Data Integration and Policy Frameworks for Bridging the Gap: (Public - Private Partnerships)

Steven Hagan
Vice President Engineering
Oracle Database Server Technologies
July 2018

PRIORITY ONE = ??

HOW FAST DO YOU WANT TO MEET SDG GOALS ?

DO MORE PARTNERING WITH PRIVATE INDUSTRY!

Private Industry People want the SAME GOALS!

You Enhance Integration & Sharing By Using **STANDARDS**

e.g. – The Spatial / Semantics/ Statistics Data Domains

- **ISO**
  - **TC 211; TC 204, 19115**
- **Open Geospatial Consortium**
  - **Simple Features; GML; Web Services**
- **De-facto Standards**
  - **SHP, MGE, DXF, KML**
- **Professional Standards**
  - **ISPRS, FIG, WMO,DDI, SDMX**
- **Java, .NET, Flash**
- **W3C: RDF,OWL, SPARQL, GeoSPARQL**
- **TAGGED METADATA – agree on tags**

PLATFORM TRENDS: ENHANCING INTEGRATION & NEW POLICIES:

Hardware - **EVOLUTIONARY** – Moore's law still holding

- New possibilities at Research Level – not yet proven – DNA, Quantum, Holography, Graphene …

- Software – **DISRUPTIVE** – **Parallelism** => clusters of **10,000+** computers: Enabling

  - **CLOUD, Machine Learning, Artificial Intelligence**

- Software: AVAILABLE NOW - Supporting all Data types in Databases

  - Databases/persistent stores: *POLYGLOT PERSISTENCE* now can handle *ALL* types of data

  - Software – *GRAPH STORAGE, SEMANTICS, ONTOLOGIES, STATISTICS*

    - – Add all types of data, build *NEW* relationships

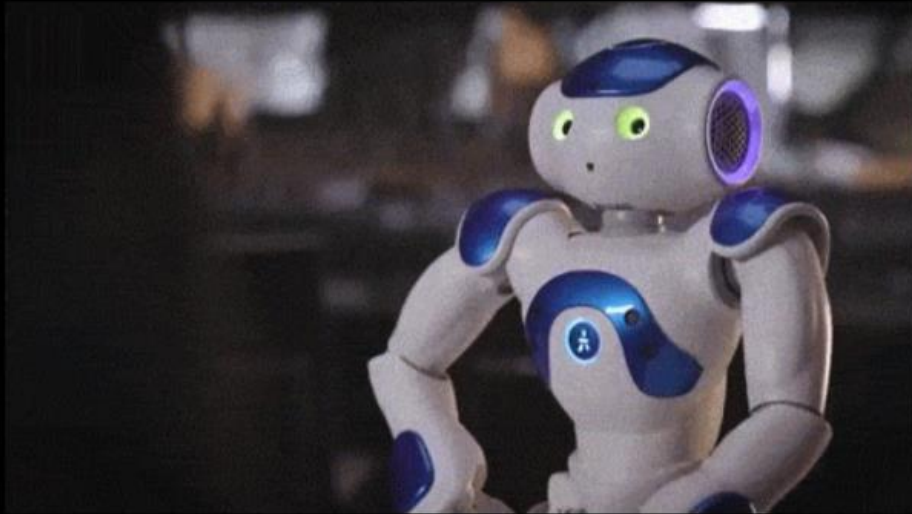 - Enables *MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE* (ML, AI)

  - Stream data arriving; Filter the data; ML: Keep what matches your requirements; aggregate it, make it accessible for **ALL SEVENTEEN (17)** goals.

  - SECURITY – PRIVACY – Encryption improvements

ORACLE®

Modern Infrastructure for Artificial Intelligence

ORACLE®

Fueling all industries

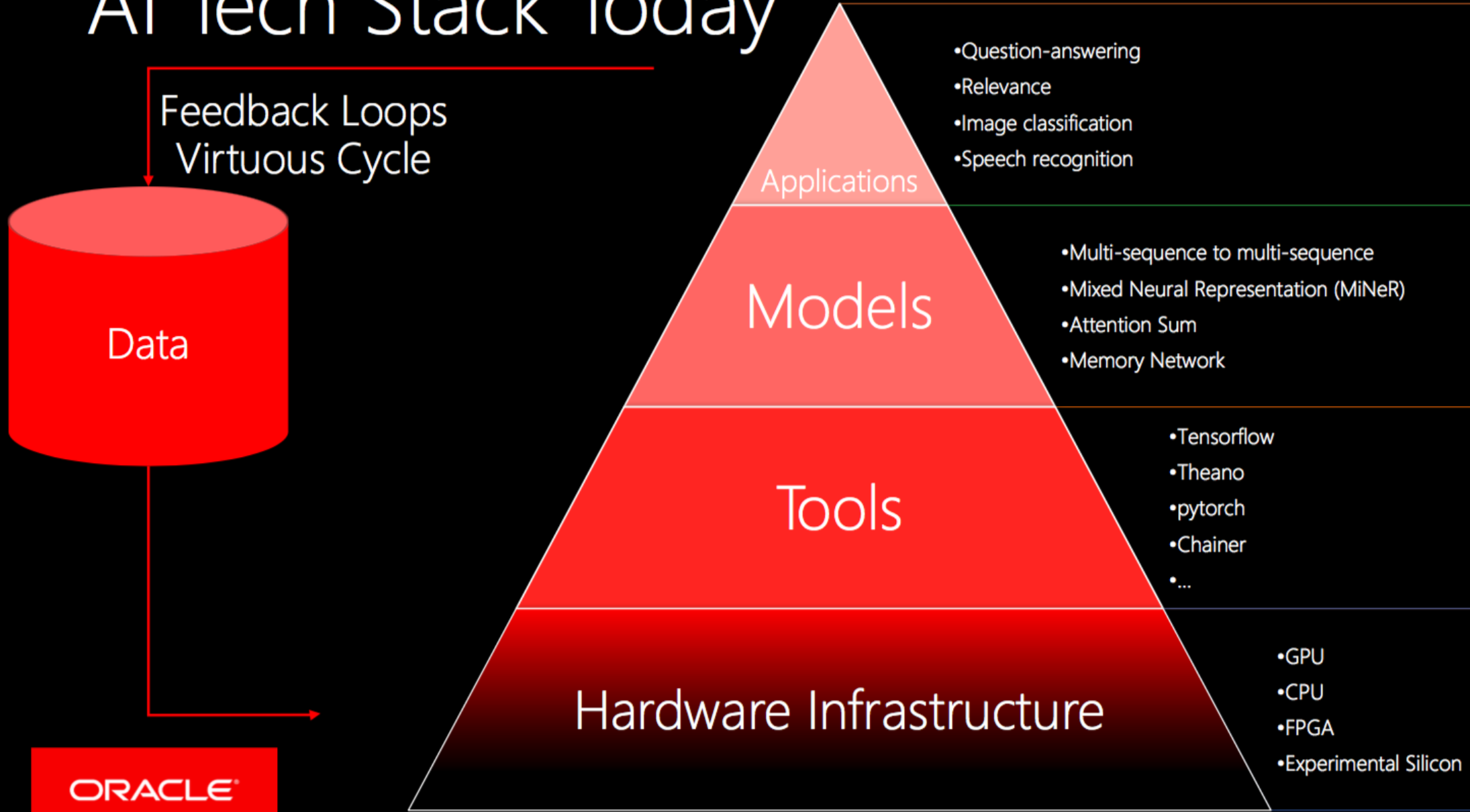
Providing intelligent services in hotels, banks and stores

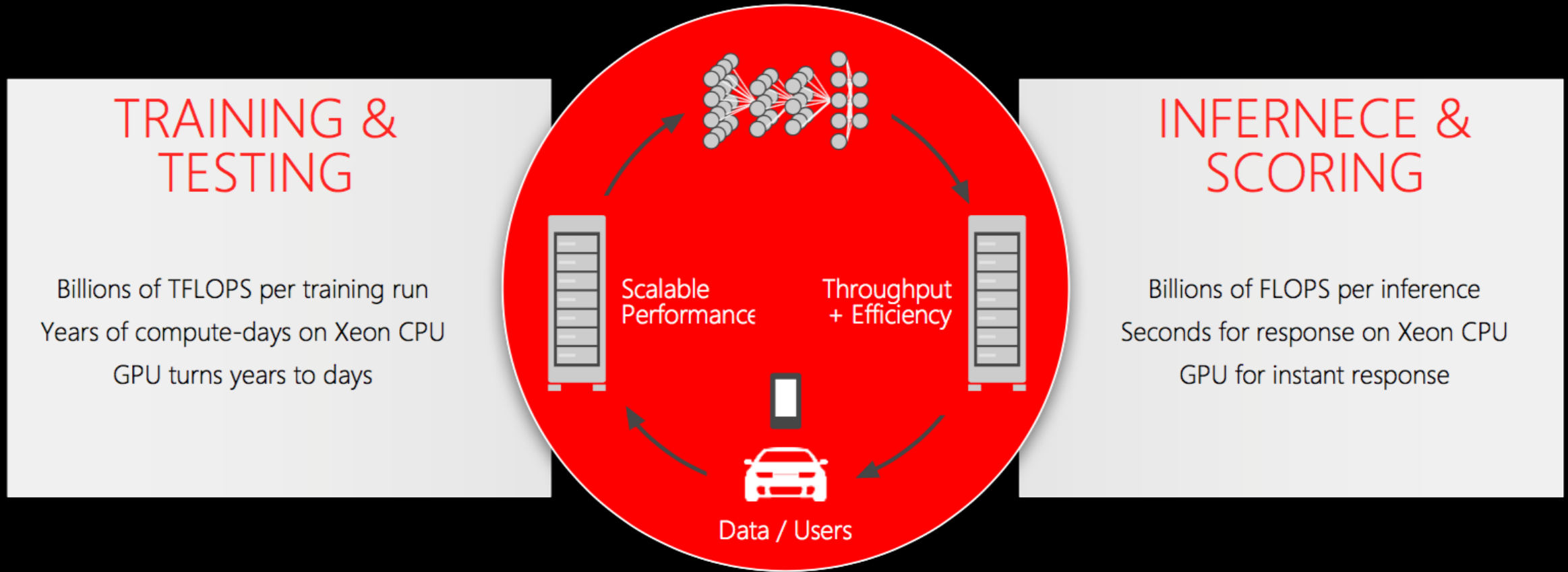Separating weeds as it harvests, reduces chemical usage by 90%

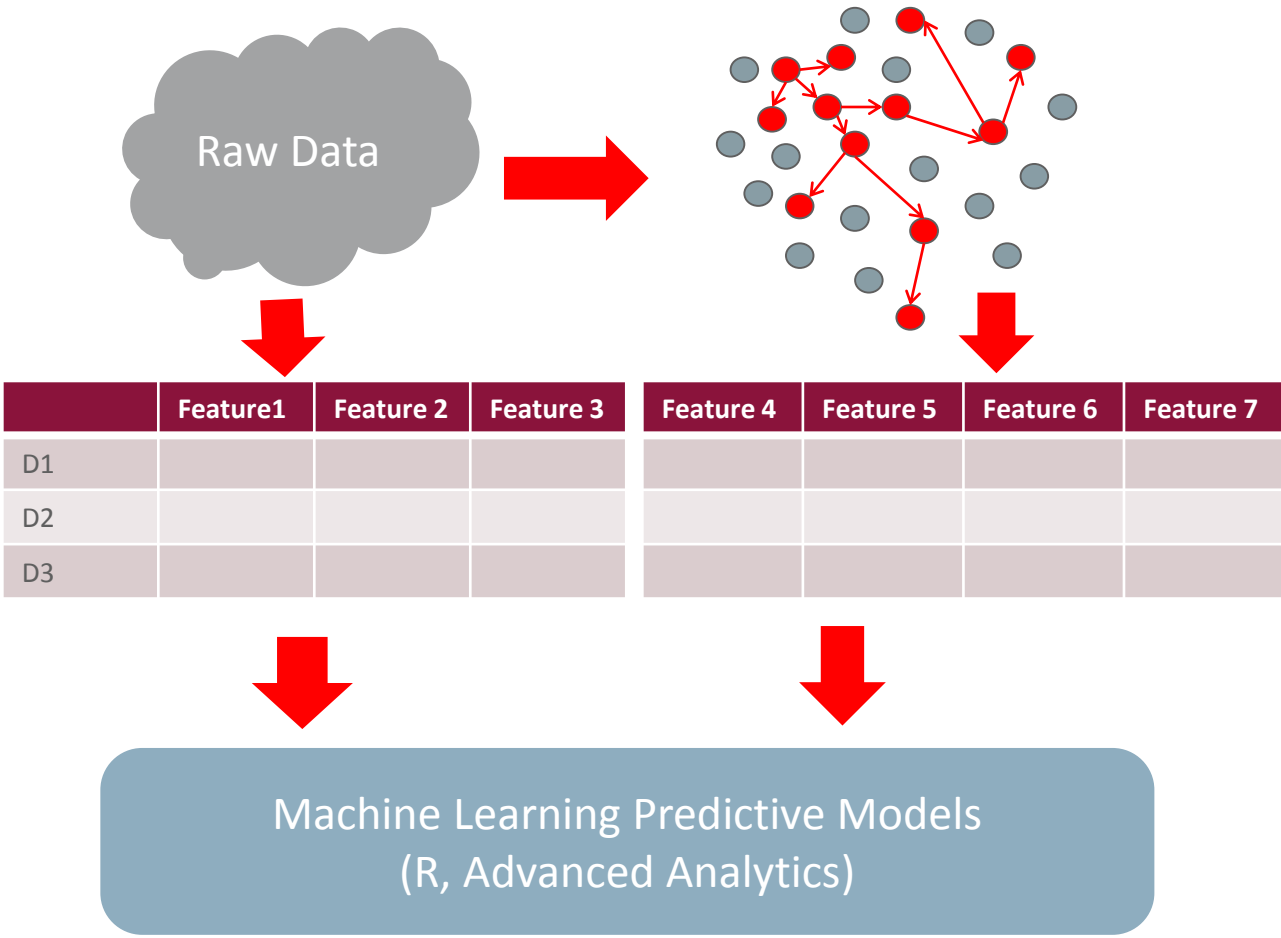Increasing public safety with smart video surveillance at airports & malls

ORACLE®

AI Tech Stack Today

Feedback Loops
Virtuous Cycle

Data

Applications
- Question-answering
- Relevance
- Image classification
- Speech recognition

Models
- Multi-sequence to multi-sequence
- Mixed Neural Representation (MiNeR)
- Attention Sum
- Memory Network

Tools
- Tensorflow
- Theano
- pytorch
- Chainer
- ...

Hardware Infrastructure
- GPU
- CPU
- FPGA
- Experimental Silicon

ORACLE®

Enhancing ML and Data Analytics with Graphs

- Graph analysis can enhance the quality of ML and data analytics

- Graph representation helps discover hidden information about the data
  - Multi-hop relationship between data entities

- This can be used to further improve predictive models in R, Advanced Analytics, machine learning

	Feature1	Feature 2	Feature 3
D1			
D2			
D3			

	Feature 4	Feature 5	Feature 6	Feature 7
D1				
D2				
D3				

Machine Learning Predictive Models
(R, Advanced Analytics)

How do they connect?

Graph(+)ML World

Graph Types
1. Standard graphs
2. Property graphs
3. Knowledge graphs

Extracted graph features can **boost** standard ML predictive models by providing structural info

	Feature 4	Feature 5	Feature 6
D1			
D2			
D3			

Raw Data

	Feature 1	Feature 2	Feature 3
D1			
D2			
D3			

ML Tasks on Graphs
1. Node/GraphLet Classification
2. Node similarity
3. Graph similarity

Oracle Statistics / Analytics Machine Learning Algorithms

Classification
- Logistic Regression
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Naïve Bayes
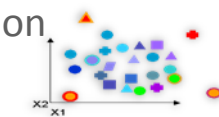- Explicit Semantic Analysis
- Gaussian Mixture Models

Clustering
- Hierarchical K-Means
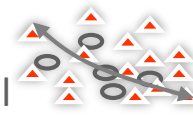- Hierarchical O-Cluster
- Expectation Maximization
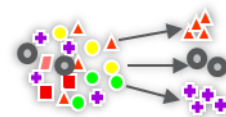
Anomaly Detection
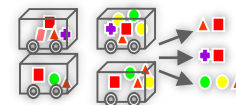- One-Class Support Vector Machine

Regression
- Generalized Linear Model
- Support Vector Machine
- Random Forest
- Linear Model
- Stepwise Linear regression
- LASSO

Association Rules
- A priori

Attribute Importance
- Minimum Description Length
- Principal Component Analysis
- Unsupervised Pairwise KL Divergence

SQL Predictive Queries

Statistical Functions

Algorithm Text Support
- Algorithms support text type
- Tokenization and theme extraction
- Document similarity

Feature Extraction
- Principal Component Analysis
- Non-negative Matrix Factorization
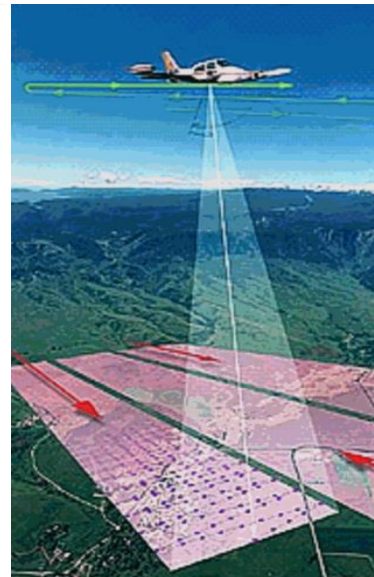- Singular Value Decomposition

Time Series
- Single Exponential Smoothing
- Double Exponential Smoothing

Open Source ML Algorithms
- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLlib algorithm integration

Acquiring/Keeping Data for _17_ Sustainable Development Goals: Need Platforms for **ALL** Variety, Velocity, Volume of Data

- VIDEO: UAVs, DRONES, SURVEILLANCE

- IMAGERY/Raster: (Satellites, Medical)

- Sensors (IOT), LIDAR, 3D, RFID, Wearables

- Social Media, Web Scraping, Mobile Phones

- New data products for: Land and Water mgmt, Agriculture, Environment Transportation, Terrain and City Models, SDIs for planning, maintenance, Emergency response, Defense, Intelligence, Consumers , Healthcare

- Genomics (DNA Sequencing)

- Semantics , Ontologies →

- Machine Learning, AI, Statistics

- **Location is a Powerful Organizing Principle**
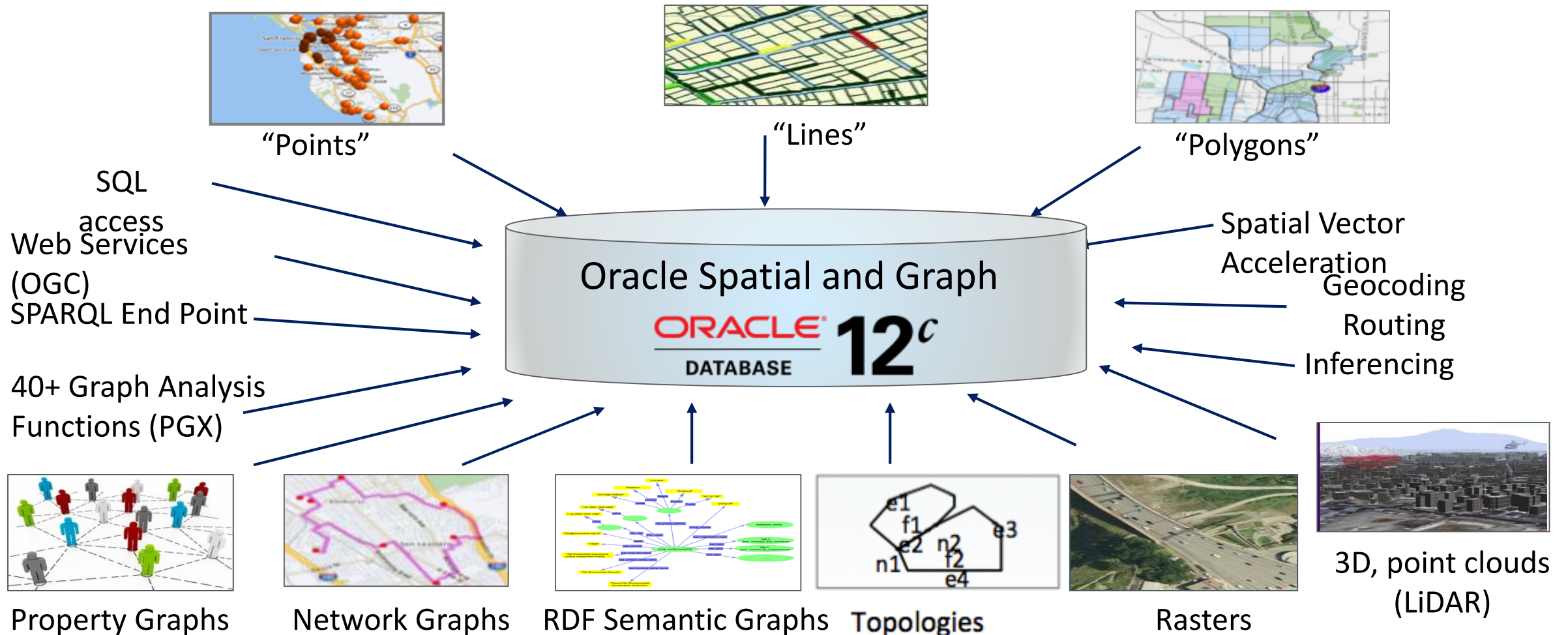
- **MULTIPLE VERSIONS OF THE ABOVE**

Success Examples of Public - Private Partnerships
Tell Industry what you want – They implement / Maintain it

- U.S. Census Taught Oracle Spatial how to do Planar Topology
  - Also extensive work on the fundamental spatial data structures

- Netherlands – TU Delft - 3D point Clouds for entire country >500B
  - Also – we worked together on building a Private Cloud for entire country

- Kegg Japan – Explained need for and how to do RDF – Semantics

- Ireland Ordnance Survey – Needed Linked Open Data – Now Public

- Digital Globe – Extremely Large Polygons - done

- Ordnance Survey – England, Ireland – needed Workspace Manager
  - multiple scenarios for what-if analyses or multiple editions of data for publication

- Many – needed Parallelism for Spatial Operators – use 1000s of Cores

- Navteq Nokia – Location Services

Managing All Spatial, Graph, Statistic Data – in One Store

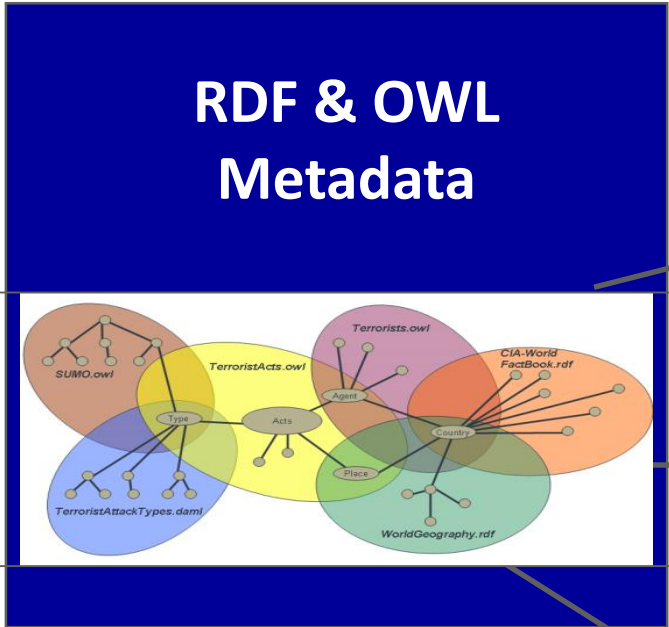## Location and Statistics analysis with Secure, scalable storage for enterprise data

"Points"

"Lines"

"Polygons"

SQL
access
Web Services
(OGC)
SPARQL End Point

40+ Graph Analysis
Functions (PGX)

Oracle Spatial and Graph

ORACLE 12c
DATABASE

Spatial Vector
Acceleration
Geocoding
Routing
Inferencing

Property Graphs

Network Graphs

RDF Semantic Graphs

Topologies

Rasters

3D, point clouds
(LiDAR)

Meet Sustainable Goals: Repurposing Data: Ontology-driven Enable Shared, Actionable Knowledge

Application Ontologies

National Mapping Private Cloud
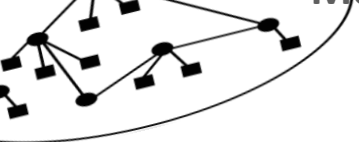
Spatial Data

Geographic Names

Raster Data
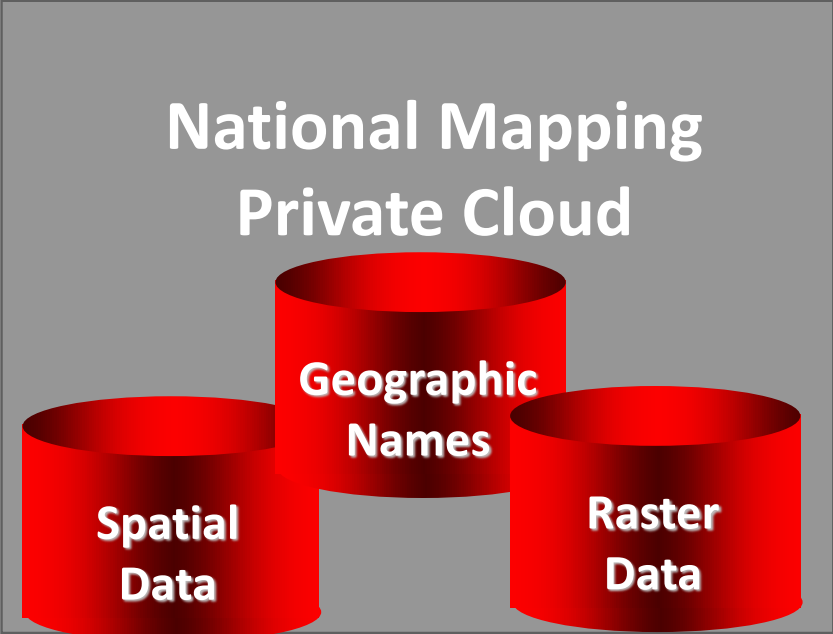
- Simple Features
- GeoRaster
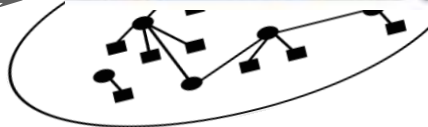- Topology
- Networks
- Gazetteers
- ...

RDF & OWL Metadata

- Data Integration
- National Map schemas
- Geographic names
- Temporal
- Naïve Geography
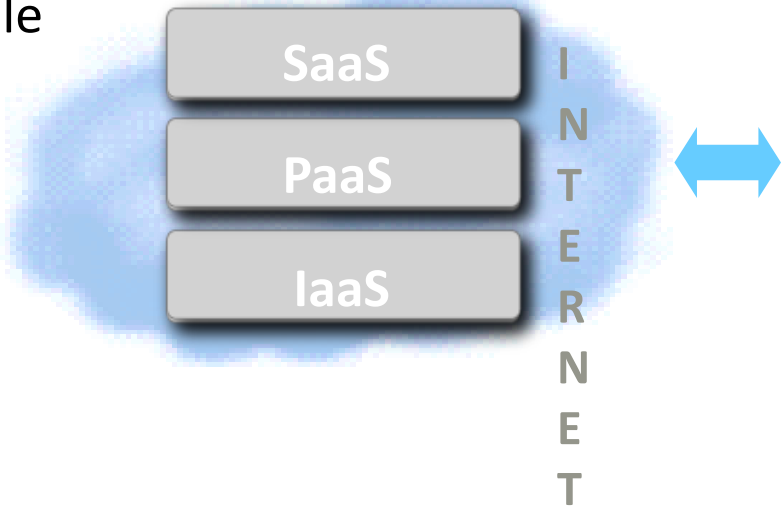- ...

Environmental Monitoring

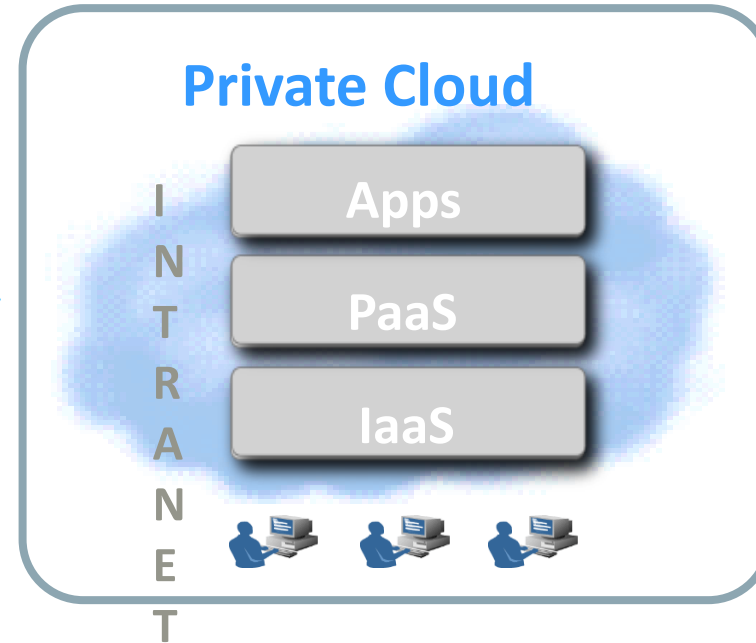**Disaster Recovery**

Healthcare Biotech

Public Clouds, Private Clouds & WEB SERVICES

Public Clouds

SaaS

PaaS

IaaS

INTERNET

Private Cloud

Apps

PaaS

IaaS

INTRANET

- Used by multiple tenants on a shared basis
- Hosted and managed by cloud service provider

- Exclusively used by a single organization
- Controlled and managed by in-house IT

Trade-offs

Lower *upfront* costs ⟷ Lower *total* costs

Outsourced management ⟷ Greater control over **SECURITY, COMPLIANCE, QOS**

OpEx ⟷ CapEx & OpEx

ELASTICITY is key value of Clouds

Oracle Technology Supplies both Public and Private clouds

YOU MAY NEED A CLOUD IN EACH COUNTRY ---DEPENDS ON THEIR LAWS

To Meet 2030 Goals: Do NOT Build Your Solutions From Scratch
Long Term Cost of Ownership rises with custom construction & Open Source

Time to Build

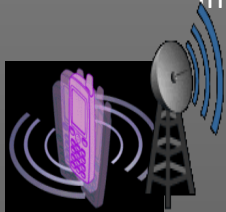**Optimizations**

Maintenance

UN-GGIM: "train the individuals is at least five years"

Sustainable Goals: All Data Types /Statistics/ ML / AI Bases: Success Enhanced with **MULTI-MODEL DATABASE PLATFORM**
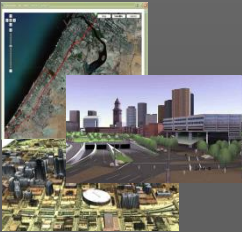
Big & Fast Data

Volunteered Geographic Statistical Information

Sensors Streaming Data

Geo-referenced Video, 3D, LiDAR Satellites

Simplify Statistics IT

Support for Open Standards

Spatial Database, Application Server, BI, tools
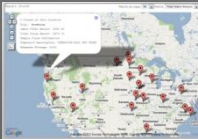
Support by Leading Partner solutions

Multi-Model Engineered Systems

Deep Analytics

Real-time Complex Event Processing

Dense Visualization

Spatial Analysis Graph Analytics

On Premise, On Cloud, Shared Services

Shared GeoSpatial Services Location Aware Everything

Fully Parallel and SECURE