Data and AI
be ready for AI with trusted data
—
Arkadiusz Wiśniewski
UG&I Leader
IBM Analytics
Central & Eastern Europe
What organizations really want

1. Grow revenue
2. Reduce cost
3. Mitigate risk
4. Innovate
What organizations really want

- **REVENUE INCREASE**
  - POS VISITS
  - PURCHASE FREQUENCY
  - AVERAGE VALUE OF BASKET
  - (NET) CUSTOMER ACQUISITION
  - CHURN
  - SHARE OF WALLET

- **MARGIN INCREASE**
  - PROMO CONVERSION RATE
  - CAMPAIGN CONVERSION RATE

- **COST OPTIMIZATION**
  - DEMAND FORECASTING
  - PLANNING & BUDGETING
AI is a Science
computers performing
cognitive functions associated
with human minds.

Machine learning, deep
learning, neural networks,
predictive analytics,
prescriptive analytics, etc. are
all technologies that enable AI.
There is no AI without an IA (information architecture)

80% of data is either inaccessible, untrusted, or unanalyzed.

81% do not understand the data required for AI.

"No amount of AI algorithmic sophistication will overcome a lack of data [architecture] ... bad data is simply paralyzing."
4 critical steps to operationalizing AI

Business-Ready data

1. What AI teams Want to Focus on
2. Where 40-80% of their time spent on
3. Where organizations see business results
4 critical steps to operationalizing AI

Business-Ready data

1. What AI teams Want to Focus on

Organize

Build

Deploy

Manage

2. Where 40-80% of their time spent on

Discover
Integrate
Assess Quality
Profile
Curate
Catalog
Govern

3. Where organizations see business results
4 critical steps to operationalizing AI

Business-Ready data

1. What AI teams Want to Focus on

2. Where 40-80% of their time spent on

3. Where organizations see business results
The AI Ladder
A prescriptive approach to the journey to AI

- **COLLECT** - Make data simple and accessible
- **ORGANIZE** - Create a business-ready analytics foundation
- **ANALYZE** - Build and scale AI with trust & explainability
- **INFUSE** - Operationalize AI throughout the business

Unlock the value of data for an AI in multicloud world

One Platform, Any Cloud
The big from the small. The simple from complex. The structured from ...
The IBM analytics portfolio

**Collect**
 Hybrid Data Management

Write once, access everywhere
with a common access layer to foster application independency

**Organize**
 Unified Governance & Integration

Prepare, publish, integrate and protect
data to obtain better insights and to avoid regulatory incompliance

**Analyze**
 Data Science & Visualization

Descriptive, predictive, prescriptive
To understand the now and to predict the future in order to influence events

Automation through Machine Learning

Hybrid Cloud Foundation
Data must be complete, applicable and accessible everywhere.

Discover, classify and understand all types of data.

Data must be secure, clean and easy to find to encourage trusted self-service access.

Understand where data came from and its quality.

Ability to drive self-service discovery and automate decision making to evolve the business.

Provide a view of all information to those that need it and allow them to access it.
A data catalog ensures capabilities that enable any user, from business as well as from IT, to easily discover, understand, associate and consume data that is trusted and meaningful.

A data catalog additionally supports crowdsourcing that allows every user collaborate and contribute their knowledge, as well as benefit from suggestions and completeness.
IBM Unified Governance & Integration Strategy

One Catalog for IT Users & Knowledge Works - Powered by Machine Learning

Data Sources
- Systems of Record
- IOT
- Systems of Insights
- Cloud
- Hadoop
- Social Media
- Unstructured
- Other
- External
- Logs

Data Integration & Replication
- Data Integration (DataStage) *
- Data Stage + Real-Time Replication **
- Multi-Cloud Data Integration **
- Data Replication *

Automated Data Curation
- Auto Discover & Classify Data *
- Auto Detect Sensitive Data *
- Auto Analyze Data Quality *
- Auto Assign Business Terms *
- Unstructured Curation (StoredIQ) *

Core Governance Services
- Data Lineage *
- Business Glossary Management *
- Policy Management & enforcement *
- Entity & Consent Management *
- Model Governance & Bias Reporting **
- Data Quality & Exception Mgt *
- Data Archival & Disposal
- Reference Data Management **

Metadata Management & Data Catalog

Open Metadata Integration (Egeria)

Self-Services Interaction *
- Search & Find Relevant Data
- Tagging, Comments Annotation & Share
- Basic Data Preparation
- Recommendations & Data Insights
- Advanced Data Preparation **

Data Integration & Replication

Industry Models & Accelerators

Unstructured Curation (StoredIQ) *

Advanced Data Preparation **

Deployment: OnPrem, Public Clouds (Azure, AWS, IBM, Google) and IBM ICP for Data

* Evolving capability
** Future roadmap

Users
- CDO
- Governance Officers
- Data Quality Analyst
- Data Steward
- Data Scientist
- Business Users
- Data Engineer

IBM Cloud / © 2019 IBM Corporation
IBM Unified Governance & Integration Strategy

One Catalog for IT Users & Knowledge Works - Powered by Machine Learning

Data Sources
- Systems of Record
- IOT
- Systems of Insights
- Cloud
- Hadoop
- Social Media
- Unstructured
- Other
- External
- Logs

Data Integration & Replication
- Data Integration (DataStage) *
- Data Stage + Real-Time Replication **
- Multi-Cloud Data Integration **
- Data Replication *

Automated Data Curation
- Auto Discover & Classify Data *
- Auto Detect Sensitive Data *
- Auto Analyze Data Quality *
- Auto Assign Business Terms *
- Unstructured Curation (StoredIQ) *

Metadata Management & Data Catalog

Open Metadata Integration (Egeria)

Knowledge Catalog

Self-Services Interaction
- Search & Find Relevant Data
- Tagging, Comments Annotation & Share
- Basic Data Preparation
- Recommendations & Data Insights
- Advanced Data Preparation **

Core Governance Services
- Data Lineage *
- Business Glossary Management *
- Policy Management & enforcement *
- Entity & Consent Management *
- Data Quality & Exception Mgt *
- Data Archival & Disposal
- Reference Data Management **

Industry Models & Accelerators
- Model Governance & Bias Reporting **
- Open Metadata Integration (Egeria)
- ML Infused Everywhere
- Embedded Workflow

Deployment: OnPrem, Public Clouds (Azure, AWS, IBM, Google) and IBM ICP for Data

Users
- CDO
- Governance Officers
- Data Quality Analyst
- Data Steward
- Data Scientist
- Business Users
- Data Engineer

* Evolving capability
** Future roadmap
IBM Unified Governance & Integration Strategy

**One Catalog for IT Users & Knowledge Works - Powered by Machine Learning**

- Automated Data Curation
  - Auto Discover & Classify Data *
  - Auto Detect Sensitive Data *
  - Auto Analyze Data Quality *
  - Auto Assign Business Terms *
  - Unstructured Curation (StoredIQ) *

- Metadata Management & Data Catalog
  - Open Metadata Integration (Egeria)

- Data Integration & Replication
  - Data Integration (DataStage) *
  - Data Stage + Real-Time Replication **
  - Multi-Cloud Data Integration **
  - Data Replication *

- Core Governance Services
  - Data Lineage *
  - Business Glossary Management *
  - Policy Management & enforcement *
  - Entity & Consent Management *

- Data Quality & Exception Mgt *
- Data Archival & Disposal
- Reference Data Management **

- Self-Services Interaction *
  - Search & Find Relevant Data
  - Tagging, Comments, Annotation & Share

- Advanced Data Preparation **
- Recommendations & Data Insights

- Basic Data Preparation

- Deployment: OnPrem, Public Clouds (Azure, AWS, IBM, Google) and IBM ICP for Data

- Users
  - CDO
  - Governance Officers
  - Data Quality Analyst
  - Data Steward
  - Data Scientist
  - Business Users
  - Data Engineer

- Technology
  - Cloud
  - Hadoop
  - Social Media
  - IOT
  - Systems of Insights
  - Systems of Record
  - Logs
  - Other External
  - Unstructured

- IBM Cloud / © 2019 IBM Corporation
IBM Unified Governance & Integration Strategy

One Catalog for IT Users & Knowledge Works - Powered by Machine Learning

Data Sources
- Systems of Record
- IOT
- Systems of Insights
- Cloud
- Hadoop
- Social Media
- Unstructured
- Other
- External
- Logs

Data Integration & Replication
- Data Integration (DataStage) *
- Data Stage + Real-Time Replication **
- Multi-Cloud Data Integration **
- Data Replication *

Automated Data Curation
- Auto Discover & Classify Data *
- Auto Detect Sensitive Data *
- Auto Analyze Data Quality *
- Auto Assign Business Terms *
- Unstructured Curation ( StoredIQ ) *

Metadata Management & Data Catalog

Open Metadata Integration (Egeria)

Self-Services Interaction *
- Search & Find Relevant Data
- Tagging, Comments Annotation & Share
- Basic Data Preparation
- Recommendations & Data Insights
- Advanced Data Preparation **

Core Governance Services
- Data Lineage *
- Business Glossary Management *
- Policy Management & enforcement *
- Entity & Consent Management *
- Model Governance & Bias Reporting **
- Data Quality & Exception Mgt *
- Data Archival & Disposal
- Reference Data Management **

Industry Models & Accelerators

Deployment: OnPrem, Public Clouds ( Azure, AWS, IBM, Google ) and IBM ICP for Data

Users
- CDO
- Governance Officers
- Data Quality Analyst
- Data Steward
- Data Scientist
- Business Users
- Data Engineer

* Evolving capability
** Future roadmap
IBM's Data Catalog Product Direction 2019

One Scalable Catalog
Available Multi-Cloud

Embeddable Catalog
Micro-Services

Open Metadata

Persona Specific Experiences
To be successful with AI, you need the right approach

IBM AI enabling technologies
Thank you

More Information

https://www.ibm.com/analytics/cloud-pak-for-data
https://www.ibm.com/analytics/