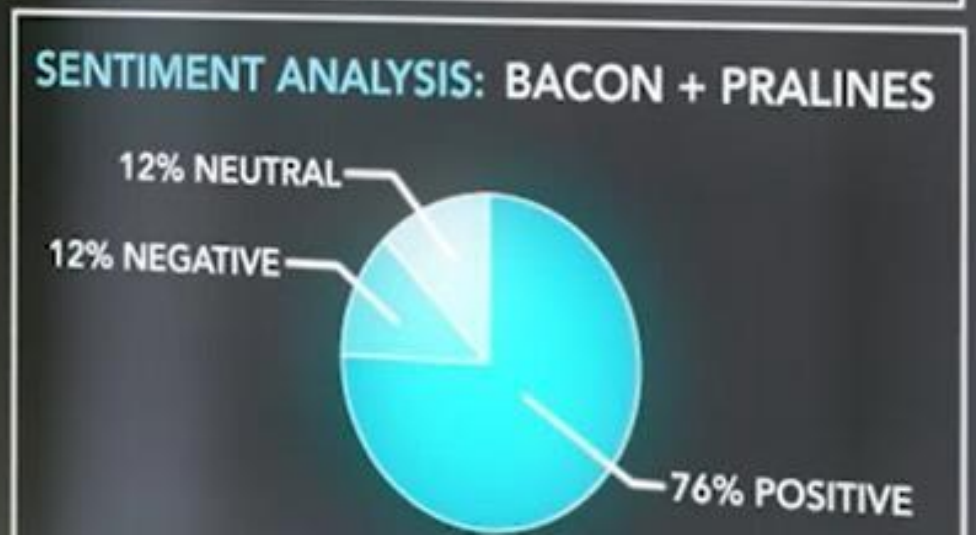


Infosys®

POWERED BY INTELLECT
DRIVEN BY VALUES

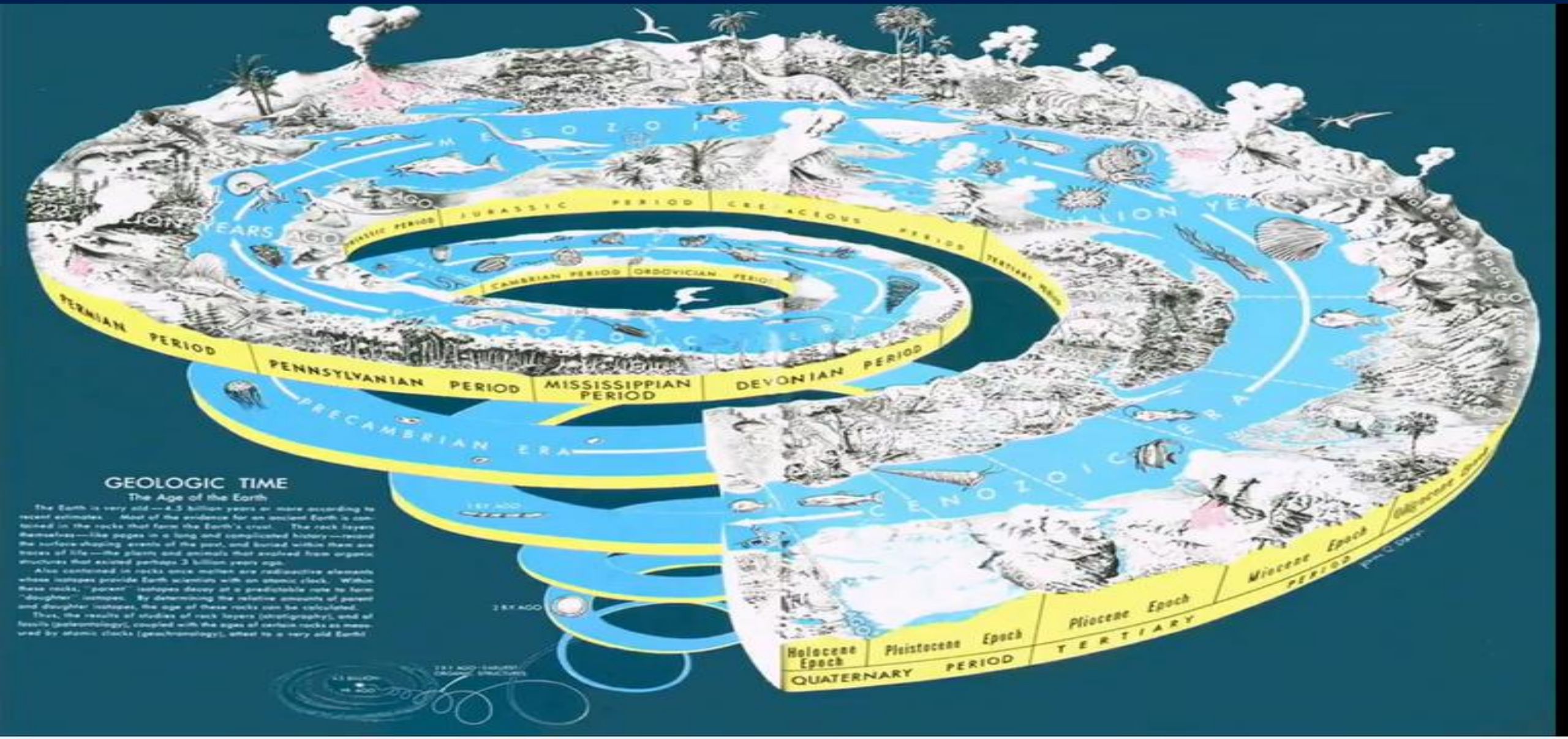
Morning *Infy* Teams!

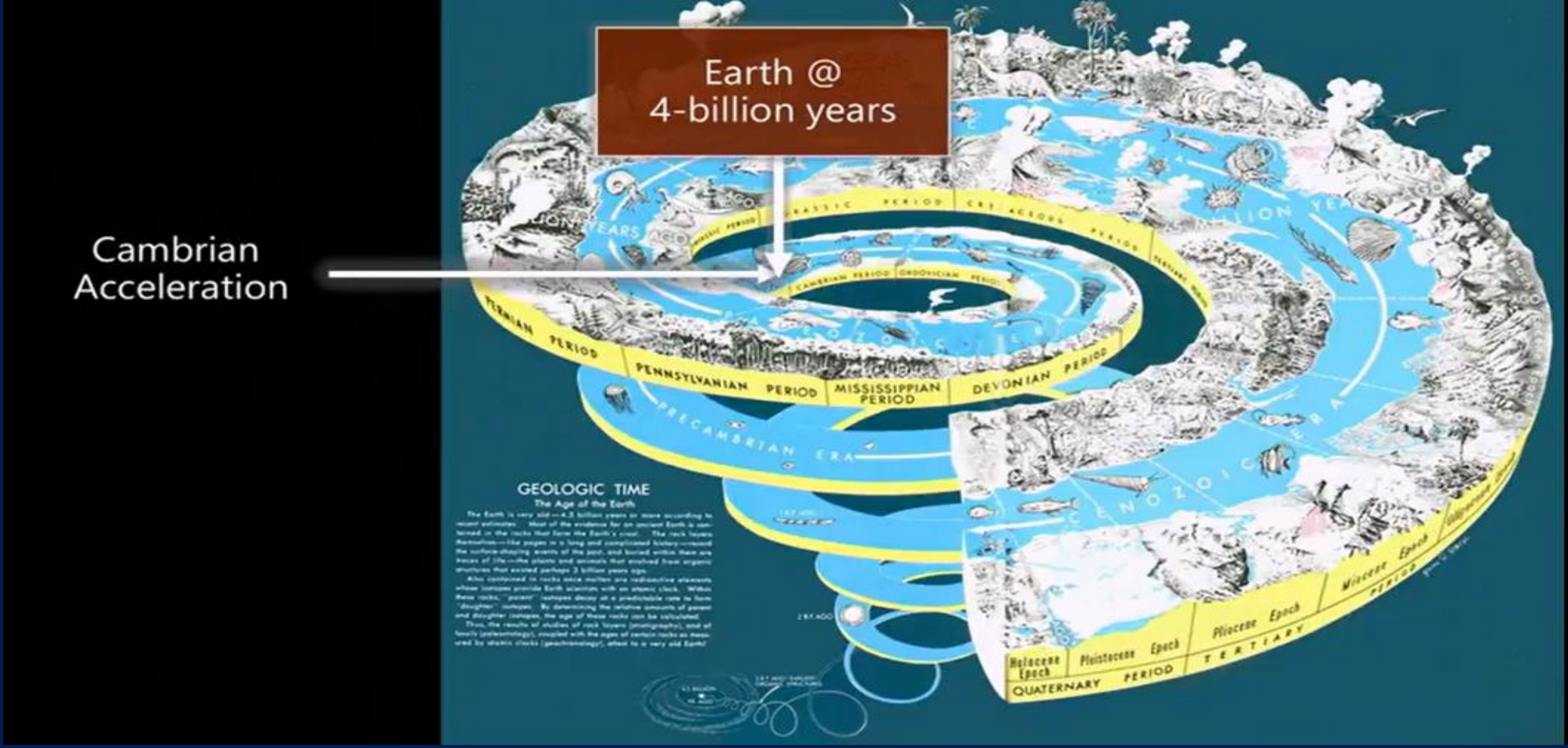
Mrinal Chakraborty
Data & AI Team
Domain Sol. Architect
PEAT
Microsoft India
mrchakra@microsoft.com

- ✓ **Programing:** SAS, CNTK, TensorFlow R-Server and Scala
- ✓ **Big-Data:** Cloudera Hadoop certification and Spark Ecosystem
- ✓ **Machine learning**: Logistic Regression, Neural Networks, Support vector machines, XGBoost, Classification and Association rules
- ✓ **Allied Analytics skills:** Visualisation, Marketing & Web analytics
- ✓ **Certifications:** PMP, Design Thinking, Certified Scrum Master & Certified in Business analytics from Indian School of Business http://www.isb.edu/cba/

Microsoft

Introductions – The AI & Cognitive Journey

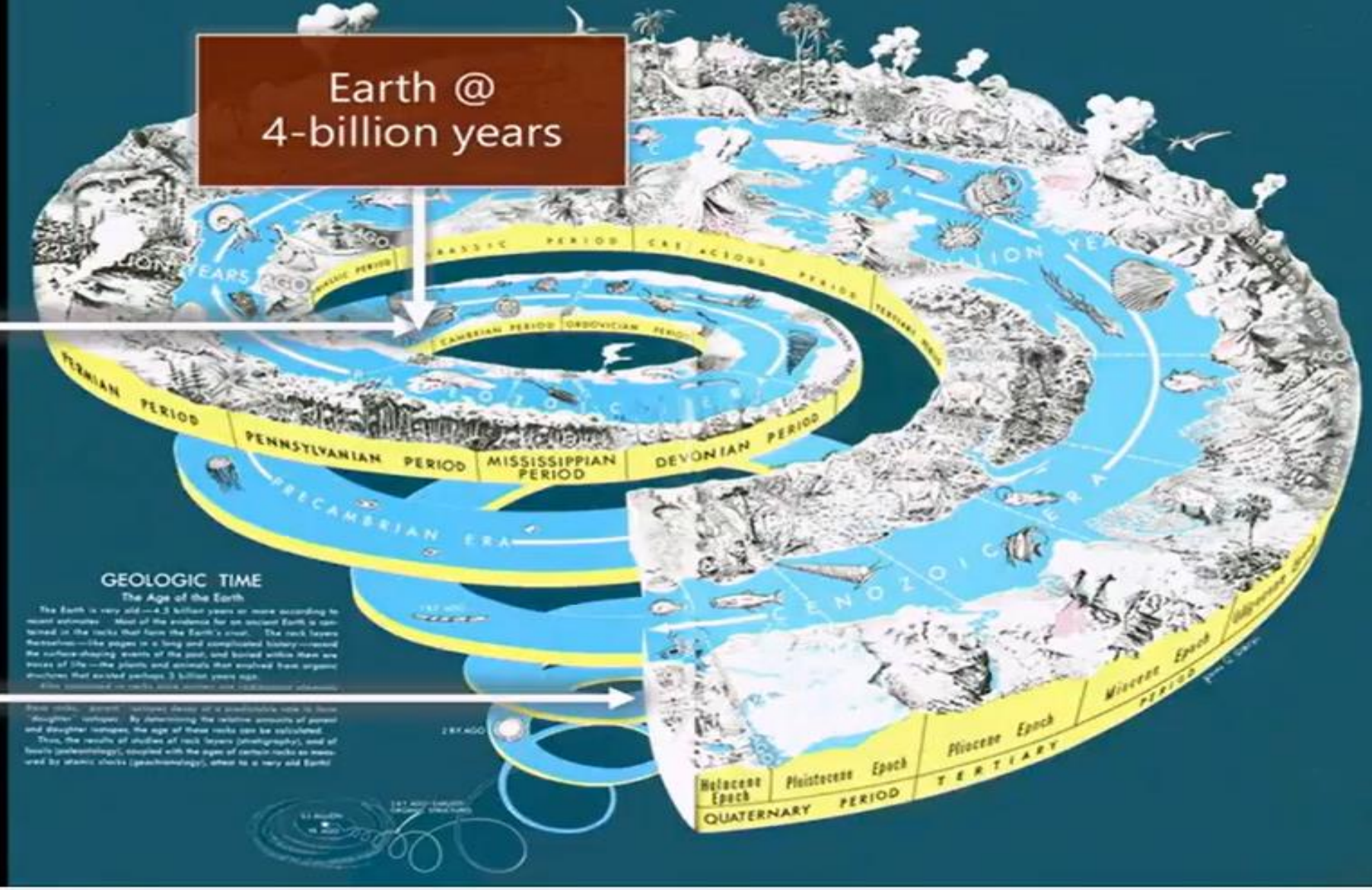




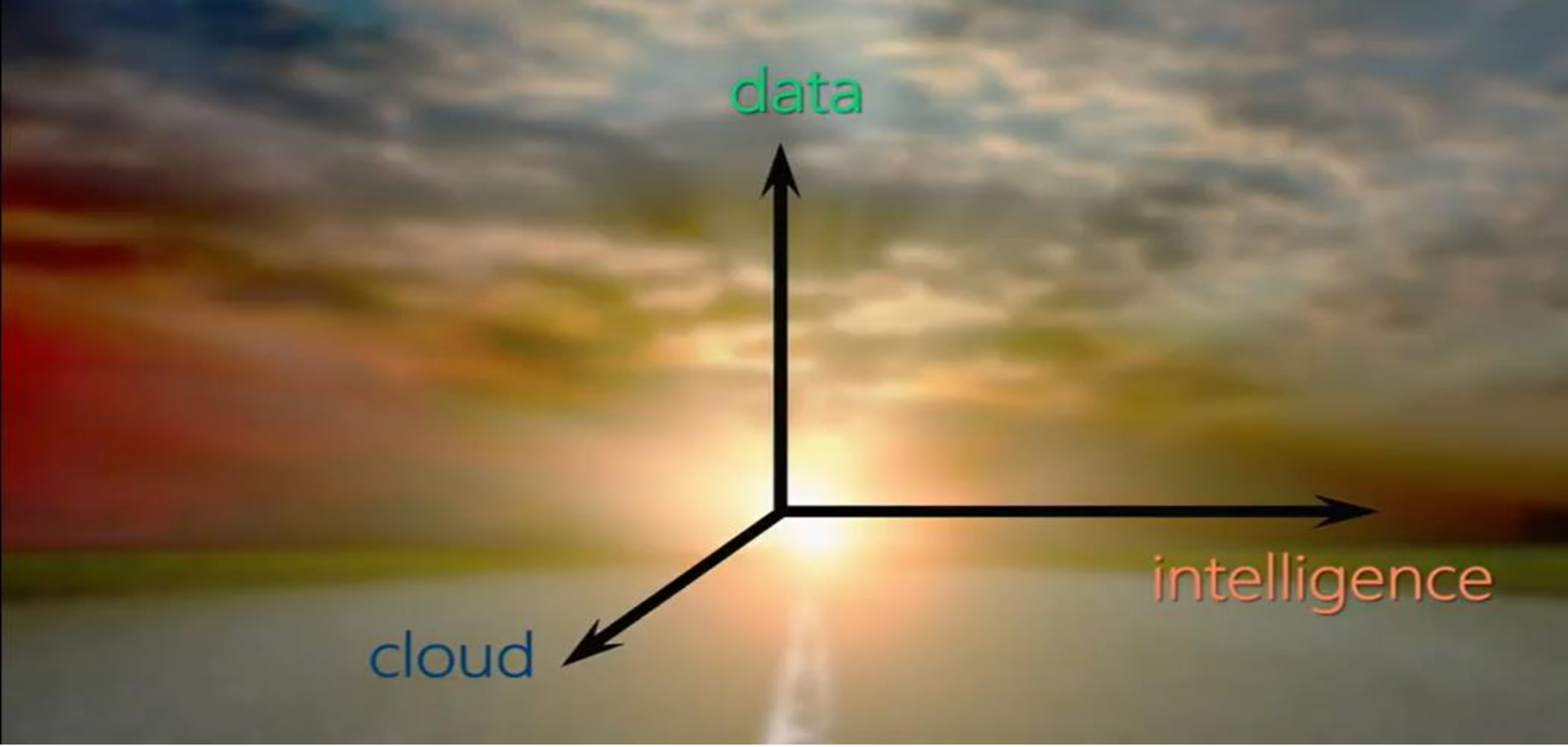
Cambrian Acceleration – Emergence of Vision

Cambrian Acceleration

AI Acceleration



AI Acceleration is destined to change our lives!



Not just software, our Economies are getting Powerful

Why cloud computing ?

https://azure.microsoft.com/en-us/suites/cortana-intelligence/

Key Customer Learnings

Systems of Intelligence are hard to build and operate

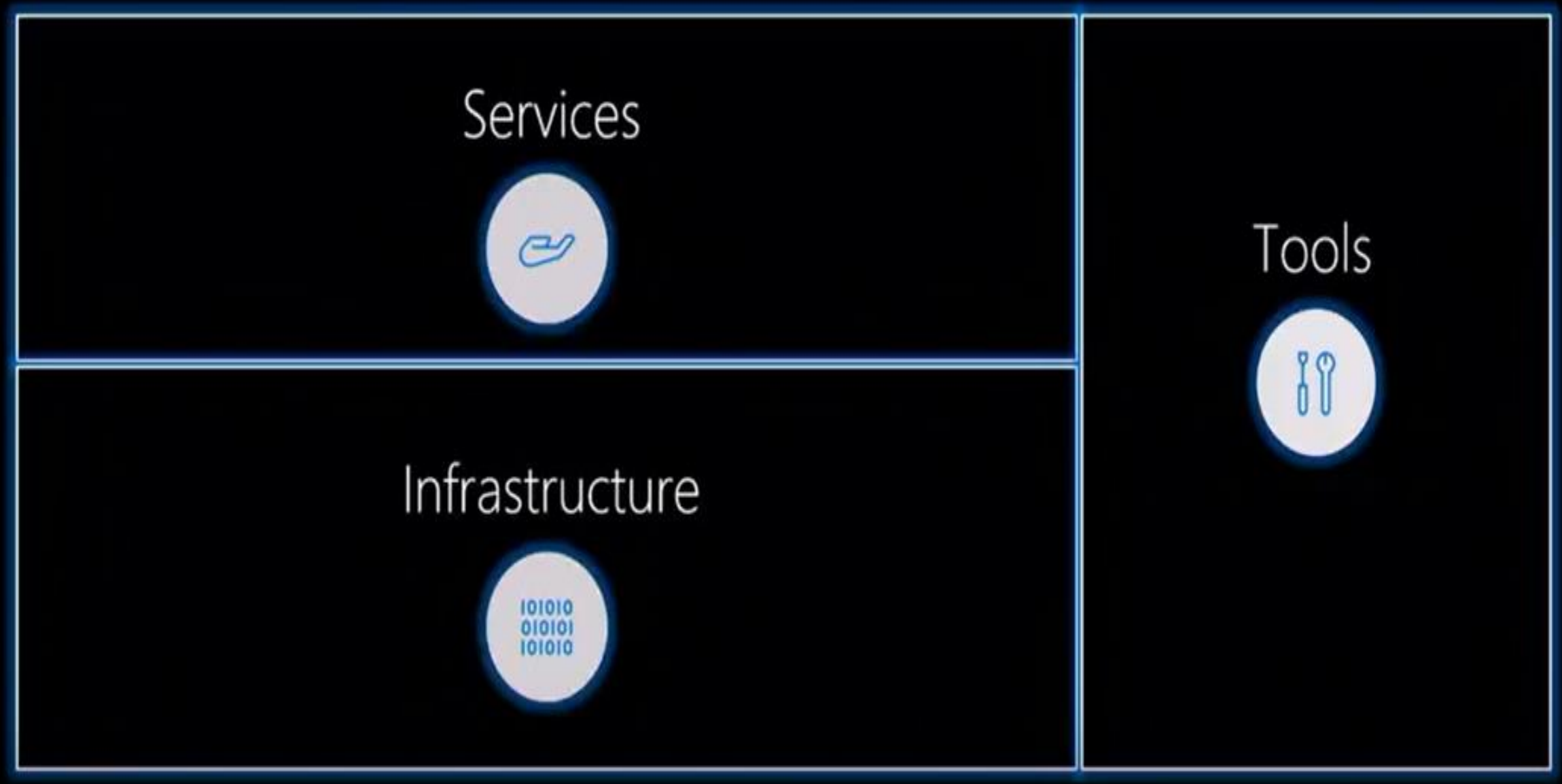
- E2E integration of infrastructure, data, AI, actions, feedback

Customers want:

- A **strategic partner** for their Systems of Intelligence.
- Well-engineered platforms, not one-off consultant built solutions.

The Microsoft AI platform: Azure+AI

Cloud-powered AI for every developer



The Microsoft AI platform: Azure+AI

Cloud-powered AI for every developer

Services

CONVERSATIONAL AI

Bot Framework

TRAINED SERVICES

Cognitive Services

CUSTOM SERVICES

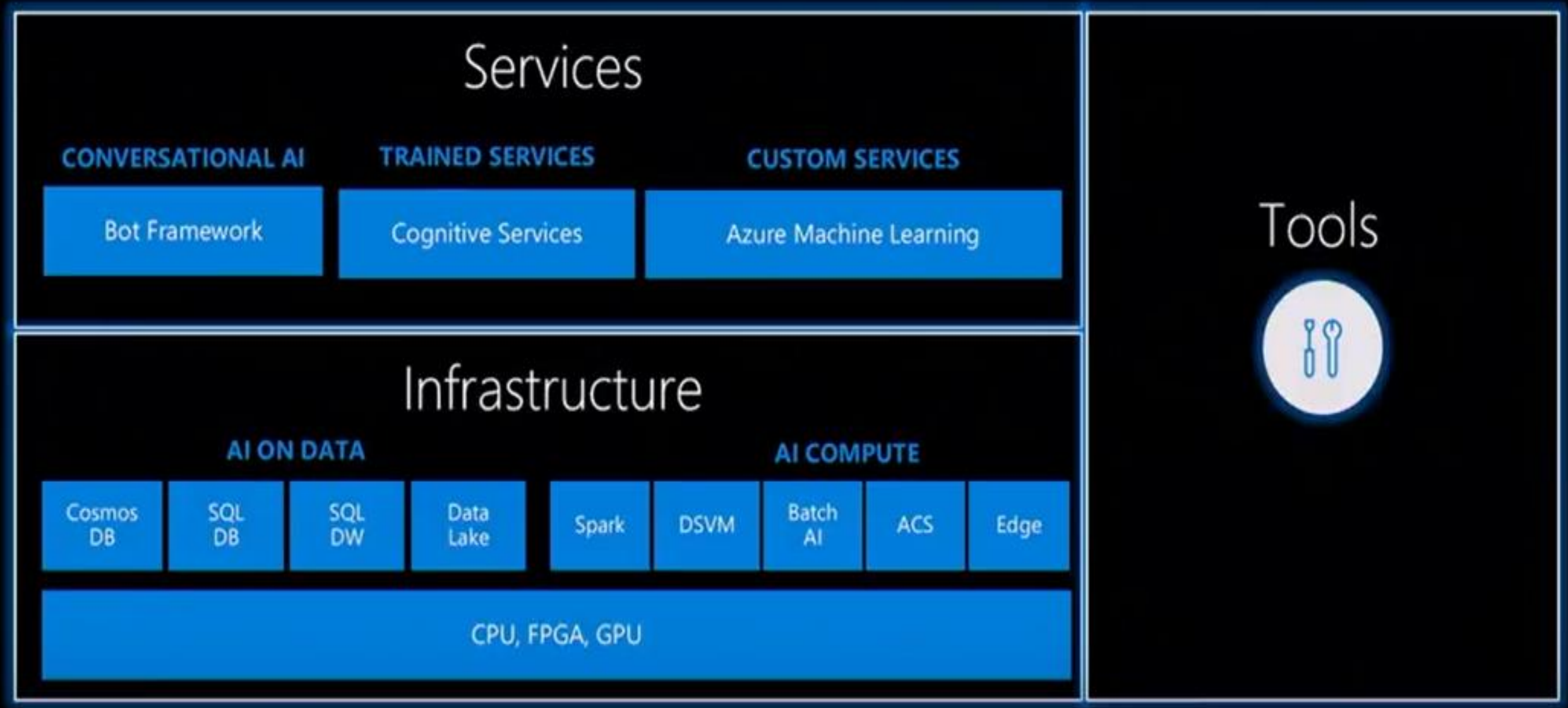
Azure Machine Learning

Tools

Infrastructure

The Microsoft AI platform: Azure+AI

Cloud-powered AI for every developer



The Microsoft AI platform: Azure+AI

Cloud-powered AI for every developer

Services

CONVERSATIONAL AI

- Bot Framework

TRAINED SERVICES

- Cognitive Services

CUSTOM SERVICES

- Azure Machine Learning

Infrastructure

AI ON DATA

- Cosmos DB
- SQL DB
- SQL DW
- Data Lake

AI COMPUTE

- Spark
- DSVM
- Batch AI
- ACS
- Edge

CPU, FPGA, GPU

Tools

CODING & MANAGEMENT TOOLS

- VS Tools for AI
- Azure ML Studio
- Azure ML Workbench

Others (PyCharm, Jupyter Notebooks...)

DEEP LEARNING FRAMEWORKS

3rd Party

- Cognitive Toolkit
- TensorFlow
- Caffe

Others (Scikit-learn, MXNet, Keras, Chainer, Gluon...)

Azure Machine Learning

AI-powered Data Wrangling
Program Synthesis

+

E2E ML Dev Productivity
SPARK, GPU, Open Source
Lifecycle Management

+

Deploy Anywhere
Docker, Spark, IOT Edge,
On prem, AWS/GCP...

=

E2E Tooling for AI Development

For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights

Forbes

Tech / #BigData

Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says

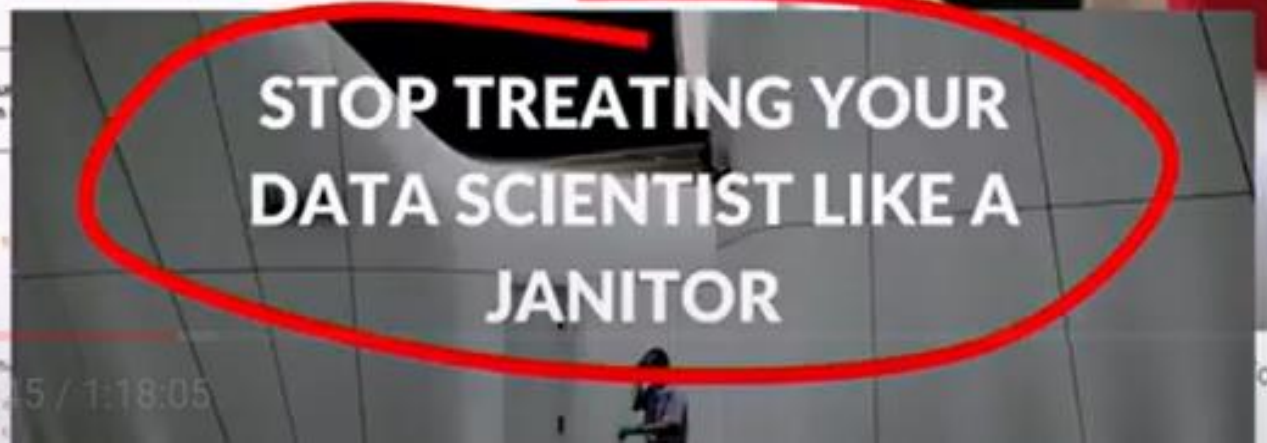
Gil Press,

InfoWorld

Hottest job? Data scientists say they're still mostly digital 'janitors'

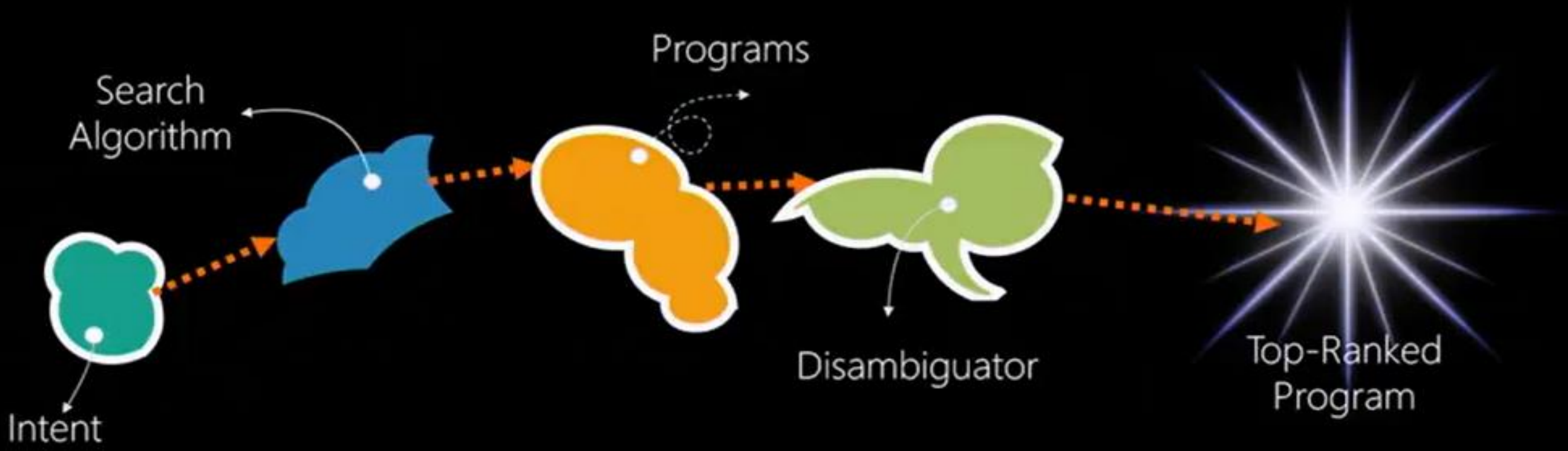
analysis

By Katherine Noyes

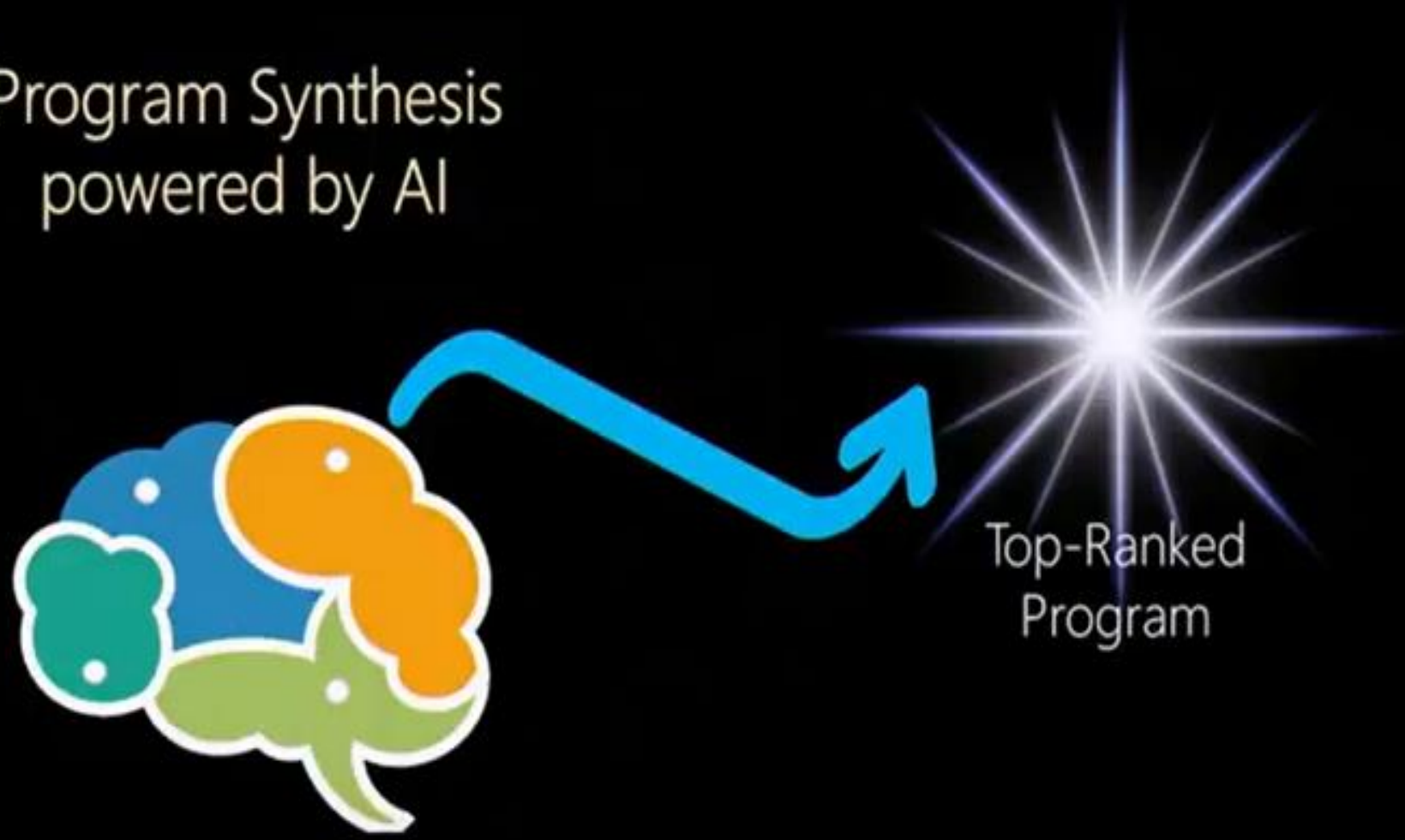


or data science, with Brian Wilt, a senior data scientist.

Less Cleaning, More Exploring with Program Synthesis



Less Cleaning, More Exploring with Program Synthesis



Azure Machine Learning Workbench

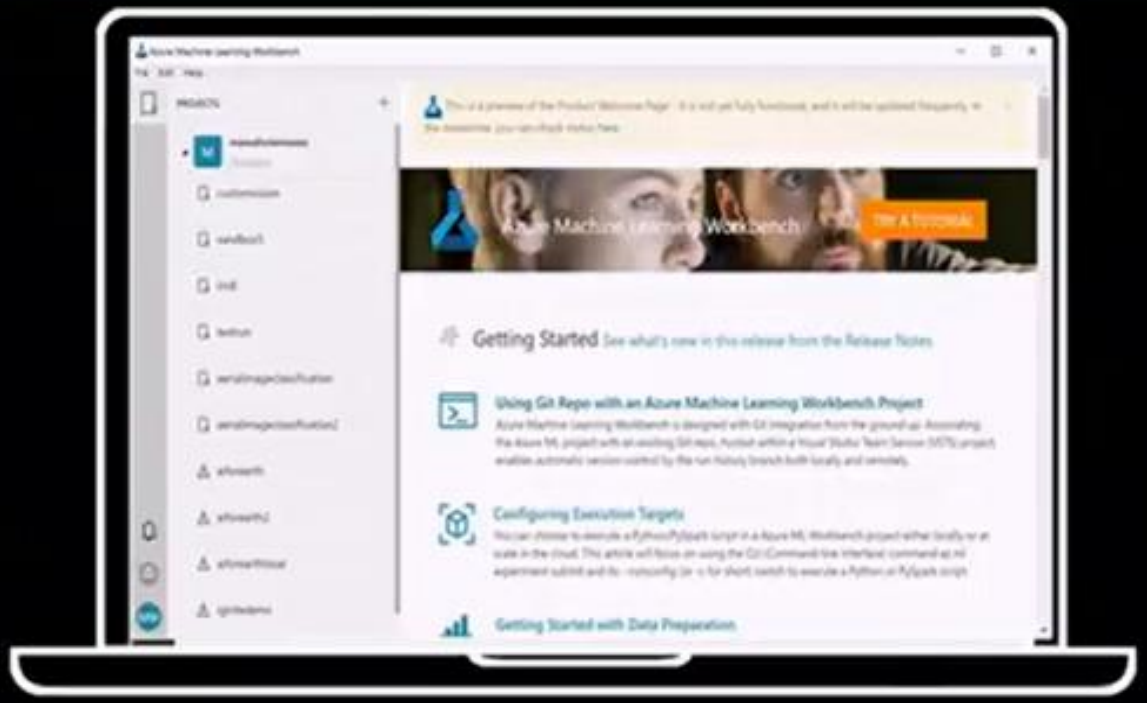
Boost ML Development Productivity

Built-in AI-powered Data Wrangling

Collaboration with notebooks & Git

Version control & reproducibility

Metrics, lineage, run history, asset management, and more



Check it out yourself:

https://blogs.msdn.microsoft.com/uk_faculty_connection/2017/09/29/azure-machine-learning-workbench/

DNNs, Big Data, Open

Build on any ML framework or library

Distributed learning with Apache Spark

Scale out GPU Training in the Cloud

Flexible AI Deployment

- Azure Container Service
  (scale out with Kubernetes clusters)
- Spark on HDInsight
- Azure IoT Edge
- On-prem, AWS, GCP....

Microsoft's finance organization digitally transforms forecasting with innovative machine learning solution

The machine-learning forecasts from Cortana Intelligence enable us to combine core financial data with additional sources of information including macroeconomic factors, product launches, promotions and Bing search trends. With the improving machine-learning forecast accuracies it is becoming an integral part of our financial planning and budgeting process.

Amy Hood: EVP & CFO

Microsoft

Employees: 114,074
Annual Revenue: $US85.32 billion

Microsoft Cortana Intelligence Suite | United States | Professional services—Software engineering | Business critical

What's in for today ?

Cognitive Services

- Our Journey into buying goods and services
- Does carpet-bombing of Ads and Print media still do the job?
- Lets try to up-sell and cross-sell with Cognitive services
- Demo? (I'm Leaving the technical deep-dive for your enthusiasm)

Image Recognition: Vision API

- Image recognition: Why is it gaining importance?
- Deep-Learning: How is it shaping our world?
- Demo? (Yes, I will cover the technical deep-dive)

Flight Delay Prediction

- Microsoft R-Server with DSVM
- Solution walk-through

Find all the resources @ https://github.com/Mrinal4Github/Infosys-PEAT-Repository

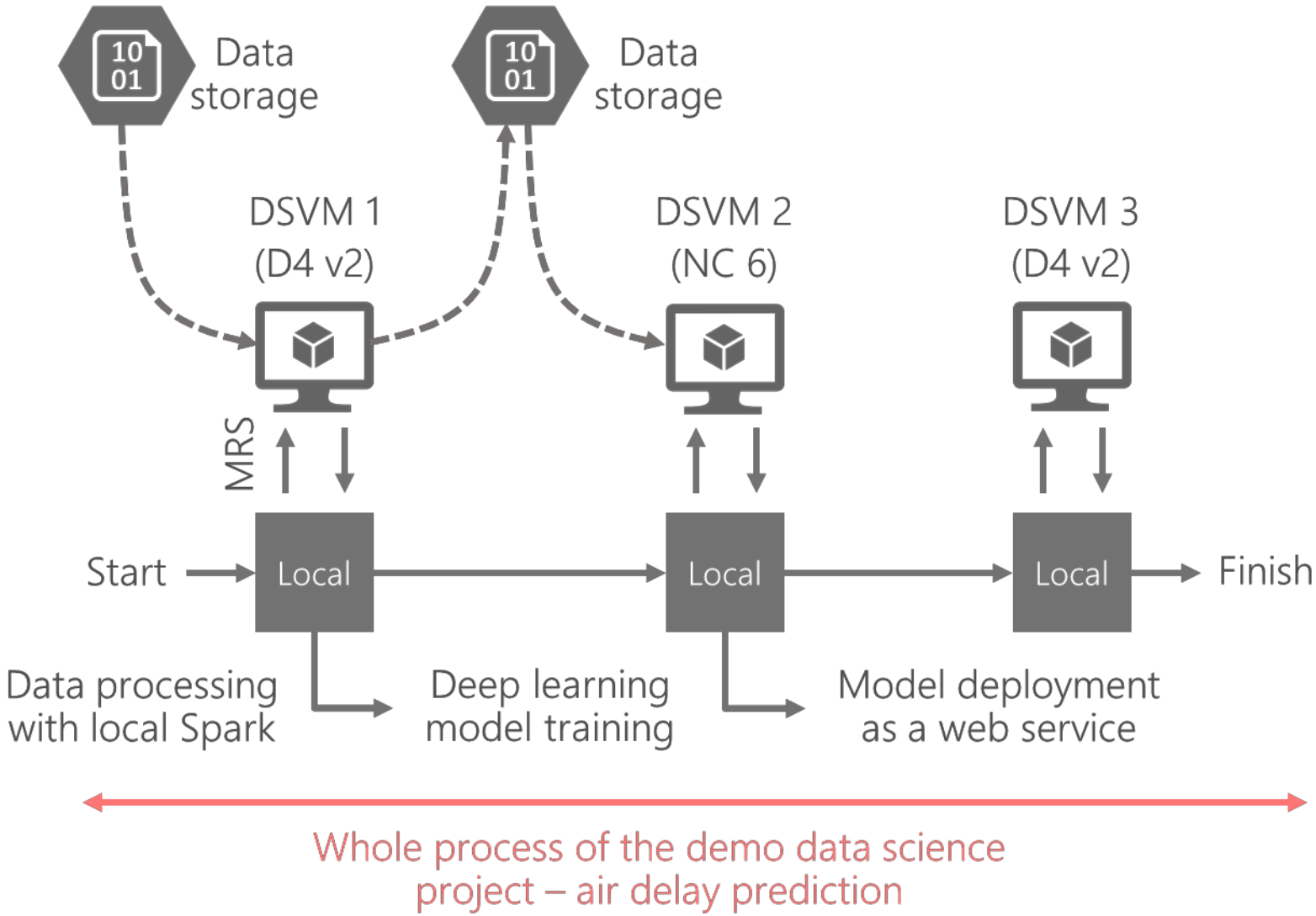
Use case demo flight delay prediction

- Flight delay prediction
  - Problem statement: predict flight delay given fleet information.
  - Data.
    - Size: ~1.4 G
    - Features: day of month, day of week, origin, destination, etc.
    - Prediction target: whether or not the flight is delayed.
- Assume we are going to build a pipeline that
  - Uses a sub-sampled (1%) and aggregated version of the original data.
  - Applies Spark for data pre-processing.
  - Trains a neural network model with GPU acceleration.
  - Publishes the model as a web based service.

- Computing resource planning.
- Data pre-processing on Spark, model training with GPU acce leration, and web-based service deployment.

DSVM name	DSVM size	OS	Description	Price
Spark	Standard D4 v2 – 8 cores with 28 GB m emory	Linux	Local standalone m ode Spark for data preprocessing and f eature engineering.	$0.585/hr
Deeplearning	Standard NC6 – 6 c ores with 56 GB me mory, and Nvidia Te sla K80 GPU	Windows	Train deep neural n etwork model with GPU acceleration.	$0.9/hr
Webserver	Standard D4 v2 – 8 cores with 28 GB m emory	Linux	Server host where web based model s ervice is published and run.	$0.585/hr

Demo – Flight Delay use case with Microsoft Data-Science Virtual Machine



Can we do everything in R?

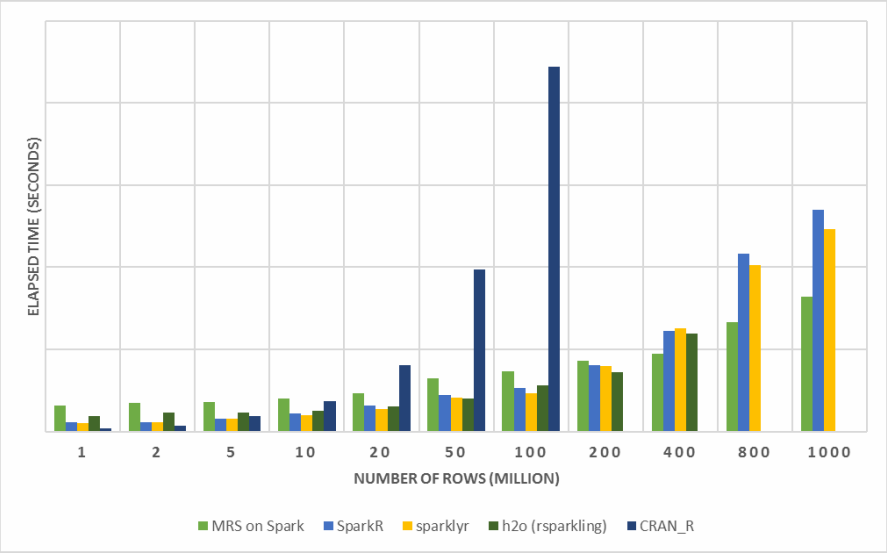
Demo – Flight Delay use case with Microsoft Data-Science Virtual Machine

Yes!

- Azure resource management in R.
  - AzureSMR
    - Managing a selection of Azure resources such as storage blobs, HDInsight, etc.
  - AzureDSVM
    - Deployment and operation on an Azure DSVM with specified size, OS, and user credentials.
    - Remote execution of script and file transfer with a Linux DSVM.
    - Retrieval of cost and expense information of using DSVM.
- Prerequisites
  - Azure subscription.
  - Initial setup for Azure Active Directory.

https://github.com/Microsoft/AzureSMR
https://github.com/Azure/AzureDSVM

- DSVM supports local standalone mode Spark.
  - For experimental and debugging purpose.
  - Up-scale code to Spark cluster in Azure HDInsight.
- R frontend for Spark
  - Microsoft R Server.
  - SparkR.
  - sparklyr.



E2E Process:

- Load Data from .csv
- Transform Features
- Split Data: Train + Test
- Fit Model: Logistic Regression (no regularization)
- Predict and Write Outputs

Configuration:

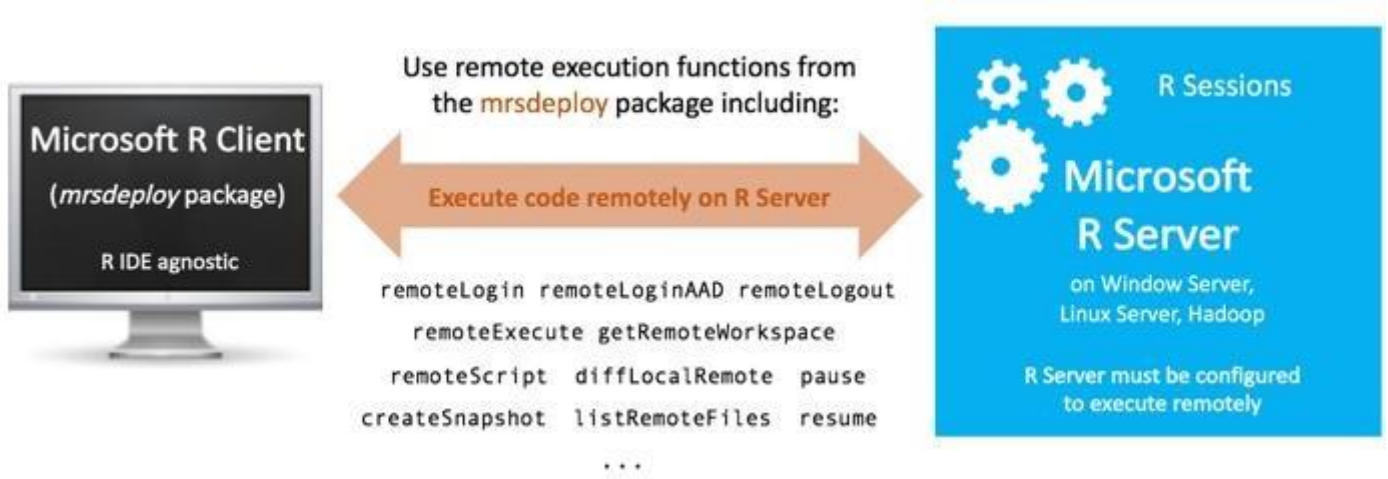
- 1 Edge Node: 16 cores, 112GB
- 4 Worker Nodes: 16 cores, 112GB
- Dataset: Duplicated Airlines data (.csv)
- Number of columns: 26

tinyurl.com/Strata2017R/Performance_Comparison

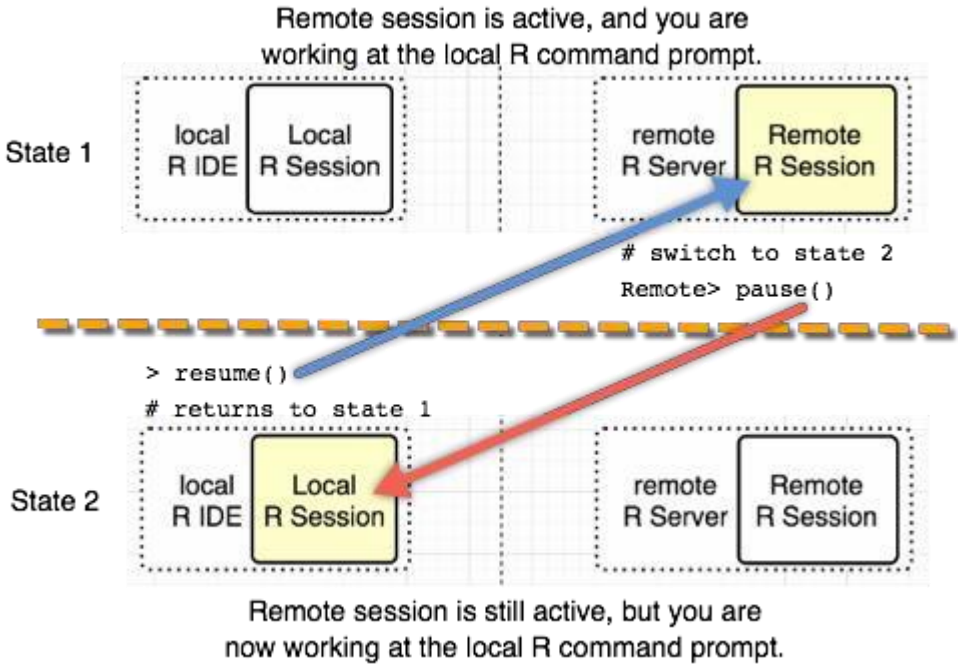
- Deep neural network in MicrosoftML package
  - rxNeuralNet() function.
  - GPU acceleration.
  - NET# language to customize network.
- Web service deployment in mrsdeploy package.
  - remoteLogin(), publishService(), getService(), etc.
  - Supports script-based and realtime based.
  - Supports Swagger.

rxNeuralNet() - https://msdn.microsoft.com/en-us/microsoft-r/microsoftml/packagehelp/neuralnet
NET# - https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-azure-ml-netsharp-reference-guide
Web service with mrsdeploy - https://msdn.microsoft.com/en-us/microsoft-r/operationalize/data-scientist-manage-services#realtime

Demo – Flight Delay use case with Microsoft Data-Science Virtual Machine



- mrsdeploy() package.
- Remote execution.
- One-box configuration.
- Access control via AAD.



- The demo can be found at https://github.com/Microsoft/acceleratoRs/tree/master/flightDelayPredictionWithDSVM
- Prerequisites:
  - Azure subscription (**free for trials**).
  - R 3.3.x.
  - Microsoft R Server 9.x.
  - Microsoft R Client.
  - R packages
    - AzureSMR – http://github/Microsoft/AzureSMR
    - AzureDSVM – http://github/Azure/AzureDSVM