Waterford Institute *of* Technology

BIG DATA I

SAMITHA SOMATHILAKA

Department of Computing & Mathematics, WIT

BIG DATA DEFINITION

- No single standard definition…

"Big Data" is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it…

WHO'S GENERATING BIG DATA

- The progress and innovation is no longer hindered by the ability to collect data.
- But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion.
- **Social media and networks**
  - (all of us are generating data)
- **Scientific instruments**
  - (collecting all sorts of data)
- **Mobile devices**
  - (tracking all objects all the time)
- **Sensor technology and networks**
  - (measuring all kinds of data)

THE MODEL HAS CHANGED…

- **The Model of Generating/Consuming Data has Changed**

Old Model: Few companies are generating data, all others are consuming data

New Model: all of us are generating data, and all of us are consuming data

WHAT IS BIG DATA

Handling > 10 TB of data

Data with a changing structure or with no structure at all

Very high throughput systems

Business requirements differ from relational database model

Massive processing

3 Vs OF BIG DATA

Velocity: Frequency of generation is too high to be managed traditionally

Volume: The growth of world data is exponential

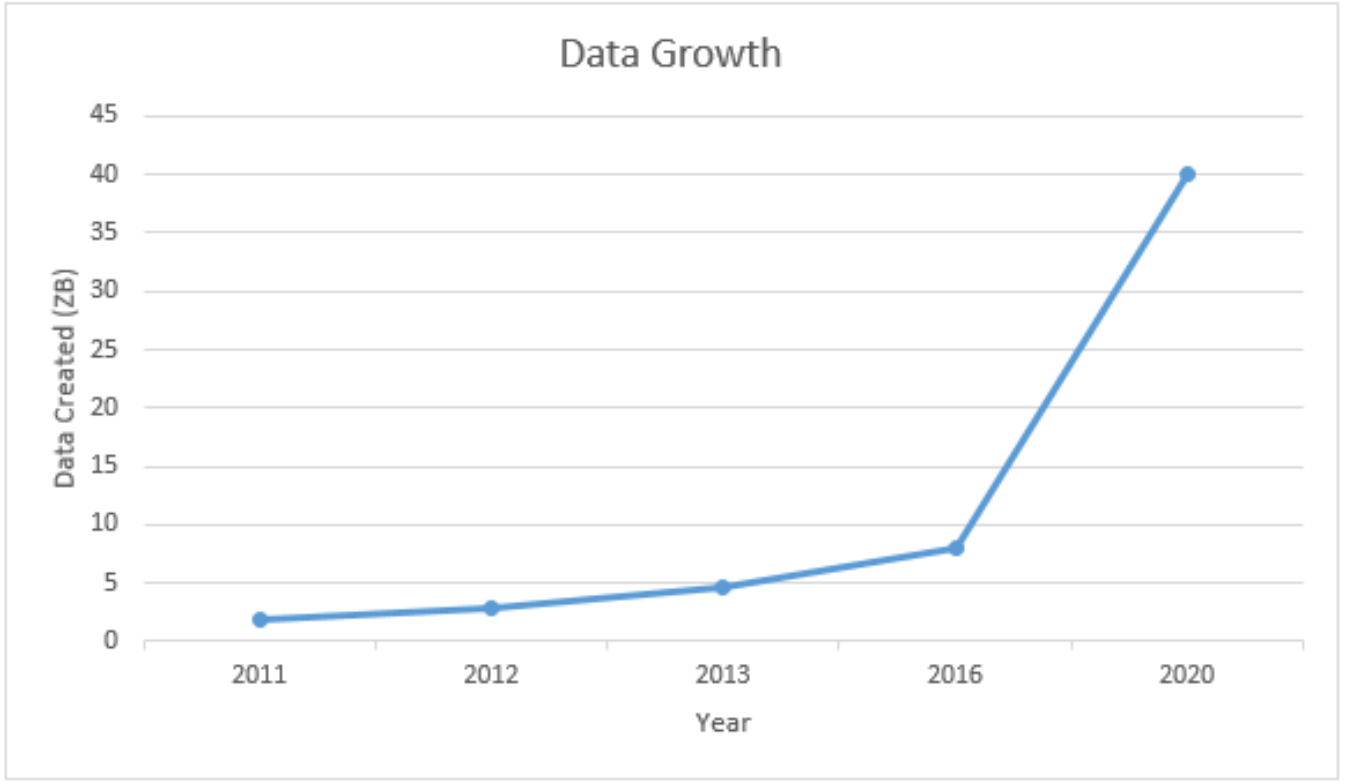
Variety: Big Data can be structured and unstructured

CHARACTERISTICS OF BIG DATA: VELOCITY

- Data is begin generated fast and need to be processed fast
- Online Data Analytics
- Late decisions ➔ missing opportunities
- **Examples**
  - **E-Promotions:** Based on your current location, your purchase history, what you like ➔ send promotions right now for store next to you
  - **Healthcare monitoring:** sensors monitoring your activities and body ➔ any abnormal measurements require immediate reaction

CHARACTERISTICS OF BIG DATA: VOLUME

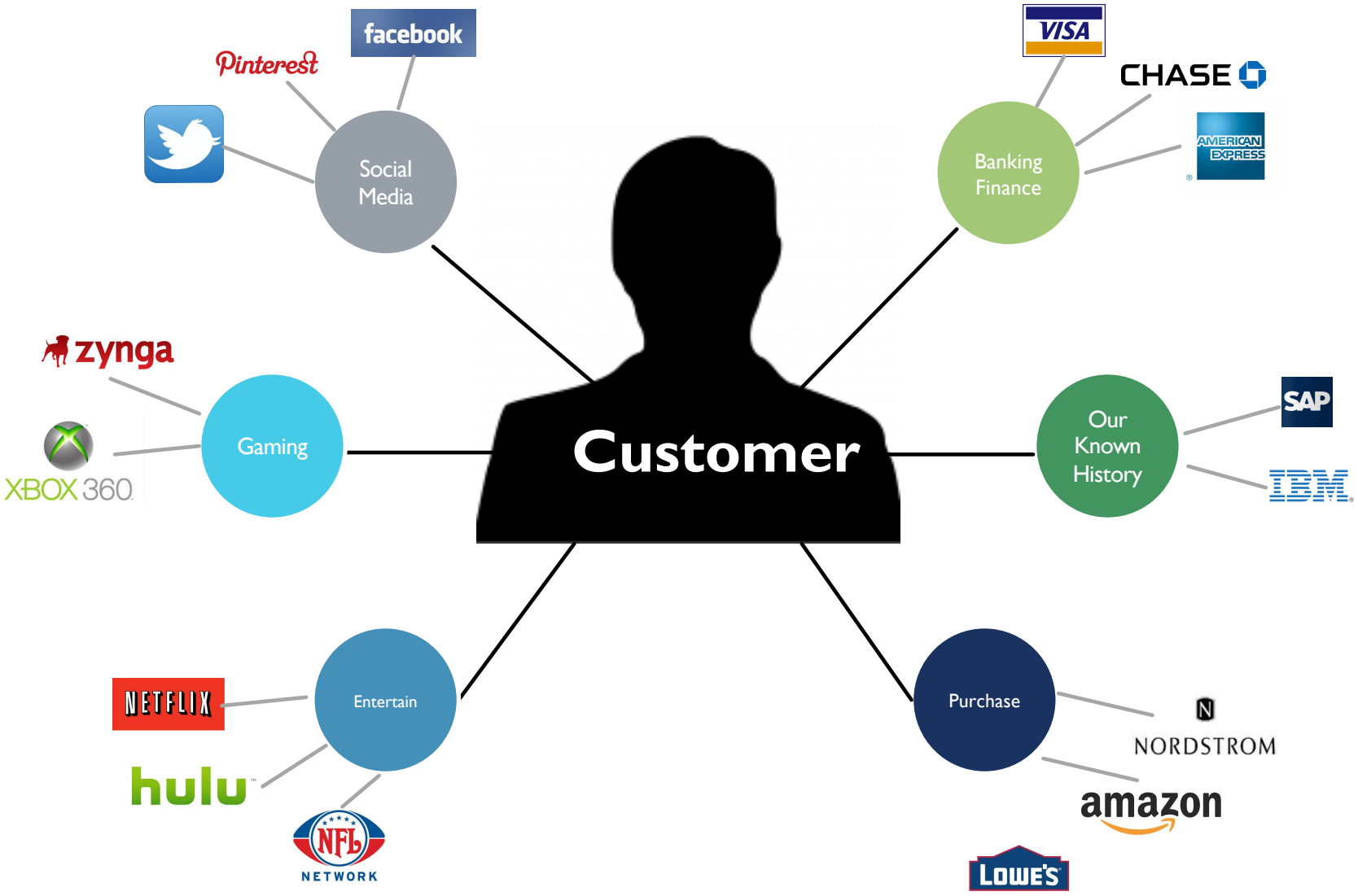
- **Data Volume**
  - 44x increase from 2009 2020
  - From 0.8 zettabytes to 35zb
- Data volume is increasing exponentially



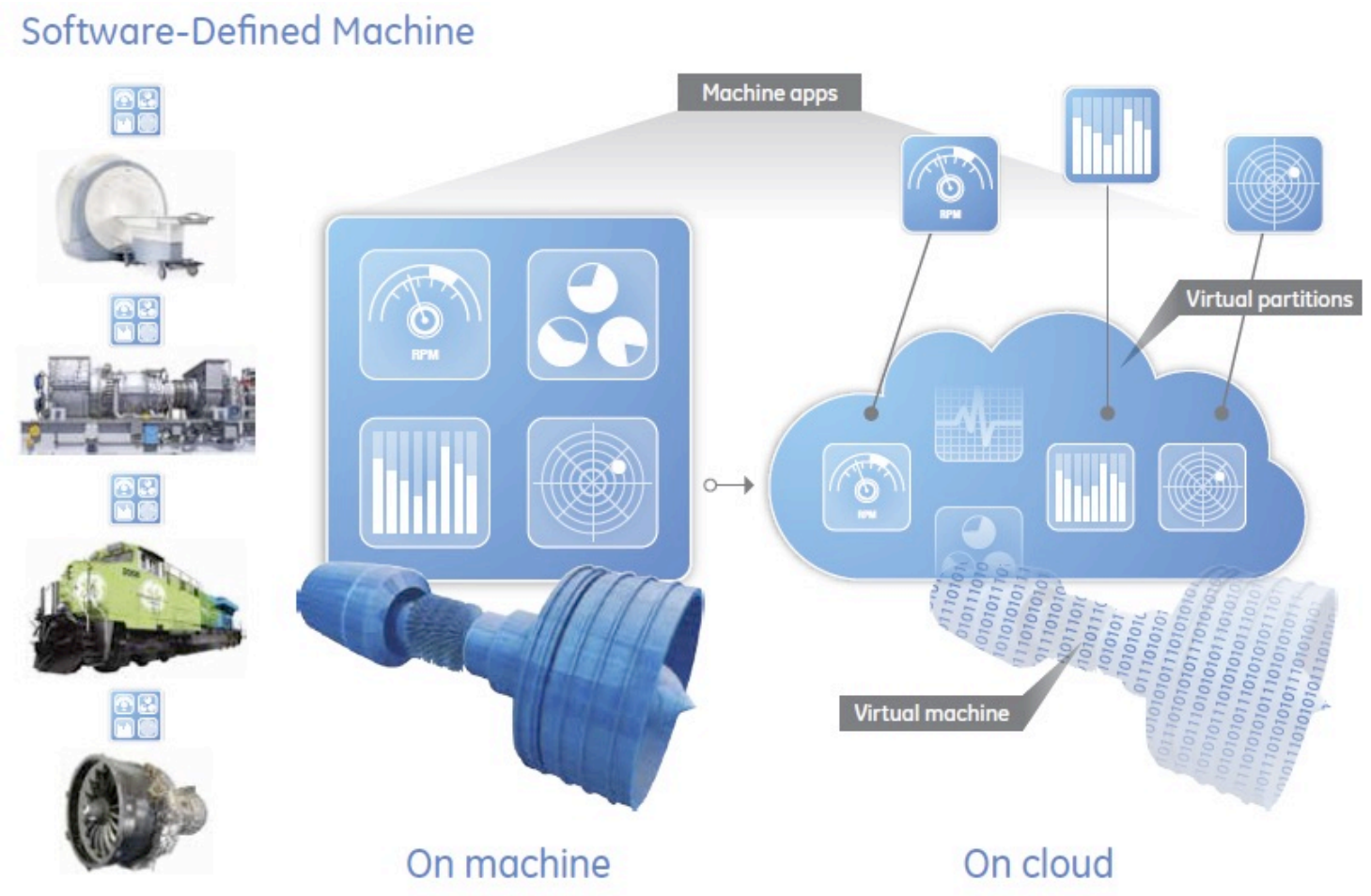
CHARACTERISTICS OF BIG DATA: VARITY

- Various formats, types, and structures
- Text, numerical, images, audio, video, sequences, time series, social media data, multi-dim arrays, etc…
- Static data vs. streaming data
- A single application can be generating/collecting many types of data

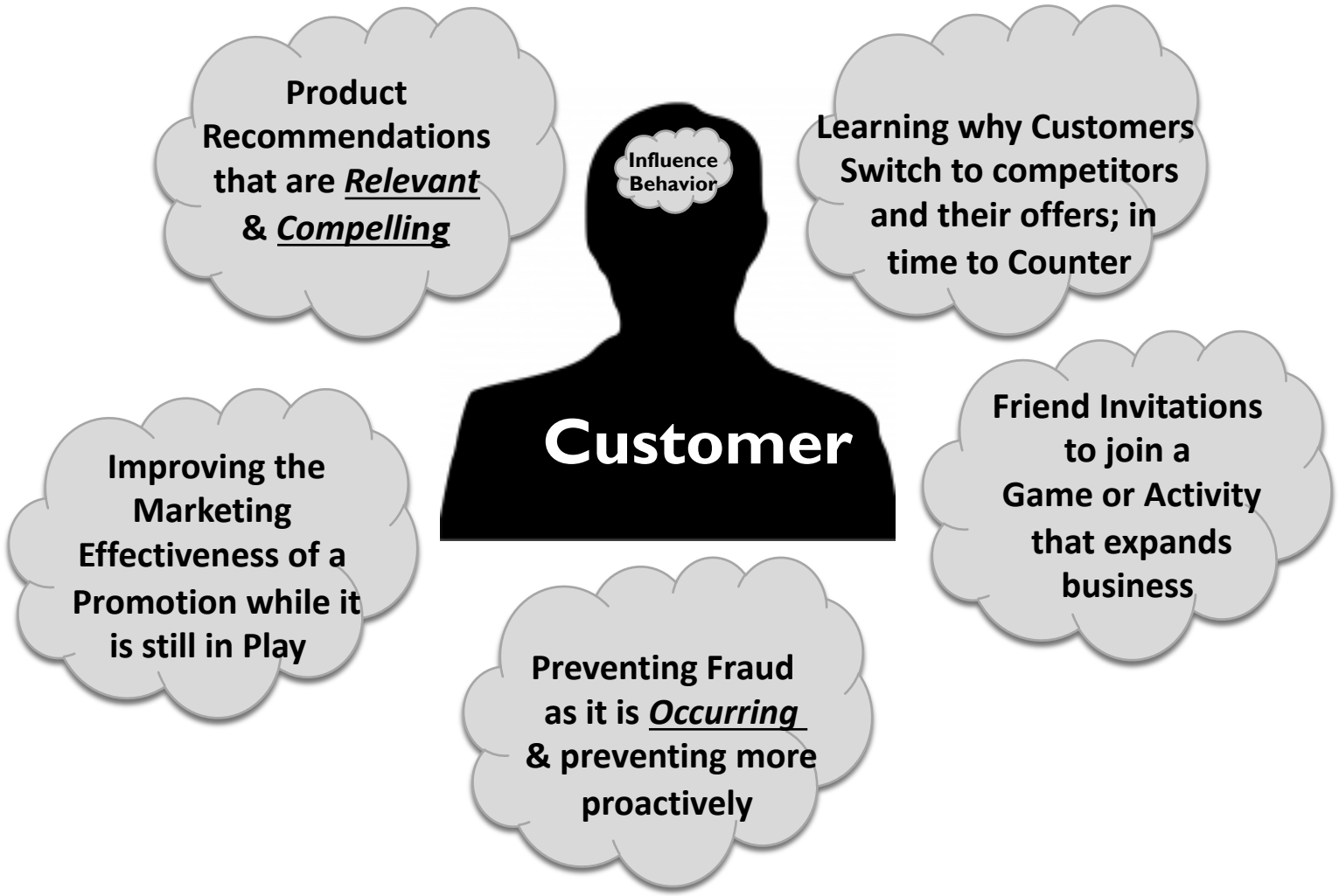
A Single View to the Customer



SOFTWARE DEFINED MACHINES WITH IOT



Real-Time Analytics/Decision Requirement



2 MORE Vs

Veracity: Establishing trust in data

Viability: Relevance and Feasibility

CHARACTERISTICS OF BIG DATA: VERACITY

- Data veracity, in general, is how accurate or truthful a data set may be.
- When it comes to the accuracy of big data, it's not just the quality of the data itself but how trustworthy the data source, type, and processing of it is.
- Removing things like bias, abnormalities or inconsistencies, duplication, and volatility are just a few aspects that factor into improving the accuracy of big data.

CHARACTERISTICS OF BIG DATA: VIABILITY

- The level of consistency between the content of data and the user's areas of interest.
- In other words: the extent to which data answers of gives insight into the question of the individual user.
- Relevant data creates strong strategies
- Relevant data is necessary for optimization

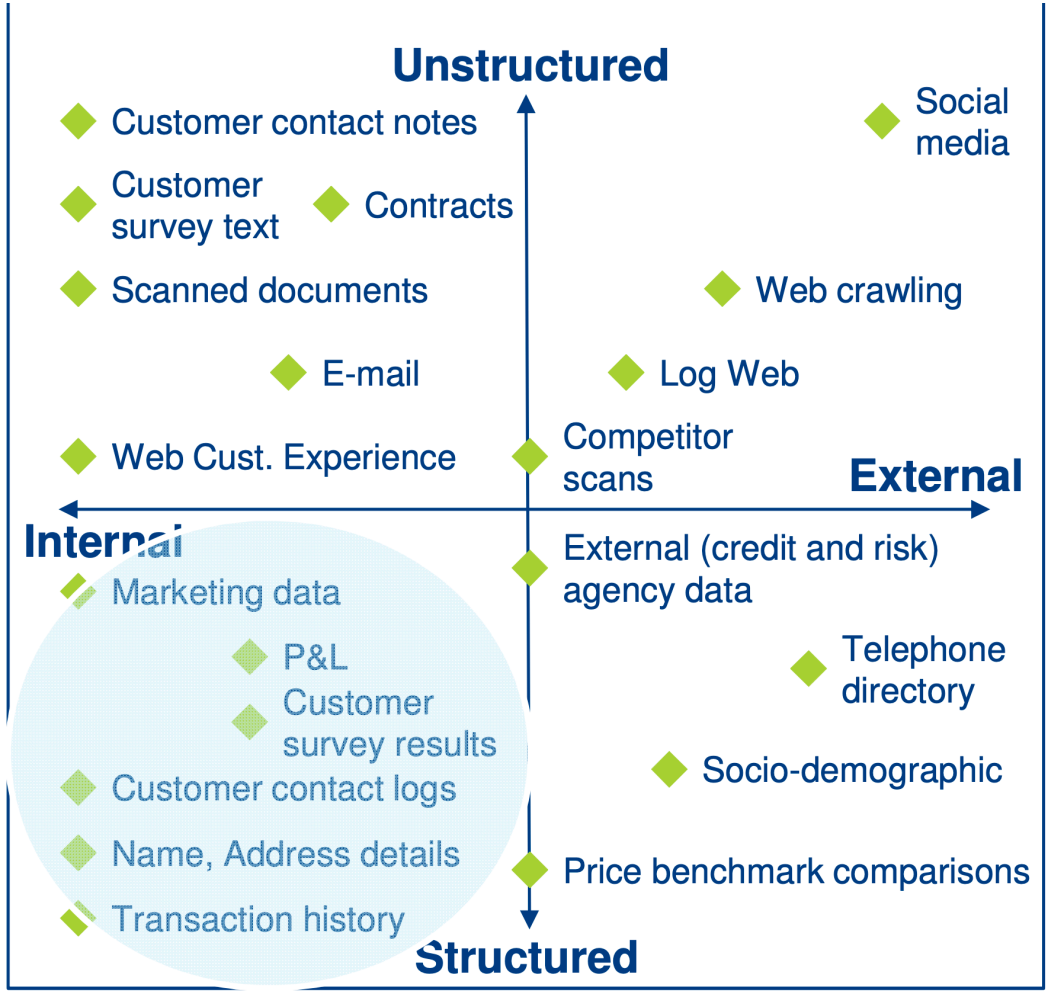
Velocity +Volume +Variety +Veracity +Viability =Value

VALUE OF DATA

Costs – there is a serious risk of simply creating Big Costs without creating strong value

Insights – Sophisticated queries, counterintuitive insights and unique learning

4 DIMENSIONS

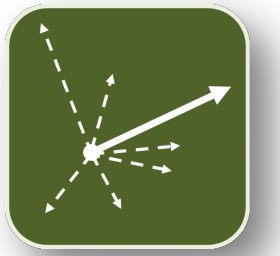


USE OF BIG DATA

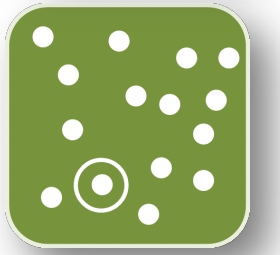
WHAT IS BIG DATA? CONTINUED

- Companies leverage data to adapt products and services to:
  - Meet customer needs
  - Optimize operations
  - Optimize infrastructure
  - Find new sources of revenue
  - Can reveal more patterns and anomalies

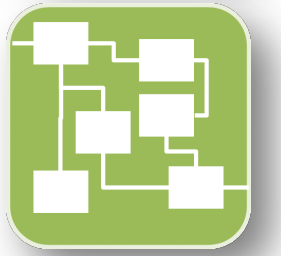
BIG DATA OPPORTUNITIES



Making better informed decisions
e.g. strategies, recommendations



Discovering hidden insights
e.g. anomalies forensics, patterns, trends



Automating business processes
e.g. complex events, translation