# ANALYTICS PROCESSES

# genAI

- ChatGPT
  - Predict the next word
    - Reinforcement learning takes it beyond just predictions
  - Powerfull but not so much
- Alignment
  - Surprising in some tasks
  - Renown researchers very worried
- Government licensing

# ENABLERS

- Humans with brains
- Hardware
  - Cheap sensors
  - Cheap data storage
  - Efficient data storage
- Software
  - Advancement in databases
    - NoSQL, NewSQL
  - Data analytics
    - Machine learning
    - Deep learning

# JOBS REPLACED BY AI

**Blue collar jobs**

- Manufacturing workers
- Clerks
- Customer service agents
- Drivers
- Food preparation

**White collar jobs**

- Translators
- Financial advisors
  - Roboadvisors
- Business assistants
  - Google Now
  - Many other start-ups

**60-70% of jobs**

# SAFE JOBS

**Education**

- Interpersonal activities

**Healthcare**

- AI assisting but not replacing

**Construction**

- Requires dexterity
- Robots still cannot fold cloths

**Advisory boards?**

**Lawyers?**


https://www.thecollegefix.com/post/27773/

# BI

# WHAT IS BUSINESS INTELLIGENCE?

**Business Intelligence is about**

- Getting the right information, to the right decision makers, at the right time.

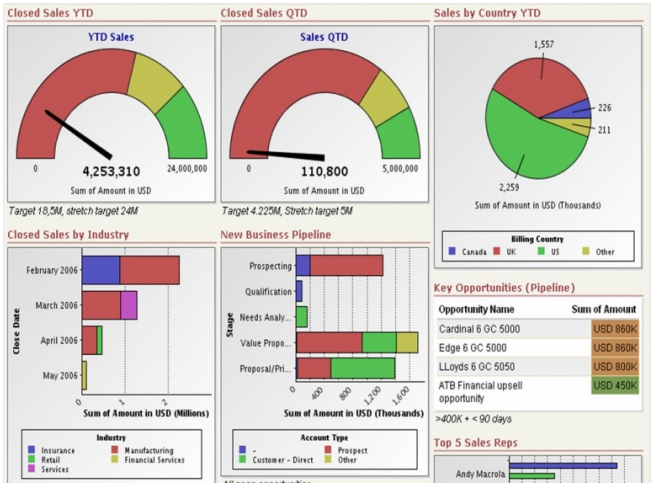
**BI is an enterprise-wide platform that supports reporting and decision making.**

**BI leads to**

- Fact-based decision making
- “Single version of the truth”
- Visibility

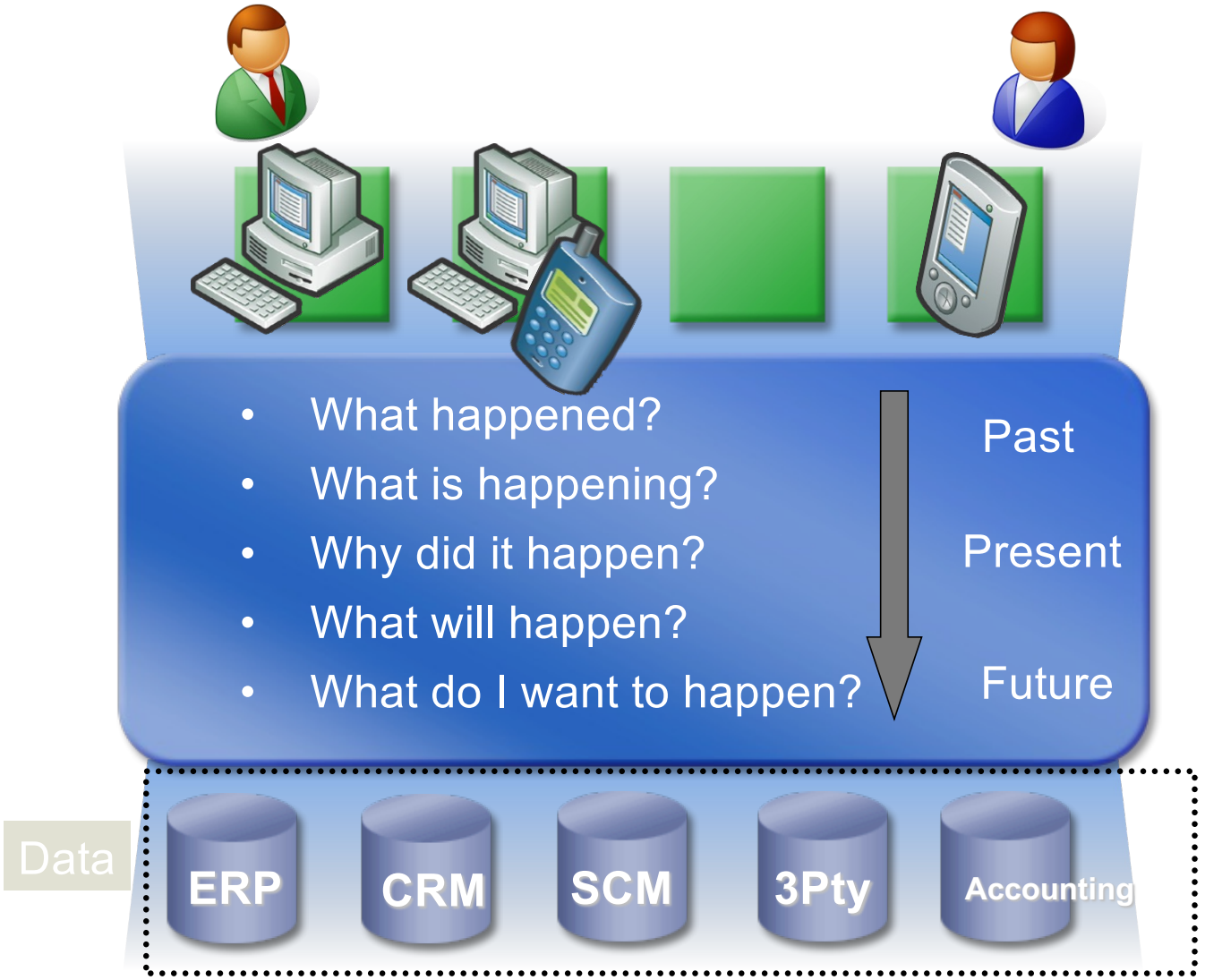
**BI includes reporting**

- Dashboards
- Trend line analysis
- Scorecards (KPI)
    - Goal, actual, trend



# SCOREBOARDS

# THE FIVE QUESTIONS OF BI



# SOFTWARE SOLUTIONS

**Cognos – IBM**

**Business Objects – SAP**

**Oracle**

**Microstrategy**

**Open source**

- JasperSoft
- Pentaho

**Moving to cloud?**

- Snowflake

# ANALYTICS

# ANALYTICS VS. BI

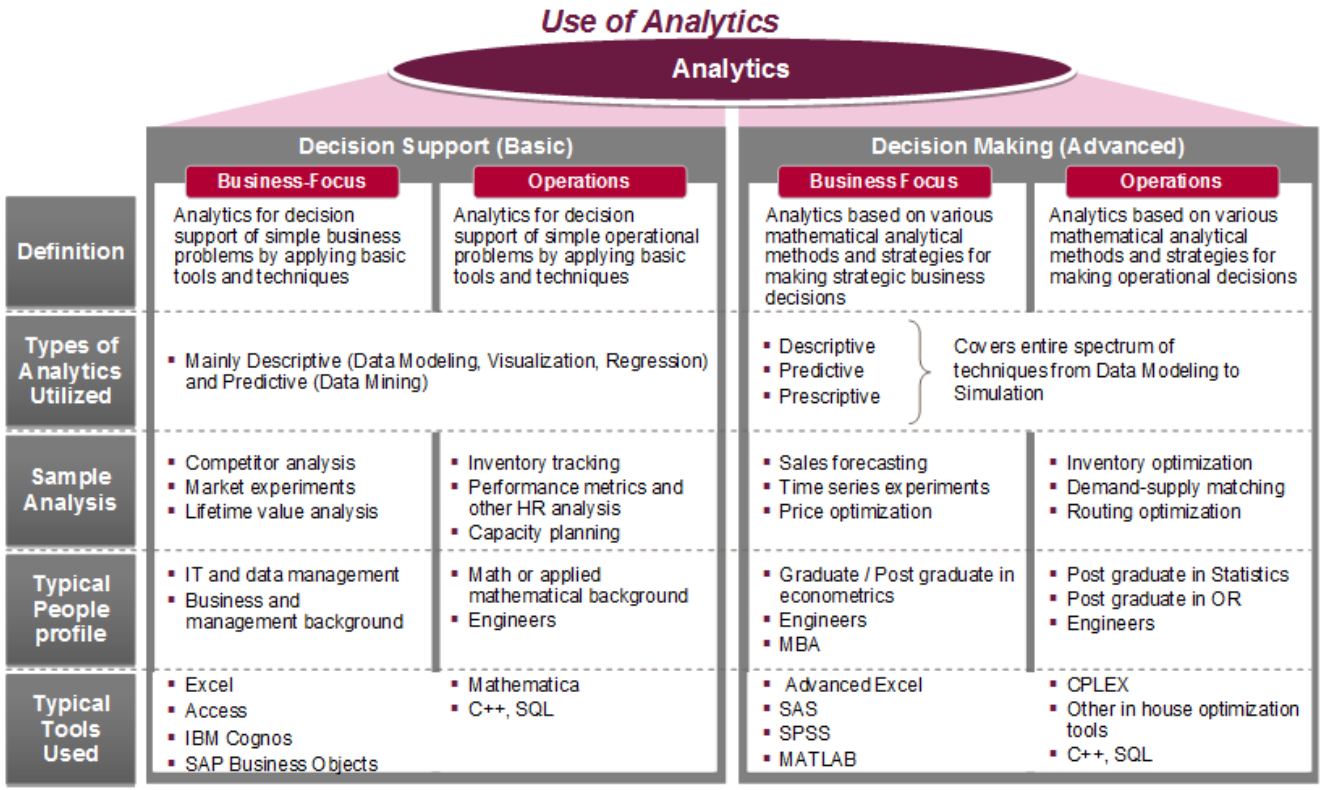
**Running reports is NOT analysis.**

**Both reports and analysis are critical to success!**

| REPORTING | | ANALYSIS |
|---|---|---|
| Provides Data | → | Provides Answers |
| Provides What is Asked For | → | Provides What is Needed |
| Typically Standardized | → | Typically Customized |
| No Person Involved | → | Involves People |
| Fairly Inflexible | → | Extremely Flexible |

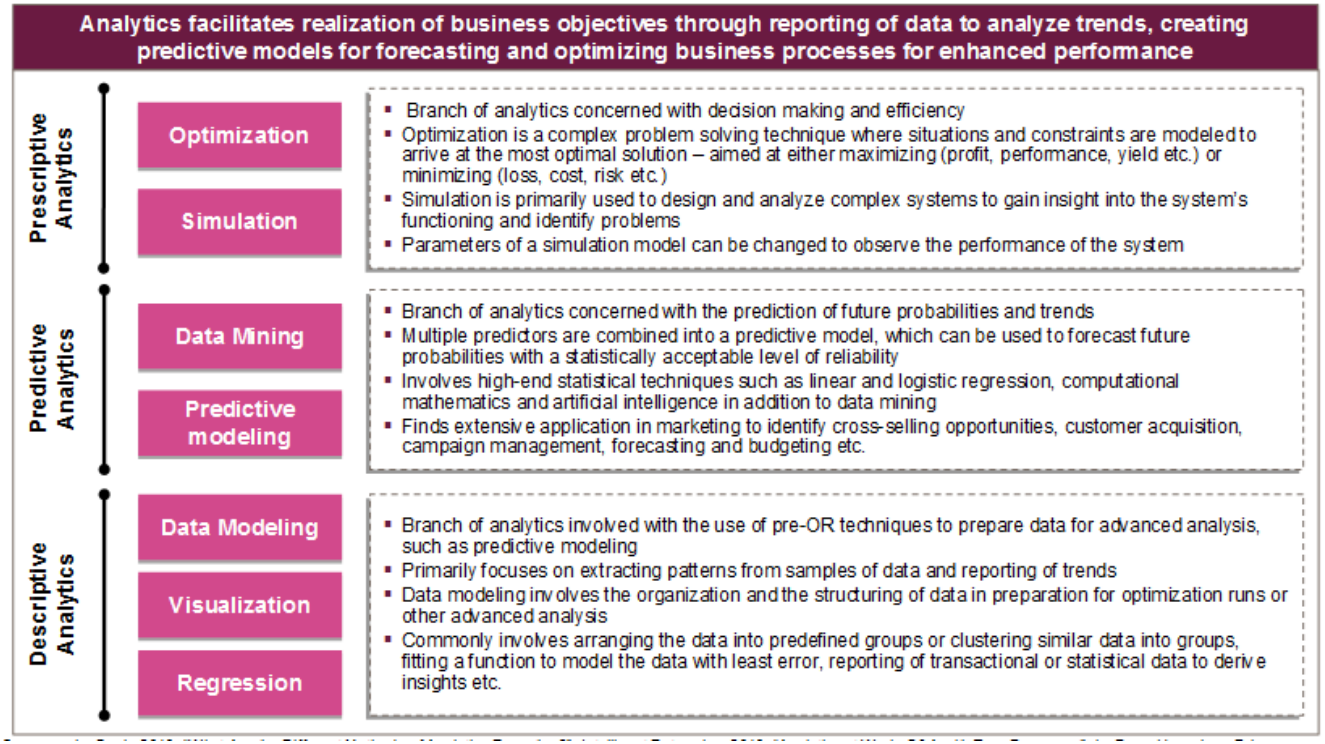
→

# THE BIG PICTURE




Capgemini Consulting 2010

# TOOLS




Capgemini Consulting 2010

# SOFTWARE SOLUTIONS

**SAS**

**Watson Analytics**

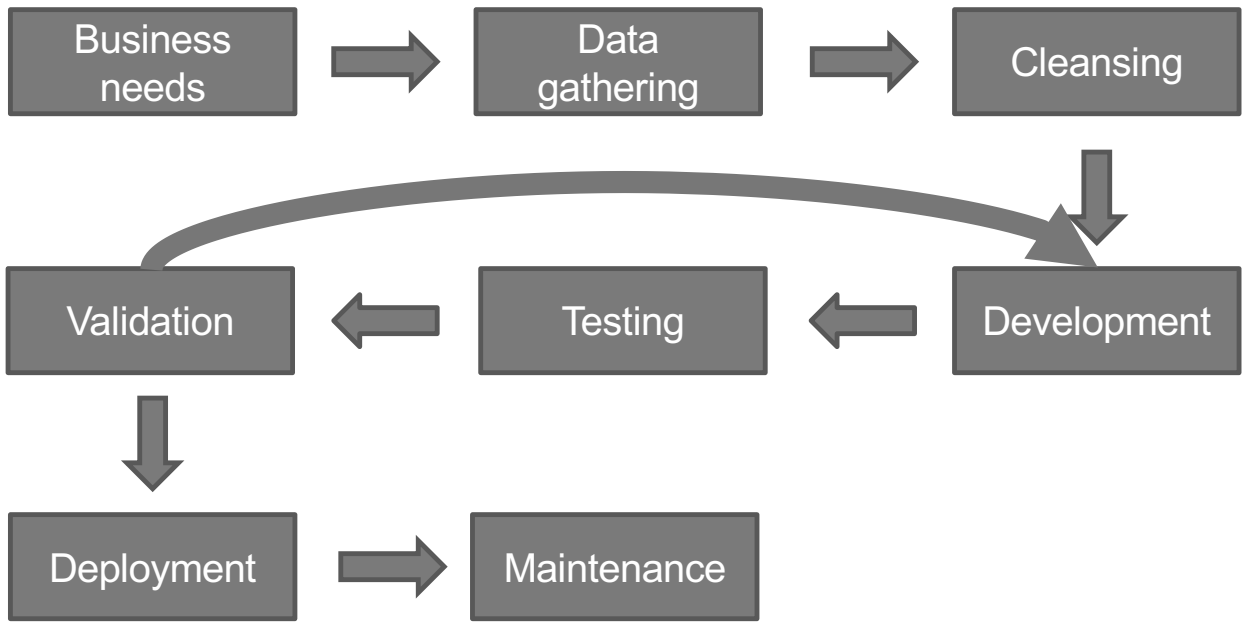
**R/Python**

**RapidMiner**

**Others**

- Salford Systems – data mining
- Many, many other vendors

# STEPS


Business needs
Data gathering
Cleansing
Development
Testing
Validation
Deployment
Maintenance

# DATA GATHERING

# DATA GATHERING

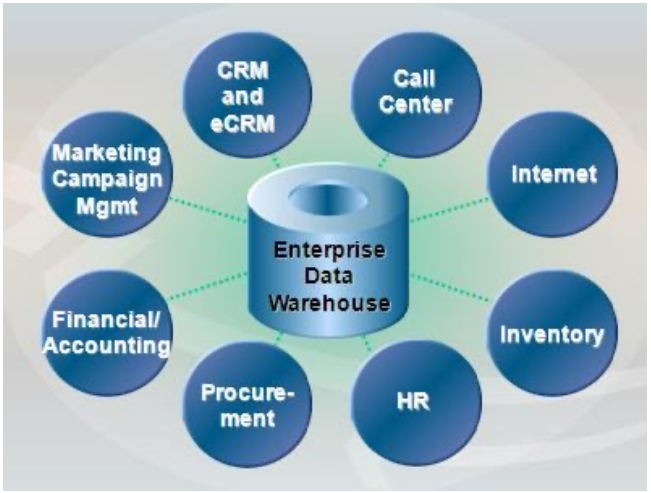
**Internal data sources**

**External data**

- Social media
- Consumer data

**Who is the data owner**

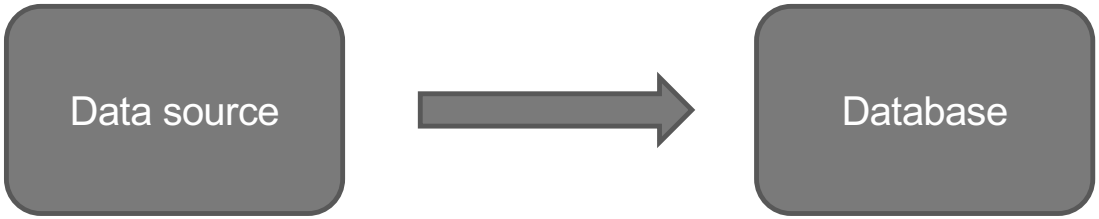
- IT
- Business user



# DATA CLEANSING

**80/20 rule**

- 80% of time data cleansing
- 20% of time analytics




ETL = extract, transform, load

# DATA CLEANSING

# DATA CLEANSING

**Data is never 100% clean**

- Missing values
- Text when number expected
- Extremely large values
- Negative values (e.g., weight)
- Misspelled
- Address
    - Av – Avenue
    - Il = Illinois
    - 5 digit zip vs. 9 digit zip

# DATA CLEANSING

**Corrections**

- Replace by average
- Interpolate (e.g., temperature)
- Often talk with data owner

**Text**

- Encode phonetically
- Count swaps in sounds
    - If low, likely to be the same

**Advanced**

- Data imputation
    - Regression for monotonically missing data
- Develop model (probabilistic)

# WHY IS DATA DIRTY?

**Incomplete data comes from:**

- Non available data value when collected
- Different criteria between the time when the data was collected and when it is analyzed.
- Human/hardware/software problems

**Noisy data comes from:**

- Data collection: faulty instruments
- Data entry: human or computer errors
- Data transmission

**Inconsistent (and redundant) data comes from:**

- Different data sources, so non uniform naming conventions/data codes
- Functional dependency and/or referential integrity violation

# SOFTWARE AS A SERVICE

# LAWS OF ON-PREMISE SOFTWARE

**Substantial up-front licensing cost**

- Millions of dollars

**Months or years to deploy**

**Substantial maintenance fees**

- Up to 50% of revenue of a software powerhouse comes from maintenance fees

**Vendor not engaged until renegotiating the licensing agreement**

# ON-PREMISE

**Very hard to switch vendor**

- Locked and dependent on the mercy of the vendor
- Capital
- Time

**Training**

- Always required

# SOFTWARE-AS-A-SERVICE

**On-demand**

**Pay-as-you-go**

**Basic characteristics**

- Pay-per-use or monthly subscription
  - $10 per-user per-month
  - Per resource
    - $20 per each employee
- $100,000 to $200,000 set-up installation cost
  - Drastically lower than on-premise

# SOFTWARE-AS-A-SERVICE

**Easy to switch solution/vendor**

- Low up-front licensing cost
    - Spill over to per-user charge
- Low deployment cost
    - How to be a profitable vendor?

**Maintenance**

- Let the vendor take care of it
- Hosted solution

# SOFTWARE-AS-A-SERVICE

**Delivered online over Internet**

- Hosted solution
- Infrastructure resources provided by the vendor

**Maintenance**

- Vendor responsible

**Upgrades transparent**

- User options on upgrades

# CLOUD COMPUTING 101

**Computing on demand**

**Rent hardware or infrastructure**

- Pay per CPU cycle
- Pay per hard drive usage

**Long-term not beneficial**

**Spike in demand for an application**

- Instead of investing into new resources
    - Rent cycles in the cloud

# EC2 PRICING

| Standard on-demand | Linux (per hour) | Windows (per hour) |
|---|---|---|
| Small | $0.065 | $0.115 |
| Large | $0.130 | $0.230 |
| Extra-large | $0.520 | $0.920 |
| **High-CPU on-demand** | **Linux (per hour)** | **Windows (per hour)** |
| Medium | $0.165 | $0.285 |
| Extra-large | $0.660 | $1.140 |

**A year large Linux $875**

**One year reserved $800**

# ROI CASE

| | On-premise | SaaS |
|---|---|---|
| Hardware | $100,000 | |
| License | $200,000 | $10,000 per month |
| Maintenance per year | $25,000 | |
| Internal maintenance per year | $200,000 | |
| Database | $50,000 | |
| **Total over 5 years** | **$1,475,000** | **$600,000** |

**Performance, compliance, integration neglected**

# ROI

**Integration**

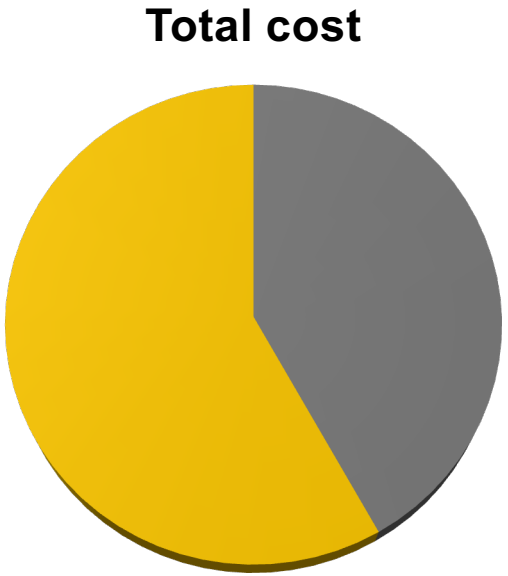
- $5,000 per month

**Bandwidth not an issue**

- Moving GBs and not TBs

**Compliance**

- Privacy regulations
- Login accounts
- $3,000 per month




| | On-premise | SaaS |
|---|---|---|
| **Total** | **$1,475,000** | **$1,080,000** |

# MYTH OR BUST

**Too much data outside the walls**

- Build trust and relationship

**Security**

- High security level can be reached
    - Cloud providers offer better security
- Devote attention, do not neglect

**Tactical and strategic planning tools**

- Operational tools need special care
    - Latency might be a problem

# KEY POINTS FROM ENTERPRISE'S PERSPECTIVE

**Critical mission applications not appropriate**

- Possible downtime
    - Cloud providers getting better than on-premise
- Reluctance to put data to a third party

**Thorough ROI analysis**

- Return depends on longevity

**Do not cross out due to perceived complexity**

# THE BIG THREE

**Amazon web services**

**Microsoft Azure**

**Google cloud platform**

**Not just cloud provider**

- One stop shop
- Analytics and machine learning tools
- AI
- Nosql databases
- Streaming