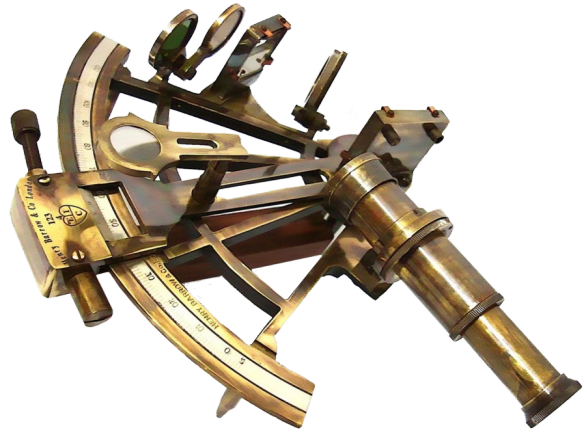


Bringing Data into Focus

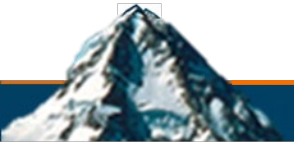
Brian F. Tankersley, CPA.CITP, CGMA

K2 Enterprises

Bringing Data into Focus

It has been said that data is the new oil, and our smartphones, computer systems, and internet of things devices add hundreds of millions of gigabytes more every day. The data can create new opportunities for your cooperative, but your team must take care to harvest and store it properly. Just as oil must be refined and separated into gasoline, diesel fuel, and lubricants, organizations must create digital processing platforms to the realize value from this new resource. This session will cover fundamental concepts including extract/transform/load, big data, analytics, and the analysis of structured and unstructured data. The materials include an extensive set of definitions, tools and resources which you can use to help you create your data, big data, and analytics strategy so you can create systems which measure what really matters in near real time.

Stop drowning in data! Attend this session to learn techniques for navigating your ship on ocean of opportunity provided by digital exhaust, and set your course for a more efficient and effective future.



About Brian Tankersley @BFTCPA

CPA, CITP, CGMA with over 25 years of Accounting and Technology business experience, including public accounting, industry, consulting, media, and education.

- ***Director, Strategic Relationships***, K2 Enterprises, LLC (k2e.com) (2005-present)
  - Delivered presentations in 48 US states, Canada, and Bermuda.
- ***Author***, *2014-2019 CPA Firm Operations and Technology Survey*
- ***Director, Strategic Relationships / Instructor***, Yaeger CPA Review (2017-present)
- **Freelance Writer** for accounting industry media outlets such as AccountingWeb and CPA Practice Advisor (2015-present)
- ***Technology Editor***, The CPA Practice Advisor (CPAPracAdvisor.com) (2010-2014)
- Selected seven times as a "Top 25 Thought Leader" by *The CPA Practice Advisor*.
- 2012 and 2009 Outstanding Discussion Leader Award, *TN Society of CPAs*
- Author of numerous articles and have direct consulting experience with and for numerous organizations such as software vendors and CPA Firms.
- brian@k2e.com // @BFTCPA // +1(865) 684-4707
- www.cpatechblog.com

CPA Practice Advisor TOP THOUGHT LEADER 2018

YAEGER CPA REVIEW

CPA CITP Certified Information Technology Professional

CGMA Chartered Global Management Accountant

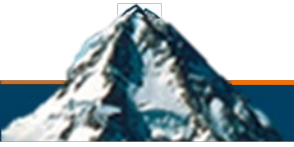
quickbooks Certified ProAdvisor Online | quickbooks Certified ProAdvisor Desktop | quickbooks Certified ProAdvisor Enterprise

CPA Practice Advisor www.CPAPracticeAdvisor.com

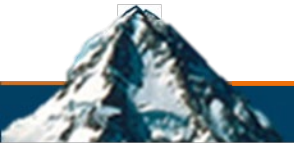
Overview of Presentation

- **Digital Exhaust, ETL, and Business Intelligence**
- **Big Data**
- **Machine Learning**

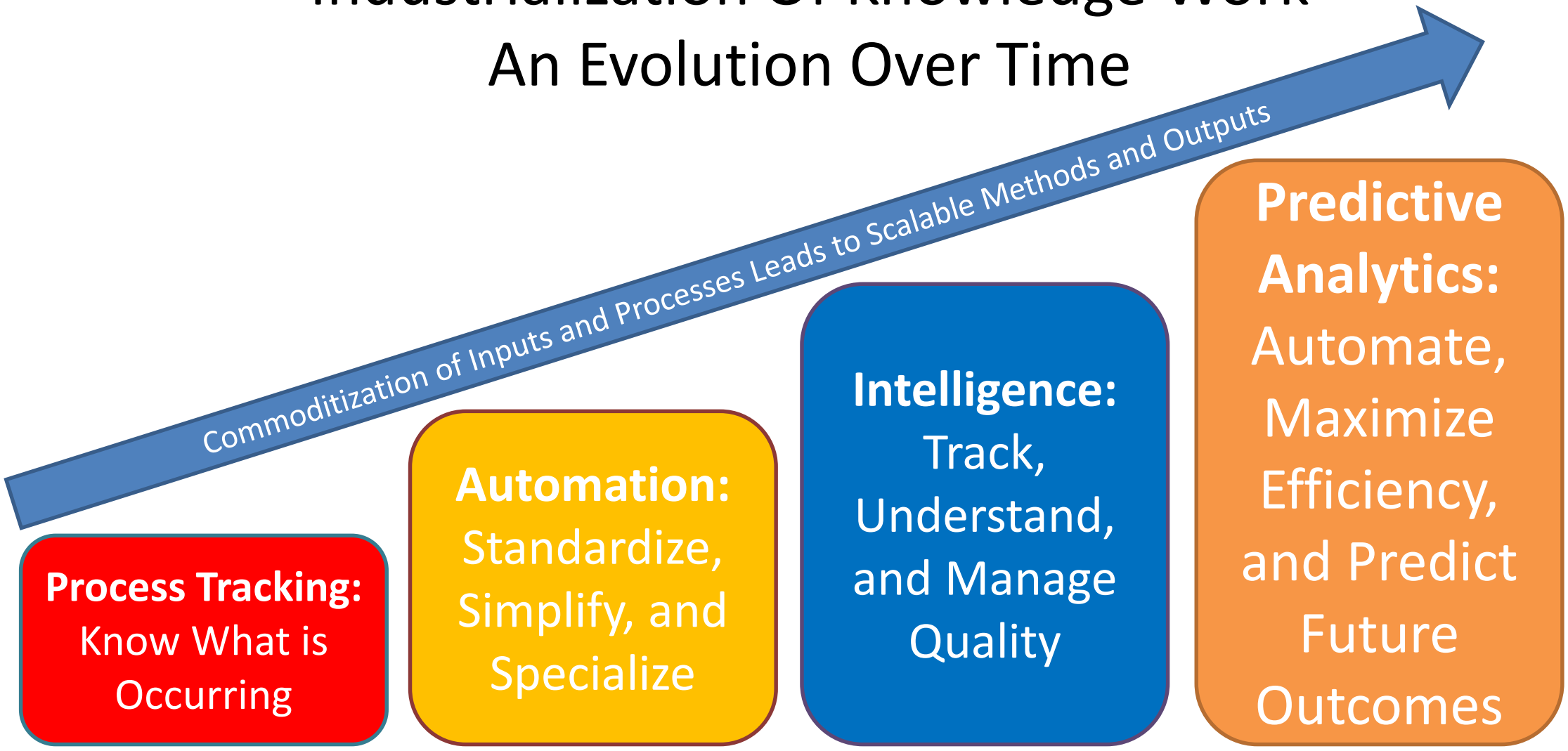
- **<u>Appendices</u>**
  - Business Intelligence Tools and Platforms
  - Big Data Tools and Platforms
  - Machine Learning Tools and Platforms



DIGITAL EXHAUST, ETL AND BUSINESS INTELLIGENCE

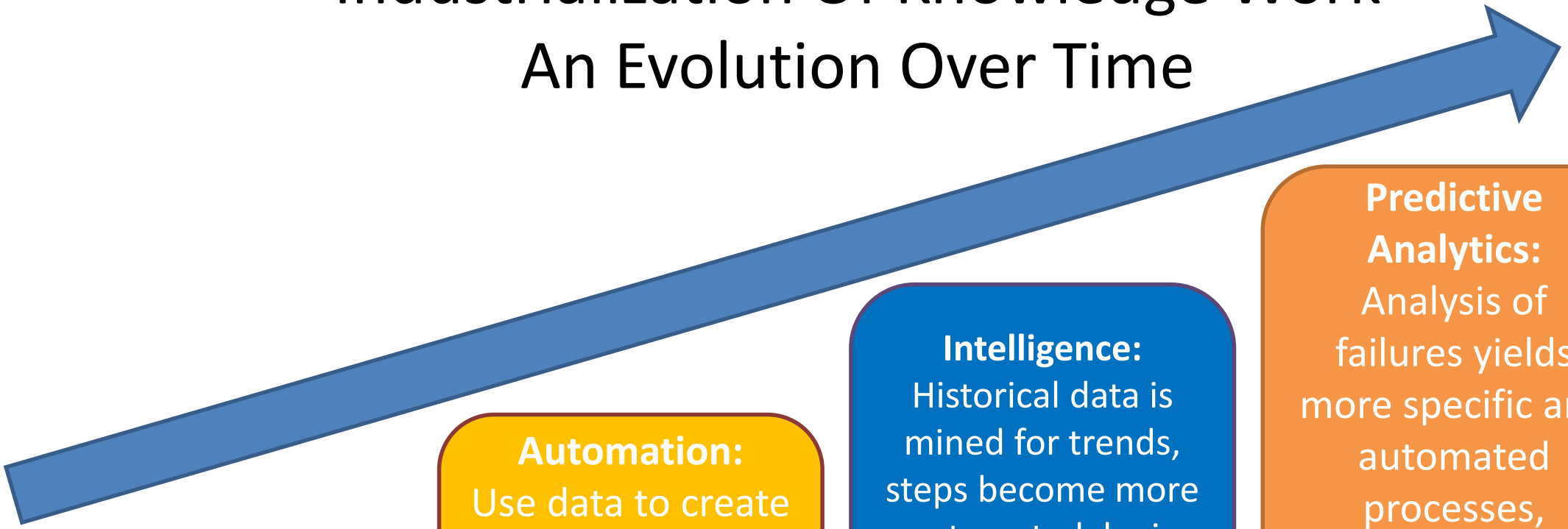


Industrialization Of Knowledge Work
An Evolution Over Time



As Work is Industrialized, More of It Can Be Performed By Less Knowledgeable Team Members at Lower Cost

Industrialization Of Knowledge Work
An Evolution Over Time



Process Tracking: Requires consistent processes, creates structured data

Automation: Use data to create alerts, develop new KPI's and automate process logic

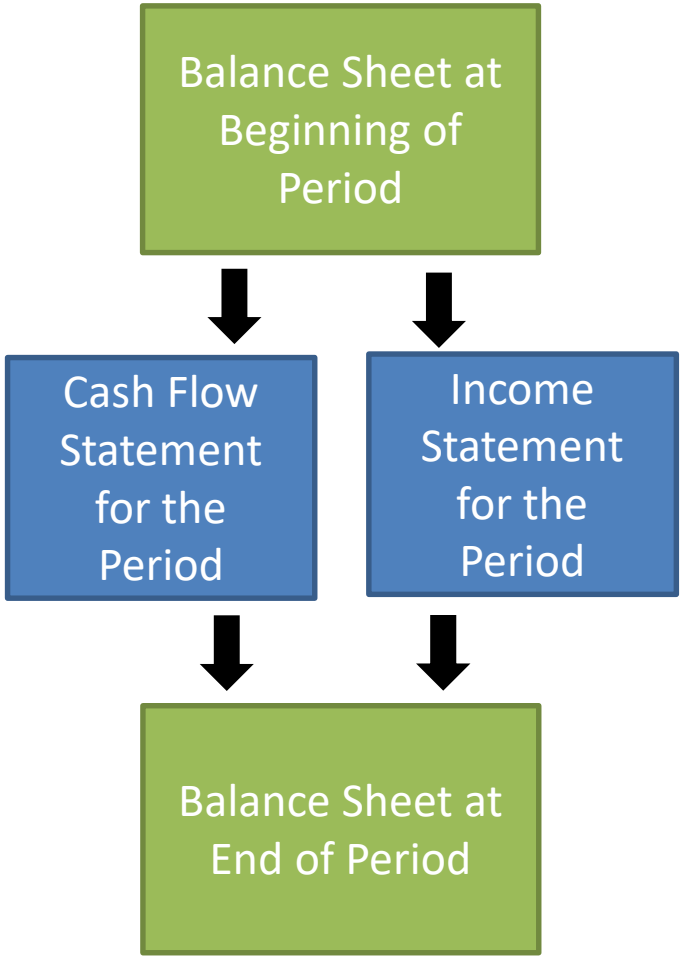
Intelligence: Historical data is mined for trends, steps become more automated, logic more sophisticated. Benchmarking becomes common.

Predictive Analytics: Analysis of failures yields more specific and automated processes, structured data used to identify "at risk" transactions

As process information is standardized, data is accumulated, analyzed to create benchmarks (phase 1) and is then used to create new measures of process performance (phase 2), followed by predictive analytics (phase 3)

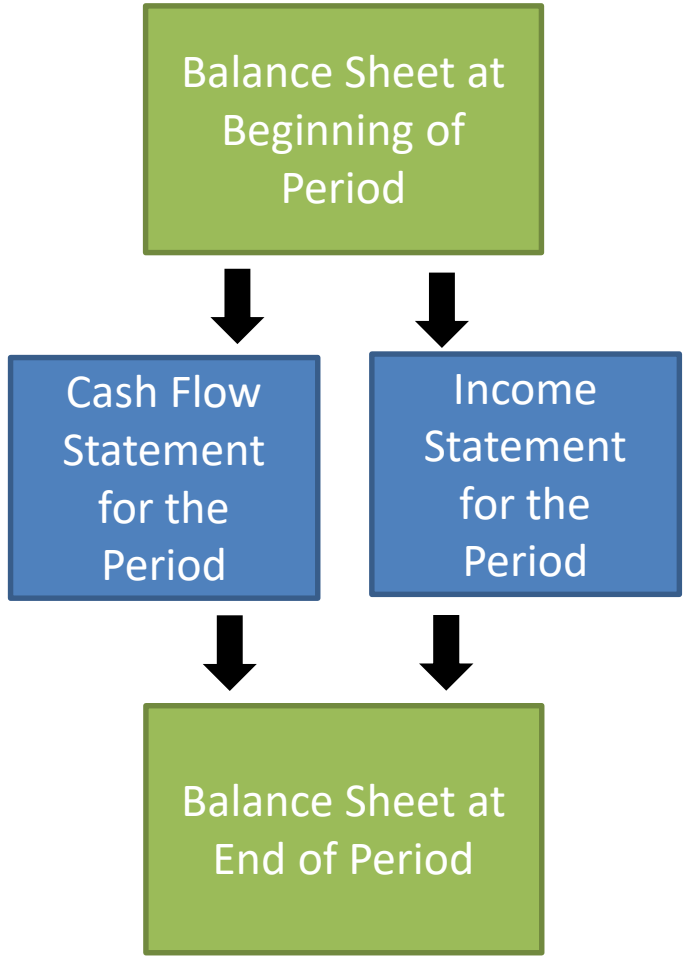
Basic GAAP Financial Reporting Model

- Requires 5 days or so after end of period to create financial statements
  - Instant report at end of period is STILL not fast enough for modern business
  - Substantive audit of records usually takes 45 days from end of period
- Third parties have little to no assurance from outsiders that transactions and assets/liabilities are valid without attest work
  - Can get paper/electronic documentation, but very difficult to link them to each other
- Both the bank and the client trust you to assess the fairness of the F/S presentation



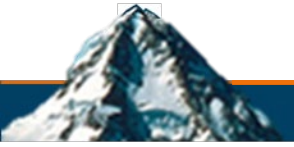
Basic GAAP Financial Reporting Model

- Balance sheet tells you what you have, what you owe, and what portion you own
- Cash flow statement and income statement are simply two different bases of accounting for the activity between the two balance sheets
- In this model, data can be either timely or reliable, **but not both.**
- Real time full I/S financial reporting is not as useful for tactical decisions as benchmarked real time operational data linked to processes

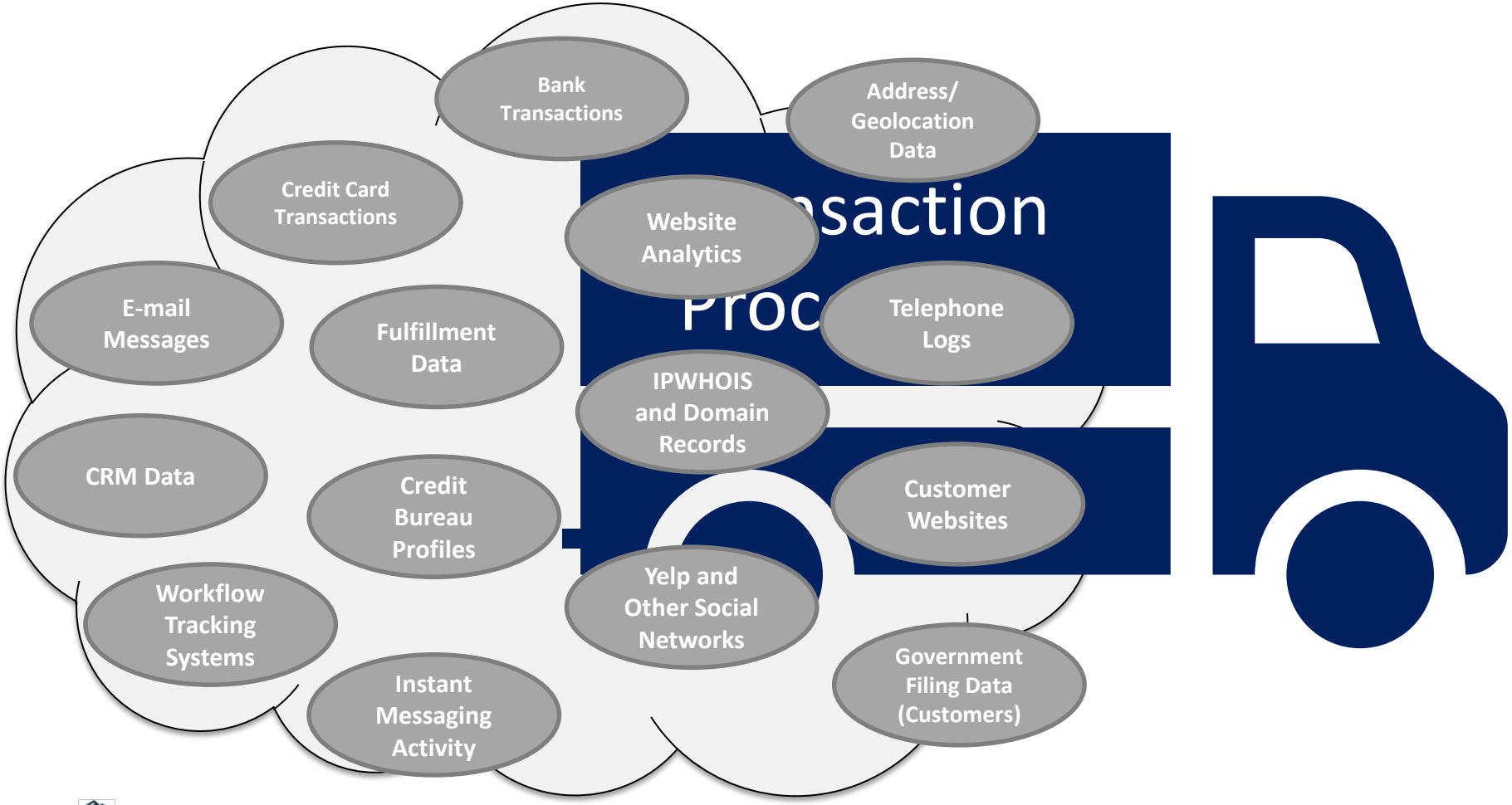


Many Great Reporting Platforms for Financials

- Sage Intelligence
- Qvinci
- BizNet BizExcelerator
- Infor's F9
- FYISoft
- Extraction/creating audit analytics from F/S data with Validis/CCH Audit Accelerator



Digital Exhaust, Workflow, and API's Can Enable Real Time Alerts, Reporting, and Data Validation

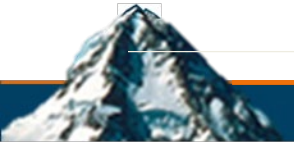


Digital Exhaust

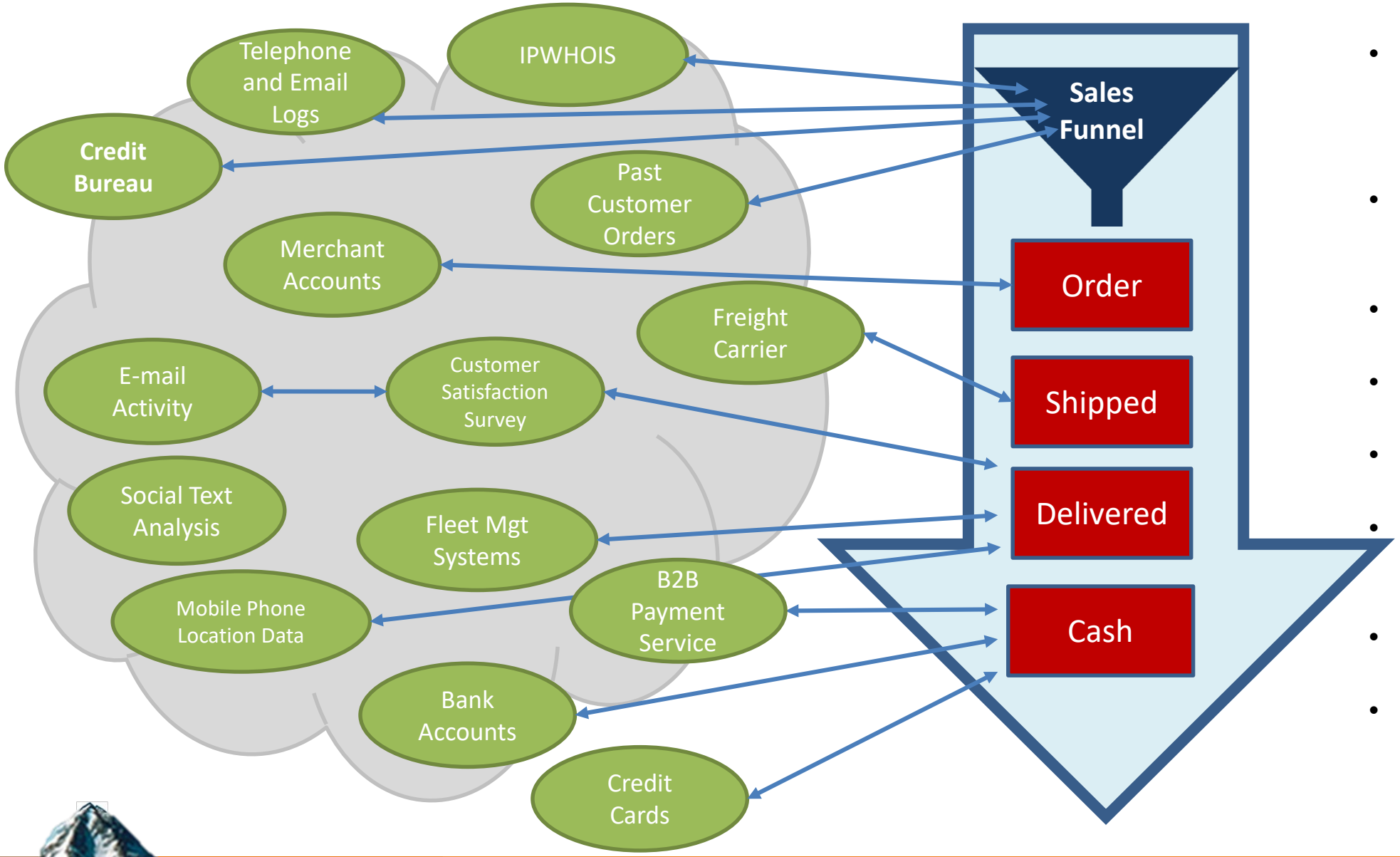
Goal of New Systems: Assess Quality of Current Activity and Signal When Quality Declines

How Do You Use The Digital Exhaust to Create Danger Warnings (Canary in a Coal Mine)?

- Standardize, optimize, automate, and benchmark processes based on data from the digital exhaust stream
- Build workflows on platforms that allow data to be exfiltrated in near real time
- Use enterprise grade process management tools to create alerts which let stakeholders know when something is out of standard through the business intelligence system
- Store exported in a SQL database, in a data warehouse, and use tools like Microsoft Power BI, Tableau, or Qlik Sense to track, monitor, and manage it
- Create rules to calculate new performance measures
  - First, track the process and analyze the related data
  - With the analysis, you can set goals/standards for processes
    - Calculate the difference in two date/time stamps on workflows
    - Set alerts when items appear likely to not meet standards
  - The resulting reporting can be used to evaluate performance and focus improvement efforts

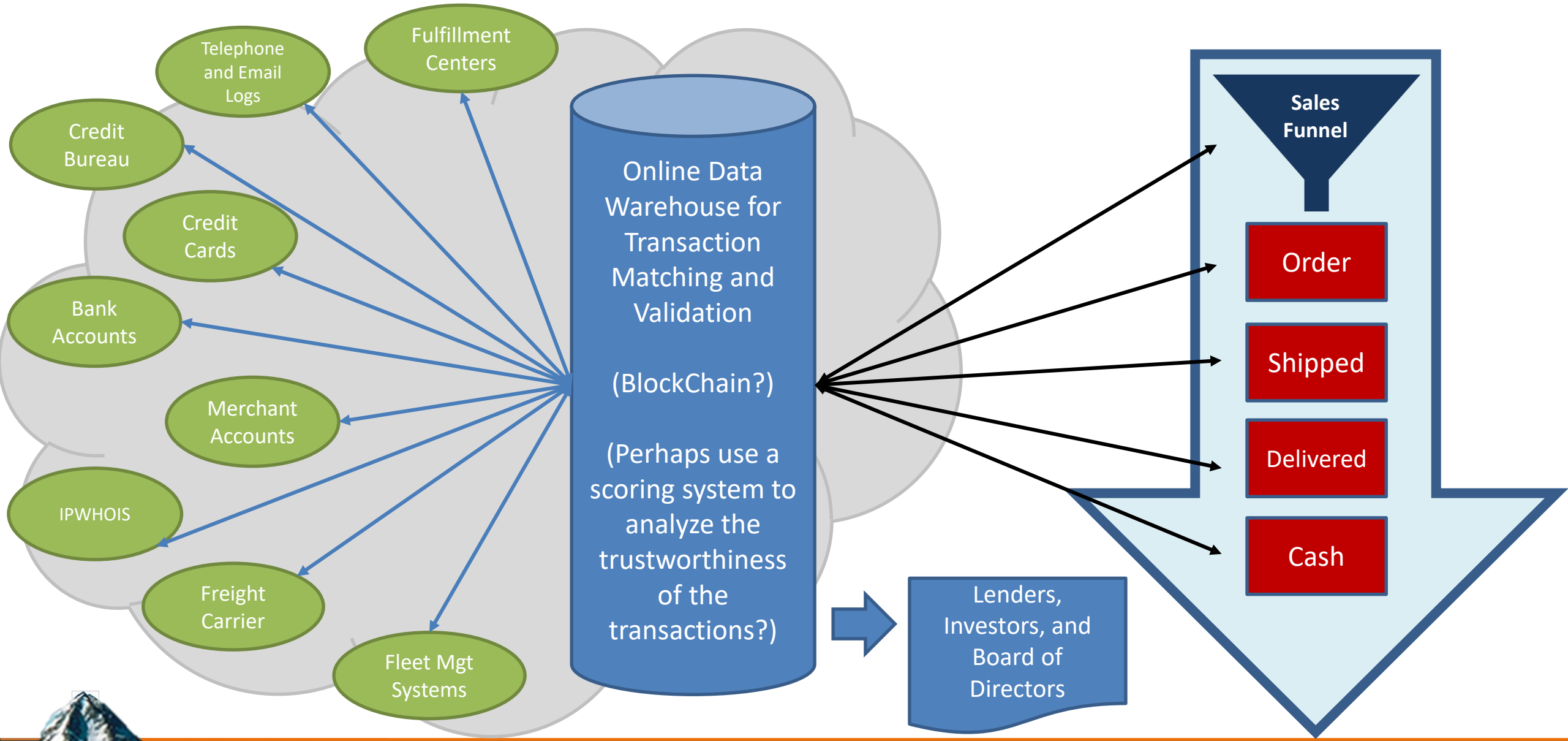


Real Time Transaction Validation Model



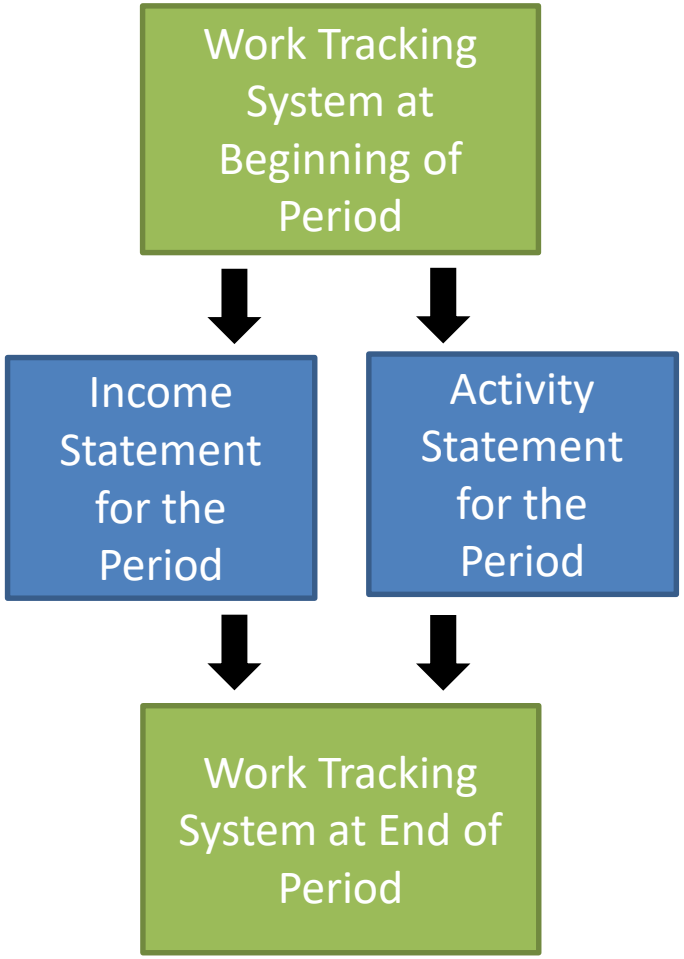
- Is this customer real?
  - Credit card
  - Credit evaluation
  - MOSAIC profile
- Is this order consistent with our history with this customer?
- Can the customer pay for this order?
- Is the address for the order legitimate?
- Was the item actually shipped and delivered?
- What is the customer's third party customer satisfaction rating for the company?
- How many chargebacks in the last year? Mo.? Wk.?
- Did company trucks and mobile phones visit customer site?

Future Real Time Transaction Validation Model



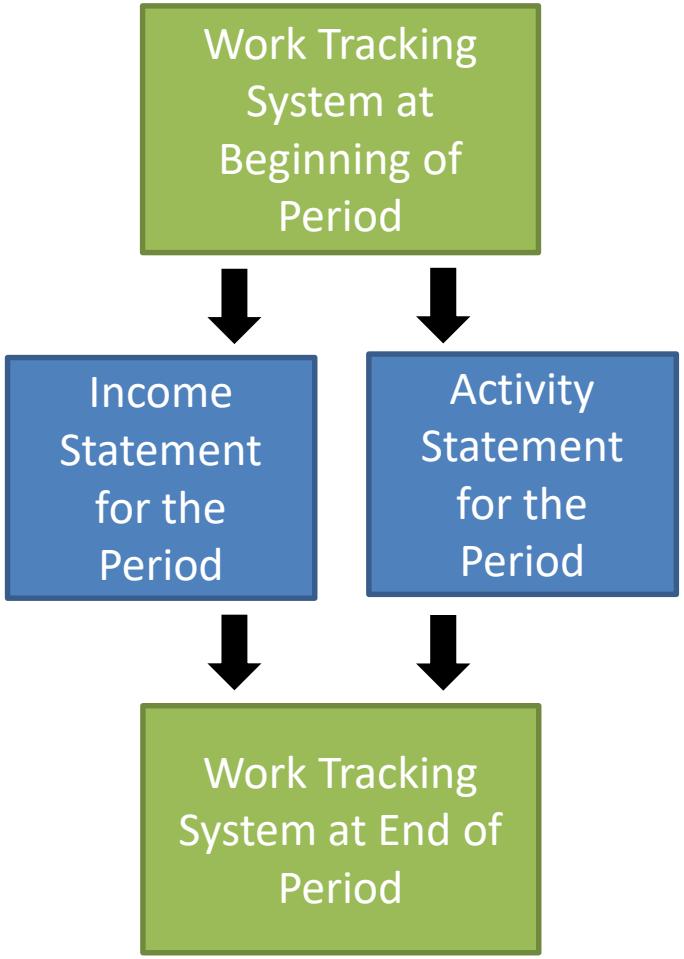
Real Time Reporting Model

- Reporting can be instant (or near-instant), and always up to date
- Processes are standardized and are first tracked, then mined for historical performance analytics, and with sufficient history, can be used to predict future activity/performance timeline
- Integrations with third party applications (e.g. UPS, FedEx, Amazon, eBay, merchant accounts, bank accounts) give real time insight into the quality of the activity in the systems
  - Tracking numbers and activity for shipments can be verified online by third parties
  - GPS and mapping data from third parties like Google Street View can be used to validate businesses at an address



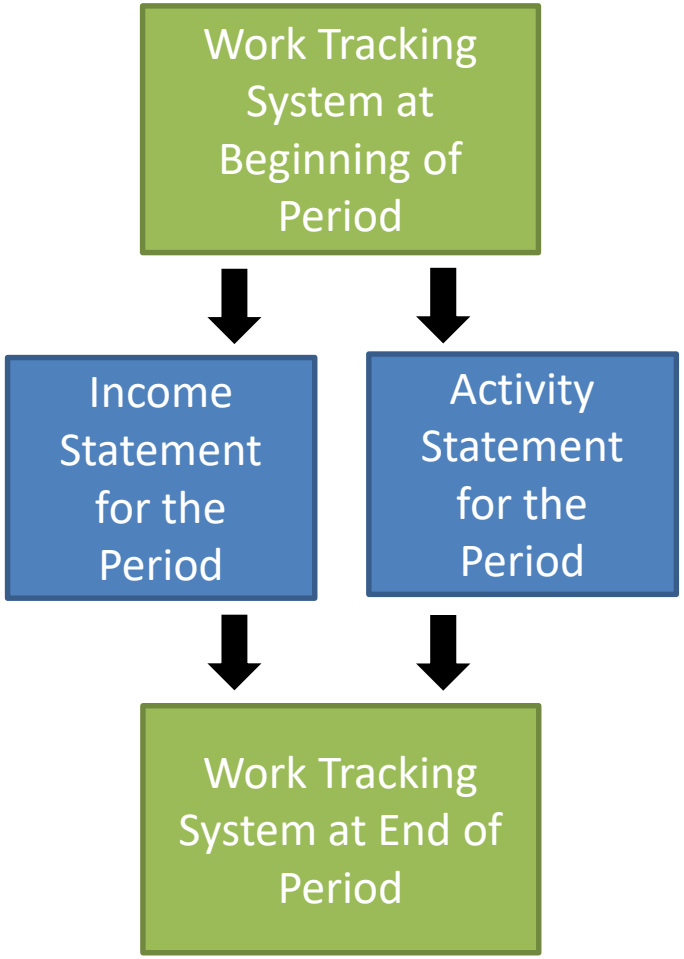
Real Time Reporting Model

- Third parties have little to no assurance from outsiders that transactions and assets/liabilities are valid without attest work – which can be automated
  - Orders can be tracked back to IP address, validated against IPWHOIS databases (is the IP registered to a business with a fixed IP address, a public VPN services, or a residence)
  - Customer credit cards used to pay for services can be validated against credit bureau information and compared to data sources like
  - Merchant account history can be scanned for chargebacks/credits as a percentage of charges as an assessment of validity
- Work tracking system tells you what work is in the sales process, what has been ordered, and what is currently being processed
- Third party customer satisfaction surveys can be used to assess the health of the customer relationships



Real Time Reporting Model

- Third party customer satisfaction survey services can be used to assess the health of the customer relationships
- Income statement and activity statement are simply two different bases of accounting for the activity between the two balance sheets
- In this model, data can be both timely and reliable, but the basis of reporting will necessarily be industry/company specific – so it will require more focus on a smaller number of industries and integration into industry best practices
- **Financial institutions can get assurance on history which gives them a basis for lending based on forward-looking information like orders (retail/wholesale), patient appointments (medical), and automatic renewing subscriptions.**



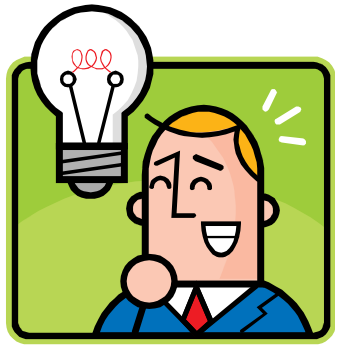
Cooking vs. Business Intelligence

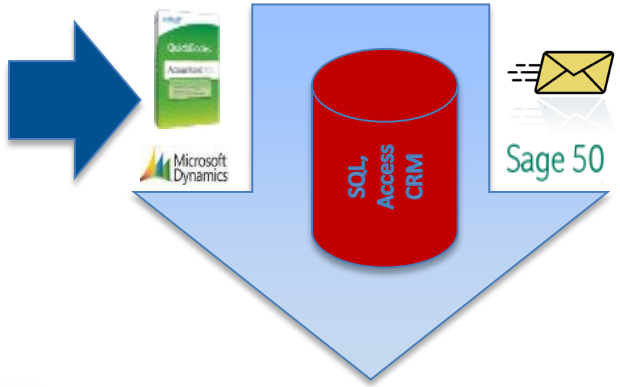
Cooking

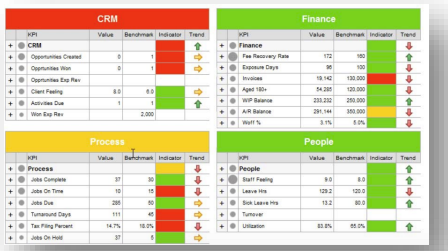
- Decide on a menu
- EXTRACT ingredients from the grocery
- Chop, cook, season and TRANSFORM the food
- LOAD the food into serving dishes
- Serve the food with an attractive garnish (REPORT)
- Enjoy the meal

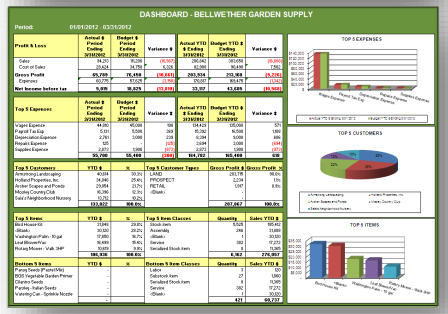
Like cooking, there are steps to creating analysis for real-time use of data analytics

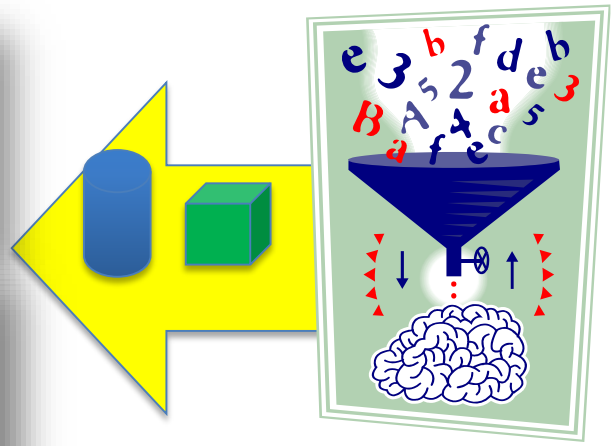
Cooking vs. Business Intelligence









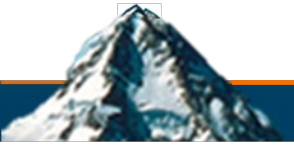


Business Intelligence

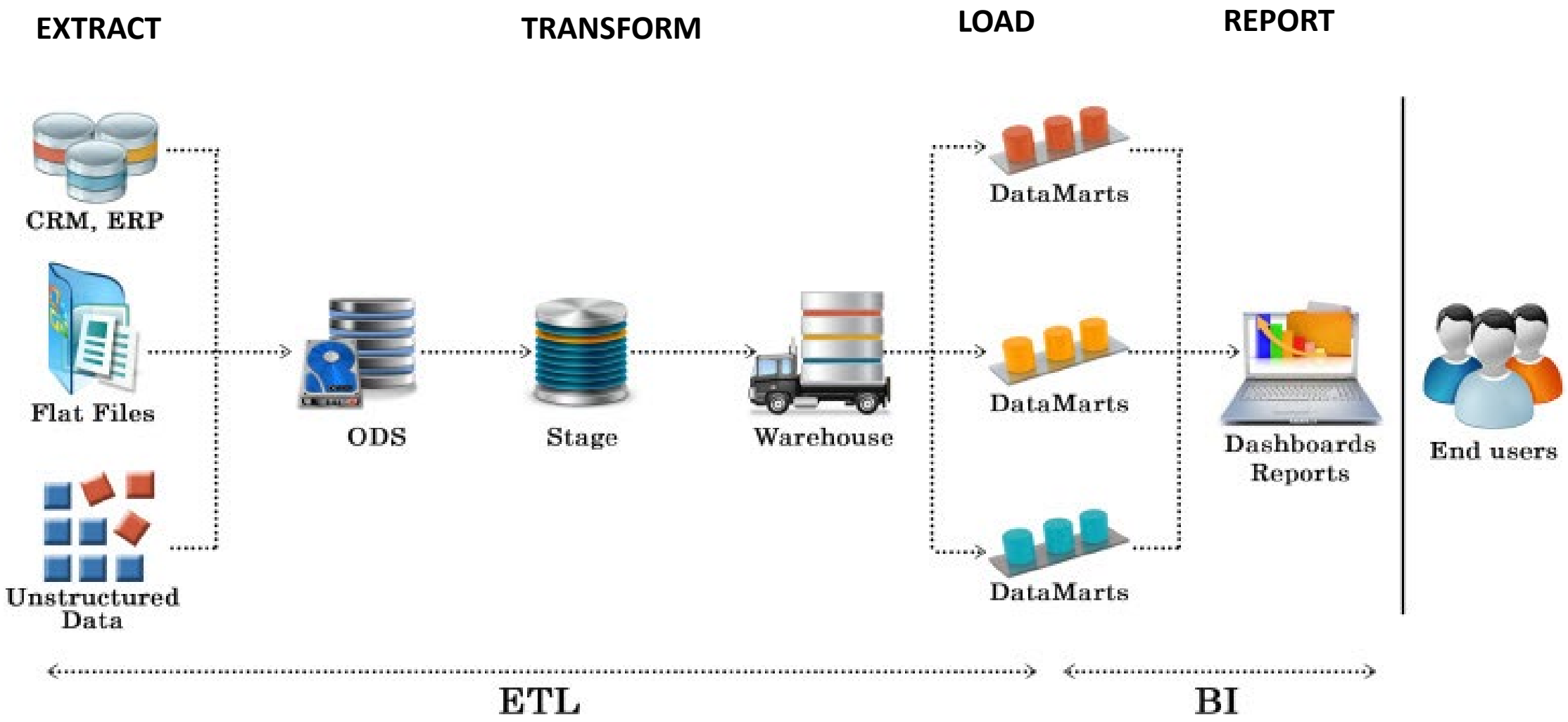
- Decide what you should measure and why it should be measured
- Look for real time tracking data in your existing systems
- Gather, and summarize data (EXTRACT)
- Reformat and reorganize data (TRANSFORM)
- Put the data into a database or OLAP cubes (LOAD)
- Serve up the data in a variety of formats (web, paper, mobile, etc.) (REPORT)

ETLR

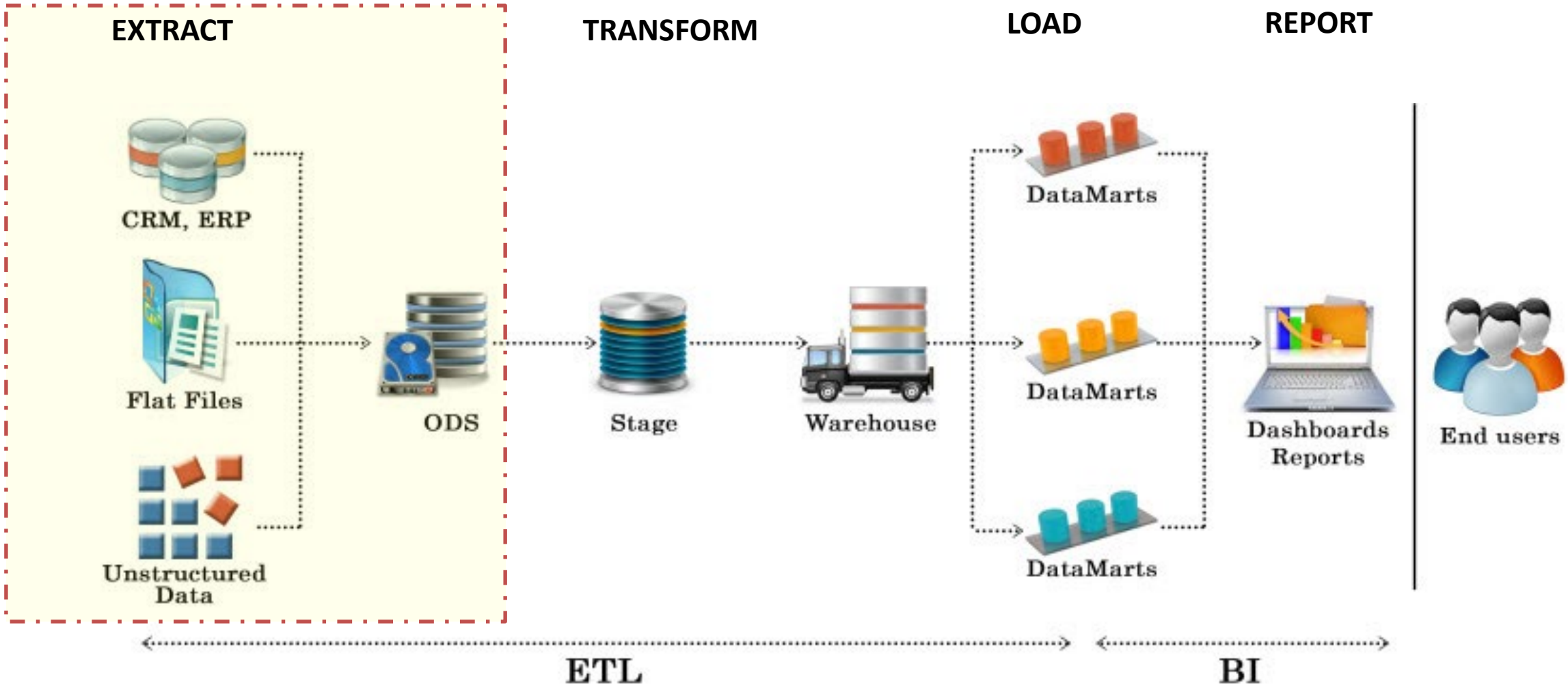
Extract, Transform, Load, Report:
The Core of Data Analysis Systems



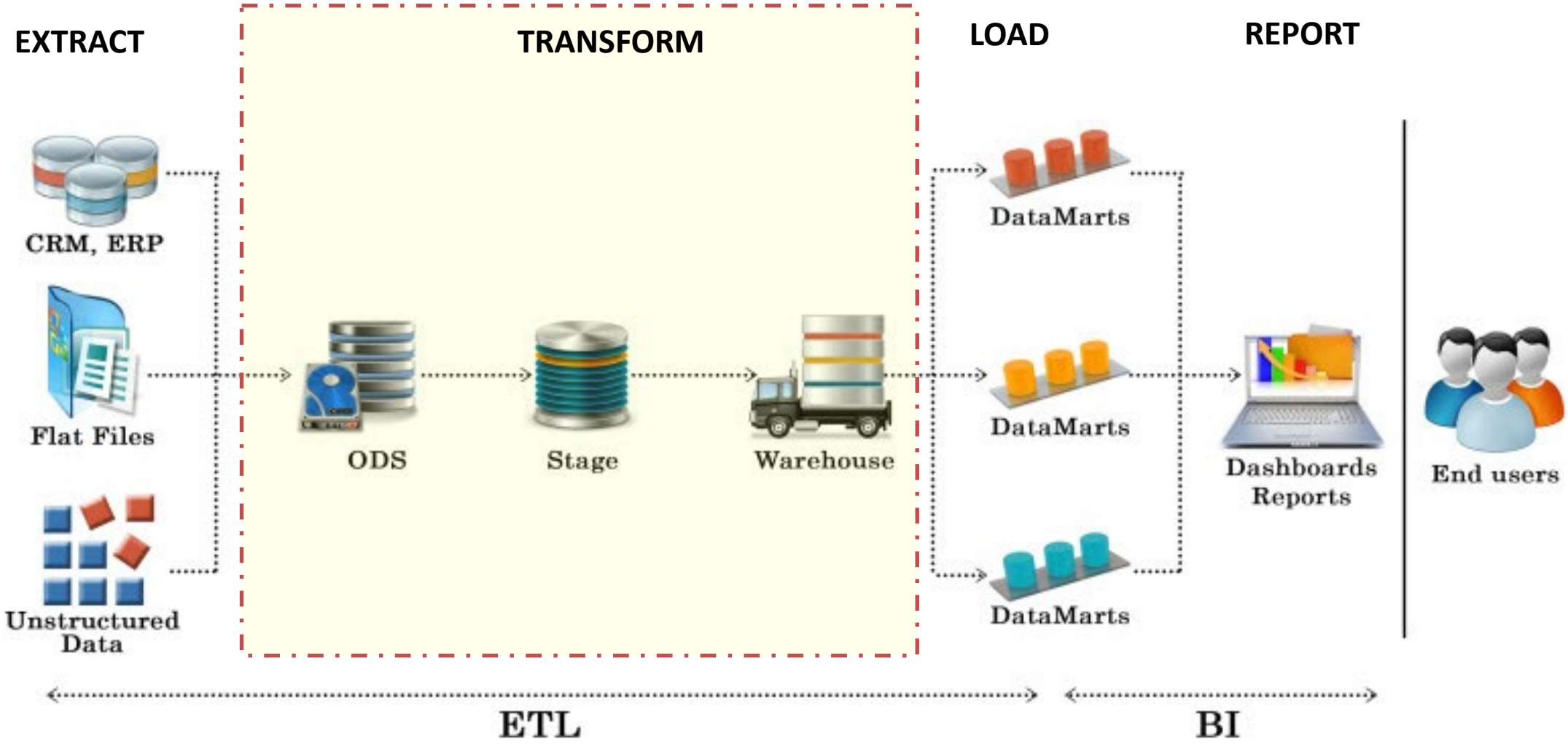
Extract, Transform, Load, Report (Legacy)



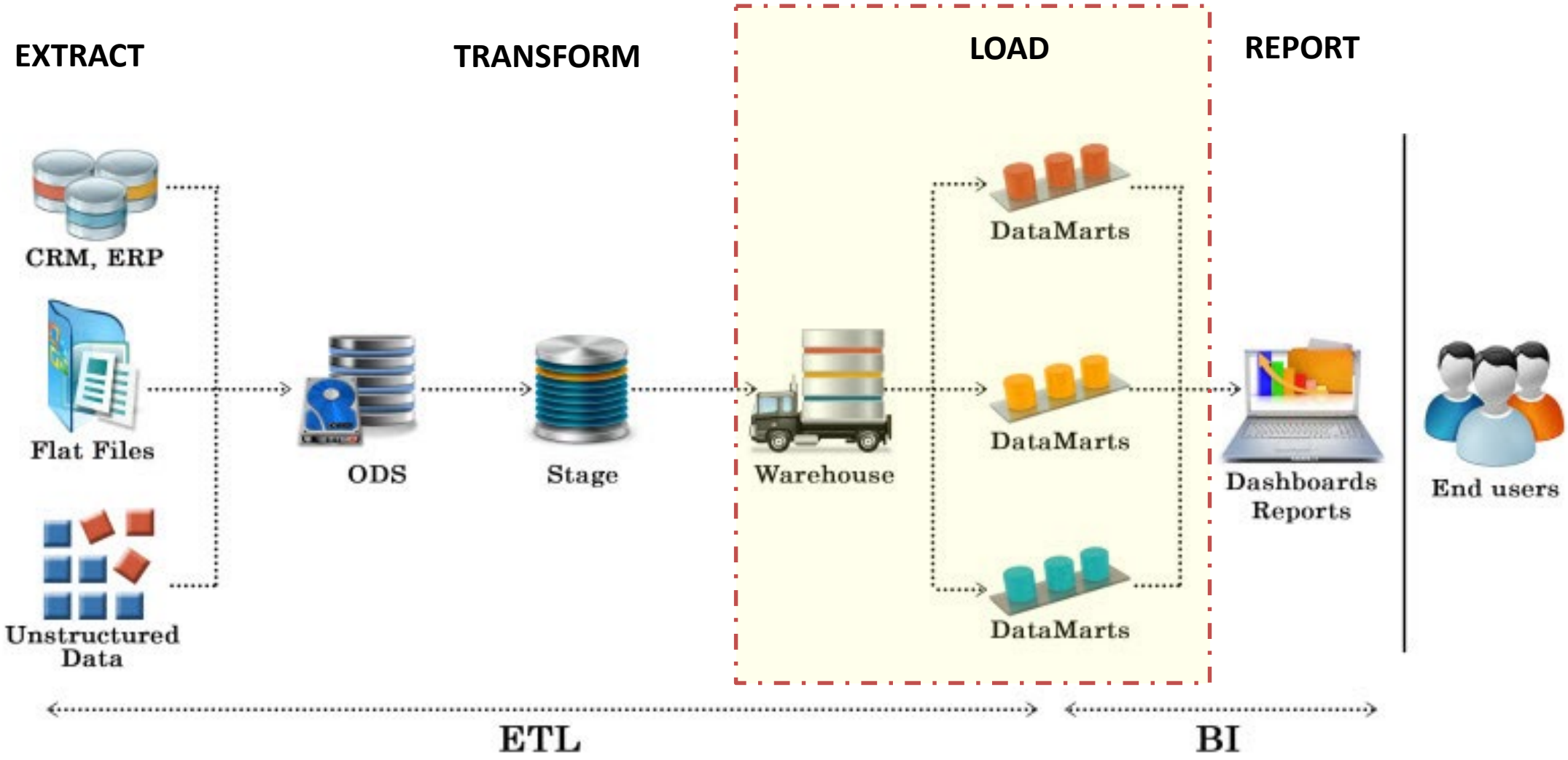
Traditional Extract, Transform, Load Steps



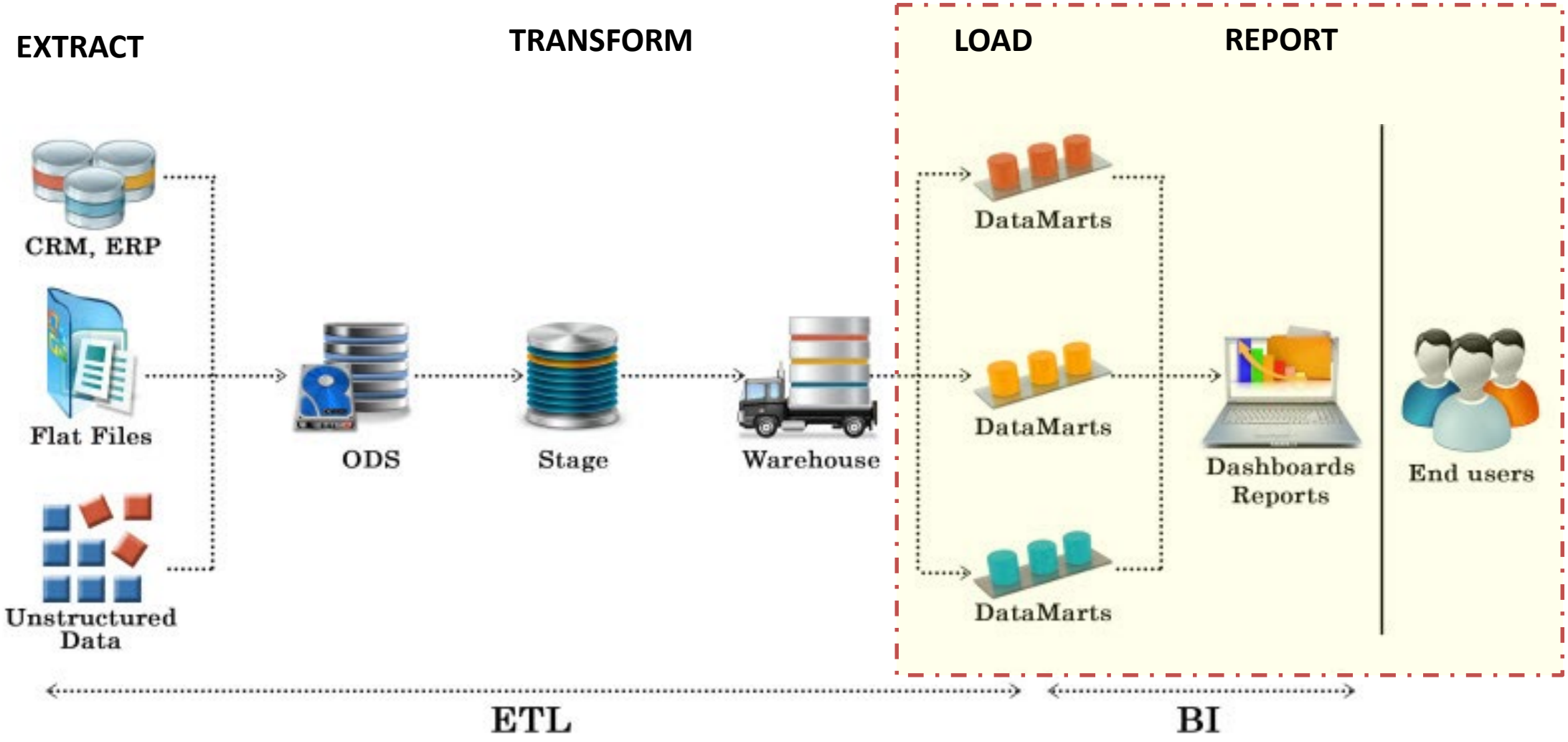
Traditional Extract, Transform, Load Steps



Traditional Extract, Transform, Load Steps



Traditional Extract, Transform, Load Steps



Five Characteristics Of Successful BI

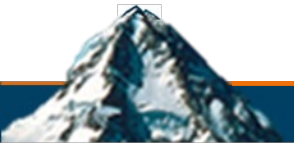
Real-time data

Mission critical info

Graphical and interactive

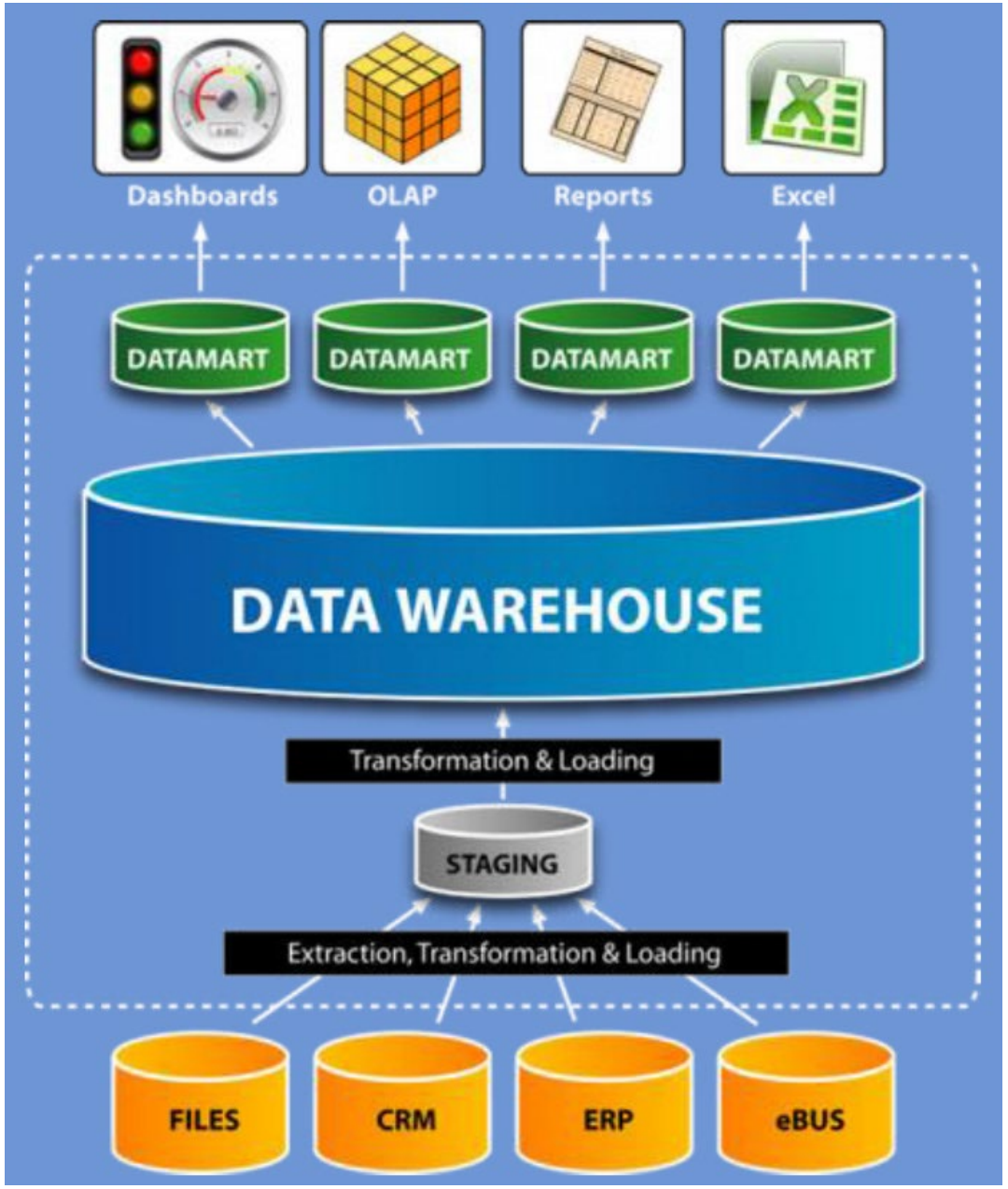
Goal-oriented

Visibility



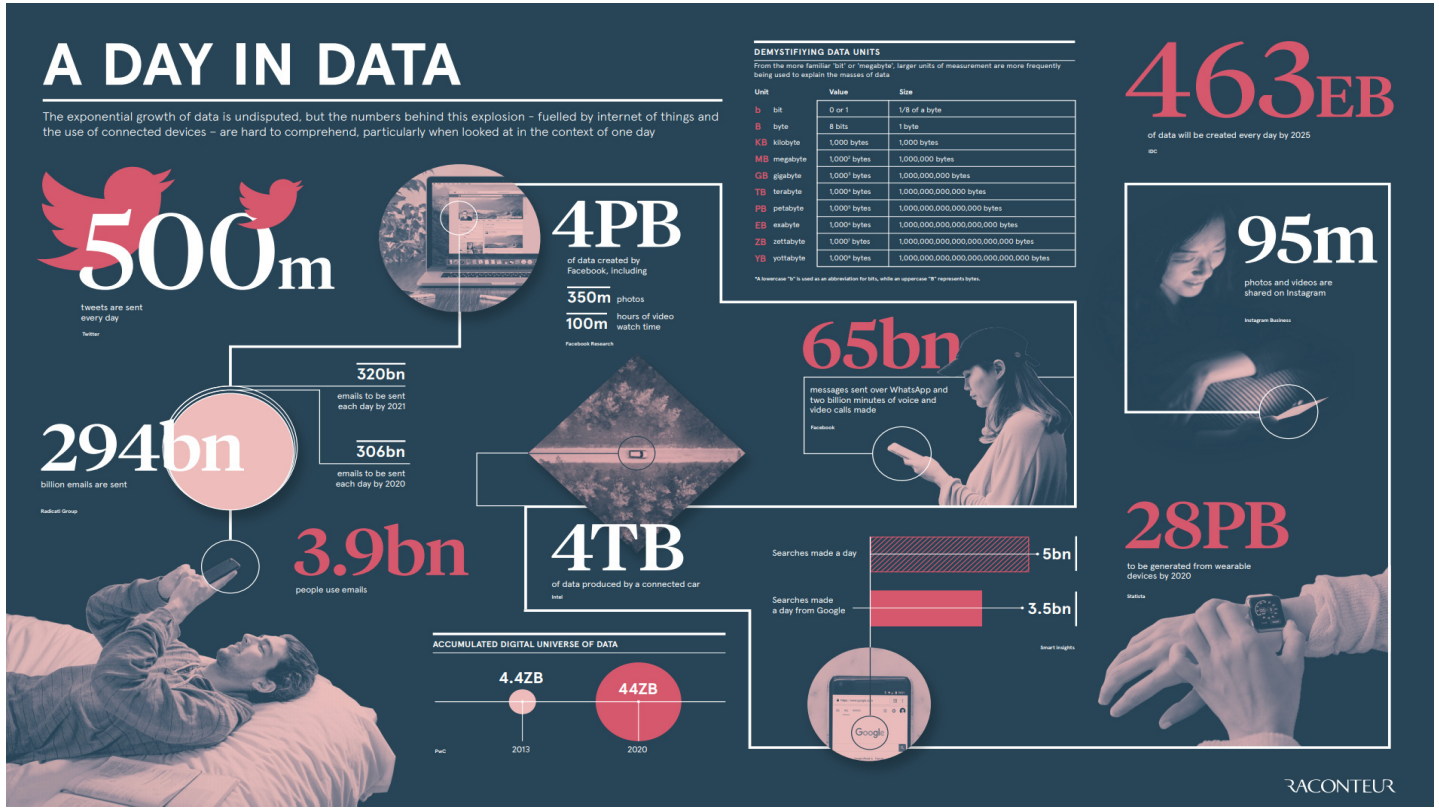
The Importance Of DIYBI

- *Traditional reporting processes are too slow, too complex, and too expensive as shown in the accompanying Gartner graphic*



The Importance Of DIYBI, According To Logi Analytics

- ✓ Over 90% of respondents to a recent survey indicated that it is important for business users to access the information they need without IT
- ✓ Only 22% of business users have access to and use self-service BI tools when needed
- ✓ Eighty-four percent of IT organizations plan to invest in self-service BI in the next 24 months
- ✓ Twenty-four percent of businesses have already purchased self-service BI tools without the approval of their IT staffs, and this trend is growing
- ✓ On average, self-service BI reduces IT requests by 37%
- ✓ Ninety-two percent of business users report that it is “very important” or “somewhat important” to access data and information without asking IT

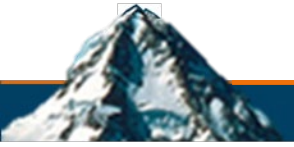


BIG DATA

Big Data

+ Extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions
+ Analysis of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and so on." The Economist

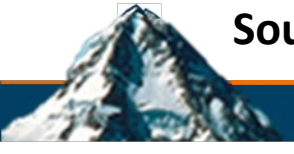
- Challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating and information privacy
- Data sets grow rapidly - in part because they are increasingly gathered by cheap and numerous information-sensing Internet of things devices such as mobile devices, aerial (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers and wireless sensor networks
- The work may require "massively parallel software running on tens, hundreds, or even thousands of servers“ according to Adam Jacobs in an ACM article



Remember Digital Exhaust? Try This…

1. Every day, we send 294 billion emails, generate 4 petabytes of Facebook content, send 500 million tweets, and upload 95 million videos and photos to Instagram
   - One petabyte (PB) is a million gigabytes (GB), or a thousand terabytes (TB)
2. By 2020, it is estimated that wearable devices like the Apple Watch and the FitBit will create 28 petabytes of data – every day
3. A single self-driving car creates four terabytes of data per day
   - 1 TB is 1,000 GB or a million megabytes (MB)
4. The total size of all stored data by 2020 is estimated to be 44 zettabytes (ZB)
   - 1 ZB is a trillion terabytes, a million million GB, or 10^{21} bytes
5. In 2011, Gizmodo reported that the NSA collected 296 TB per day, which was four times as much data as the digital archives of the Library of Congress at that time
   - The largest commercially deployed hard disks hold 16 TB of data, so the entire contents of the Library of Congress in 2011 can be stored on 19 of these supermassive drives

Sources (1-4): Raconteur **–** "A Day in Data" (2019 edition)
Source (5): Gizmodo



A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion – fuelled by internet of things and the use of connected devices – are hard to comprehend, particularly when looked at in the context of one day

500m

tweets are sent every day

Twitter

4PB

of data created by Facebook, including

350m photos

100m hours of video watch time

Facebook Research

294bn

billion emails are sent

Radicati Group

320bn

emails to be sent each day by 2021

306bn

emails to be sent each day by 2020

3.9bn

people use emails

4TB

of data produced by a connected car

Intel

DEMYSTIFIYING DATA UNITS

From the more familiar 'bit' or 'megabyte', larger units of measurement are more frequently being used to explain the masses of data

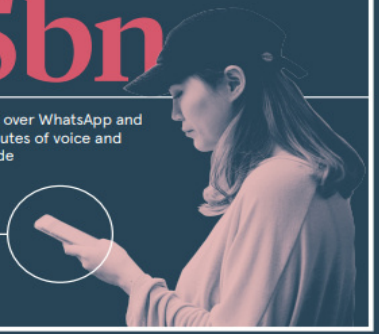
Unit		Value	Size
b	bit	0 or 1	1/8 of a byte
B	byte	8 bits	1 byte
KB	kilobyte	1,000 bytes	1,000 bytes
MB	megabyte	$1,000^2$ bytes	1,000,000 bytes
GB	gigabyte	$1,000^3$ bytes	1,000,000,000 bytes
TB	terabyte	$1,000^4$ bytes	1,000,000,000,000 bytes
PB	petabyte	$1,000^5$ bytes	1,000,000,000,000,000 bytes
EB	exabyte	$1,000^6$ bytes	1,000,000,000,000,000,000 bytes
ZB	zettabyte	$1,000^7$ bytes	1,000,000,000,000,000,000,000 bytes
YB	yottabyte	$1,000^8$ bytes	1,000,000,000,000,000,000,000,000 bytes

*A lowercase "b" is used as an abbreviation for bits, while an uppercase "B" represents bytes.

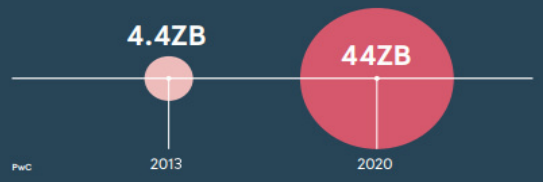
65bn

messages sent over WhatsApp and two billion minutes of voice and video calls made

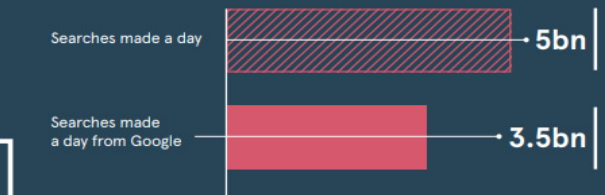
Facebook



ACCUMULATED DIGITAL UNIVERSE OF DATA



PwC



Smart Insights

463EB

of data will be created every day by 2025

IDC

95m

photos and videos are shared on Instagram

Instagram Business

28PB

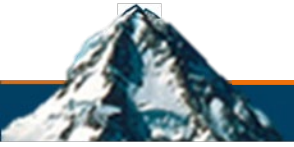
to be generated from wearable devices by 2020

Statista

RACONTEUR

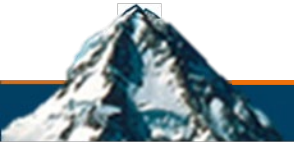
Big Data Facts

- The term has been in use since the 1990s, with some giving credit to computer scientist John Mashey, formerly of Bell Labs, for coining or at least making it popular
- Big Data philosophy encompasses unstructured, semi-structured and structured data, however the main focus is on large, unstructured data sets
- Big data very often means "dirty data" or "big bad data" and the fraction of data inaccuracies increases with data volume growth
- Five dimensions to big data are: Volume, Variety, Velocity and the recently added Veracity and Value



Big Data Characteristics

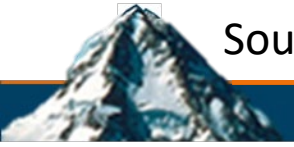
- **Volume**: big data doesn't sample; it just observes and tracks what happens
- **Variety**: big data draws from text, images, audio, video; plus it completes missing pieces through data fusion
- **Velocity**: big data is often available in real-time
- **Veracity**: the data quality of captured data can vary greatly, affecting the accurate analysis
- **Value**: Technology and Analytical Methods for big data transformation as well as usefulness



Big Data Components And Ecosystem

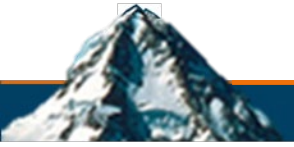
- Main Components
  - Techniques for analyzing data, such as A/B testing, machine learning and natural language processing
  - Big data technologies, like business intelligence, cloud computing and databases
  - Visualization, such as charts, graphs and other displays of the data
- Machine learning: big data often doesn't ask why and simply detects patterns
- Digital footprint: big data is often a cost-free byproduct of digital interaction, “digital exhaust” from IOT and other computing efforts

Source: 2011 McKinsey Global Institute



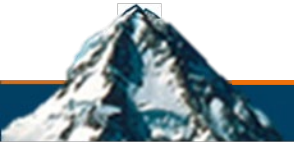
Big Data Vs. Business Intelligence

- Business Intelligence uses descriptive statistics with data with high information density to measure things, detect trends, etc.
- Big data uses inductive statistics (statistical inference and descriptive statistics) and concepts from nonlinear system identification to infer laws (regressions, nonlinear relationships, and causal effects) from large sets of data with low information density to reveal relationships and dependencies, or to perform predictions of outcomes and behaviors
- Big data is generally considered to be data sets larger than 16 TB (16,000 GB), and database operations may need to be performed by a cluster of servers instead of a single server



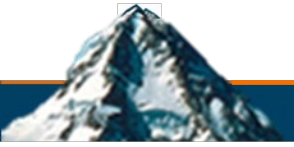
Big Data Outputs

- Big data continuously evolves according to Kryder's Law, the assumption that disk drive density, also known as areal density, will double every thirteen months according to Mark Kryder, former senior vice president of research and CTO of Seagate
- Teradata Corporation
  - In 1984 marketed the parallel processing DBC 1012 system
  - Installed the first petabyte class RDBMS based system in 2007
  - As of 2017, there are a few dozen petabyte class Teradata relational databases installed, the largest of which exceeds 50 PB
- Multidimensional big data can also be represented as tensors, that is objects that describe linear relations between geometric vectors & scalars



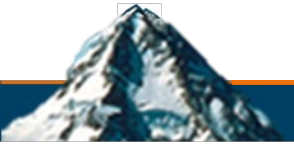
Big Data Approaches

- In 2000, Seisint Inc. (now LexisNexis Group) developed a C++-based distributed file-sharing framework for data storage and query
- In 2004, Google published a paper on a similar architecture called MapReduce that uses a parallel processing model.
  - With MapReduce, queries are split and distributed across parallel nodes and processed in parallel (the Map step).
  - The results are then gathered and delivered (the Reduce step)
- The MapReduce framework was adopted by an open-source project named Hadoop in 2006
- In 2003, Google published a white paper on the Google File System, the file system used by Google to store and archive the world's information
  - GFS uses the clusters of servers and storage devices in a data center as a massive multi-tenant, multi-server software defined computing environment
- Apache Spark was developed in 2012 in response to limitations in the MapReduce approach



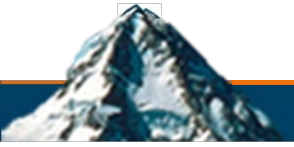
Big Data Hadoop

- Hadoop Common – contains libraries and utilities needed by other Hadoop modules
- Hadoop Distributed File System (HDFS) – a distributed file-system that stores data on commodity machines, providing very high aggregate bandwidth across the cluster
- Hadoop YARN – a platform responsible for managing computing resources in clusters and using them for scheduling users' applications
- Hadoop MapReduce – an implementation of the MapReduce programming model for large-scale data processing



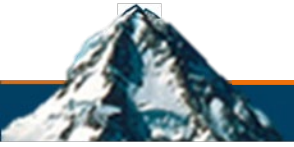
Which Leads To Data Lakes

- The data lake allows an organization to shift its focus from centralized control to a shared model to respond to the changing dynamics of information management
- A data lake is a method of storing data within a system or repository, in its natural format, that facilitates the collocation of data in various schemata and structural forms, usually object blobs or files. The concept is fuzzy and arbitrary
- A data lake refers to any tool or data management practice that does not fit into the traditional data warehouse architecture

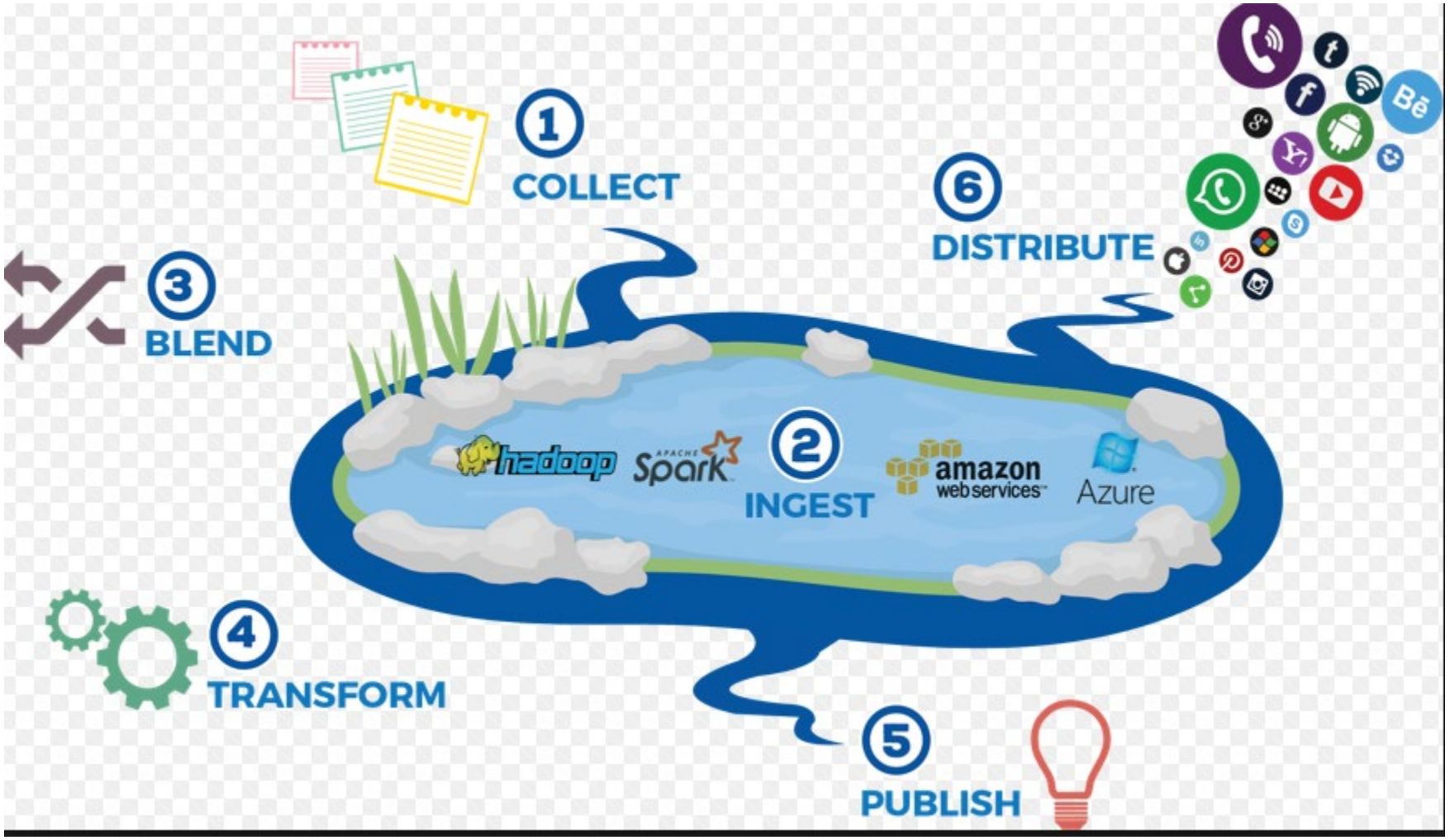


Big Unstructured Data

- The idea of a data lake is to have a single store of all data in the enterprise ranging from raw data (which implies exact copy of source system data) to transformed data which is used for various tasks including reporting, visualization, analytics and machine learning
- The data lake includes structured data from relational databases (rows and columns), semi-structured data (CSV, logs, XML, JSON), unstructured data (emails, documents, PDFs) and even binary data (images, audio, video) thus creating a centralized data store accommodating all forms of data

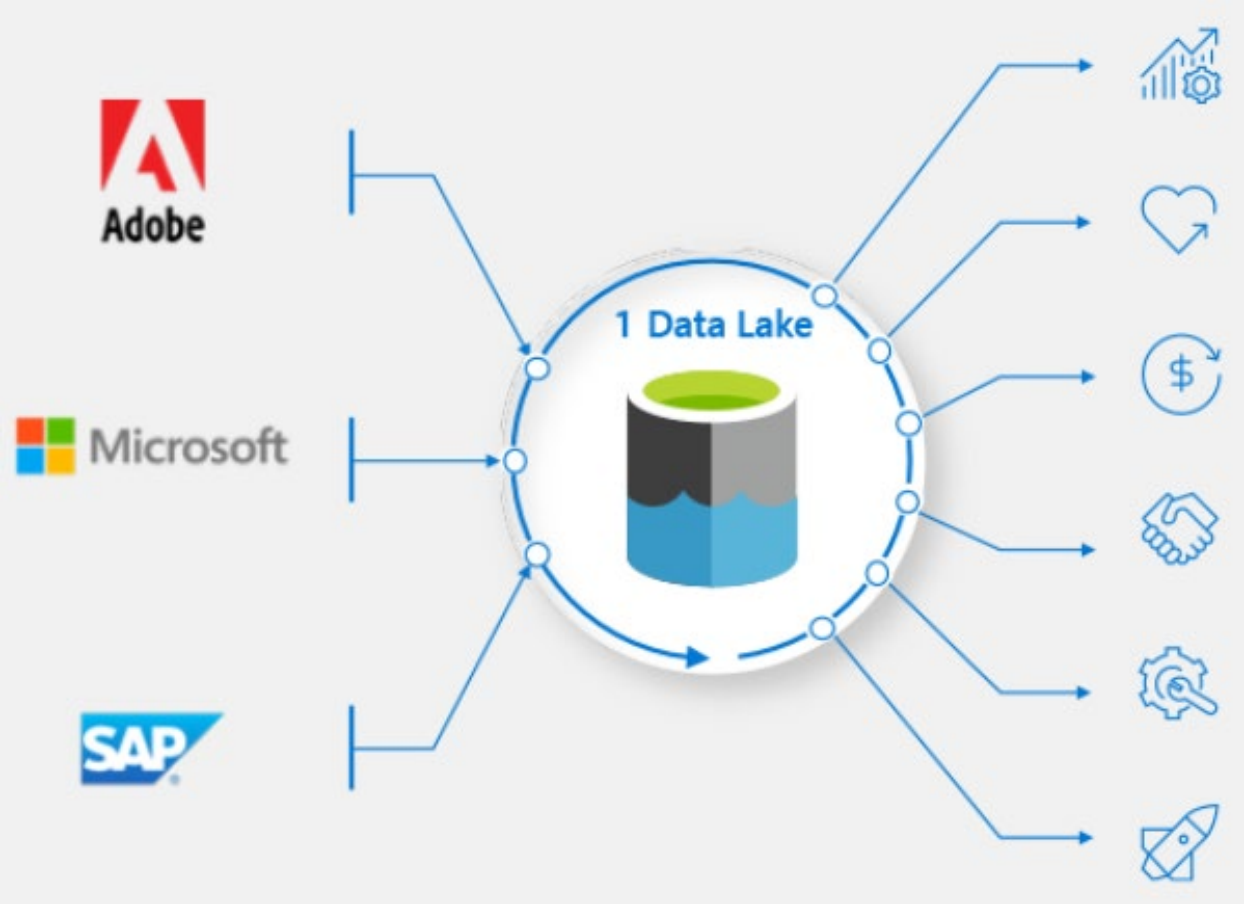


How Does Big Data Fit Together?



Open Data Initiative – Adobe, Microsoft, SAP

- Goal is making apps and platforms more interoperable through a common data model
- Data from all three platforms can be stored in a customer-chosen unified data lake
- This will facilitate integrating data from the entire organization with a common set of tools and breaking down the silos between data sets
- Will likely facilitate simpler continuous auditing of transactions in Fortune 500 companies



The Common Data Model

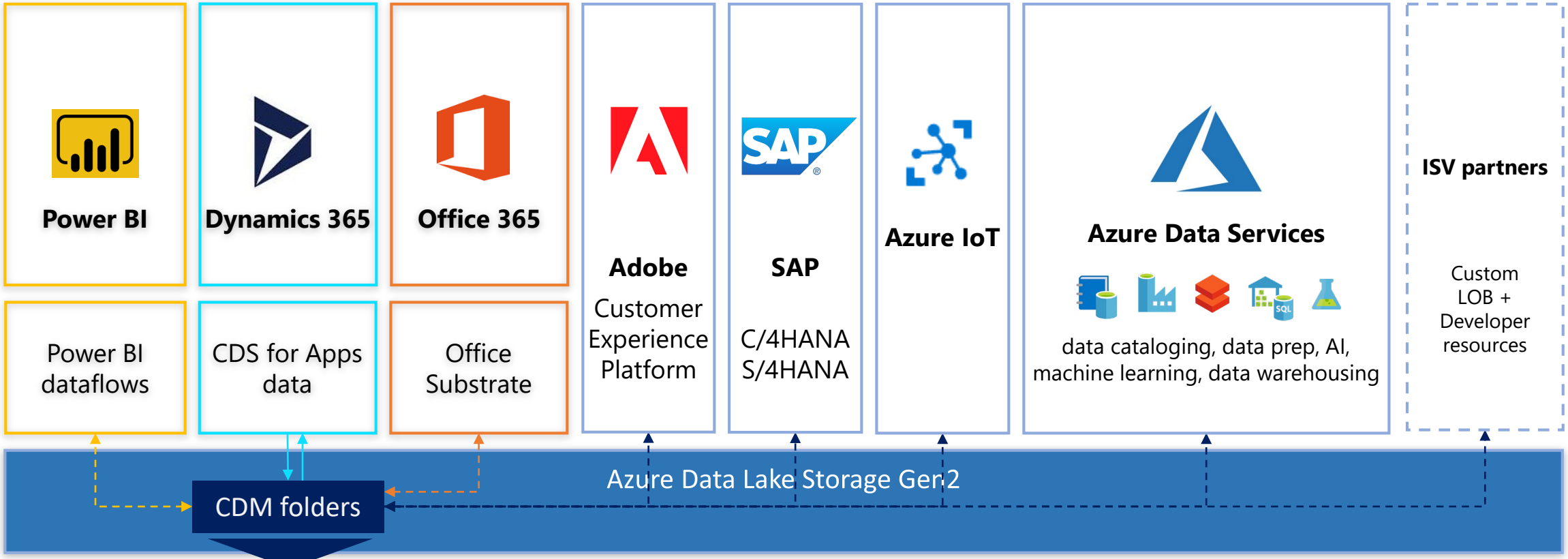
- Started from Dynamics 365
- Open-sourced in GitHub with over 250 entities
  - Ecosystem of partners including industry-specific releases
- Rich metadata system including relationships, hierarchies, traits and more
- Multiple systems and platforms implement CDM today

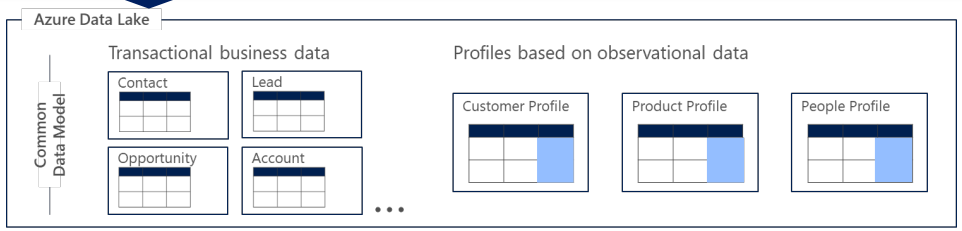
Source: Microsoft Ignite 2018 Presentation BRK2052

Enabling Low Friction Collaboration Among Data + AI Professionals

Business analysts ⟷ Data scientists, Data engineers

Low/no code — Low to high code

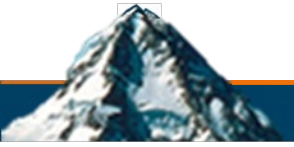




Source: Microsoft Ignite 2018 Presentation BRK2052

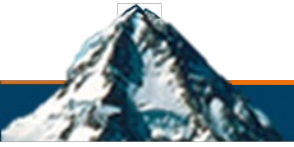
Fields Related To Big Data

- **Predictive analytics** – the result of data analytics devising complex models and algorithms, allowing data scientists, engineers and analysts to produce reliable, repeatable results
- **User behavior analytics (UBA)** – cybersecurity process about detection of insider threats, targeted attacks & financial fraud
- **Distributed Computing** – Massively parallel-processing (MPP) databases, search-based applications, data mining, distributed file systems, distributed databases, cloud and HPC-based infrastructure (applications, storage and computing resources)



What Does This Mean To Accountants?

- Working examples
  - The video, audio, and textual information made available via Big Data can provide for improved managerial accounting, financial accounting, and financial reporting practices
  - In managerial accounting, Big Data will contribute to the development and evolution of effective management control systems and budgeting processes
  - In financial accounting, Big Data will improve the quality and relevance of accounting information, thereby enhancing transparency and stakeholder decision making
  - In reporting, Big Data can assist with the creation and refinement of accounting standards, helping to ensure that the accounting profession will continue to provide useful information as the dynamic, real-time, global economy evolves
  - In the press, C-Span is using Amazon's vision system to compile a database of politicians so they can name them quickly when they appear on screen



Is Facial Recognition Just A Big Data Question?

FACIAL RECOGNITION

Face-scanning technology is evolving rapidly and is increasingly employed in high-security facilities such as airports and government offices. Now some stores are even using it to identify returning customers or shoplifters.

1 Finding a face

Systems extract patterns from an image and compare them to a model of a face. When patterns start to resemble the model, the system signals it has homed in on a face.

Personal devices

Smartphones use face recognition for apps and security, such as unlocking the phone.

Checkpoint cameras

Faces are recorded at customs and security checkpoints, and the images are archived.

Other cameras

Laptop, video, and thermal cameras used in some security systems can capture face images.

CCTV cameras

Systems can isolate and track individuals by face, gait, and clothing color and pattern.

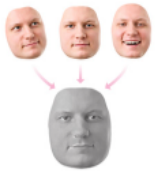
Face imagery captured when a person poses for the camera, such as at security checkpoints, is easier to analyze; imagery captured from CCTV cameras may require advanced methods and detailed analysis.

2 Creating a face template

Algorithms build more informative and accurate digital representations—called face templates—using thermal, geometric, and other data, either separately or combined.

Geometric

Spatial relationships between facial features, such as the center of the eyes and tip of the nose, are calculated.



Photometric

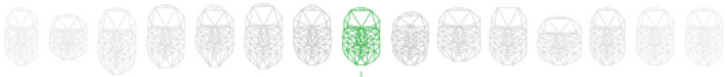
Algorithms can build a face even if an image is obscured by poor lighting or distorted by odd angles or expressions.

Skin-texture analysis

Pores, wrinkles, and spots are mapped and analyzed; the technology can even differentiate between twins.

Thermal sensors

This technology can provide further information despite obstacles such as heavy makeup or disguises.



3 Identifying a face

Once a face template is created, it can be compared with databases (such as for mug shots) to verify a person's identity or recognize an individual in CCTV footage.

Identity confirmed

Source: They Are Watching You—and Everything Else on the Planet, National Geographic, Feb 2018, pp. 30-65

Solving The 2005 London Terrorist Bombings…

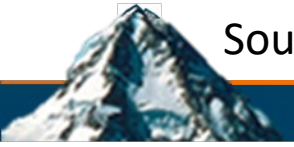
JASON TREAT AND RYAN T. WILLIAMS, NGM STAFF. SOURCES: BIG BROTHER WATCH; ISLINGTON COUNCIL; ORDNANCE SURVEY, U.K.

Source: They Are Watching You—and Everything Else on the Planet, National Geographic, Feb 2018, pp. 30-65

For Safety, Numbers Say We Are OK With This

- 54 percent of Americans plan to use face recognition to protect their personal data or already own a device that uses face recognition
- Nearly two thirds (64 percent) of Americans think security personnel guarding airports, concerts, sporting events and other public areas should be allowed to use face recognition to help recognize terrorists and prevent crime
- 77 percent of Americans think that security guarding airports and tourist attractions are not likely to remember the names and faces of potential terrorists on a watch list without face recognition
- “In the case of one of our retailers, we’ve seen criminal acts decrease by 91 percent” according to Peter Trepp, the CEO of FaceFirst

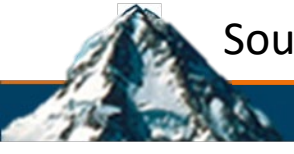
Source: Security InfoWatch “Facial recognition crosses both security, consumer applications” by Steve Lasky



Surveillance And Predictive Policing

- Automatic number/license plate readers (ANPRS/ALPRS) can be used to track all automobile movements – 7Bln datapoints
- Body cameras effectively do the same…would you notice if a policeman took pictures of every parked car and pedestrian?
- Logical extraction of phones: text messages, emails, pictures and IM, including those deleted…tablets for digital triage
- PredPol, CompStat and others use algorithms w/o humans
- Andrew Ferguson suggests five questions: risks, accurate data, community relations, accountability, autonomy

Source: The Economist, Technology Quarterly, June 2-8, 2018, pp. 3-12



Xinjiang Totalitarian Police State

- Uighurs (46%) are quite distinct from the Han Chinese (40%)
- Hundreds of “re-education camps” with 800,000+ people
- Police stations every 300 meters which require photographs, fingerprints, iris recognition and DNA samples
- 10,000+ teams go house to house to compile dossiers: 15-55 (military age), Ughur, unemployed, untrustworthy=camp
- 8-10 cameras every 100-200 meters, with facial recognition
- “Becoming kin” so police can live in homes for periods of time

Source: The Economist, Apartheid with Chinese characteristics, June 2-8, 2018, pp. 19-22

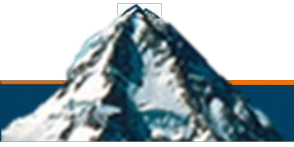
Facial Technology

- Face2Gene from FDNA in Boston – app can diagnose 2,500 diseases
- Minerva & Me – Christopher Nellaker of Univ. of Oxford – Healthy and diseased people can upload pictures for research
- Smartphone app diagnoses children with more than 90% accuracy from Maximillian Muenke of National Human Genome Research in Bethesda and Mariuis Linguraru of the Children's National Health System in Washington, D.C.
- Researchers produce images of people's faces from their genomes from Human Longevity, San Diego, Dr. Craig Venter
- Could match 8 out of 10 real photographs using unknown Genomes. Police could see faces based on DNA/genetic evidence.
- Can also predict appearance base on given age
- 23andMe public info enough to do this
- Privacy concern

Source: The Economist (September 9th-15th, 2017, pp. 73-75)

Big Data Applications

- Banking
- Education
- Government/Public Sector
- Healthcare
- Information Technology
- Insurance
- International development
- Internet of Things (IOT)
- Manufacturing
- Media
- Science
- Sports
- A to Z



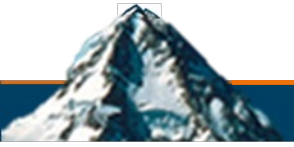
Technology Decision Grid – Big Data

KEY INFORMATION	TECHNOLOGY: Big Data
Why is the new technology better?	Analyze extremely large data sets to reveal patterns, trends, and associations, especially relating to human behavior and interactions
How can you do this today?	Amazon, Cloudera, Dell, HP, IBM, MapR, Microsoft, Oracle, SAP, and Software AG
Risks	Expensive processing on bad data can lead to incorrect strategic conclusions
Where/when to use	To find trends in large amounts of data
How much?	Can be thousands to start, or free on open source
When expected in mainstream	Three to five years
Displaced technology or service	Data Warehouse
Other resources	Accounting Today, CPAPA

Machine Learning

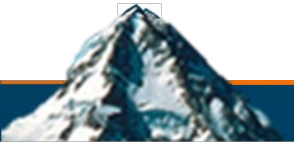
+ Strategic technology, of the six leading companies, four are Chinese (other two are Google and Facebook)
+ Requires lots of data and input to work effectively
+ GPU's or specialized chips (Google's TPU3, NVIDIA or MSFT FPGA's)

- Threatens many knowledge worker jobs in the long run
- Leads to a natural concentration of intellectual/computing skills in a small number of very large firms
- Is Big Data the new Big Brother?



Machine Learning Facts

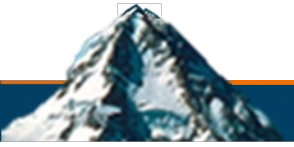
- Term coined by Arthur Samuel of IBM in 1959
- In the 1990's shifted from AI to practical, solvable problems
- Algorithms that can learn from and make predictions on data – typically 1 million records are needed today
- Two types
  - Supervised learning – example inputs and desired outputs
    - Semi-supervised – incomplete training with some target outputs missing
    - Active – label limited instances, often by users
    - Reinforcement – training data is dynamic like playing a game or driving
  - Unsupervised learning – no labels discovering hidden patterns



Machine Learning Outputs

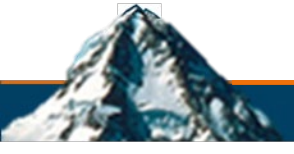
(Note The Similarity To Data Science Algorithms)

- Classification – inputs divided into two or more classes (spam)
- Regression – outputs are continuous rather than discrete
- Clustering – divided into groups (unsupervised)
- Density estimation – finds the distribution of inputs
- Dimensionality reduction – simplify inputs by mapping them to similar topics



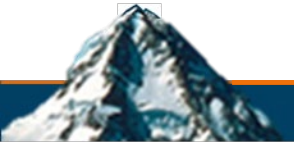
Machine Learning Approaches

- Decision tree learning – go from observations to conclusions
- Association rule learning – relationships between variables
- Artificial neural networks (ANN) – non-linear statistical data models
- Deep learning – GPUs used to find multiple hidden layers using ANN, particularly for computer vision and speech recognition
- Inductive logic programming – rule learning with logic programs
- Support vector machines – SVMs for classification and regression
- Clustering – assignment of a set of observations into subsets
- Bayesian networks – belief network that performs inference

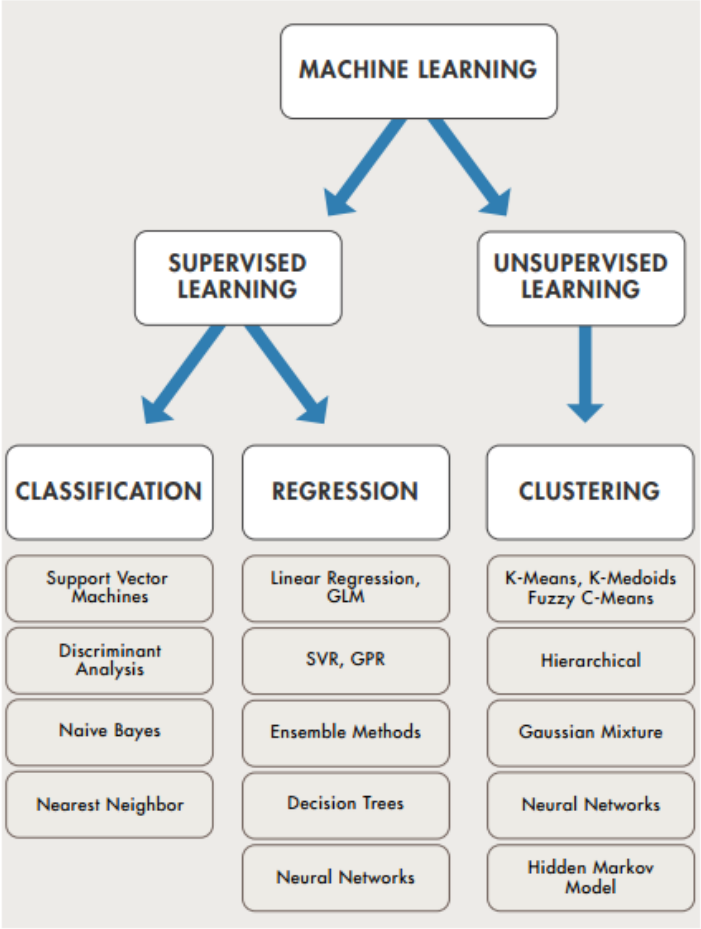


Machine Learning Approaches

- Reinforcement learning – agents taking actions in an environment to maximize rewards
- Representation learning – also called “feature learning” allows a system to automatically detect features or classify from raw data
- Similarity and metric learning – pairs of similar/dissimilar objects
- Sparse dictionary learning – linear combination of basis functions
- Genetic algorithms – search heuristic that mimics natural selection
- Rule-based machine learning – identifies, learns or evolves rules to store, manipulate or apply knowledge
- Learning classifier system (LCS) – discovery and learning component to generate a set of context dependent rules



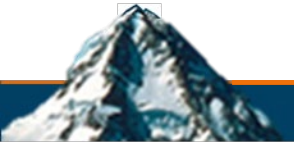
Machine Learning Approaches From DSC



- **Top DSC Resources**
- Article: Difference between Machine Learning, Data Science, AI, Deep Learnin...
- Article: What is Data Science? 24 Fundamental Articles Answering This Question
- Article: Hitchhiker's Guide to Data Science, Machine Learning, R, Python
- Tutorial: Data Science Cheat Sheet
- Tutorial: How to Become a Data Scientist - On Your Own
- Categories: Data Science - Machine Learning - AI - IoT - Deep Learning
- Tools: Hadoop - DataViZ - Python - R - SQL - Excel
- Techniques: Clustering - Regression - SVM - Neural Nets - Ensembles - Decision Trees
- Links: Cheat Sheets - Books - Events - Webinars - Tutorials - Training - News - Jobs
- Links: Announcements - Salary Surveys - Data Sets - Certification - RSS Feeds - About Us
- Newsletter: Sign-up - Past Editions - Members-Only Section - Content Search - For Bloggers
- DSC on: Ning - Twitter - LinkedIn - Facebook - GooglePlus

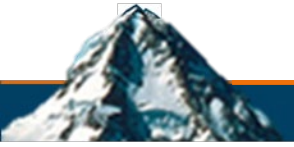
Toronto Declaration On Machine Learning

- The Toronto Declaration: Protecting the **rights to equality** and **non-discrimination** in machine learning systems was launched on May 16, 2018 at RightsCon Toronto May 16-18
- These types of technologies can exacerbate discrimination at scale
- At the time of the launch, the Declaration prepared by Amnesty International and Access Now and it has been endorsed by Human Rights Watch and Wikimedia Foundation
- States have obligations to promote, protect and respect human rights; the private sector, including companies, has a responsibility to respect human rights at all times



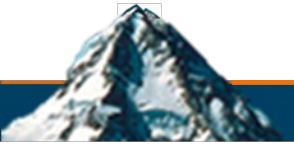
Toronto Declaration On Machine Learning

- “The rights to equality and non-discrimination are only two of the human rights that may be adversely affected through the use of machine learning systems: privacy, data protection, freedom of expression, participation in cultural life, equality before the law, and meaningful access to remedy are just some of the other rights that may be harmed with the misuse of this technology.”
- “Whilst this Declaration is focused on machine learning technologies, many of the norms and principles included are equally applicable to artificial intelligence more widely, as well as to related data systems. The declaration focuses on the rights to equality and non-discrimination. Machine learning, and artificial intelligence more broadly, impact a wider array of human rights, such as the right to privacy, the right to freedom of expression, participation in cultural life, the right to remedy, and the right to life.”
- “This Declaration focuses on the rights to equality and non-discrimination, critical principles underpinning all human rights.”



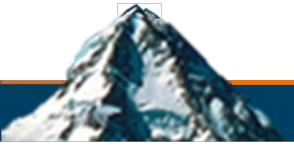
Fields Related To Machine Learning

- Data mining – employ the same methods and overlap significantly, but focuses on discovery rather than prediction
- Predictive analytics – the result of data analytics devising complex models and algorithms, allowing data scientists, engineers and analysts to produce reliable, repeatable results
- Computational statistics – focuses on prediction making through the use of computers and mathematical optimization
- Computational learning theory – branch of computer science



What Does This Mean To Accountants?

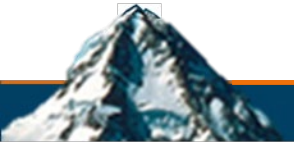
- Working examples
  - Xero – 4 transactions from any user, and they can post it. Banks are using this to establish credit
  - Receipt Bank – classify transactions with 100% accuracy after being classified once
  - Intuit QuickBooks Online bank reconciliation can learn from your transactions as well as that of others
  - CyberSecurity – Darktrace, Email filtering, OCR, detection of network intruders
  - Plantly – Vertical farming using camera recognition and ML



What does this mean to accountants?

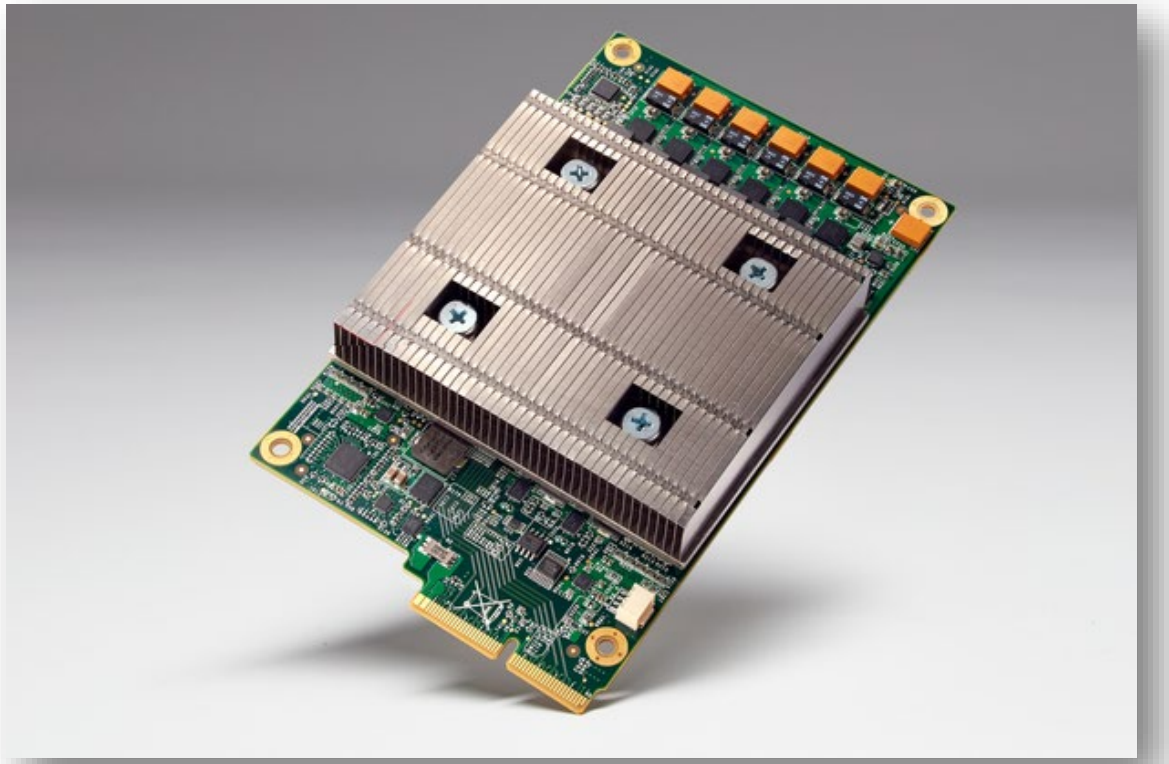
What's coming soon for machine learning in accounting

- Automated mapping of accounts in forthcoming CCH trial balance tool, with percentage confidence for each possible statement line grouping for each G/L account
- Automated classification of incoming client responses to requests and automated filing of PBC requests in engagement software and portal applications
- Automated classification of transactions in cloud accounting software
- Data mining to identify clients with tax planning opportunities
- Underwriting of asset-based small business loans with data from bank-required cloud accounting software – so the bank may do its own analytics on loans collateralized by A/R, inventory, and equipment



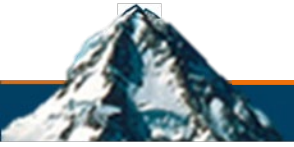
Machine Learning Will Require New Chips And Massive Investment – Cloud TPUs

- Accelerators for key tasks
- Google's Tensor Processing Unit or TPU, in development 2 years
- Replaced in May 2017 as TPU2 with four processors – 45 Tflops, 64 boards in a TPU pod for 11.5 Petaflops
- Order of magnitude better performance per watt for machine learning tasks
- Used on the paid RankBrain search result sorting service and paid Google Voice
- Microsoft has similar strategy using Field Programmable Gate Arrays (FPGA's) in its Azure data centers
- Jeff Dean, Director of Google Brain, thinks 10 million businesses can use ML



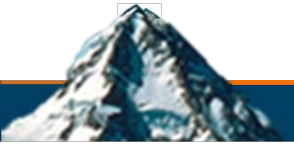
Machine Learning Applications

- Automated theorem proving
- Adaptive websites
- Affective computing
- Bioinformatics
- Brain–machine interfaces
- Cheminformatics
- Classifying DNA sequences
- Computational anatomy
- Computer Networks
- Computer vision, including object recognition
- Detecting credit-card fraud
- Economics
- General game playing
- Information retrieval
- Insurance
- Internet fraud detection
- Linguistics
- Marketing
- Machine learning control
- Machine perception
- Medical diagnosis



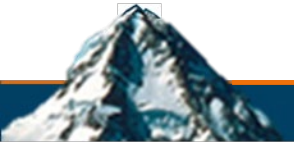
More Machine Learning Applications

- Natural language processing
- Natural language understanding
- Optimization and metaheuristic
- Online advertising
- Recommender systems
- Robot locomotion
- Search engines
- Sentiment analysis (or opinion mining)
- Sequence mining
- Software engineering
- Speech and handwriting recognition
- Financial market analysis
- Structural health monitoring
- Syntactic pattern recognition
- Time series forecasting
- User behavior analytics
- Translation



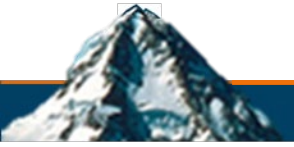
Technology Decision Grid – Machine Learning

KEY INFORMATION	TECHNOLOGY: Machine Learning
Why is the new technology better?	It is a method of statistical data analysis that automates analytical model building
How can you do this today?	AWS ML, Azure FPGA ML, Google Prediction AI, IBM SPSS Modeler, R, SAS Miner
Risks	Insufficient data, conclusion unguided
Where/when to use	When lots of data can answer a specific question
How much?	Can be up to $10K per hour, or free on open source
When expected in mainstream	Arriving now, more within three years
Displaced technology or service	Repetitive or analytical human labor
Other resources	Accounting Today, CPAPA



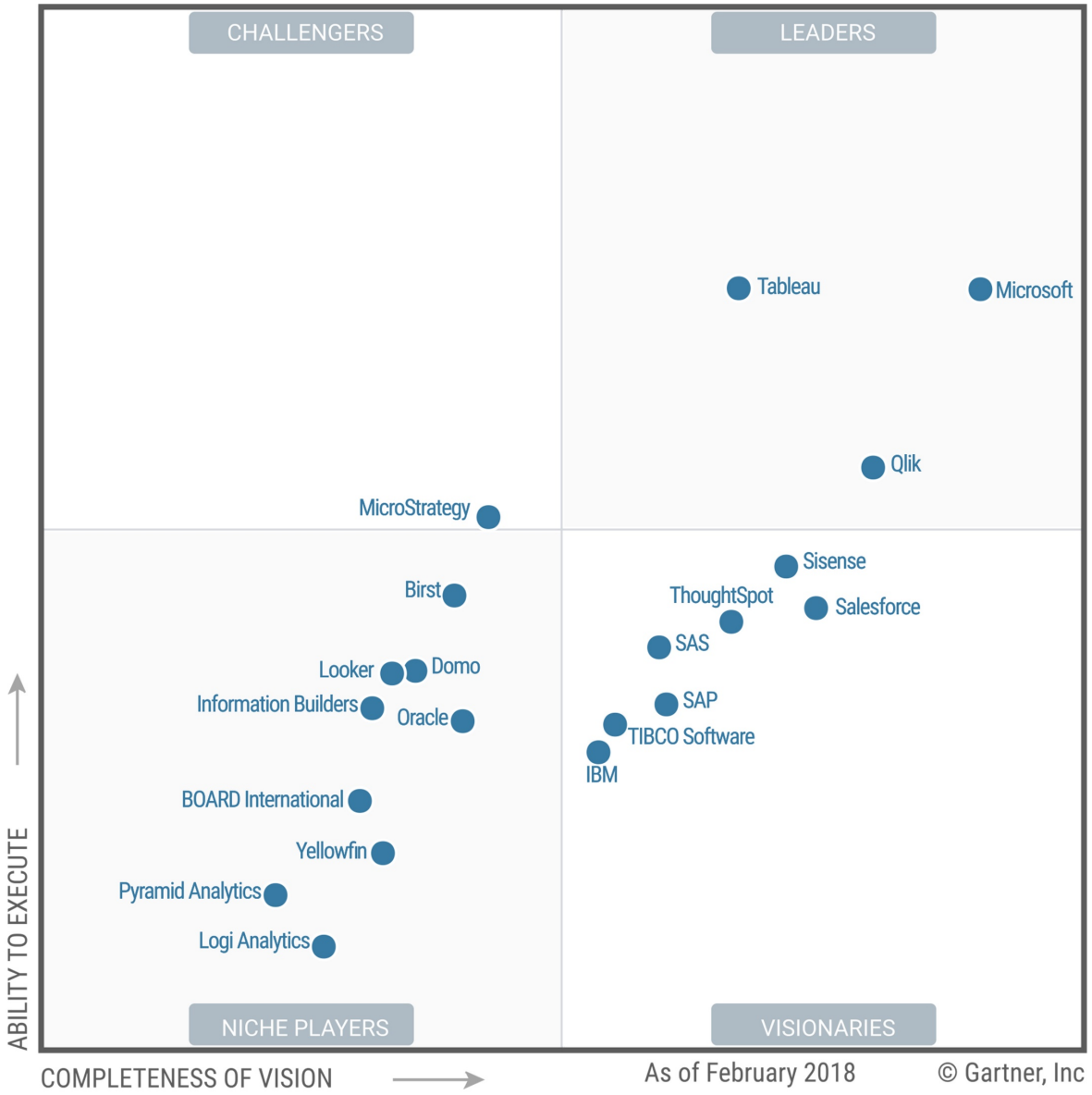
QUESTIONS?

APPENDIX A: BUSINESS INTELLIGENCE PLATFORMS AND TOOLS



Today's Leading BI Players

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner (February 2018)

Today's Leading BI Platforms

Visual-based data discovery, a defining feature of the modern business intelligence (BI) platform, began in around 2004 and has since transformed the market and new buying trends away from IT-centric system of record (SOR) reporting to business-centric agile analytics. ***Modern BI and analytics platforms are characterized by easy-to-use tools that support a full range of analytic workflow capabilities and do not require significant involvement from IT in order to predefine data models upfront as a prerequisite to analysis*** *(including at enterprise-scale deployment).*

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms

Source: Gartner (February 2018)

Excel As A BI Tool

- The ubiquitous presence of Excel means that it is the *de facto* BI tool for many individuals and organizations
- This is not an altogether “bad” thing, as Excel provides numerous features that are needed in a BI platform
    - Open Database Connectivity
    - PivotTables and PivotCharts
    - Tables, sorting, filtering
    - Charting and graphing
    - Integration with Power BI, Data Model
- But limitations, particularly is distributing/sharing dashboards limit Excel as a dashboarding tool

Tableau

- Tableau is one of the leaders in the BI market, according to Gartner, offering a comprehensive set of tools for creating and sharing dashboards across businesses of all sizes
- Further, the Company's recent conversion to a subscription pricing model makes Tableau solutions quite affordable for most companies

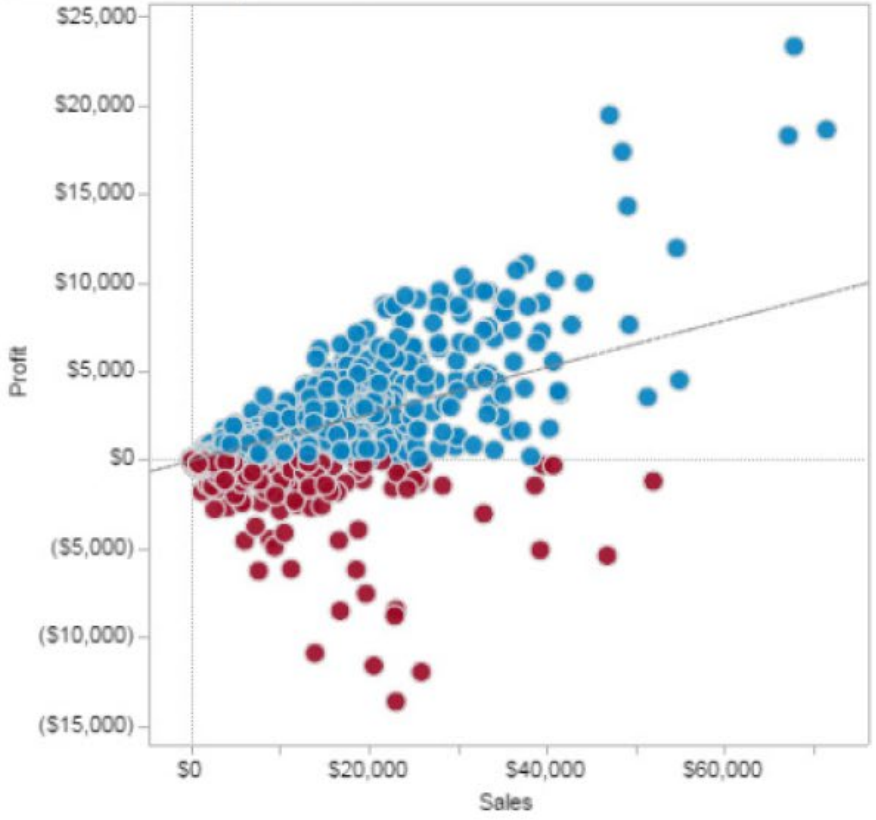
Executive Dashboard | Product Dashboard | Customer Dashboard

Customer Dashboard

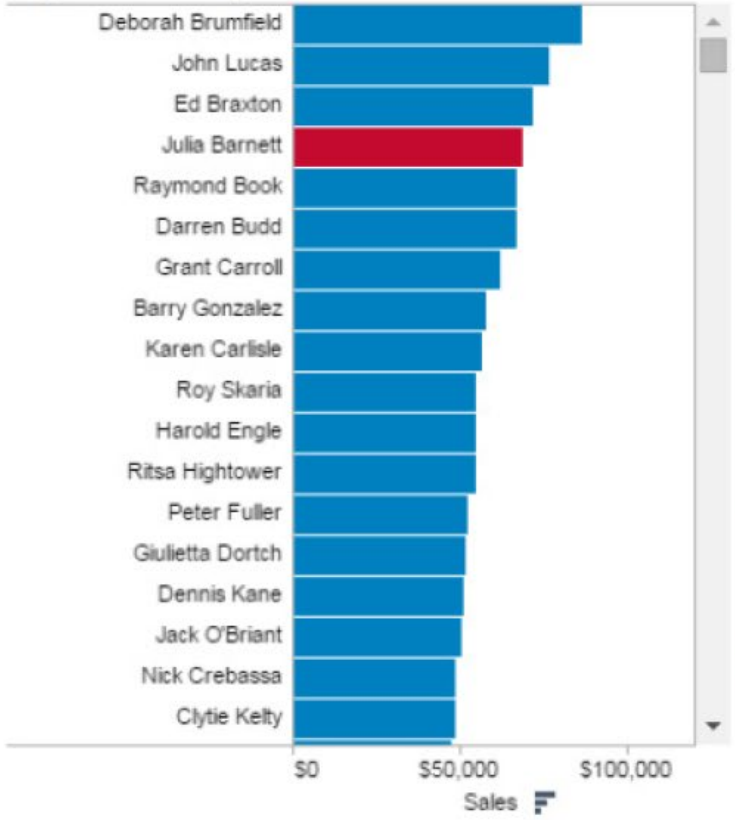
Customer Account Overview

Region	# of Accounts	Orders per Account	Avg. Order Quantity	Sales	Profit	Profit Ratio
South	215	8	25	$2,895,619	$353,859	12.22%
West	215	8	26	$2,697,148	$292,931	10.86%
East	167	8	25	$2,384,550	$235,360	9.87%
Central	192	8	26	$2,739,813	$228,417	8.34%

Sales and Profit by Customer



Top Customers by Sales



Select Year

- (All)
- 2008
- 2009
- 2010

Customer Segment

- (All)
- Consumer
- Corporate
- Home Office
- Small Business

Find Customer

Profit

($27,264) $27,264

Undo Redo Reset

+ableau

Share Download

Qlik Sense

- Like Tableau, Qlik offers a comprehensive set of tools for businesses of all sizes to use in their BI efforts
    - For example, the ***free*** Qlik Sense Desktop tool can be used in small businesses or in workgroups to create powerful BI dashboards in a matter of minutes
- Qlik's pricing ranges for its Cloud-based solutions ranges from free to $25 per user, per month
- Qlik's pricing for its on-premise solutions ranges from free to $1,500 per token
- Additionally, Qlik sells subscription packages to data

Dashboard | Pipeline | Forecast | Data Management | Planning | Details

Search

Pipeline

Select a Salesperson: Alfredia M. Casada (Ala)

Sales Stage: I | II | III | IV | V

Sales Status: Open | Closed Won | Closed Lost

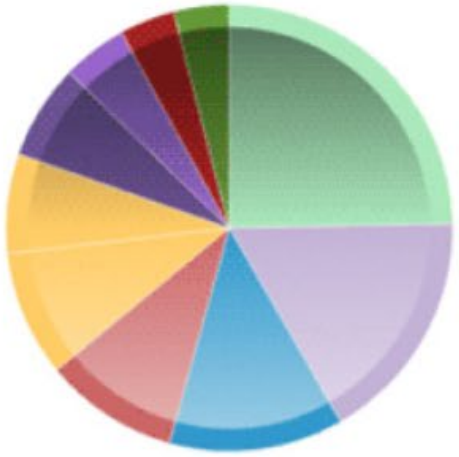
Opportunity Close by: 2008 | 2009 | 2010 | 2011 | 2012 | Q1 | Q2 | Q3 | Q4

Pipeline

Opportunities Ratio | Opportunity Details

Opportunities Ratio

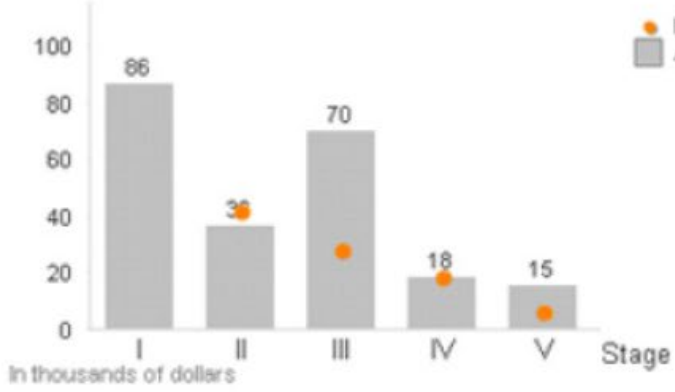
A look at opportunities by number of opps and value of opps



Account Industry	
Retail & Wholesale Distri...	$35,803
Mill Products	$24,611
Services	$18,288
Consumer Products	$14,068
Retail Banking	$13,228
Industrial Machinery	$10,448
Engineering, Construction...	$9,991
High Tech	$6,831

Sales Pipeline

A look at the total deal sizes in each stage of the process.



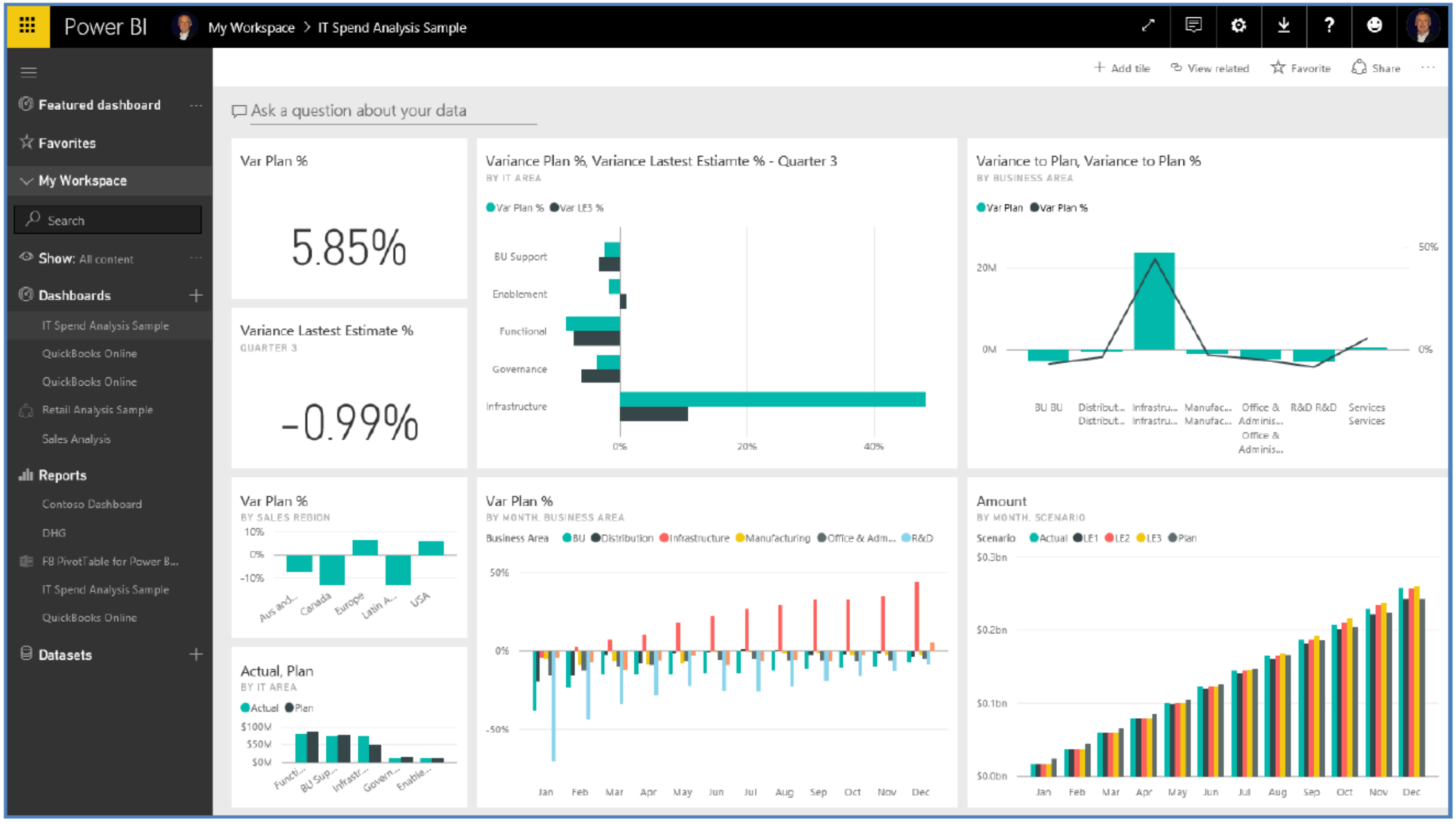
Opportunities

Forecast Category	
Commit	1
BestCase	4
Closed	0
Pipeline	17
Omitted	6

Deal Age	
Less than 6 months	23
7-18 months	5
19-36 months	0
More than 36 months	0

Power BI

- **Power BI Desktop** and **PowerBI.com** are Microsoft's answers to the limitations of Excel in a BI environment
- These tools provide outstanding capabilities, very modest pricing, and in many ways resemble Excel – making the transition to building BI dashboards in these tools easier
- Further, the ability to publish completed dashboards to the web overcomes the distribution problem found with using Excel as a business intelligence platform


Power BI
My Workspace > IT Spend Analysis Sample
Add tile
View related
Favorite
Share
Featured dashboard
Favorites
My Workspace
Search
Show: All content
Dashboards
IT Spend Analysis Sample
QuickBooks Online
QuickBooks Online
Retail Analysis Sample
Sales Analysis
Reports
Contoso Dashboard
DHG
F8 PivotTable for Power B...
IT Spend Analysis Sample
QuickBooks Online
Datasets
Ask a question about your data
Var Plan %
5.85%
Variance Lastest Estimate %
QUARTER 3
-0.99%
Variance Plan %, Variance Lastest Estiamte % - Quarter 3
BY IT AREA
Var Plan %
Var LE3 %
BU Support
Enablement
Functional
Governance
Infrastructure
0%
20%
40%
Variance to Plan, Variance to Plan %
BY BUSINESS AREA
Var Plan
Var Plan %
20M
0M
50%
0%
BU BU
Distribut...
Infrastru...
Manufac...
Office & Adminis...
R&D R&D
Services
Var Plan %
BY SALES REGION
10%
0%
-10%
Aus and...
Canada
Europe
Latin A...
USA
Var Plan %
BY MONTH, BUSINESS AREA
Business Area
BU
Distribution
Infrastructure
Manufacturing
Office & Adm...
R&D
50%
0%
-50%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Amount
BY MONTH, SCENARIO
Scenario
Actual
LE1
LE2
LE3
Plan
$0.3bn
$0.2bn
$0.1bn
$0.0bn
Actual, Plan
BY IT AREA
Actual
Plan
$100M
$50M
$0M

Two Parallel Development Environments

- **Excel-based development environment**
  - Requires Excel 2013/2016 Professional Plus or the standalone version of Excel
  - Using the four Power BI add-ins (PowerPivot, Power View, Power Maps, Power Query)
  - Can Publish analyses and queries to PBI
- **Power BI development environment**
  - Using the cloud-based tools at PowerBI.com or the free Power BI Desktop application (available in 32-bit and 64-bit Windows versions)
  - The interfaces of these tools are nearly identical in look, feel, and functionality
- Both use data models, which are explained in detail in a different session

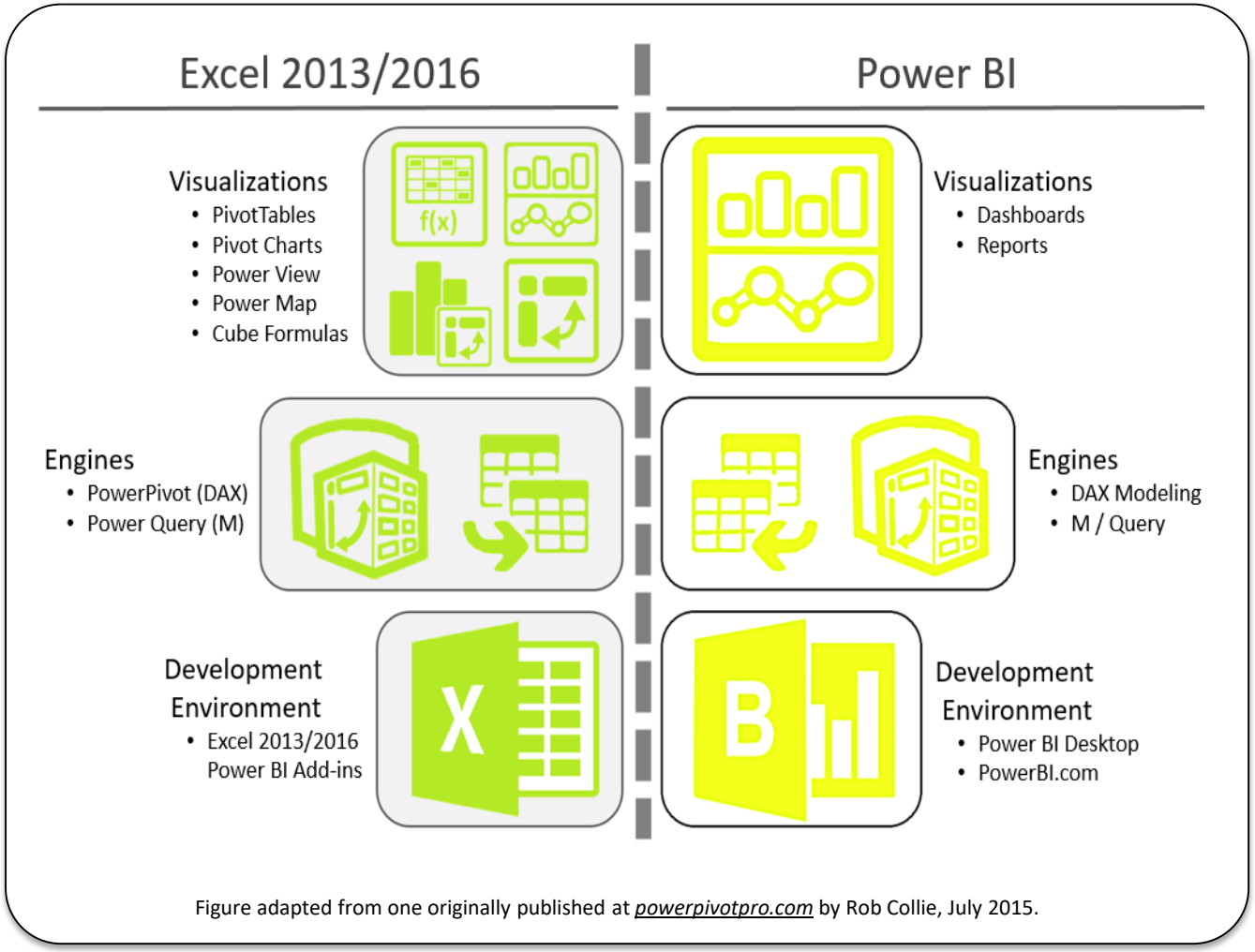


Figure adapted from one originally published at *powerpivotpro.com* by Rob Collie, July 2015.

Parallel Development Environments

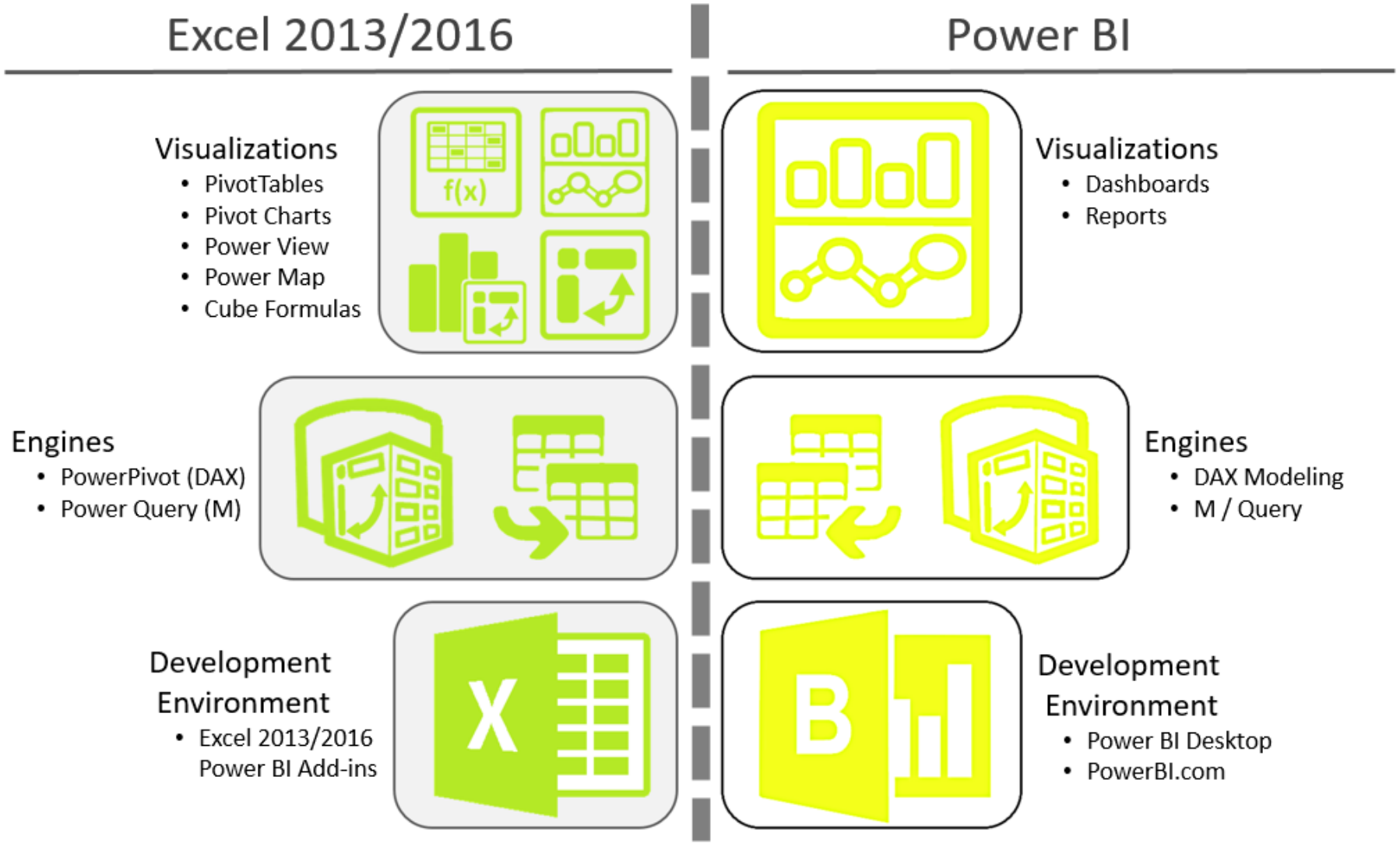
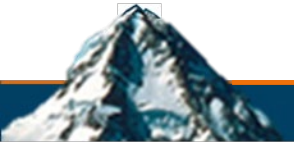


Figure adapted from one originally published at *powerpivotpro.com* by Rob Collie, July 2015.

Differences in the Environments

- The Excel-based environment can be used to produce PivotTables, Power View dashboards, and 3D maps with Power Map
  - Power Query is used to access, manipulate, and transform data
- PowerBI.com and Power BI Desktop can be used to produce dashboards similar to those produced by Power View, but with more visualization choices
  - Query is used to access, manipulate, and transform data
  - Will require a Data Management Gateway for real-time access to local data



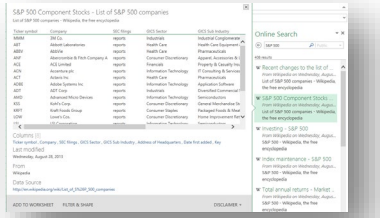
Why Two Environments?

- The Excel-based add-ins leverage the large number of Excel users in the marketplace and build on their knowledge and acceptance of Excel as an analytical tool
- PowerBI.com allows Microsoft to overcome the limitations of Excel
  - Long development cycles in a rapidly changing market (users want new features, and they want them faster)
  - Memory management issues with large data sets
  - Difficulty in publishing large workbooks to mobile devices
  - Delivery of BI tools to those using older versions of Excel or who use non-Windows platforms

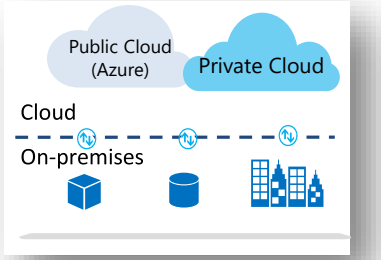
Democratizing ETLR in Power BI

EXTRACT

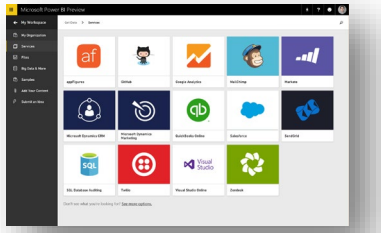
Power Query



DB Connectors



Power BI Content Packs

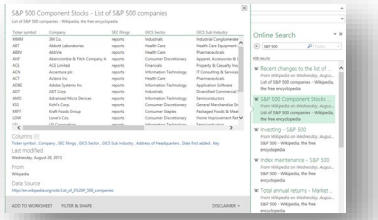


TRANSFORM

Clean up Data

Excel

Using M within Power Query

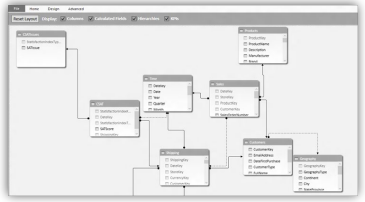


Access/SQL Transforms (Hard)

Microsoft SQL Server

LOAD

*Excel Data Model**



SQL/SSRS Data Warehouse

Microsoft SQL Server

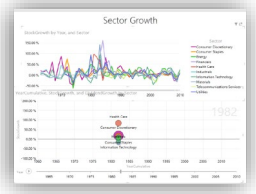
OLAP Cubes/ Data Mart

REPORT

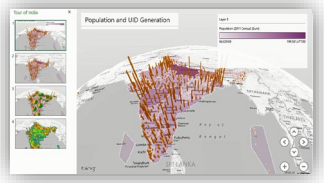
PivotTables/Charts

Excel

Power View



Power Map



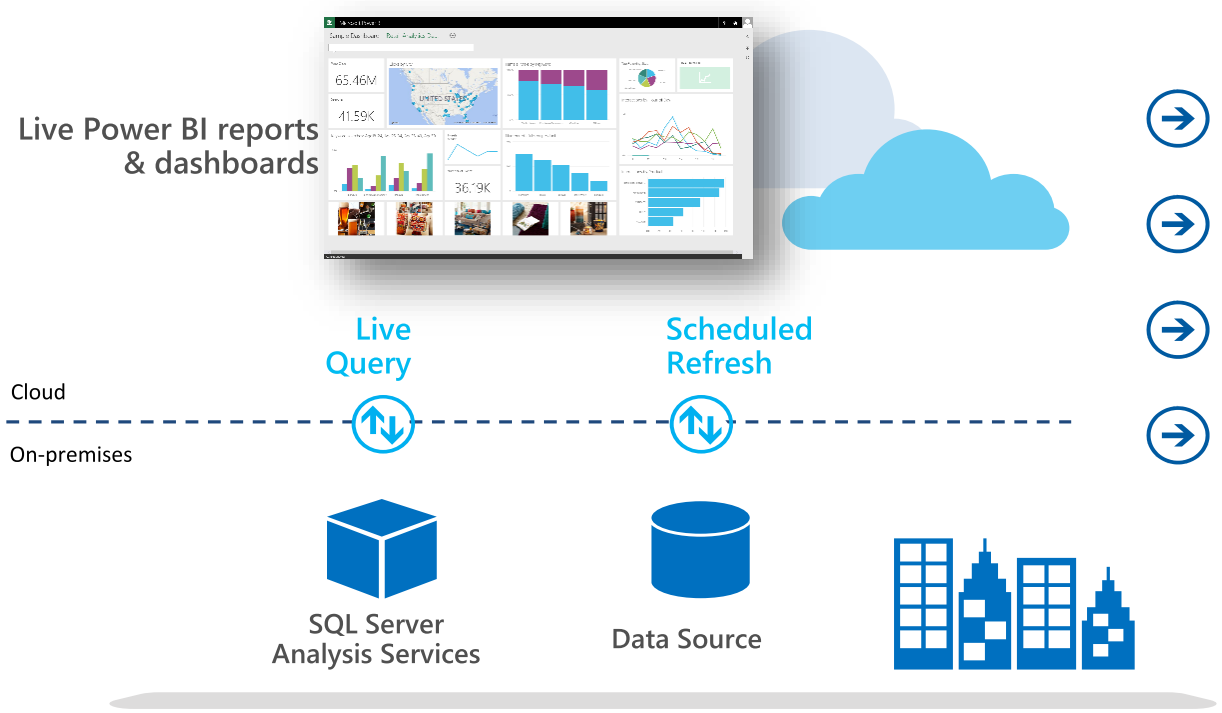
PBI Desktop/Sites

Power BI

* *Formerly known as "Power Pivot"*

Connect live to your on-premises data

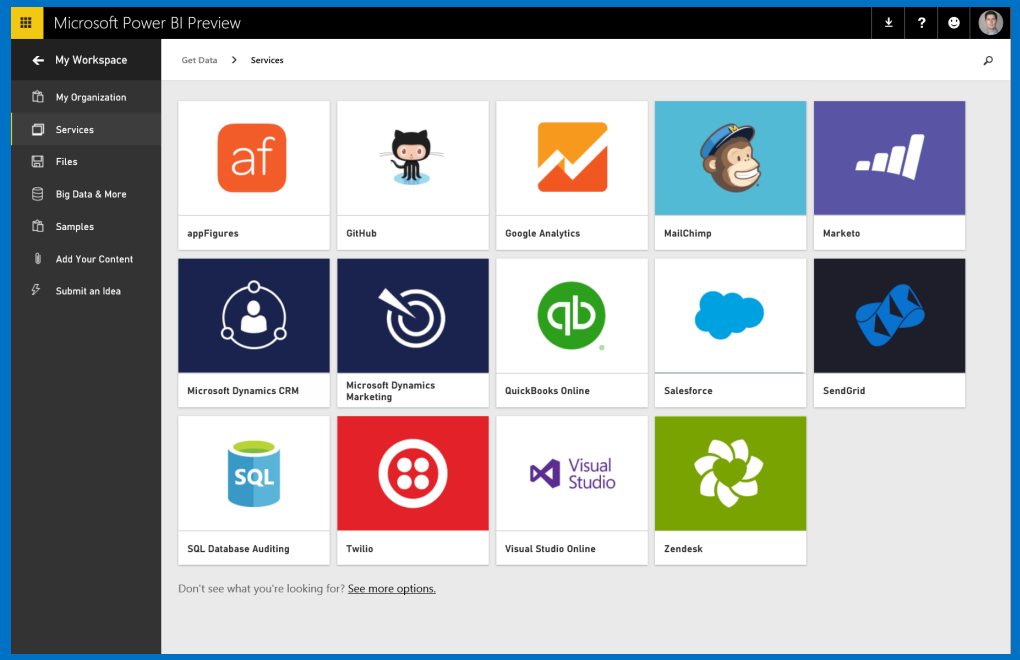
Live Query & Scheduled Data Refresh



- Deploy quickly with a hybrid BI solution
- Manage and secure data on-premises with SQL Server Analysis Services
- Optimized query performance for interactive data exploration
- Scheduled refresh from on-premises data sources

Power BI Content Packs

Prebuilt Power BI experience for your service



TODAY

Marketo
QuickBooks
salesforce
SendGrid
Microsoft Dynamics CRM
Visual Studio Online
GitHub
twilio
Dynamics Marketing
appfigures
Google Analytics
zendesk
MailChimp
SweetIQ
Acumatica THE CLOUD ERP
uservoice

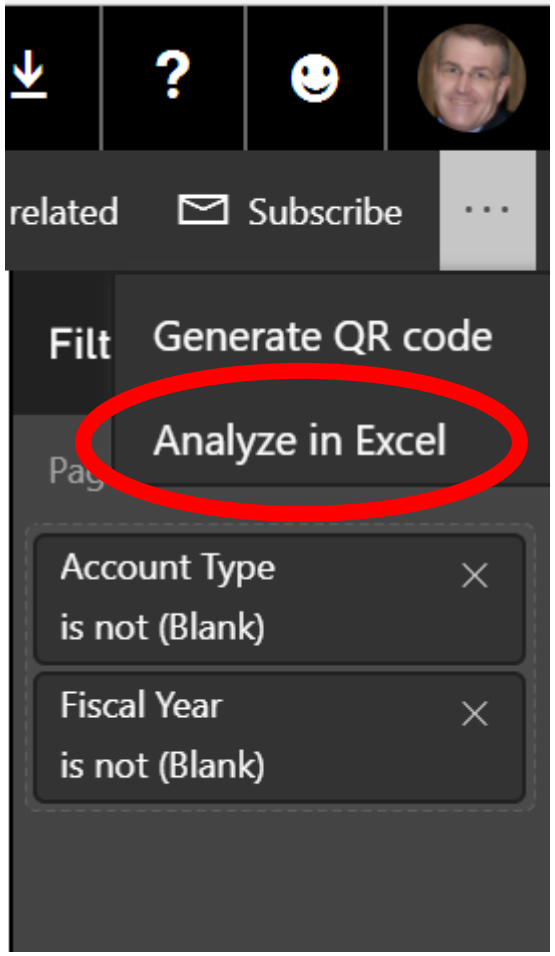
COMING SOON

zuora
Office 365
sage
sumologic
SQLSentry
comScore
inkling MARKETS
Adobe Analytics
SpaceCurve Spatial at the Speed of Reality
Azure Mobile Engagement
Troux by planview
tyGraph
CIRCUIT ID™
PLANVIEW
alteryx
insightly

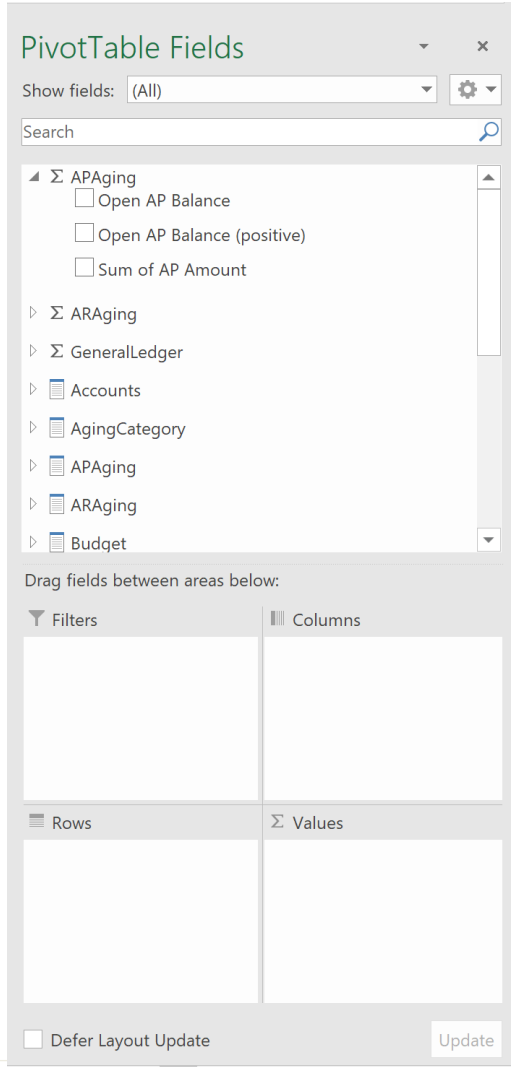
Added Benefit: Publishing to PBI Allows Excel Analysis of Content Pack Data in PivotTables

- QB Online
- CCH Axcess (with API connection)
- Xero
- Acumatica
- SalesForce
- Google Analytics
- Dynamics 365

… and many more



*.ODC
File Export



PowerBI.com

- Power BI subscriptions are available in two versions:
  - Power BI – FREE
  - Power BI Pro – $9.99 per month, per user
- Office 365 subscription or an Excel license NOT REQUIRED
- Provides a host of functionality
  - PowerBI.com and Power BI Desktop development environment
  - Site to publish Power BI Desktop and Power View reports
  - Data Catalogs (in Pro only)
  - Q&A natural language queries of reports

PowerBI.com

- A Power BI subscription is *required* to author and *consume* content
- Free version will meet the needs of most users
- An organization-based email account is required to register
  - Consumer accounts at Gmail, Yahoo, Outlook.com, etc., are *not* allowed
  - Email addresses ending in *.gov* are now supported

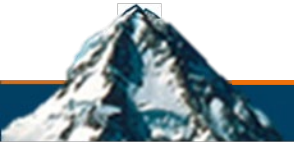
Power BI Desktop

- Standalone application that is free for download and use from Microsoft
- Available in 32-bit and 64-bit versions
- Does *not* require Excel
- Used to produce interactive dashboards from local or cloud data
- Can save and exchange dashboard files (.*pbix*)
- Reports can be published for consumption by others on PowerBI.com

Power BI Desktop

- Use 32-bit version when:
  - Need integration with 32-bit applications and/or 32-bit ODBC drivers
  - Using small datasets
- Use 64-bit version when:
  - Performing sophisticated data analysis
  - Using large datasets

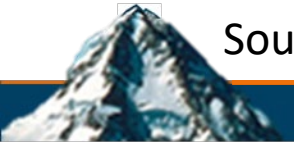
APPENDIX B: BIG DATA PLATFORMS AND TOOLS



Five Open Source Big Data Tools

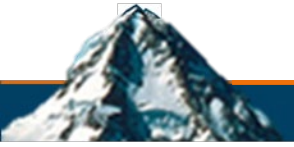
- Distributed data processing using Hadoop, which is pretty much the standard for processing large data sets across distributed systems
- Processing data streams using Spark or Flink, and then graduate to Beam
- Machine learning using Google's TensorFlow
- Big Data tool chain integration using Talend Open Studio
- Data Lakes using Kylo

Source: 5 Big Data Tools to Master in 2018 by Max Maxfield, Electronic Products, February 2018, page 8



Top Proprietary Suites With Big Data

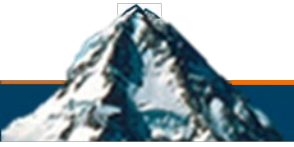
1. MapR Data Converged Platform
2. IBM Big Data
3. Amazon Web Service
4. Cloudera Enterprise Bigdata
5. Hortonworks Data Platform
6. Microsoft Azure
7. HP Bigdata
8. Intel Bigdata
9. DataStax Bigdata
10. Informatica PowerCenter Big Data Edition
11. Daatmeer
12. GoodData
13. Kognito Analytical Platform
14. Syncsort
15. Actian Analytics Platform
16. 1010data
17. SAP Bigdata Analytics
18. Oracle Bigdata Analytics
19. Teradata Bigdata Analytics



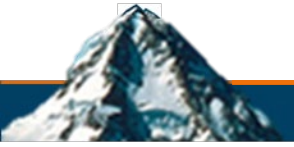
Top Trending Suites With Big Data

(Bold = Was On Top List, Too)

1. Talend Open Studio
2. Arcadia Data
3. **Informatica PowerCenter Big Data Edition**
4. **GoodData**
5. **Actian Analytics Platform**
6. Attivio Active Intelligence Engine
7. Google Bigdata
8. Wavefront
9. Opera Solutions Signal Hubs
10. **Daatmeer**
11. FICO Big Data Analyzer
12. **IBM Big Data**
13. **Amazon Web Service**
14. DataTorrent
15. **Oracle Bigdata Analytics**
16. Palintir Bigdata
17. **Cloudera Enterprise Bigdata**
18. Amdocs Insight
19. Splunk Bigdata Analytics
20. **Syncsort**

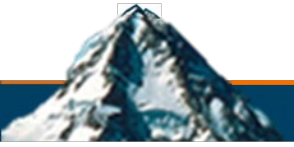


APPENDIX C: MACHINE LEARNING PLATFORMS AND TOOLS



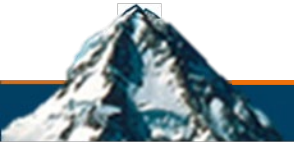
Free Software Suites With Machine Learning

- CNTK
- Deeplearning4j
- dlib
- ELKI
- GNU Octave
- H2O
- Mahout
- Mallet
- MEPX
- mlpy
- MLPACK
- MOA (Massive Online Analysis)
- MXNet
- ND4J: ND arrays for Java
- NuPIC
- OpenAI Gym
- OpenAI Universe
- OpenNN
- Orange
- R
- scikit-learn
- Shogun
- TensorFlow
- Torch
- Weka
- Yooreeka



Free Software Suites With Machine Learning

- CNTK
- Deeplearning4j
- dlib
- ELKI
- GNU Octave
- H2O
- Mahout
- Mallet
- MEPX
- mlpy
- MLPACK
- MOA (Massive Online Analysis)
- MXNet
- ND4J: ND arrays for Java
- NuPIC
- OpenAI Gym
- OpenAI Universe
- OpenNN
- Orange
- R
- scikit-learn
- Shogun
- TensorFlow
- Torch
- Weka
- Yooreeka



Proprietary Suites With Machine Learning

- Amazon Machine Learning
- Angoss KnowledgeSTUDIO
- Ayasdi
- IBM Data Science Experience
- Google Prediction API
- IBM SPSS Modeler
- KNIME
- KXEN Modeler
- LIONsolver
- Mathematica
- MATLAB
- Microsoft Azure Machine Learning
- Neural Designer
- NeuroSolutions
- Oracle Data Mining
- Oracle AI Platform Cloud Service
- RapidMiner
- RCASE
- SAP Leonardo
- SAS Enterprise Miner
- SequenceL
- Splunk
- STATISTICA Data Miner