outthink limits

Winning the Hearts & Minds of the Data Scientist in the Cognitive Era

Gaurav Rao

Director, Advanced Analytics

IBM Analytics

gaurarao@us.ibm.com

IBM

Data Is The Basis Of Competitive Advantage

97% ACCURACY

THE WALL STREET JOURNAL.

CIO JOURNAL.

Bad Election Day Forecasts Deal Blow to Data Science

Prediction models suffered from narrow data, faulty algorithms and human foibles

The New York Times

How Data Failed Us in Calling an Election

"The election prediction business is one small aspect of a far-reaching change across industries that have increasingly become obsessed with data, the value of it and the potential to mine it for cost-saving and profit-making insights" - Steve Lohr, New York Times

3

New digital technologies are disrupting the business landscape

Contextual mobility

3hrs per day

spent by every US adult per day on a mobile device on average[1]

Internet of Things

20.8bn

connected things will be in use worldwide by 2020[2]

Cloud technologies

$419bn

expected global opportunity for *cloud computing by* 2019[3]

Cognitive analytics

$12.6bn

expected global market size of cognitive computing by 2019 (38% CAGR)[4]

Digital businesses are disrupting industries and professions

"The 'Uber syndrome' – where a competitor with a completely different business model enters your industry and flattens you."
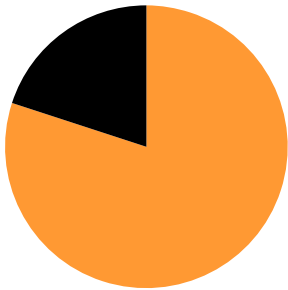
CIO, Transportation, United States

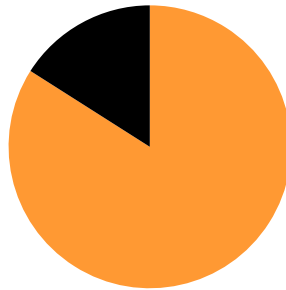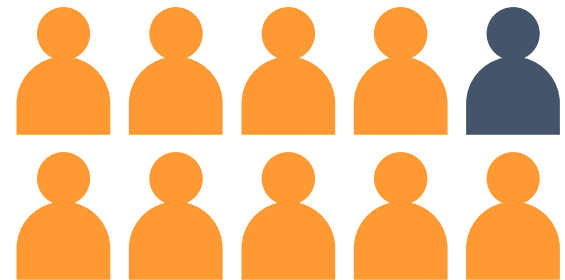**72%** are vulnerable to disruption within *three* years

What is Holding us Back?

80%
are unable to collaborate on common data

84%
say fragmented data gets in the way
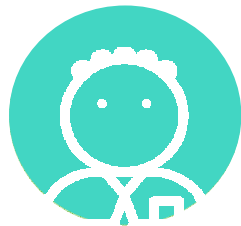
more than **9** out of **10**
require faster data and analytics to compete
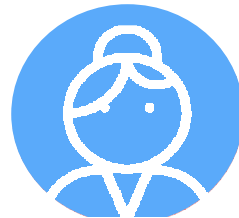
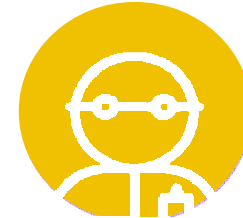# Enabling a Culture for the Data Hungry

Data engineer architects how data is organized and ensures operability

Data scientist gets deep into the data to draw hidden insights for the business

Business analyst works with data to apply insights to the business strategy

Application developer plugs into data and models and writes code to build applications

Transforming Industries & Professions: The Data Scientist

Data science is the process of discovering hidden insights from massive amounts of structured and unstructured data, using methods such as statistics, machine learning, data mining, and predictive analytics (similar to Knowledge Discovery in Databases). This multi-disciplinary area is changing the way organizations solve problems and gain a competitive advantage.

The Perfect Data Scientist
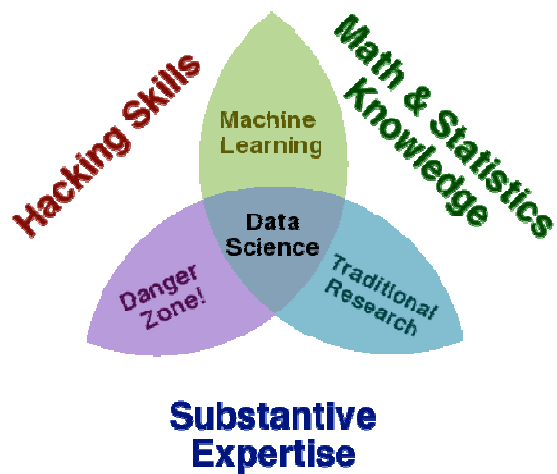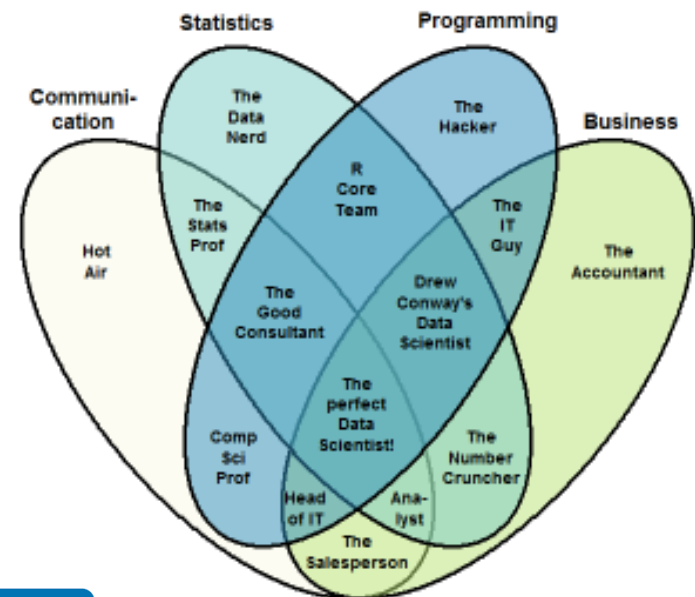
Innovation, Exploration, and Business Value

Increased operational efficiency

More efficient business processes

Better impact analyses

Cost reductions

Business outcomes:
New business models
Increased revenue opportunities
Efficient cost optimizations

Business unit 1

New ways to tackle critical business challanges

The CEO

Functional department

Collaborate

Data science team

New ideas based on discovery and exploration

Business unit 2

The lead data scientist

Source: *Mastering the art of data science*. IBM Institute for Business Value.
© 2016 IBM. ibm.biz/artofdata

What is Machine Learning?

- **Wikipedia**: subfield of computer science that "gives computers the ability to learn without being explicitly programmed" (Arthur Samuel, 1959).

"I DIDN'T LEARN TO TALK FROM A LINGUIST," SAYS GREG CORRADO. "I LEARNED TO TALK FROM HEARING OTHER PEOPLE TALK."

Context & Intelligence Matter

Features:
1. Color: **Radish/Red**
2. Type : **Fruit**
3. Shape
etc...

Features:
1. Sky Blue
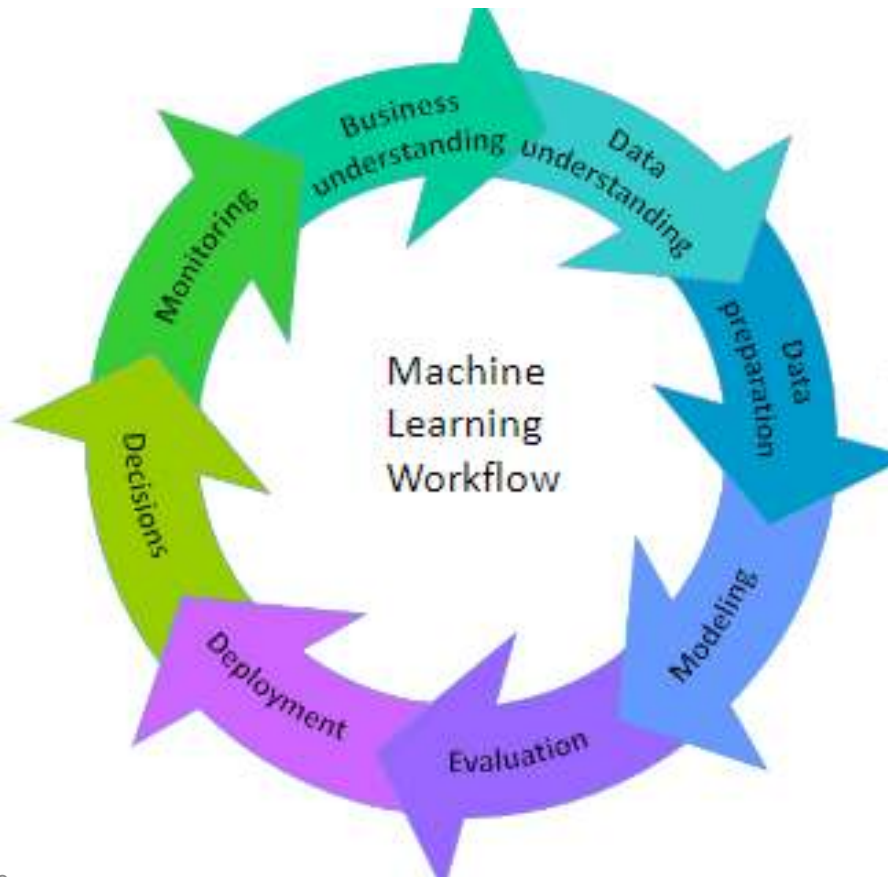2. **Logo**
3. Shape
etc...

Features:
1. **Yellow**
2. **Fruit**
3. Shape
etc...

The Machine Learning Workflow

Examples

- Classification: predict class from observations
  - E.g. Spam Email Detection
- Clustering: group observations into "meaningful" groups
  - E.g. Amazon Recommendations
- Regression (prediction): predict value from observations
  - E.g. Energy consumption prediction

Frequent Data Scientist Pain Points

- Rigid toolset
  - Have to choose one and only one approach
  - Cannot easily connect all of the capabilities needed
  - Difficult to navigate between the various tools used

- Fragmented and time consuming
  - Using multiple disjointed environments
  - Separate on-ramp/community for each tool/environment
  - Does not have meta data or data lineage

- Analytical Silo
  - Difficult to maintain and version control project assets
  - Limited means of collaborating with team
  - Results are difficult to share

Cognitive business

Create knowledge from data to outthink the needs of the marketplace by converging digital business with a new level of digital intelligence

Transformation Brings A Shift in Expectations

Collaborate across teams and roles

Open access with trust & security

"Make Data Simple"

Innovate with Open Source and the community

Drive more intelligence faster than ever before

outthink limits

IBM.
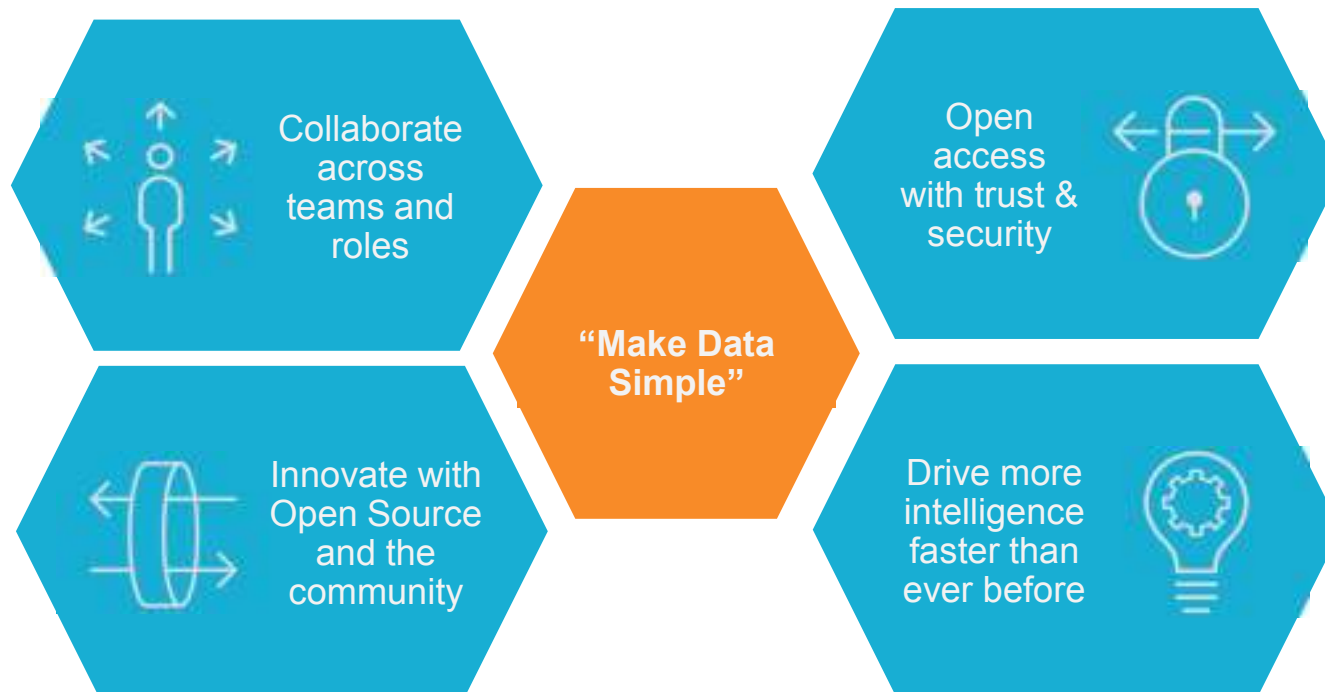
Shift to Collaboration

Feeding the Data Hungry Professionals

Data Engineering

Data Science

Business Analysis

App Development

outthink limits

IBM.
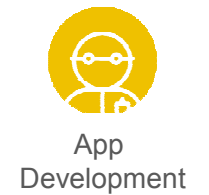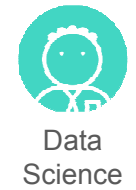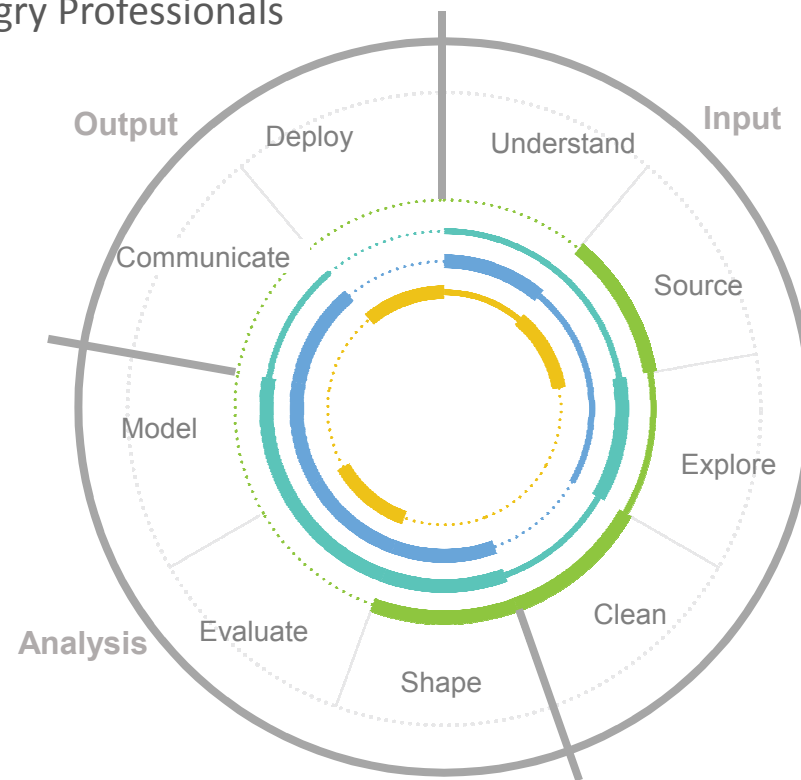
Shift to Open Access with Trust

Serve data-driven functions thru a "fit-for-purpose" sourcing approach

Find   **Share**   **Collaborate**

Data You Own   **Data In Motion**   **Data To Help You Compete**

IBM.
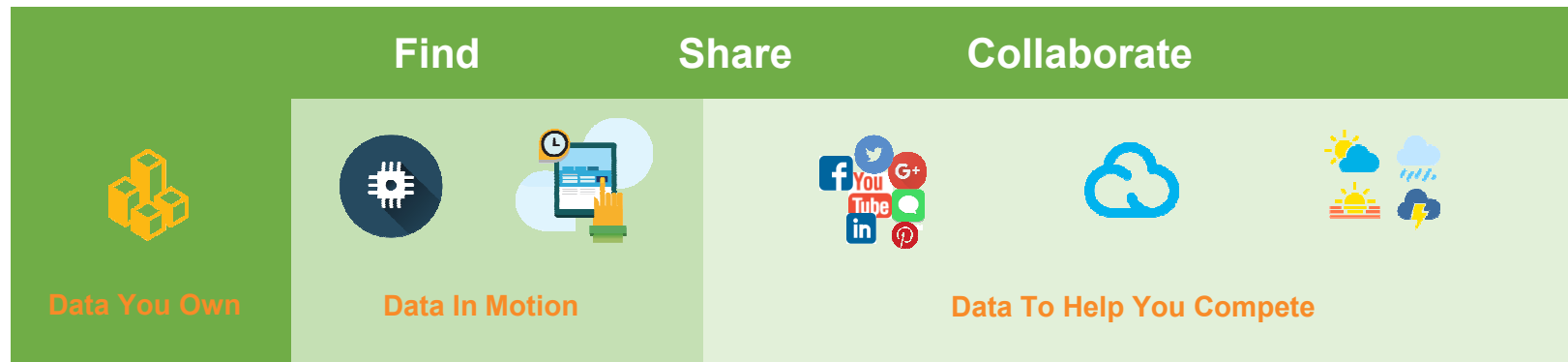
Shift to Open Standards and Open Source

Ecosystems for sustainable *and* continuous technological innovation

cassandra

PostgreSQL

elasticsearch.

etcd

python

Spark

Zeppelin

jupyter

CouchDB
relax

mongoDB

TinkerPop

hadoop

Lucene

redis

TITAN

kafka

RabbitMQ

RethinkDB

R

ODPi
odpi.org

R consortium

THE APACHE
SOFTWARE FOUNDATION

Shift to Accelerated Intelligence

Continuously infuse insight to accelerate value across the organization
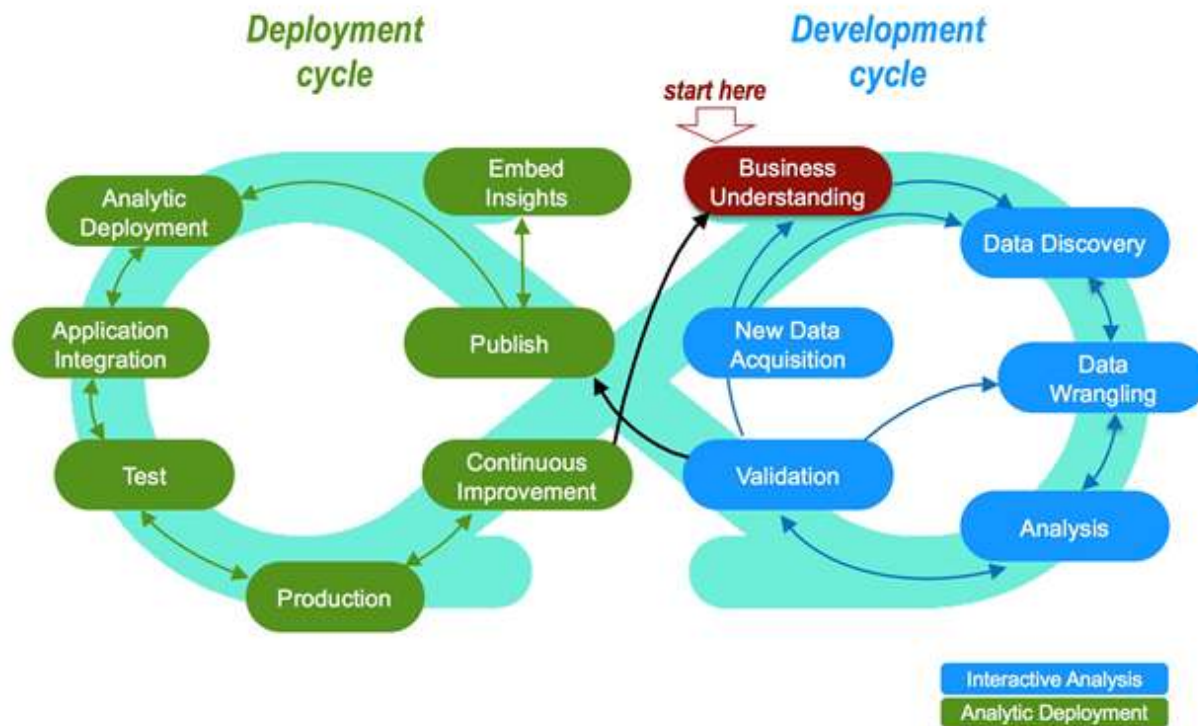
Architecting Around the Shifts

Frame your capabilities thru experiences and services

Data Engineering

Data Science

Business Analysis

App Development

Experiences
use cases, collaborative

Data and Analytics Services
comprehensive

open • intelligent • hybrid

outthink limits

IBM.

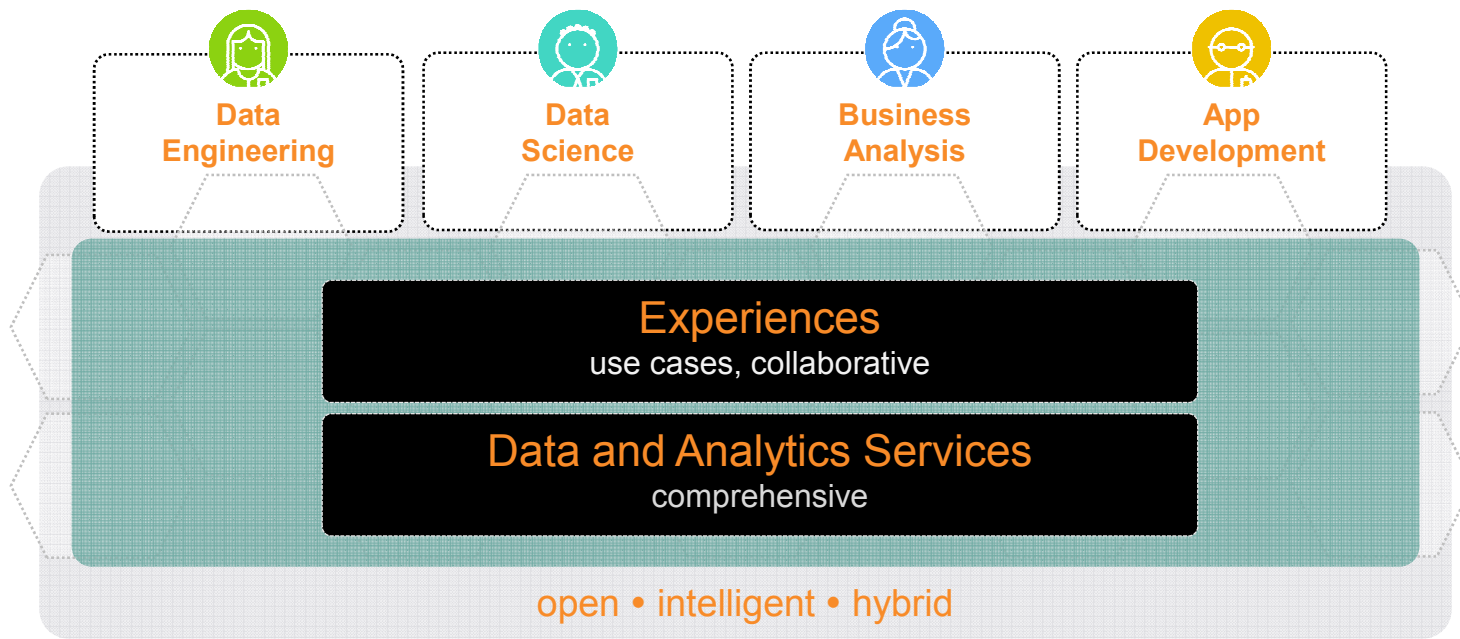# Digital Innovators Take Data Seriously

Promote a Data-Driven Culture

A **nimble**, business-driven data and analytics strategy

Push boundaries of ecosystem with technologies and software

Empower all to have **data access**

Use **data science** to disrupt competitors and delight customers

Guide **architectural** considerations for trusted data

A **"fail fast"** agile culture

Enterprise level, business-driven data and information **governance**

Optimize systems by creating algorithmic and machine-managed process

Create an **integrated, agile data foundation**

Expand the ecosystem to include non-native data yet maintain agility

Efficiency

Modernization

Democratization

Monetization

outthink limits

IBM.

outthink limits

Thank You!

Q&A

IBM