



# Applying Big Data

Graeme Seaton, Data Engineering Lead, EMC Global Services

# EMC Global Professional Services



- **13,000 Services Professionals**
- **Delivery Focus**
  - Transformation Services
  - Application and Big Data Services
  - Technical Infrastructure Services
  - Program Management
- **Industry Focus**
  - Financial Services
  - Life Sciences
  - Communications, Media & Entertainment

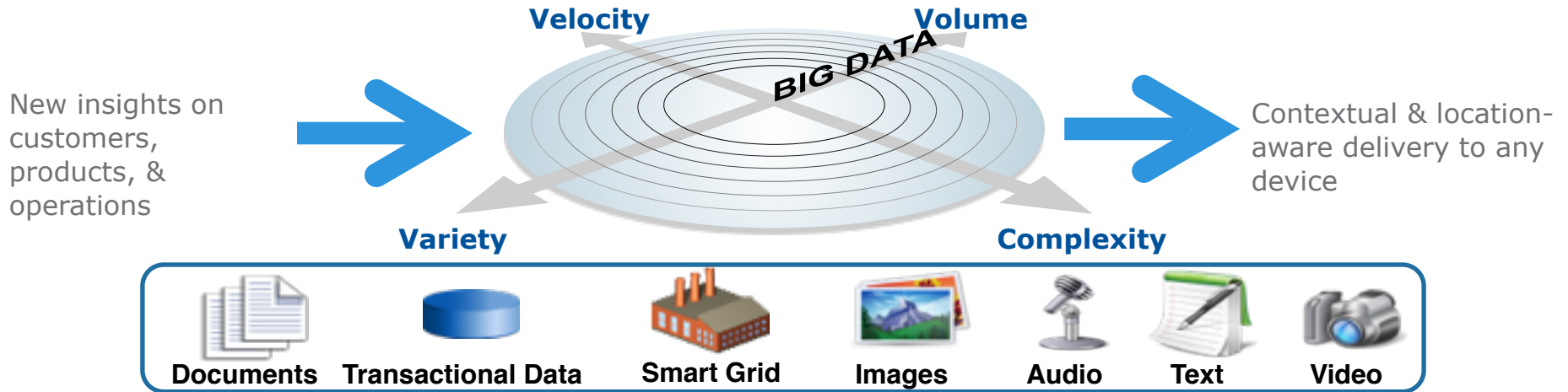
Global presence with thousands of engagements delivered to global **Fortune 1000** companies



EMC<sup>2</sup>

# What we mean by Big Data

More than just 'big,' big data is fast, varied and complex



- **Volume:** data volumes approaching multiple petabytes
- **Velocity:** data being generated & ingested for analysis in real-time
- **Variety:** tabular, documents, e-mail, metering, network, video, image, audio
- **Complexity:** different standards, domain rules, & storage formats per data type

# Structured + Unstructured



Structured Information  
In Relational Databases



Managed & Unmanaged  
Unstructured Information

## *Internet Of Things*



Telemetry, Location-Based Information (SoMoLo).

## *Non-Enterprise Information*



***The Digital Universe Is Growing By 7,600 PB / Day***

# Why Now?



Data  
Growth



Cheap  
Storage



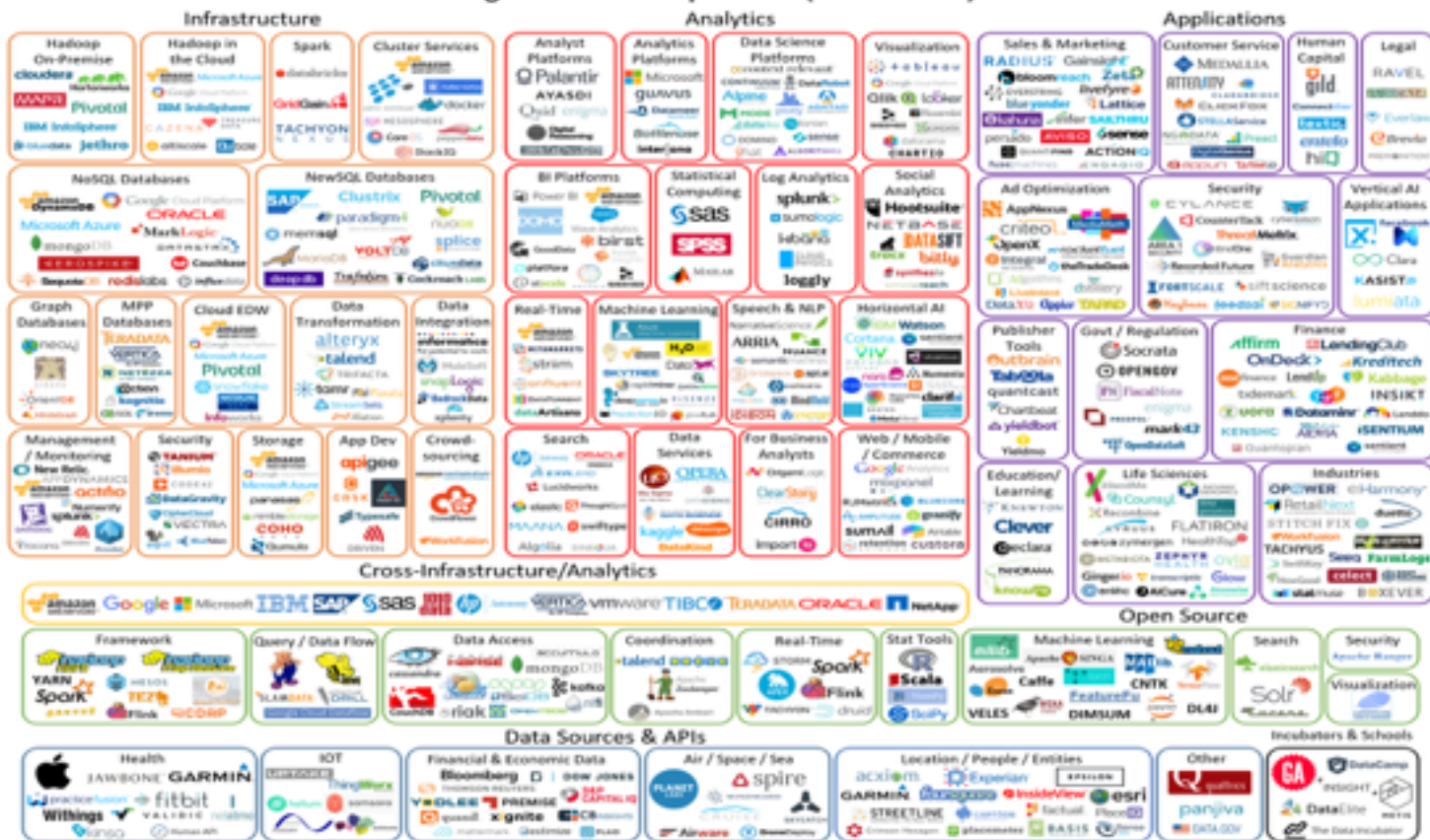
Limitless  
Compute



Real-Time  
Technologies



# Big Data Landscape 2016 (Version 2.0)



Last Updated 2/12/2016

© Matt Turck (@mattturck), Jim Hao (@jimhao), & FirstMark Capital (@firstmarkcap)

FIRSTMARK

EMC<sup>2</sup>

# Challenges

Uncertain Markets



Cost Pressures



Global Competition



7x24 Availability



Regulatory Compliance



New Business Pressures



Operational Limitations



Rapid Information Growth



Siloed Operations



Inflexible Processes



Security Threats



Lack of Transparency

# Trend: Focus on the Customer

Understanding customers is the foundation to a sustainable competitive advantage. Organizations can no longer wait to embrace the power of advanced analytics to gain insights and evaluate opportunities that will improve cross-selling, increase up-selling and enhance customer value



Source: February 2015 Digital Banking Report



# Big Data Business Model Maturity Index

Measures the degree to which the organization has integrated data and analytics into their business models



Business Monitoring



Business Insights



Business Optimization



Data Monetization



Business Metamorphosis

# The Power Of The Decisions

## Targeted Business Initiative

↳ Business Stakeholders (Personas | UEX)

↳ Decisions (Prescriptive & Recommendations)

↳ Questions and Metrics (Predictive Variables)

↳ Data Sources (Data Roadmap)

↳ Data Architecture (Data Lake)

↳ Technology (Technology Roadmap)

# Big Data Approach

# What's Important To Chipotle?



## Chipotle Business Strategy

- Continue to build a people culture that attracts and empowers top performers
- Continue to grow revenues (up 20.3% in 2012) by opening new stores (opened 183 in 2012)...
- ...and increase comparable restaurant sales growth (7.1% in 2012)
- Marketing focused on building the Chipotle brand and engaging with our customers in ways that create stronger, deeper bonds

# Chipotle 2012 Business Strategy

- Initiative #1: Continue to grow revenues (up 20.3% in 2012) by opening new stores (opened 183 in 2012)
- **Initiative #2: Increase comparable restaurant sales (by 7.1% in 2012)**
- Initiative #3: Marketing building Chipotle brand and engaging customers to create stronger, deeper bonds

## Key Entities & Key Decisions

Entity 1:

Entity 2:

Financial Drivers

|  |
|--|
|  |
|  |
|  |
|  |
|  |



# Chipotle 2012 Business Strategy

- Initiative #1: Continue to grow revenues (up 20.3% in 2012) by opening new stores (opened 183 in 2012)
- **Initiative #2: Increase comparable restaurant sales (by 7.1% in 2012)**
- Initiative #3: Marketing building Chipotle brand and engaging customers to create stronger, deeper bonds

## Key Entities & Key Decisions

Entity 1: Develop individual store profiles including traffic patterns, customer demographics and product preferences by time of day/day of week

Entity 2: Leverage local sporting and entertainment events to drive store traffic at slow stores and during slow times

- Which promotions...
- Which local events...
- What staffing and inventory...

Financial Drivers

|  |
|--|
|  |
|  |
|  |
|  |
|  |

# Chipotle 2012 Business Strategy

- Initiative #1: Continue to grow revenues (up 20.3% in 2012) by opening new stores (opened 183 in 2012)
- **Initiative #2: Increase comparable restaurant sales (by 7.1% in 2012)**
- Initiative #3: Marketing building Chipotle brand and engaging customers to create stronger, deeper bonds

## Key Entities & Key Decisions

Entity 1: Develop individual store profiles including traffic patterns, customer demographics and product preferences by time of day/day of week

Entity 2: Leverage local sporting and entertainment events to drive store traffic at slow stores and during slow times

- Which promotions...
- Which local events...
- What staffing and inventory...

## Financial Drivers

Increase store traffic (acquire new customers, frequency of repeat customers)

Increase shopping bag revenue and margins (cross-sell complementary products, up-sell)

Increase number of corporate events (catering, repeat catering events)

Improve promotional effectiveness (Halloween Boo-ritto, Christmas gift cards, graduation gift cards)

Improve new product introduction effectiveness (seasonal, holiday)

# Potential Chipotle Data Sources

- Point of Sales Transactions
- Market Baskets
- Product Master
- Store Demographics
- Competitive Stores Sales
- Store Manager Notes
- Employee Demographics
- Store Manager Demographics
- Consumer Comments
- Weather
- Traffic Patterns
- Yelp
- Zillow / Realtor.com
- Twitter / Facebook / Instagram
- Twellow /Twellowhood
- Zip Code Demographics
- EventBrite
- MaxPreps
- Mobile App
- ...

# Data Value Assessment

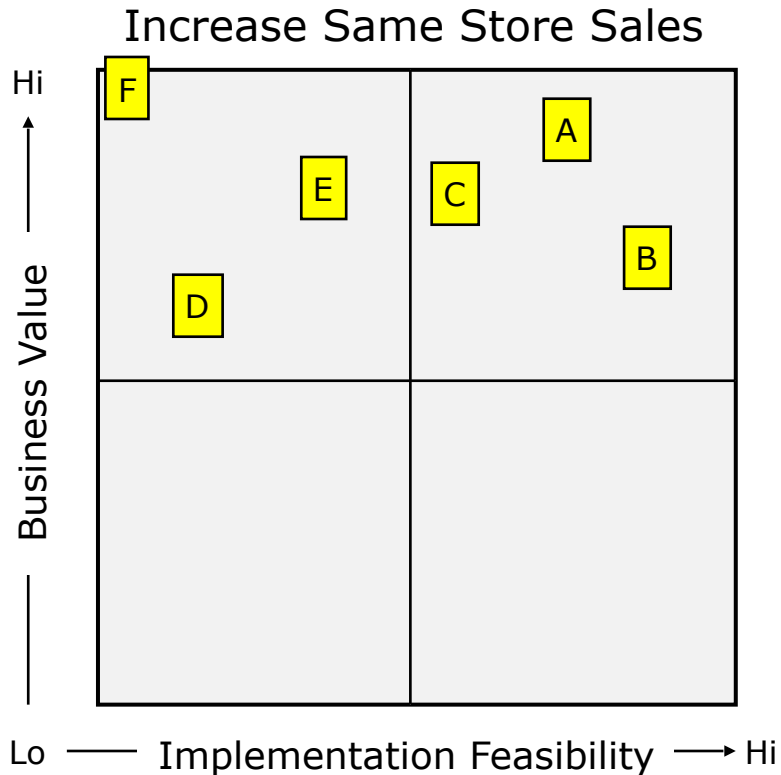
| Data Source                   | Increase Store Traffic | Increase Shopping Bag Revenue | Increase # Corporate Events | Increase Promotional Effectiveness | Improve NPI Effectiveness |
|-------------------------------|------------------------|-------------------------------|-----------------------------|------------------------------------|---------------------------|
| Point of Sales Transactions   | 4                      | 4                             | 4                           | 4                                  | 4                         |
| Market Baskets                | 4                      | 4                             | 2                           | 4                                  | 4                         |
| Store Demographics (Zip Code) | 3                      | 3                             | 1                           | 3                                  | 3                         |
| Local Competitive Stores      | 2                      | 2                             | 2                           | 2                                  | 2                         |
| Store Manager Demographics    | 1                      | 1                             | 3                           | 1                                  | 1                         |
| Consumer Comments             | 3                      | 3                             | 3                           | 3                                  | 2                         |
| Social Media                  | 2                      | 1                             | 1                           | 3                                  | 3                         |
| Weather                       | 3                      | 1                             | 1                           | 1                                  | 1                         |
| Local Events                  | 4                      | 2                             | 1                           | 2                                  | 1                         |
| Traffic                       | 3                      | 1                             | 1                           | 2                                  | 1                         |
| Zillow                        | 1                      | 2                             | 2                           | 2                                  | 2                         |

# Implementation Feasibility Assessment

| Data Source                    | Ease of Acquiring | Cleanliness | Accuracy | Granularity | Cost |
|--------------------------------|-------------------|-------------|----------|-------------|------|
| Point of Sales Transactions    | 4                 | 4           | 4        | 4           | 4    |
| Market Baskets                 | 4                 | 4           | 4        | 4           | 4    |
| Store Demographics (Zip Codes) | 4                 | 4           | 4        | 4           | 4    |
| Competitive Stores Sales       | 2                 | 2           | 2        | 2           | 1    |
| Store Manager Demographics     | 4                 | 4           | 4        | 4           | 4    |
| Consumer Comments              | 2                 | 2           | 1        | 2           | 3    |
| Social Media                   | 1                 | 1           | 1        | 1           | 1    |
| Weather                        | 3                 | 3           | 2        | 2           | 2    |
| Local Events                   | 1                 | 2           | 2        | 2           | 1    |
| Traffic                        | 2                 | 2           | 1        | 3           | 2    |
| Zillow                         | 1                 | 1           | 2        | 2           | 1    |



# Prioritization Matrix

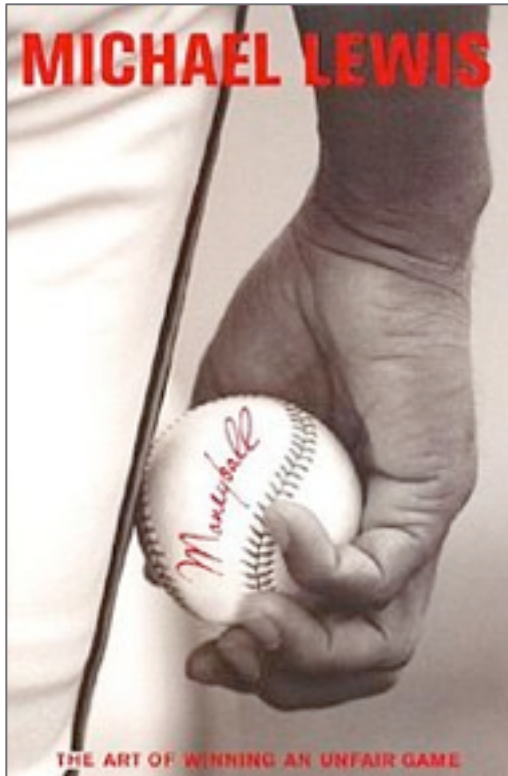


## Use Cases

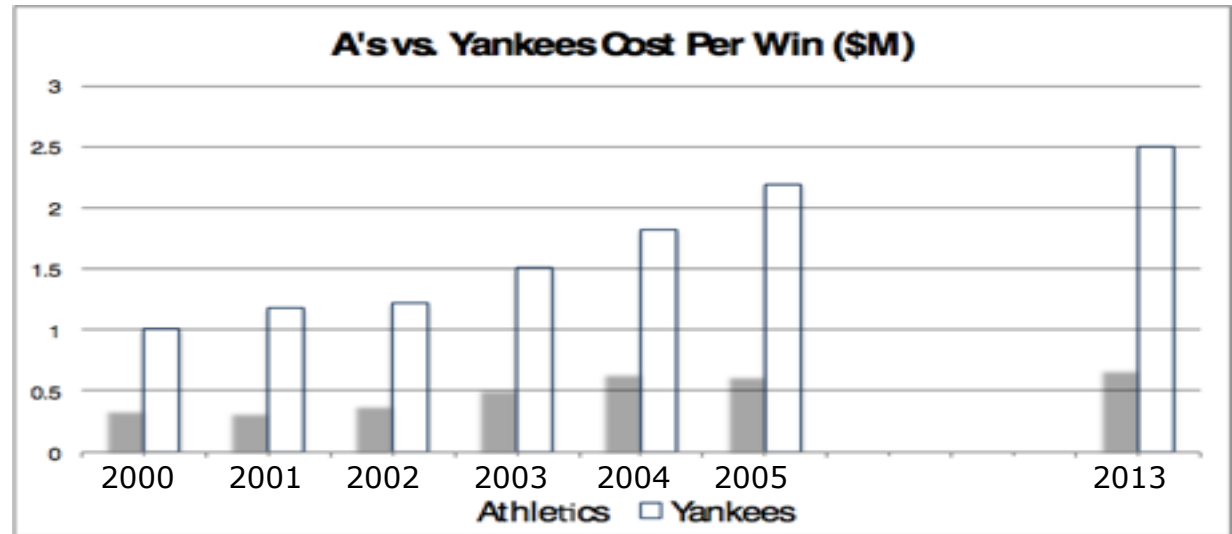
- A Increase Store Traffic
- B Increase Shopping Basket Size
- C Increase Corporate Events
- D Improve New Product Introductions
- E Improve Promotional Effectiveness
- F Increase Special Events (birthdays, parties)

# Thinking Like A Data Scientist

# "MoneyBall": Identifying The *Right* Metrics



**Data Science** is about identifying variables and metrics that are better predictors of performance



"Moneyball: The Art of Winning an Unfair Game," Michael Lewis

# Thinking Like A Data Scientist

To get full value out of big data, data scientists not sufficient...

Must get Business Users to “**Think Like a Data Scientist**”

- Identifying the right **questions** to ask, **decisions** to make, **predictions** to create and **hypotheses** to test
- Evolving from descriptive questions to **predictive** questions (What is likely to happen?) and **prescriptive** questions (What should I do?)
- Brainstorming different data sources and metrics that might yield **better predictors of business performance**
- Blending metrics and variables into predictive “**scores**” that can be used to support decision-making and process optimization
- Identifying where and how analytic results will be presented in order to **influence** customer behaviors and **empower** front-line employees

# Evolving The Business Questions

## What Happened? (Descriptive/BI)

- What are most popular products/product combinations sold at Store X?
- What products are sold at Store X after a high school football game?
- What are the age, income, educational demographics near Store X?
- What businesses, schools and shopping malls reside near Store X?

## What Will Happen? (Predictive Analytics)

- How many customers will visit the store during Sunday's Farmer's Market?
- How much extra chicken and barbacoa will we sell before and after Friday's high school football game?
- How many extra workers are needed for Saturday's college football game?
- How many customers will visit during the Halloween Boo-ritto promotion?
- How will Friday's predicted rain storm affect store traffic?

## What Should I do? (Prescriptive Analytics)

- Run Buy-One-Get-One (BOGO) Burrito promotion on Wednesday from 7:00 – 9:00pm to attract 55 - 75 more college students
- Add 2 more workers 11:00am – 2:00pm and 5:00pm - 9:00pm on Tuesday, May 5
- Increase chicken production 20% 5:00pm - 9:00pm on Tuesday, May 5
- Order 12% more beef and 20% more chicken for the Friday, May 8 high school baseball game



# Chipotle “Store” Questions

*Initiative: Increase Same Store Sales: **STORE** Questions*

## **Descriptive Analytics (What happened?)**

- What are most popular products/product combinations sold at Store X?
- What product and basket are sold at Store X after the nearby football game?
- What are the demographics of the customers living near Store X?
- What businesses, schools and shopping malls reside near Store X?

## **Predictive Analytics (What is likely to happen?)**

- How many customers will visit Store X during the Farmer’s Market?
- How many customers will visit Store X during Saturday’s football game?
- How many customers will visit Store X for the Boo-ritto promotion?

## **Prescriptive Analytics (What should I do?)**

- What promotions should Store X run to bring in more college students?
- How many more staff will I need for Friday night’s football game?



# "By" Analysis To Uncover New Metrics

**"By" Analysis technique leverages a business stakeholders natural question and query process to uncover:**

- Additional data sources and metrics
- Additional dimensional characteristics
- Additional areas of analytics exploration

**"By" Analysis examples:**

- "I want to see sales and product margin **by...** product category, store, store remodel date, day of week, store demographics, and customer demographics"
- "I want to trend hospital admissions **by...** disease category, zip code, patient demographics, hospital size, area demographics and day of week"
- "I want to compare current versus previous maintenance issues **by...** turbine, turbine manufacturer, maintenance person, date last serviced and weather conditions"

# Chipotle “By” Analysis

***Business Initiative: Increase Same Store Sales***

**What are the most popular products *by...***

- Store location
- Store size
- Store open date
- Store last remodel date
- Local demographics
- Local house values
- Local economic conditions
- Products sold
- Product add-on items
- Product configuration
- Time of day
- Day of week (weekends)
- Holidays
- Seasonality
- Weather conditions
- Traffic patterns
- Miles from high school
- Miles from mall
- Miles from business park
- Local sporting events
- Local entertainment events
- ...

# Create Actionable “Scores”

- “**Score**” is a dynamic rating or grade (e.g., IQ score) standardized to aid in comparisons, performance tracking and decision-making; ***predicts likelihood of certain actions or outcomes***
- Scores support: What decision are you trying to make? What actions or outcomes are trying to predict?

| Financial  | Credit Cards  | Manufacturing   | Gaming/Hospitality  |
|--|---|---|---|
| <ul style="list-style-type: none"> <li>• FICO</li> <li>• Retirement Readiness</li> <li>• Risk Tolerance</li> </ul> | <ul style="list-style-type: none"> <li>• Attrition Risk</li> <li>• Fraud Risk</li> <li>• Product Preferences</li> </ul> | <ul style="list-style-type: none"> <li>• Equipment Maintenance</li> <li>• Supplier Reliability</li> <li>• Supplier Quality</li> </ul> | <ul style="list-style-type: none"> <li>• Customer LTV</li> <li>• Lifestage</li> <li>• Gaming Preferences</li> </ul> |
| Education  | Healthcare  | Utilities   | Pro Sports  |
| <ul style="list-style-type: none"> <li>• Graduation Readiness</li> <li>• Cohorts Influence</li> </ul>              | <ul style="list-style-type: none"> <li>• Wellness Status</li> <li>• Stress Risk</li> </ul>                              | <ul style="list-style-type: none"> <li>• Energy Efficiency</li> <li>• Conservation Effectiveness</li> </ul>                           | <ul style="list-style-type: none"> <li>• Fatigue Factor</li> <li>• Motivation Factor</li> </ul>                     |

# Score Example: FICO

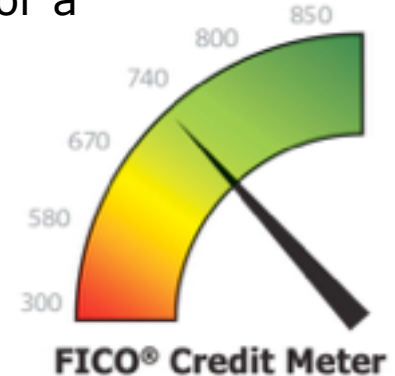
## Descriptive Metrics

- Credit card balances?
- Credit card payment history?
- Length of credit cards ownership?
- Credit utilization?
- Number of credit cards?
- Home mortgage payment?
- Student loan payments?
- Checking account balance?
- Savings account balance?

## Predictive Score

FICO Score is used by lenders to **predict your ability to repay a loan:**

- Credit worthiness in applying for credit or a loan
- Interest rate and loan terms that you receive for a home mortgage or car loan



# Creating Chipotle “Store” Scores

**Business Initiative:** Increase Same Store Sales

## What are the most popular products *by*...

- Store location
- Store size
- Store open date
- Store last remodel date
- Local demographics
- Local house values
- Local economy
- Products sold
- Product add-on items
- Product configuration
- Time of day
- Day of week (weekends)
- Holidays
- Seasonality
- Weather conditions
- Traffic patterns
- Miles from high school
- Miles from mall
- Miles from business park
- Local sporting events
- Local entertainment events
- ...

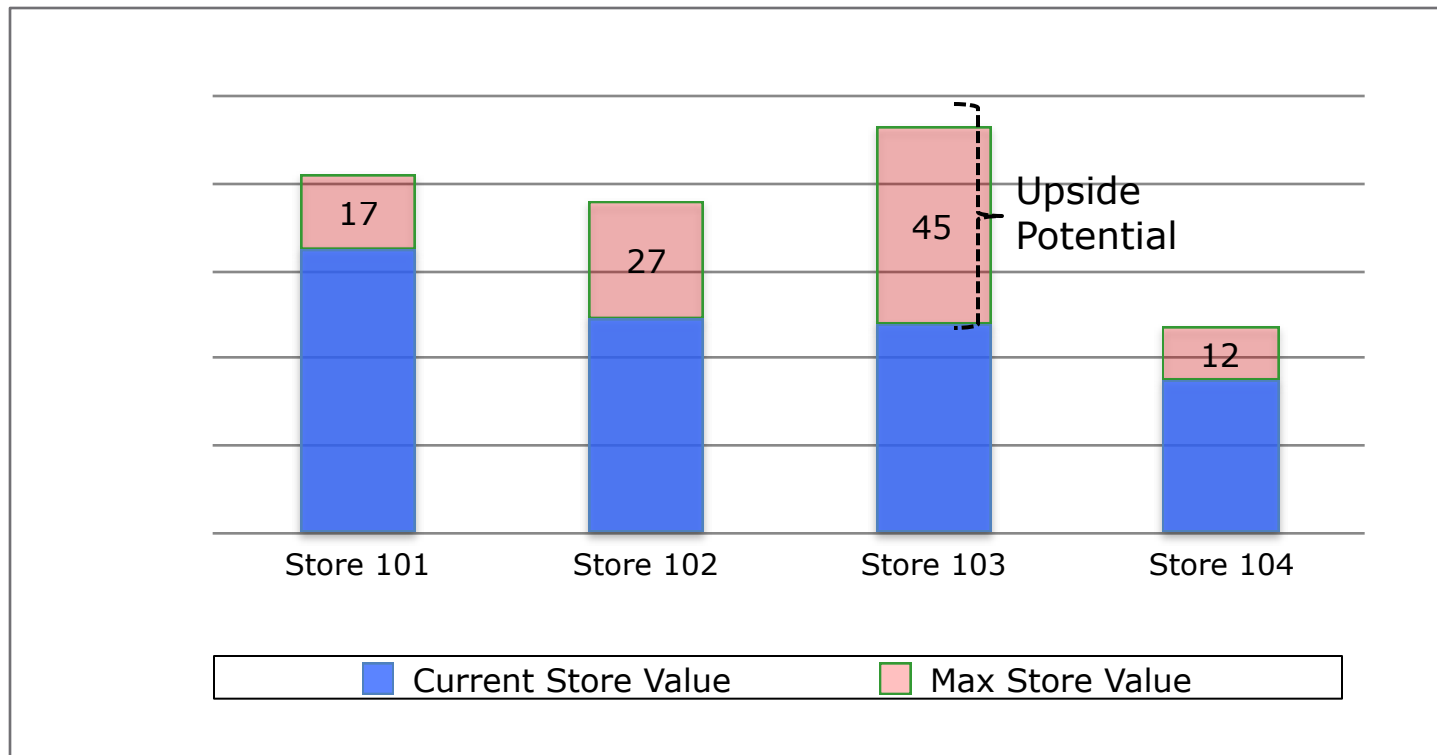
**Economic  
Potential**

**Local  
Activity**

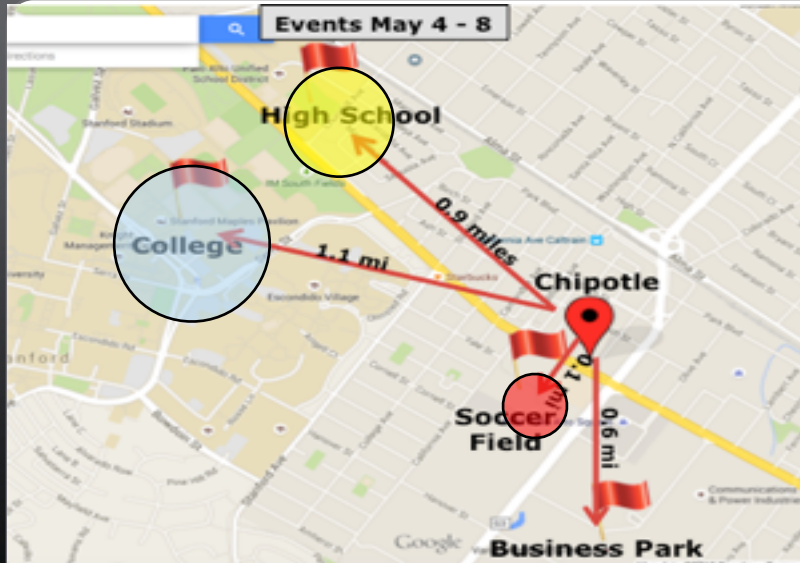
EMC<sup>2</sup>

# Leverage Score To Prioritize Spend

Create **Max Store Value** based upon **Economic Potential** and **Local Vitality** scores to prioritize spend on stores with most upside potential



# Chipotle Store Manager Dashboard



- ① Firebirds
- ② Palo Alto Sol
- ③ Jack In The Box
- ④ Phillie CheeseSteak

Top 25 Customers at Risk to Churn by VPO

| Risk Rank | VPO        | SAAS      | Stream Conflicts | Packet Errors | VESS Downstream Errors CRC | VESS Downstream Errors Seconds | VESS Downstream Severe Errors Seconds |
|-----------|------------|-----------|------------------|---------------|----------------------------|--------------------------------|---------------------------------------|
| 1         | Austin, TX | 50230115  | ↓                | ↑             | ↓                          | ↓                              | ↓                                     |
| 2         | Austin, TX | 502015396 | ↓                | ↑             | ↓                          | ↓                              | ↓                                     |
| 3         | Austin, TX | 502311910 | ↓                | ↑             | ↓                          | ↓                              | ↓                                     |
| 4         | Austin, TX | 502394844 | ↓                | ↑             | ↓                          | ↓                              | ↓                                     |
| 5         | Austin, TX | 502158475 | ↓                | ↑             | ↓                          | ↓                              | ↓                                     |
| 6         | Austin, TX | 500990201 | ↓                | ↑             | ↓                          | ↓                              | ↓                                     |
| 7         | Austin, TX | 501818861 | ↑                | ↑             | ↓                          | ↓                              | ↓                                     |
| 8         | Austin, TX | 502110481 | ↓                | ↑             | ↓                          | ↓                              | ↓                                     |
| 9         | Austin, TX | 50324624  | ↓                | ↑             | ↓                          | ↓                              | ↓                                     |
| 10        | Austin, TX | 500988177 | ↓                | ↑             | ↓                          | ↓                              | ↓                                     |
| 11        | Austin, TX | 502187750 | ↓                | ↑             | ↓                          | ↓                              | ↓                                     |
| 12        | Austin, TX | 501955345 | ↑                | ↑             | ↓                          | ↓                              | ↓                                     |

## Recommendations

- ❑ Add 2 more workers 11:00am – 2:00pm and 5:00pm – 9:00pm for Cinco de Mayo (Tuesday, May 5) [\[More\]](#) + -
- ❑ Increase chicken production 20% 5:00pm - 9:00pm for Cinco de Mayo (Tuesday, May 5) [\[More\]](#) + -
- ❑ Order 12% more chicken and 10% more beef for Friday night's high school baseball game (Friday, May 8) [\[More\]](#) + -
- ❑ Add 2 more workers 10:00 – 2:00 pm for Business Park catering (Friday, May 8) [\[More\]](#) + -

Give away free Chips (but no guacamole) Friday, May 8 evening (7:00 – 10:00 pm)

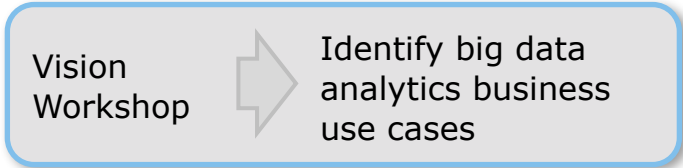
| Customer Name           | Documents | Change Orders | Analytics | Trends |
|-------------------------|-----------|---------------|-----------|--------|
| Illinois_Missouri       | 4043      | 204           | 517       |        |
| Indiana_Michigan        | 3465      | 204           | 294       |        |
| Pri_States              | 2877      | 295           | 281       |        |
| New_England             | 1261      | 123           | 176       |        |
| Greater_Midwest         | 2027      | 262           | 220       |        |
| North_Texas             | 4827      | 262           | 361       |        |
| Oklahoma_Arkansas       | 1195      | 129           | 192       |        |
| South_Texas             | 2941      | 276           | 296       |        |
| Carolina                | 279       | 32            | 29        |        |
| Georgia                 | 759       | 191           | 95        |        |
| Gulf_States             | 135       | 89            | 85        |        |
| Kentucky_Tennessee      | 122       | 9             | 19        |        |
| North_Florida           | 427       | 26            | 27        |        |
| South_Florida           | 4993      | 212           | 676       |        |
| Greater_Los_Angeles     | 3977      | 236           | 366       |        |
| Northam_California      | 3324      | 212           | 321       |        |
| San_Diego_Nevada_Hawaii | 1247      | 119           | 131       |        |



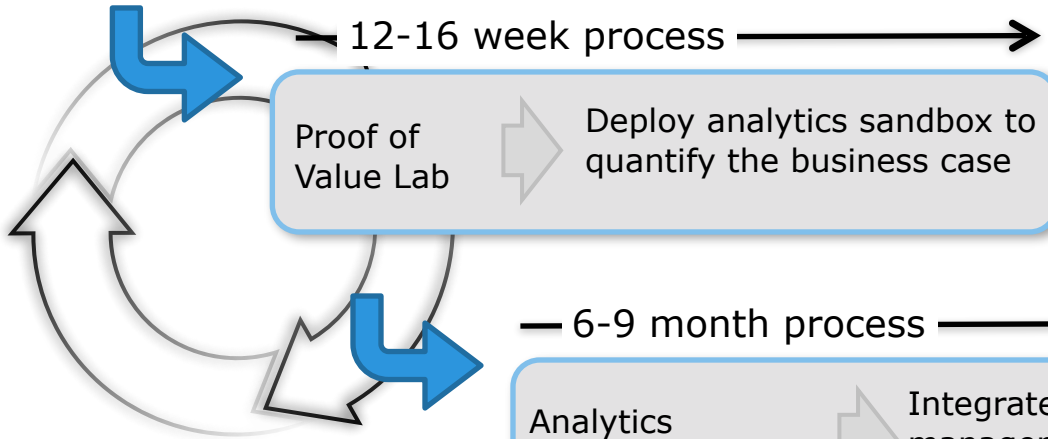
# How To Get Started...

# Prioritize Your Business Initiatives

— 2-3 week process —>

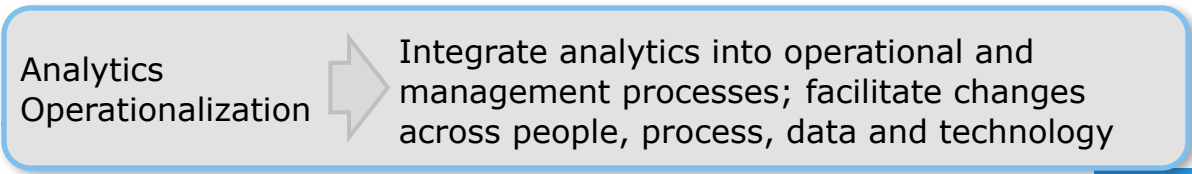


- Start with business initiative / challenge
- Drive Business – IT organizational alignment
- Brainstorm and assess potential data sources
- Build *illustrative* analytics (data science)
- Identify, “flesh out” and prioritize use cases



- For prioritized use case:
- Build business case (*high-level ROI*)
  - Prove analytic feasibility (analytic lift)
  - Create UEX mockup
  - Install big data technology
  - Build big data architecture
  - Deliver “Go Live” recommendations

— 6-9 month process —>



Repeat the process for identifying business cases

# Best Practices

- Efforts that take a technology-focused approach (vs. a use-case focused approach) often lose focus and waste resources on a technology “shoot out”
- Look for use-cases that are both valuable to mission and feasible from a data source perspective
- Start small, have initial use case success and build incrementally
- Initially look to prove value vs. large, costly data-warehouse effort
- Manage Big Data efforts “like a program”
- Take the portfolio approach; establish / compare business cases

EMC<sup>2</sup>®