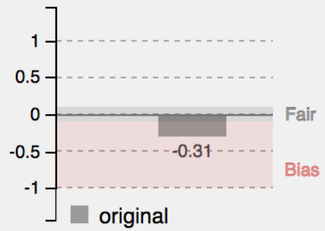
Statistical Parity Difference



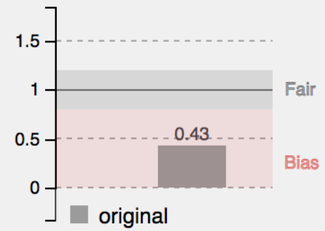
Equal Opportunity Difference



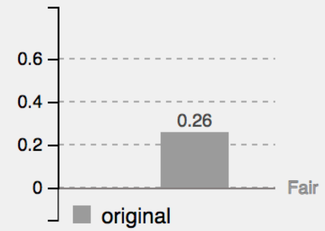
Average Odds Difference



Disparate Impact



Theil Index



Copyright © 2019 IBM Corporation.

Watson OpenScale
Supports Watson Studio & Watson Machine Learning, and 3rd Party Frameworks

(incl. Amazon SageMaker, Microsoft Azure ML Studio, and Custom Scoring Endpoints via REST APIs)

Payload logging

- Gain insights into model inference
- Logs feed a deployments data mart for monitoring and exploration

Continuous Evolution

- Intelligent re-train and data synthesis triggers for production models
- Define KPI thresholds that trigger model retraining

Operations Dashboard

- Take action on deployed models by understanding payloads and feedback data
- Ensure ongoing model health in business applications

Model Explainability

- Eliminate black box models & allow business users to understand AI outcomes in terms they understand
- Explain models with runtime explainability

Fairness Tests

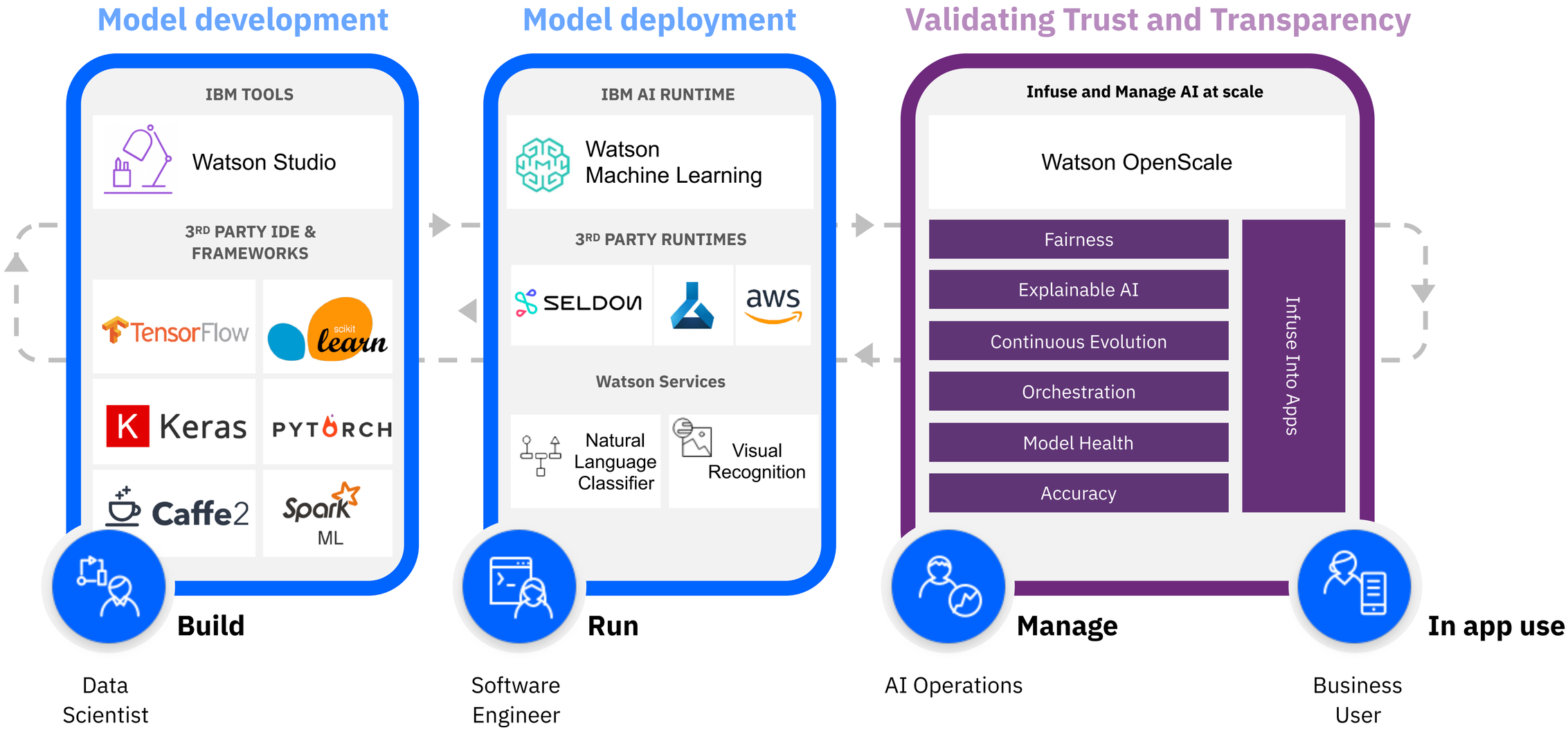
- Discover model bias through active monitoring
- Ensure models are bias free

Model Ops

- Model metrics can be integrated into common reporting tools linking AI to business and application outcomes
- AI lifecycle orchestration framework to enable AI & IT operational scale

Copyright © 2019 IBM Corporation.

Scaling AI with **Trust and Transparency** in the Enterprise



Copyright © 2019 IBM Corporation.

IBM Cloud | Search resources and offerings... | Catalog

cloud.ibm.com/catalog?category=ai

Search the catalog... | Filter

All Categories

- Compute
- Containers
- Networking
- Storage
- AI
- Analytics
- Databases
- Developer Tools
- Integration
- Internet of Things
- Security and Identity
- Starter Kits
- Web and Mobile
- Web and Application

Knowledge Catalog
Lite • IBM • IAM-enabled

Discover, catalog, and securely share enterprise data.

Knowledge Studio
Lite • IBM • IAM-enabled

Teach Watson the language of your domain.

Language Translator
Lite • IBM • IAM-enabled

Translate text, documents, and websites from one language to another. Create industry or region-specific translations via the service's customization capability.

Machine Learning
Lite • IBM • IAM-enabled

IBM Watson Machine Learning - make smarter decisions, solve tough problems, and improve user outcomes.

Natural Language Classifier
IBM • IAM-enabled

Natural Language Classifier uses advanced natural language processing and machine learning techniques to create custom classification models. Users train...

Natural Language Understanding
Lite • IBM • IAM-enabled

Analyze text to extract meta-data from content such as concepts, entities, emotion, relations, sentiment and more.

Personality Insights
Lite • IBM • IAM-enabled

The Watson Personality Insights derives insights from transactional and social media data to identify psychological traits.

Speech to Text
Lite • IBM • IAM-enabled

Low-latency, streaming transcription

Text to Speech
Lite • IBM • IAM-enabled

Synthesizes natural-sounding speech from text.

Tone Analyzer
Lite • IBM • IAM-enabled

Tone Analyzer uses linguistic analysis to detect three types of tones from communications: emotion, social, and language. This insight can then be used to drive...

Visual Recognition
Lite • IBM • IAM-enabled

Find meaning in visual content! Analyze images for scenes, objects, faces, and other content. Choose a default model off the shelf, or create your own custo...

Watson OpenScale
Lite • IBM • IAM-enabled

IBM Watson OpenScale is an enterprise-grade environment for AI infused applications that provides enterprises with visibility into how AI is being built,...

Copyright © 2019 IBM Corporation.

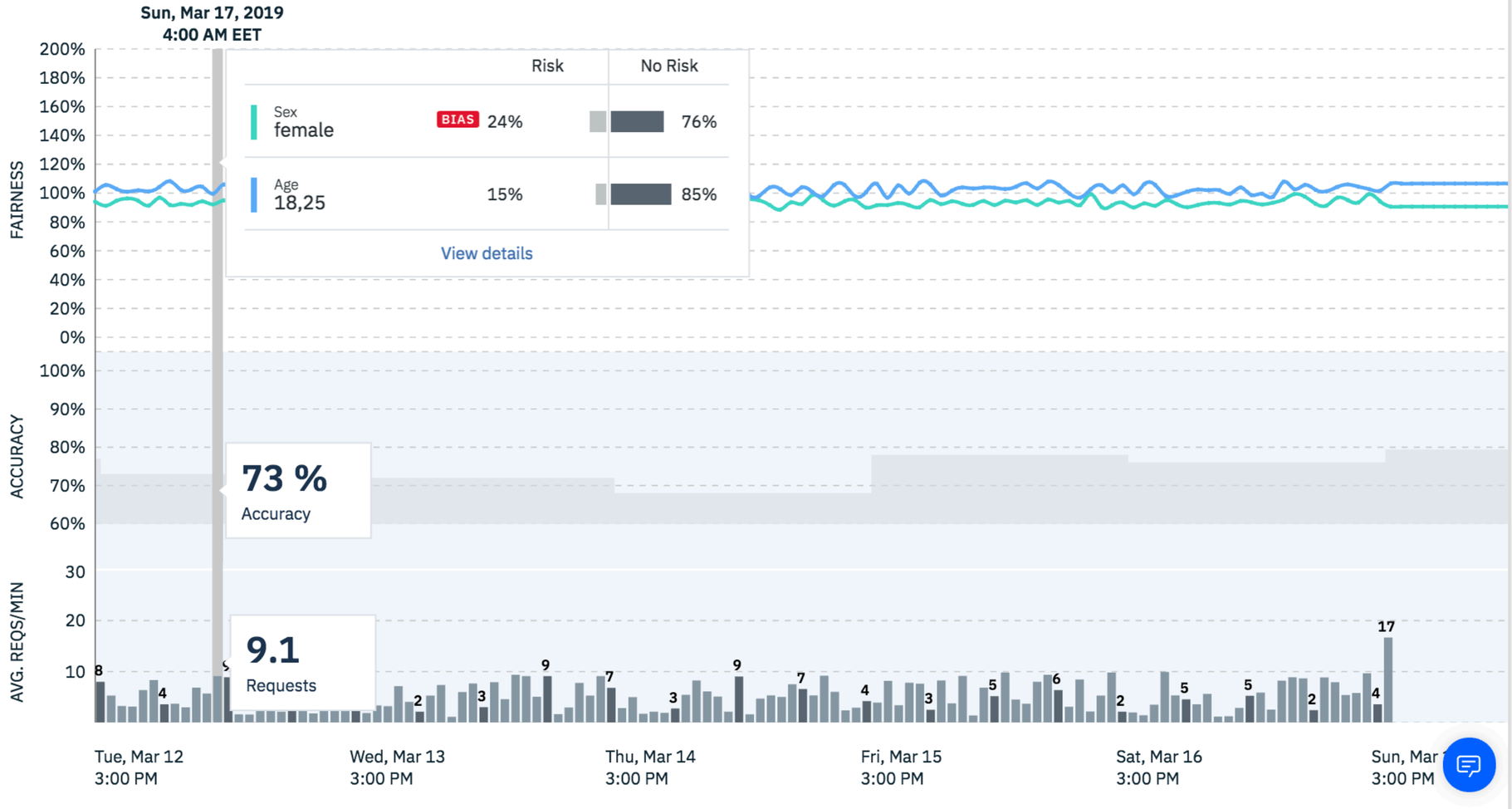
Spark German Risk Deployment - Final

Description		Fairness Evaluated	2:12 PM EET
Model ID	90cdb2fe-649c-4704-9e4d-ed9e68f86b2e	Next Fairness Evaluation	3:12 PM EET
Date Created	March 17, 2019		

Check Fairness Now

Accuracy Evaluated	1:45 PM EET
Next Accuracy Evaluation	2:45 PM EET

Check Accuracy Now



Configuration Summary

Edit

View payload logging endpoint

Fairness

Accuracy

Accuracy Threshold:
70% (Good)

Minimum Sample Size:
50

View feedback data endpoint

Add Feedback Data

Copyright © 2019 IBM Corporation.

Finally: If you couldn't pay attention, or just need to catch up again later, then I'd suggest you read this:

ibm.com/cloud/garage/files/data-analytics-field-guide.pdf

Thank you!

—

Jukka Ruponen
Senior Analytics Architect