

Building and Sustaining Predictive Analytics Capabilities

Presented by: Derek Kane

Welcome to the Predictive Analytics Workshop

- ❖ Predictive analytics has always been a sort of mystical science for many, including myself. This course is designed to break down the techniques and show you how to implement them directly into your existing business processes and show how to draw from them in practical applications.
- ❖ This training program will introduce the core concepts and technologies needed to launch predictive analytics and machine learning processes in organizations of any size and with any ERP system or BI technology.
- ❖ In this workshop, we will go through the theory and techniques used to apply predictive analytics methodologies, how to expertly evaluate the effectiveness of the tools, and then work through practical examples that fortify the methods in practice.



Welcome to the Predictive Analytics Workshop

We will cover predictive modeling topics that include:

- ❖ Getting the Work Environment Setup
- ❖ Evaluating and Preparing Data for Advanced Analytics
- ❖ Model Selection Techniques
- ❖ Time-Series Forecasting (Financial Planning / Budgeting / Forecasting)
- ❖ Regression Analysis (Sales models/ Advanced Price Models / Optimization)
- ❖ Social Media and Text Analytics (Survey Analysis, Facebook, Twitter, News Feeds)
- ❖ Market Basket Association (Product Recommendation Engine)
- ❖ Clustering and Classification techniques. (Customer / Market Segmentation)
- ❖ Artificial Intelligence systems (Neural Networks & SVM)
- ❖ Decision Tree Models (CART, Random Forests, C5.0)

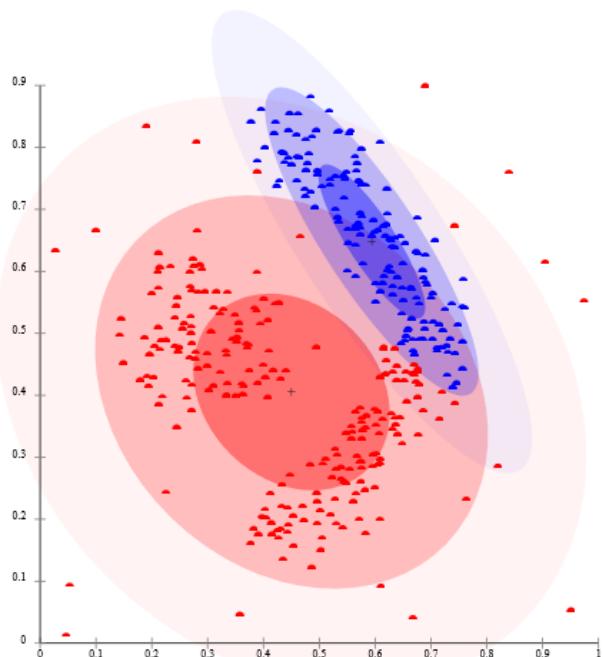
Welcome to the Predictive Analytics Workshop

Here are some of the prediction practical examples which will cover in detail:

- ❖ Robustly forecasting product sales quantities taking seasonality and trend into account.
- ❖ Identifying cross selling promotional opportunities for consumer goods.
- ❖ Identify the price sensitivity of a consumer product and identify the optimum price point that maximizes net profit.
- ❖ Optimizing product location at a supermarket retail outlet.
- ❖ Modeling variables impacting customer churn and refining strategy.



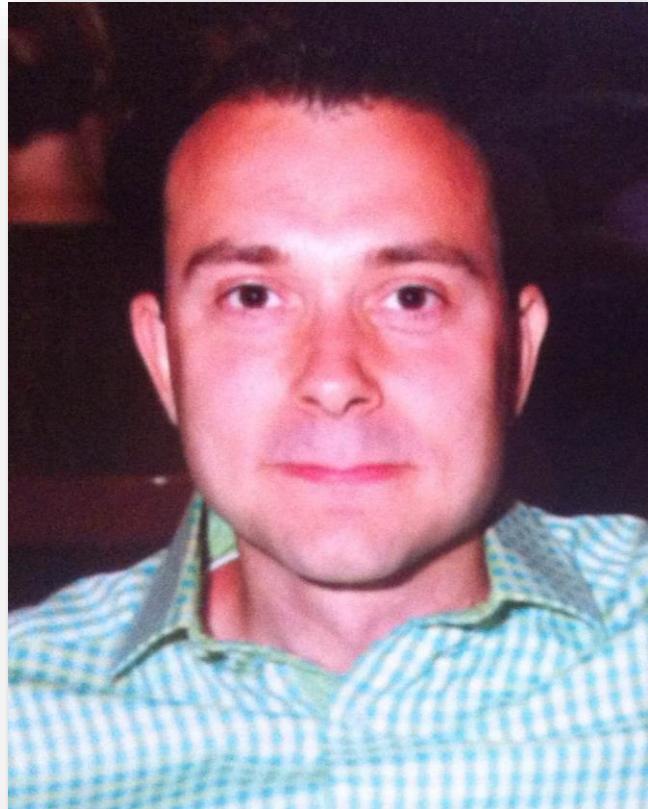
Welcome to the Predictive Analytics Workshop



- ❖ Develop an Artificial Neural Network to determine whether or not someone will default on a home mortgage.
- ❖ Understand consumer sentiment based off of unstructured text data.
- ❖ Identifying the political party based off of individual voting records.
- ❖ Forecasting women's conviction rates based off external macroeconomic factors.
- ❖ Develop a text based search engine.
- ❖ Creating a machine learning model that accurately detects breast cancer.

About Me

- ❖ Reside in Wayne, Illinois
- ❖ Active Semi-Professional Classical Musician (Bassoon).
- ❖ Married my wife on 10/10/10 and been together for 10 years.
- ❖ Pet Yorkshire Terrier / Toy Poodle named Brunzie.
- ❖ Pet Maine Coons' named Maximus Power and Nemesis Gul du Cat.
- ❖ Enjoy Cooking, Hiking, Cycling, Kayaking, and Astronomy.
- ❖ Self proclaimed Data Nerd and Technology Lover.



Overview of Topics

- ❖ Introduction
- ❖ Building an Analytics Infrastructure
 - ❖ Business Intelligence
 - ❖ Predictive Analytics
 - ❖ The Age of Big Data
- ❖ Applied Predictive Analytics Example



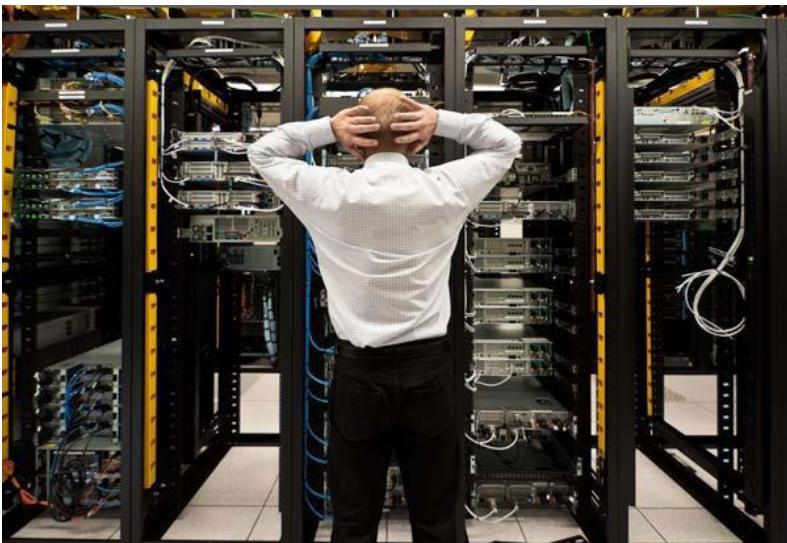
"Prediction is very difficult, especially if it's about the future." – Niels Bohr

Predictive Analytics in the Marketplace

- ❖ In the 21st century, organizations that do not focus increasing their predictive analytical capabilities will erode their competitive advantage and run the risk of failing.
- ❖ Regardless of size and market, all business organizations utilize some level of analytics for decision making. The most successful organizations have some degree of predictive analytics incorporated.
- ❖ From my experiences, there is a large volume of reports and information that flow throughout an organization.
- ❖ Most of this information doesn't effectively support the organizations operational goals and there is a real opportunity to streamline and strengthen analytics capabilities.



Predictive Analytics in the Marketplace



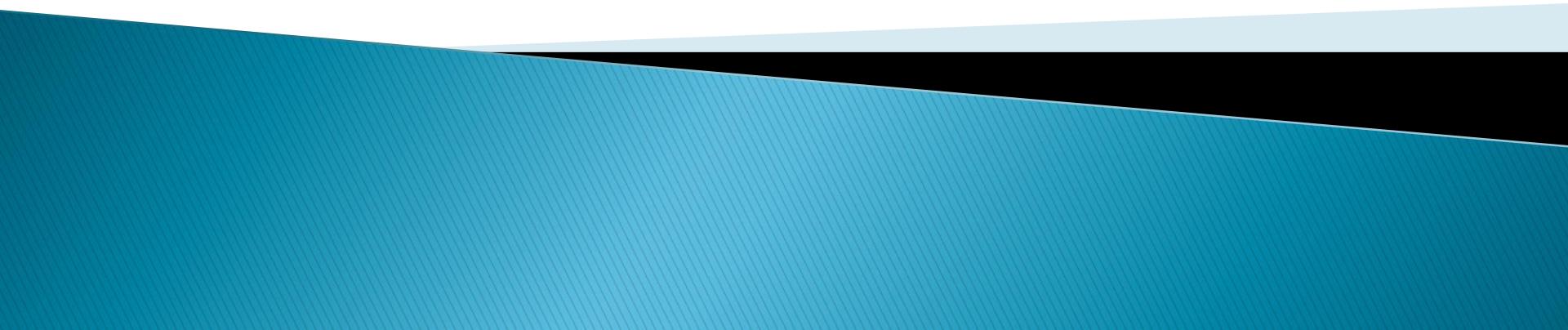
- ❖ When decision makers are considering strengthening analytics capabilities through predictive modeling, they traditionally think that this will be a costly endeavor and rely on IT managers to assess the cost/benefit.
- ❖ Ex. Large investment in technical infrastructure, human resource talent, and specialized mathematical and statistical skills.
- ❖ There is now a need for specialized talent that focuses exclusively on analytics systems and solutions which goes beyond the scope traditional IT professionals.
- ❖ This is the area that we specialize within. We can teach you how to properly build this capability without wasting time and effort on dead end, costly solutions.

Predictive Analytics in the Marketplace

- ❖ We believe that the techniques of high end predictive analytics can be easily taught to analytically focused individuals and/or organizations at any stage of development, with little to no investment in IT.
- ❖ The next focus of the presentation will explain the key technologies required for building enterprise level predictive analytics solutions that targets non-technical audiences.
- ❖ While we advocate using these technologies in tandem with each other in order to achieve a maximum effect, they are not essential to be in place in order to experience the immediate benefits that predictive analytics offers.



Introduction to Business Intelligence Systems



Challenges Facing Decision Makers

1 in 3

business leaders frequently make critical decisions without the information they need

1 in 2

don't have access to the information across their organization needed to do their jobs

19+ Hours

spent by knowledge workers each week just searching for and understanding information

What is Business Intelligence?

Multiple Definitions:

- ❖ Business Intelligence (BI) refers to skills, processes, technologies, applications and practices used to support decision making.
- ❖ Business Intelligence is a broad category of applications and technologies for gathering, storing, analysing, and providing access to data to help clients make better business decisions.
- ❖ A system that collects, integrates, analyses and presents business information to support better business decision making.
- ❖ Business Intelligence is an environment in which business users receive information that is reliable, secure, consistent, understandable, easily manipulated and timely...facilitating more informed decision making.

What is Business Intelligence?

Improving organizations by providing business insights to all employees leading to better, faster, more relevant decisions



Benefits of Business Intelligence

"Get the right information to the right people at the right time"

- ❖ Improve Operational efficiency.
- ❖ Eliminate report backlog and delays.
- ❖ Find root causes and take action.
- ❖ Negotiate better contracts with suppliers and customers.
- ❖ Identify wasted resources and reduce inventory costs.
- ❖ Sell information to customers, partners, and suppliers.
- ❖ Leverage your investment in your ERP or data warehouse.
- ❖ Improve strategies with better marketing analysis.
- ❖ Give users the means to make better decisions.
- ❖ Challenge assumptions with factual information.

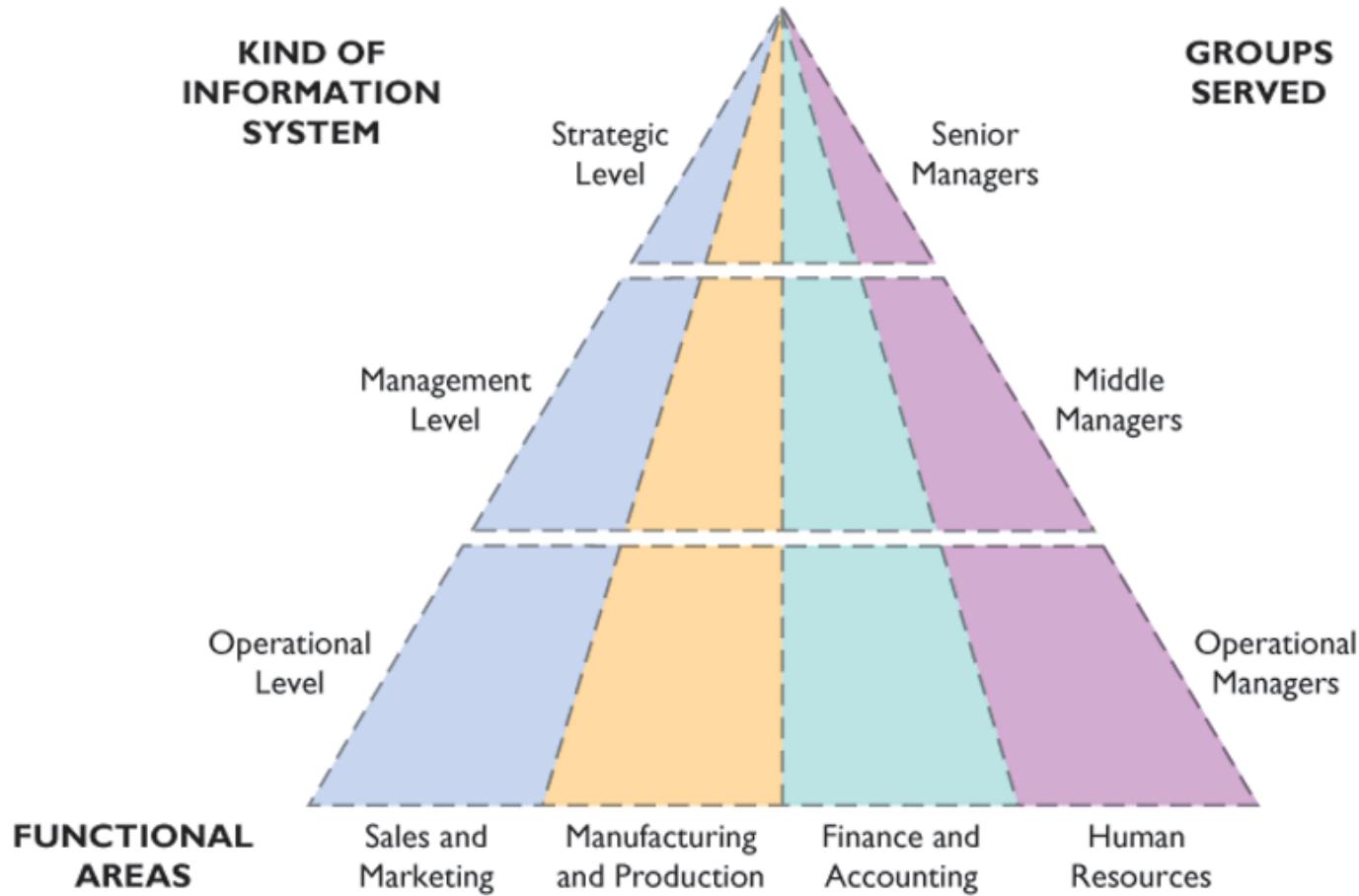
Why Do Organizations Need BI?



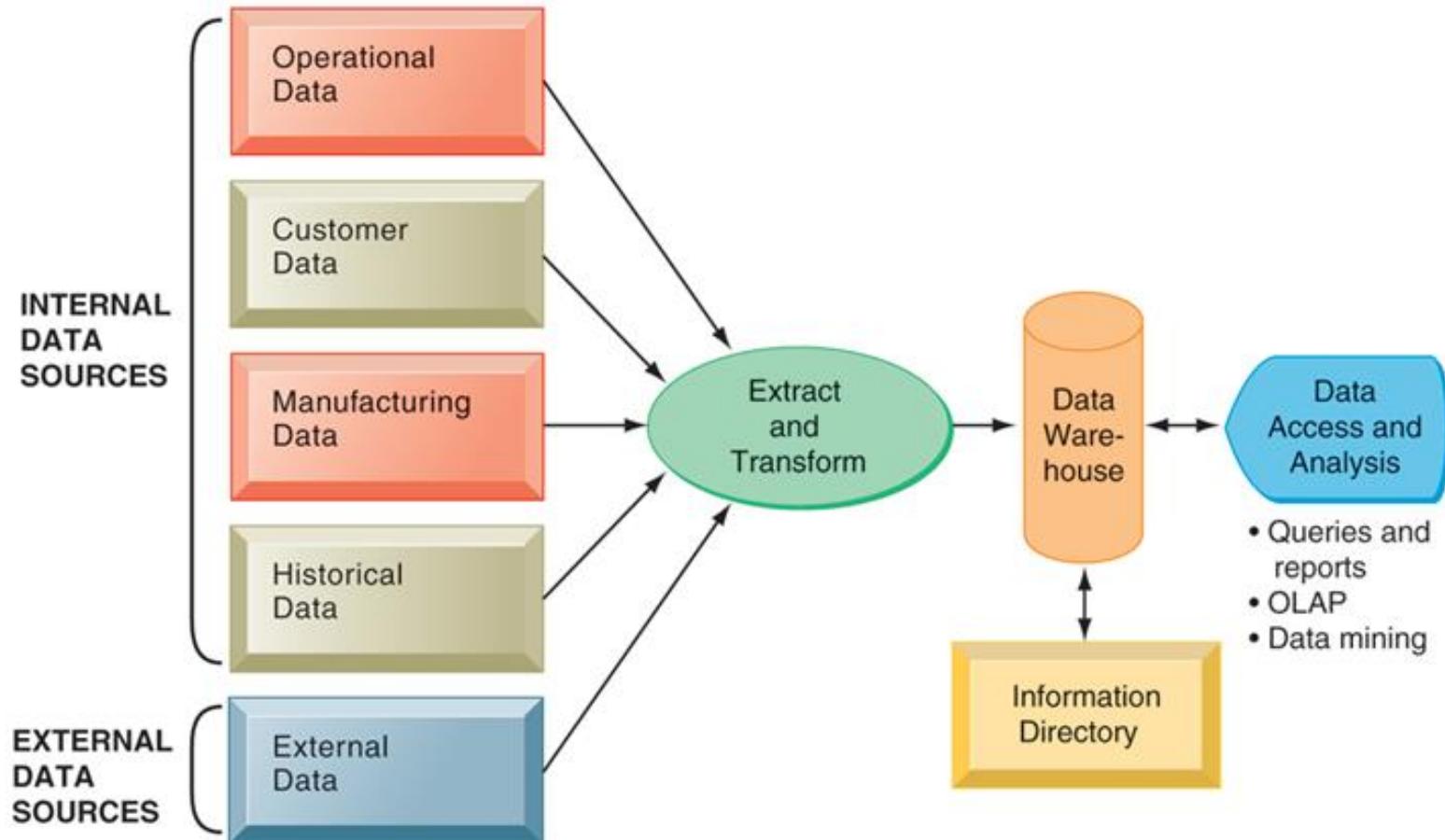
The Users of Business Intelligence

- ❖ Executives and business decision makers look at the business from a high level, performing limited analysis.
- ❖ Analysts perform complex, detailed data analysis.
- ❖ Information workers need static reports or limited analytic power.
- ❖ Line workers need no analytic capabilities as BI is presented to them as part of their job.

The Users of Business Intelligence



Data Warehousing



Challenges of Building BI Solutions

There are several issues inherent to any BI development:

- ❖ Data exists in multiple places.
- ❖ Data is not formatted to support complex analysis.
- ❖ Different kinds of workers have different data needs.
- ❖ What data should be examined and in what detail?
- ❖ How will users interact with that data?

Key Point:

Business users across all disciplines (led by a BI professional) must drive what should be in the data warehouse. Controllers will design BI systems with a financial focus, IT with an optimization emphasis but not user friendly, Engineers with a technical emphasis, etc...

Strategic Business Intelligence

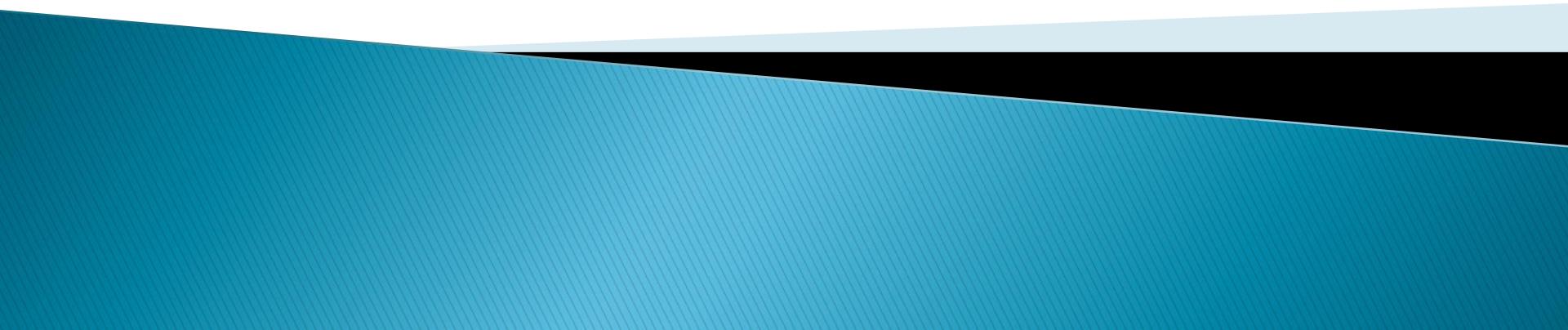
This must first come from the top of the organization as a key initiative and be integrated into the core business strategy/ culture.

The data infrastructure must first be developed, with the guidance of management, and refined to address the KPI's for the various business disciplines.

We should then develop objective performance measurements for the respective areas breaking down these KPI's to the deepest level (transactional) in the organizations ERP system.

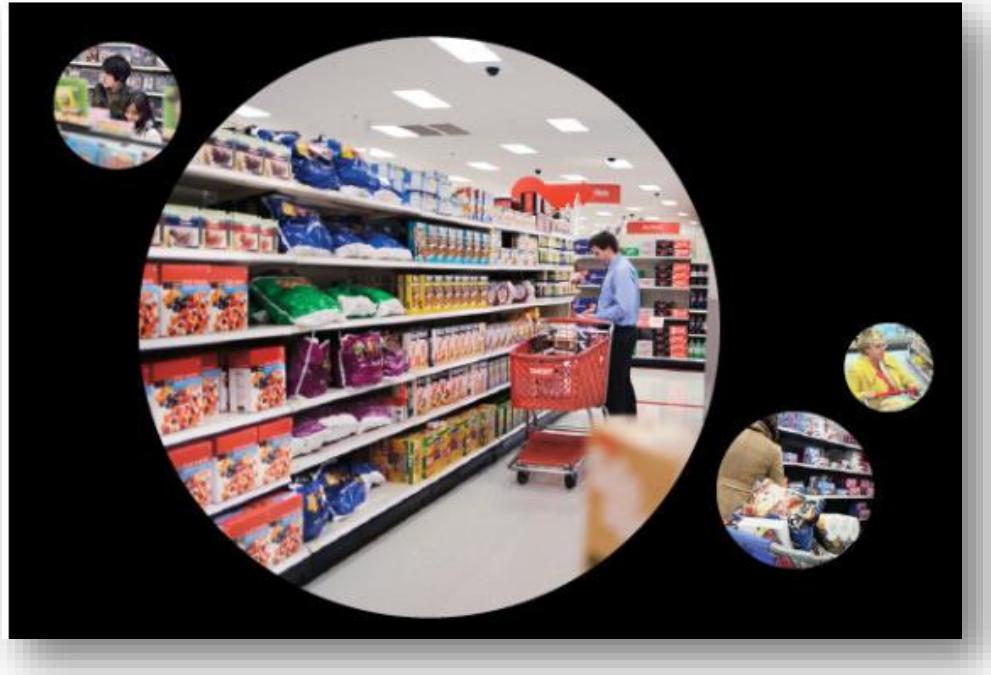
Having a bottom up approach is critical in order to identify cost saving opportunities and eliminate ineffective processes.

Introduction to Predictive Analytics Technologies



What is Predictive Analytics?

- ❖ A set of Business Intelligence technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events.
- ❖ Predictive Analytics is forward looking, using past events to anticipate the future.



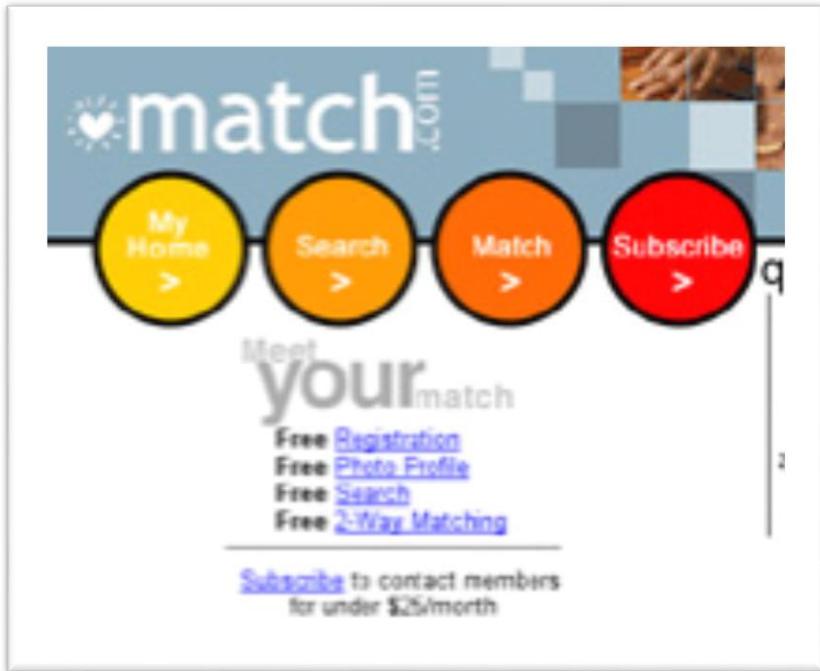
"A lot of companies want to do predictive analytics, but have yet to master basic reporting" Deloitte Consulting's Miller

Company Profile - Netflix



- ❖ Asked engineers and scientists around the world to solve what might have seemed like a simple problem: improve Netflix's ability to predict what movies users would like by a modest 10%.
- ❖ From \$5 million revenue in 1999 reached \$3.2 billion revenue in 2011 as a result of becoming an analytics competitor.
- ❖ By analyzing customer behavior and buying patterns created a recommendation engine which optimizes both customer tastes and inventory condition.

Company Profile - Match.com



- ❖ From experience with millions of singles from 24 countries since 1995, it can predict the likelihood of attractions between people.
- ❖ 95% of relationship can be predicted by analyzing as few as 10 characteristics in each profile.
- ❖ Members with accounts on Twitter, which only allows for messages of no more than 140 characters, have shorter relationships.
- ❖ People identifying themselves as Republicans are more willing to connect with Democrats than the reverse.

Source: Analyzing Love: Data Mining on Match.com – column in *AllAnalytics* - Nov 2011

What is Predictive Analytics?

“Use historical data to make certain predictions for the future”

Hindsight

“What is happening ?”

- Typical MIS or BI
- Cognos; Business Objects; Hyperion; ProClarity; etc
- Largely backward looking
- Referred to by many folks as ‘Analytics’ although it is not

Insight

“Why is it happening ?”

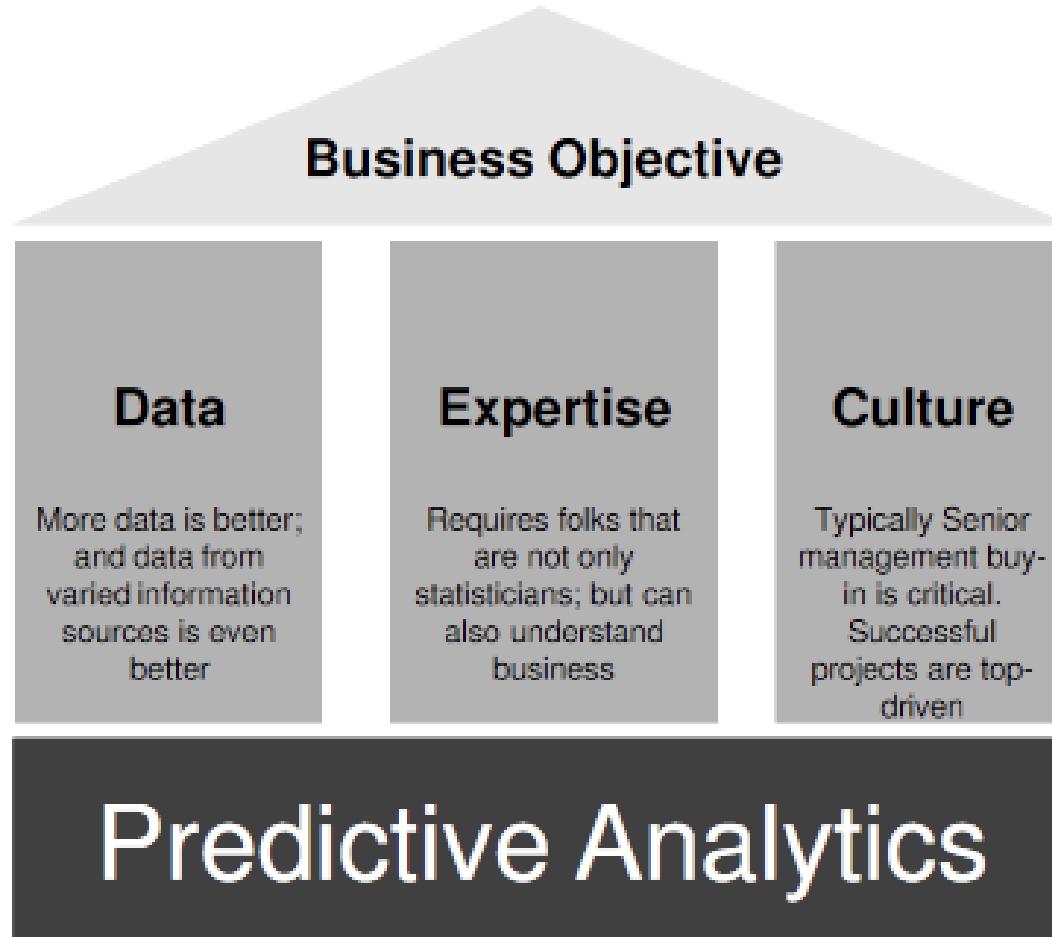
- Business analysis
- behavior analysis; trends; etc
- Gives us insights on what is happening and why

Foresight

“What will happen?”
“What should happen?”

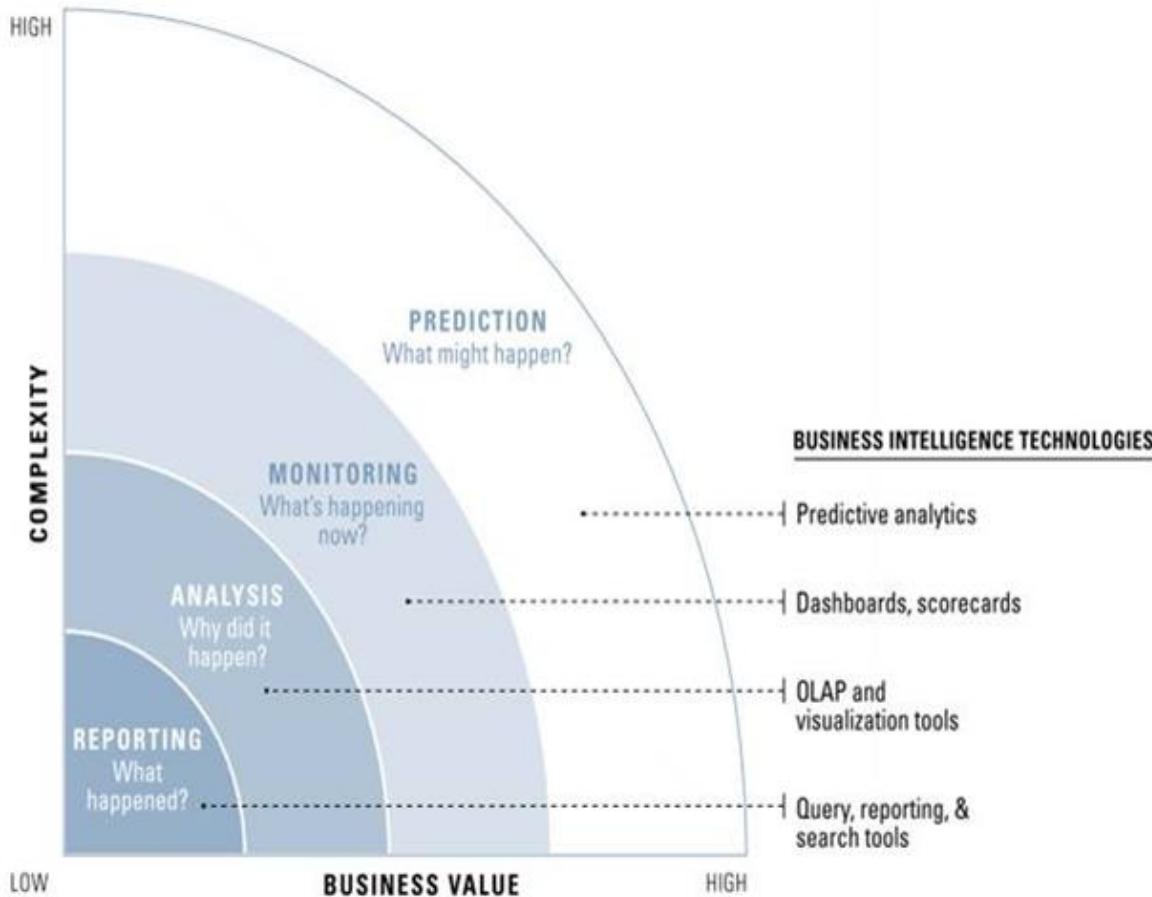
- **Predictive Analytics; forecasting; optimization, etc**
- **Uses past behavior to predict future outcomes**
- **Game changing**
- **Forward-looking**

What is Predictive Analytics?



What is Predictive Analytics?

The Spectrum of BI Technologies



Predictive Analytics Software

Industry Standard:

- ❖ SAS
- ❖ IBM/SPSS
- ❖ R

Additional Software:

- ❖ MS SQL Server, MS Excel (in combination with VBA),
- ❖ Matlab
- ❖ SAP announced BusinessObjects Predictive Analysis software
- ❖ Other vendors include Fair Isaac, Unica, Oracle, KXEN, Salford Systems, StatSoft, Insightful, Pentaho / WEKA, Quadstone and ThinkAnalytics
- ❖ WEKA
- ❖ Revolution R Enterprise (Revolution Analytics)
- ❖ Integrated workbenches provide support for Graphical modeling, automated testing, text analytics, analytical datamarts

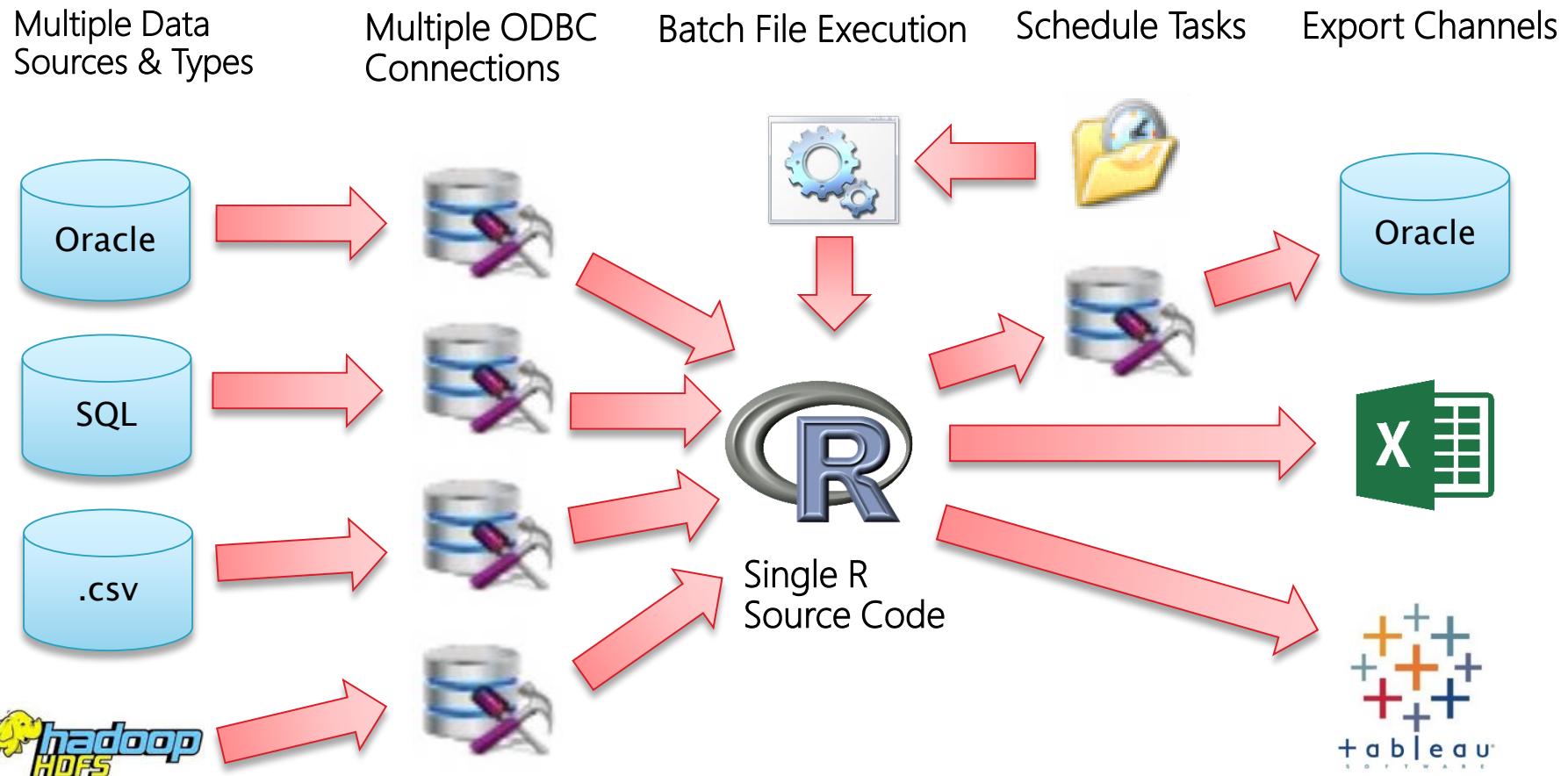
Source: TDWI – Predictive Analytics – Extending Value of Your Data Warehouse Investment by Wayne Eckerson

The Benefits of Using R

- ❖ Ideal investment cost. There is **no** upfront cost for using the technologies.
- ❖ Open Source – No black box mysteries, no proprietary lockdown into a specific tool.
- ❖ Most powerful statistical programming language.
- ❖ Easy to share across a business.
- ❖ Often works better/faster than Microsoft or Oracle products for data and analysis.
- ❖ Infinitely customizable to your problem and your products – vertical integration.
- ❖ Large support group of users worldwide. Most widely used data analysis software.
- ❖ Highly credible due to submission standards and university usage.
- ❖ Relatively easy to learn.



Predictive Analytics Data Flow

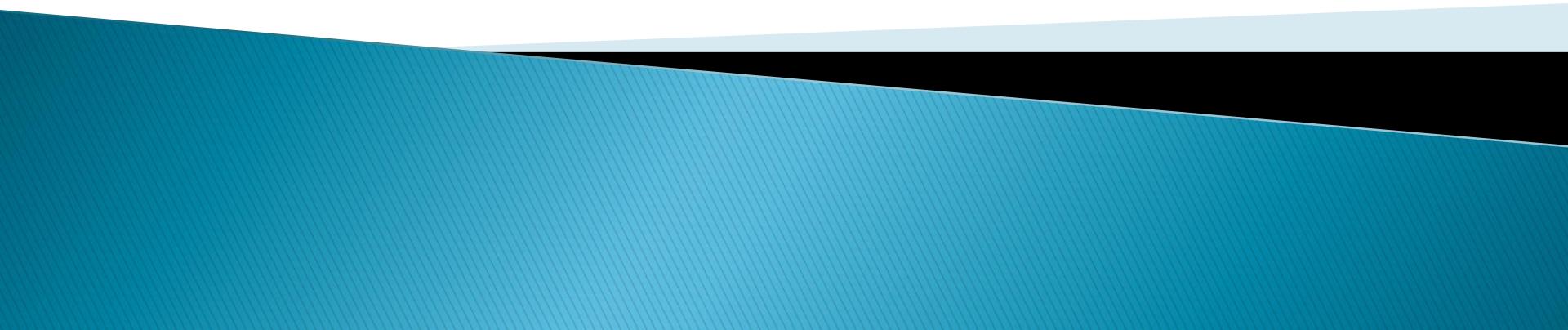


Predictive Modeling Methods

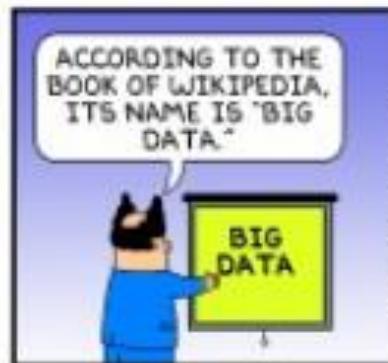
- ❖ Analysts build models using different techniques: neural networks, decision trees, linear regression, ARIMA, Naïve Bayes, etc...
- ❖ In order to create effective analytic models, the analyst needs to know which models and algorithms to use.
- ❖ Many analytic workbenches now automatically apply multiple models (prediction, classification, segmentation, association detection) to a problem to find the combination that works best.
- ❖ Advances make it possible for non-specialists to create fairly effective analytic models.



The Age of Big Data



DILBERT



BY SCOTT ADAMS

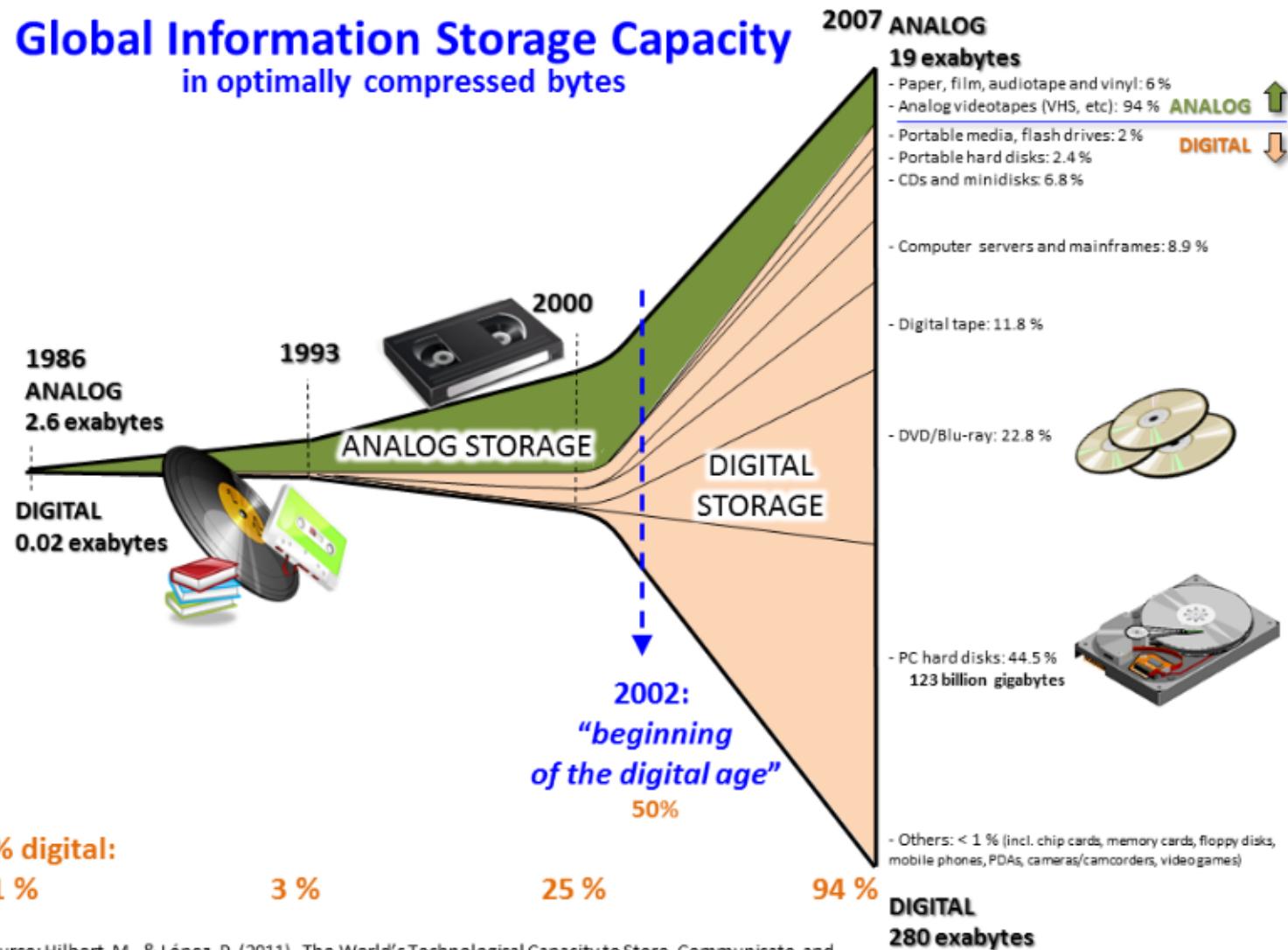
Data Explosion in the 21st Century



- ❖ Developed economies make increasing use of data-intensive technologies.
- ❖ There are 4.6 billion mobile-phone subscriptions worldwide and there are between 1 billion and 2 billion people accessing the internet.
- ❖ Between 1990 and 2005, more than 1 billion people worldwide entered the middle class which means more and more people who gain money will become more literate which in turn leads to information growth.

Global Information Storage Capacity

in optimally compressed bytes



% digital:

1 %

3 %

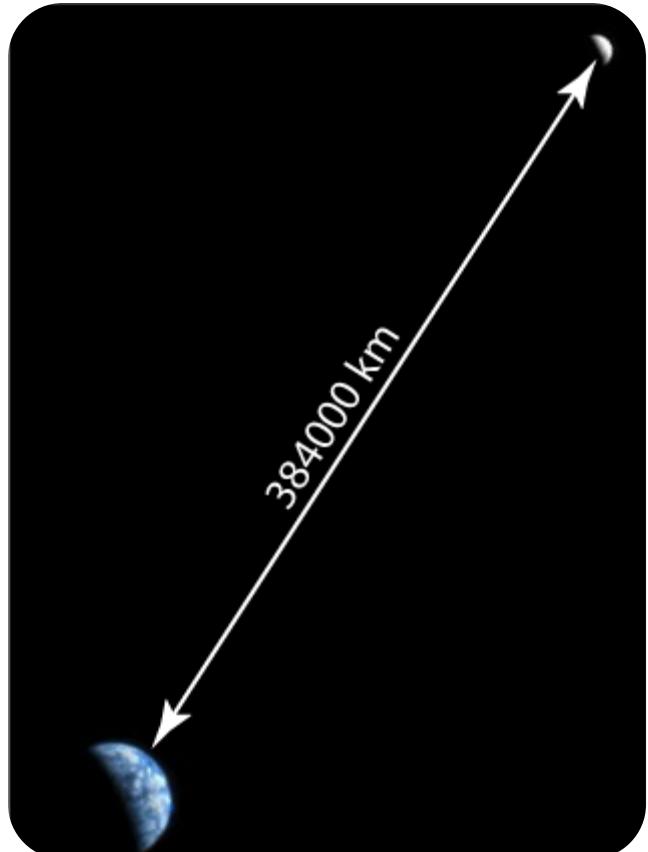
25 %

94 %

Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65. <http://www.martinhilbert.net/WorldInfoCapacity.html>

Data Explosion in the 21st Century

- ❖ In 2007, the amount of digital data represents about 61 cd's worth of information for every human on the planet.
- ❖ If we were to stack these 404 billion cd's on top of one another, this stack would reach the moon and a quarter of the distance beyond.
- ❖ According to Cisco IP traffic forecast, global mobile data traffic will grow 131% between 2008-2013 reaching 2 exabytes per month in 2013.
- ❖ A single exabyte is roughly equivalent to the entire US Library of Congress text duplicated 100,000 times.
- ❖ By the end of 2016, annual global traffic will reach 1.3 zettabytes or 110.3 exabytes per month.



Data Explosion in the 21st Century

This volume of information can be separated into 2 distinct data types: Structured and Unstructured Data.

Structured – Data is organized in a highly mechanized or manageable manner. Some examples include data tables, OLAP cubes, XML format, etc...

Unstructured – Raw and unorganized data which can be cumbersome and costly to work with. Examples include News Articles, Social Media, Video, Email, etc..

Merrill Lynch projected that 80-90% of all potential usable information exists in unstructured form. In 2010, Computer World claimed that unstructured information might account for 70-80% of all data in an organization.

The Cisco IP traffic forecast projects that video will account for 61% of the total internet data in 2015.

Data Explosion in the 21st Century



"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."

- Software AG, Oracle Corporation, IBM, Microsoft, SAP EMC, HP and Dell have spent more than \$15 billion on software firms only specializing in data management and analytics.
- In 2010, this industry on its own was worth more than \$100 billion and was growing at almost 10 percent a year: about twice as fast as the software business as a whole.

What Exactly is Big Data?

Big Data is a collection of datasets that are so large that they become difficult to process with traditional database management tools. This requires a new set of tools and techniques to manage the data.

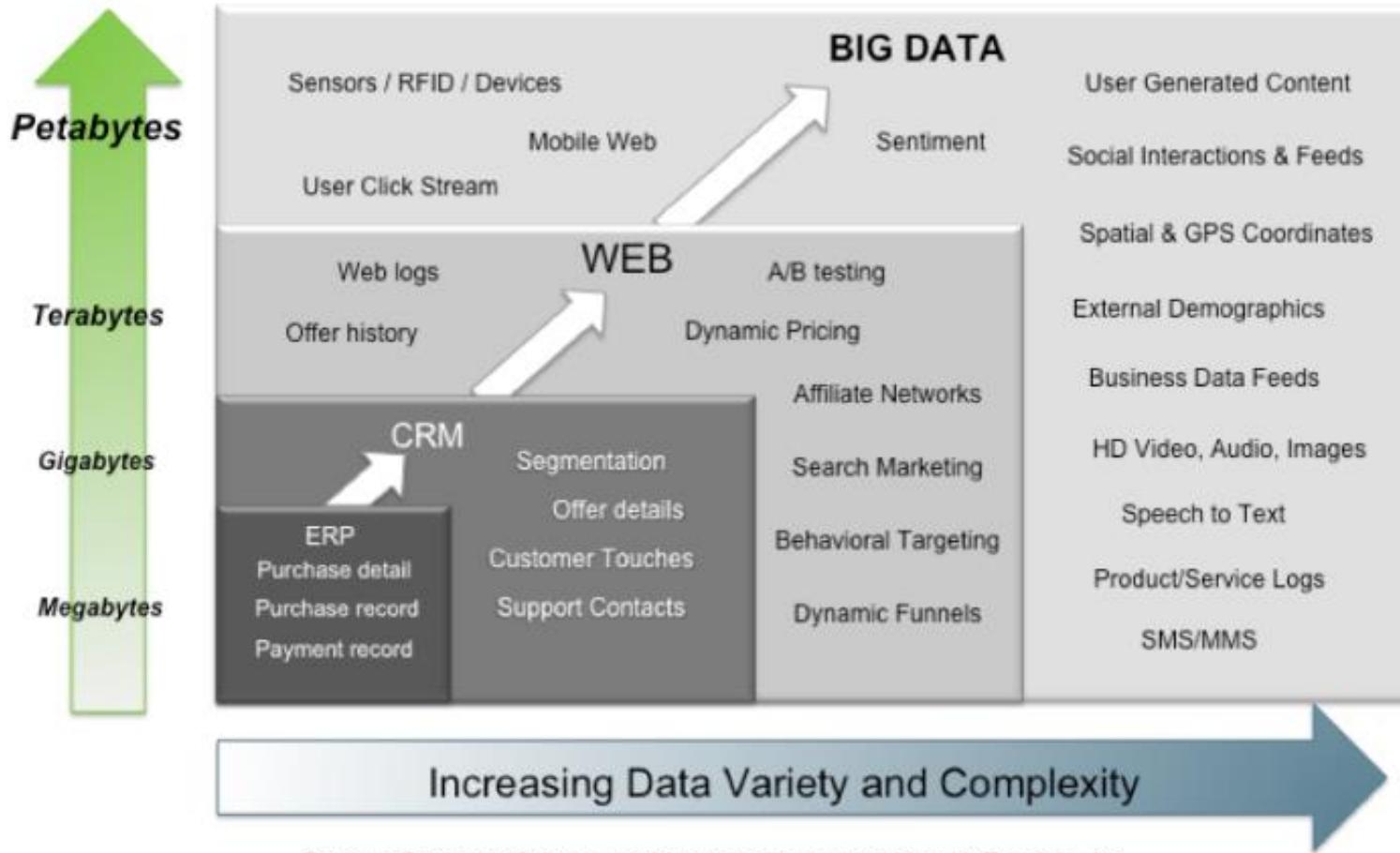
The "Big Data" concept is somewhat ambiguous and is used interchangeably with unstructured data analysis, social media/ web analytics, and growing data needs that exceeds the organizational capabilities.

A simple guideline for determining whether or not an organization is in the "Big Data" realm can be based on the size of the datasets they are working with. Generally, datasets that are many terabytes to petabytes in size would fall into the "Big Data" family.

Interesting Fact: Walmart handles more than 1 million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes (2560 terabytes) of data – the equivalent of 167 times the information contained in all the books in the US Library of Congress.

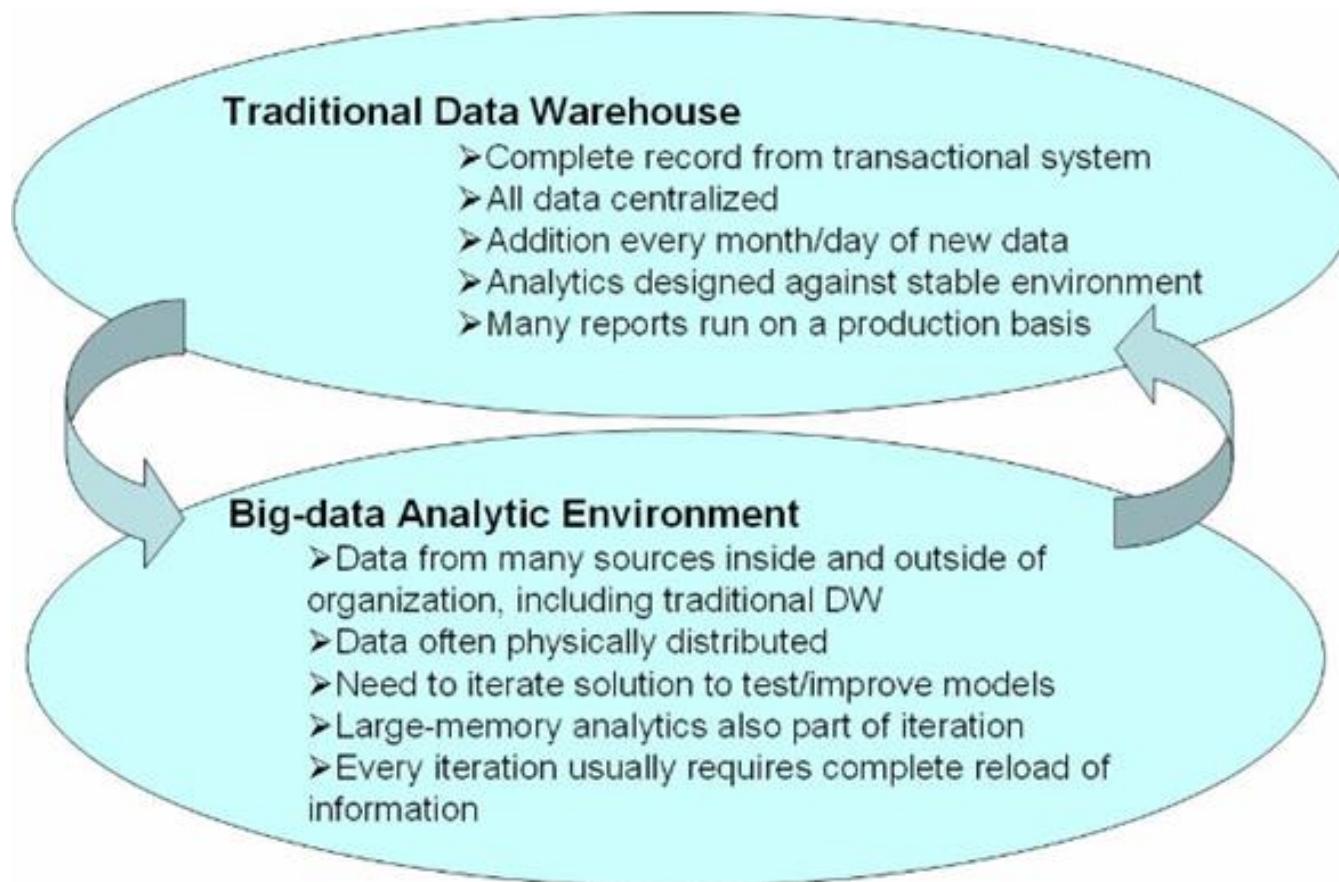
What Exactly is Big Data?

Big Data = Transactions + Interactions + Observations



Source: Contents of above graphic created in partnership with Teradata, Inc.

Big Data Analytics and Data Warehouse Complementary

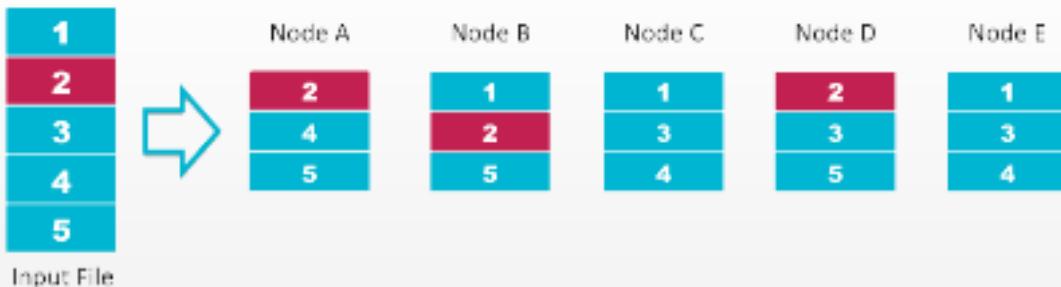


Big Data Solutions – Low Cost/High Power

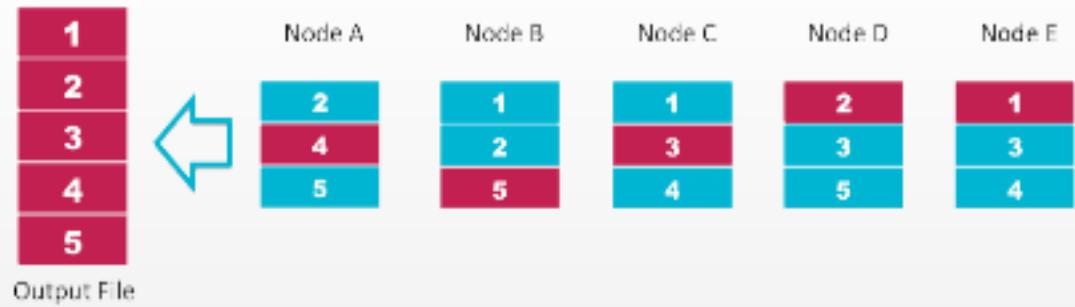


Big Data Processing Model

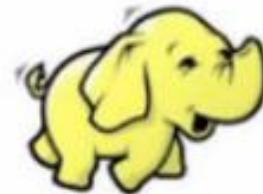
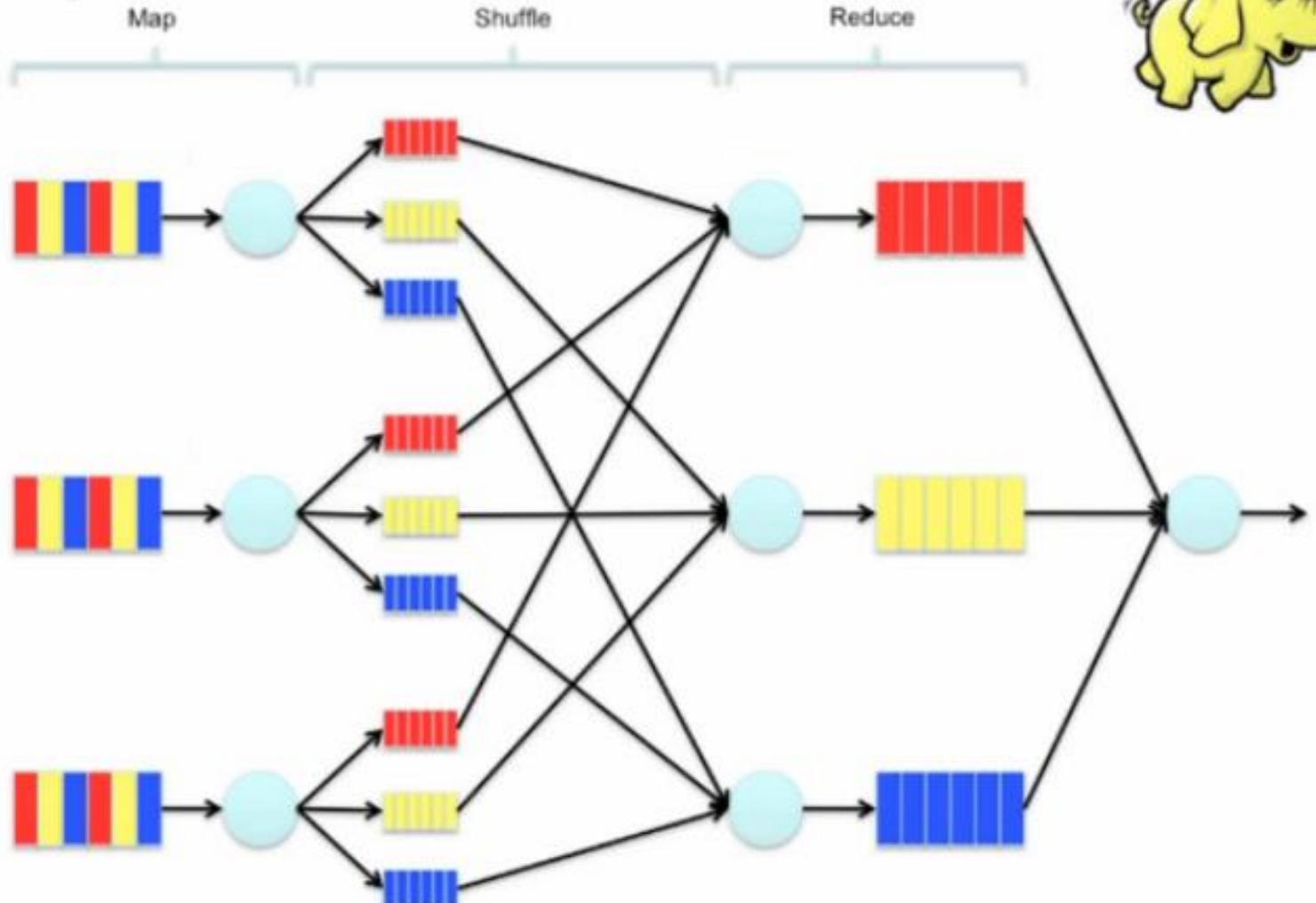
HDFS Data Distribution



MapReduce Compute Distribution

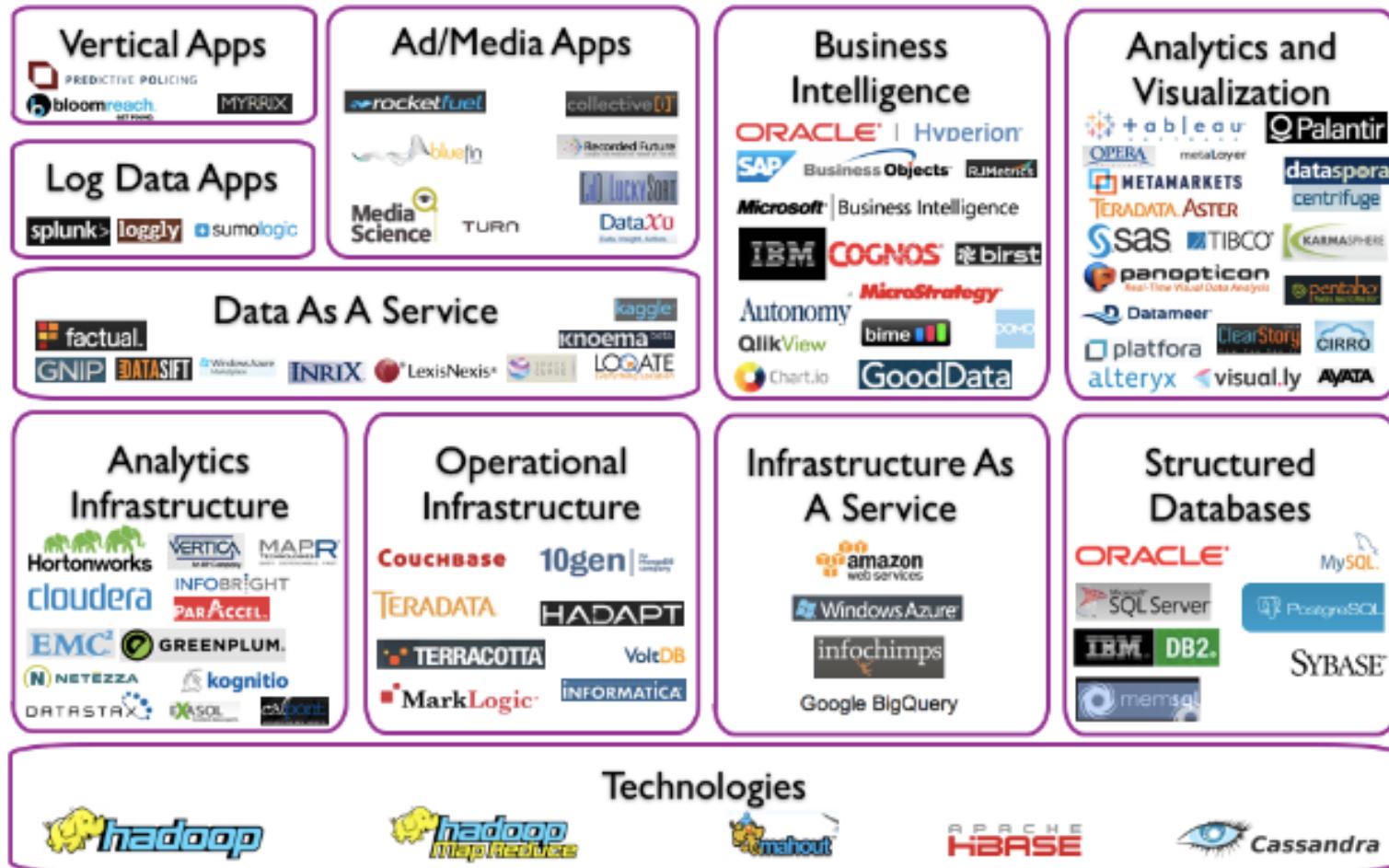


MapReduce Overview



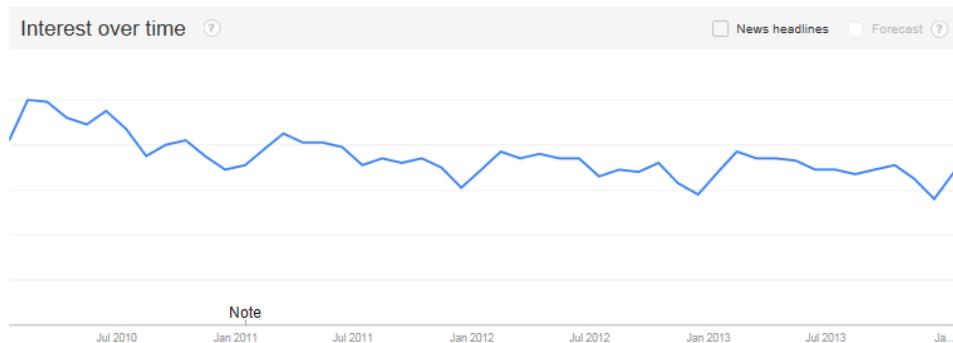
(adapted from <http://code.google.com/p/mapreduce-framework/wiki/MapReduce>)

Big Data Landscape

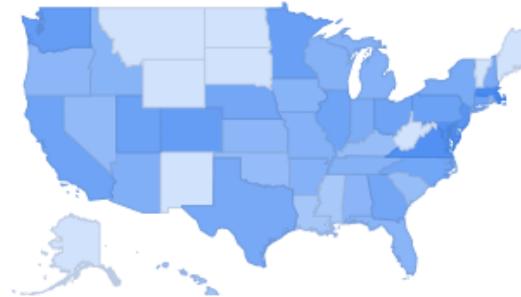


Google Trends Analysis

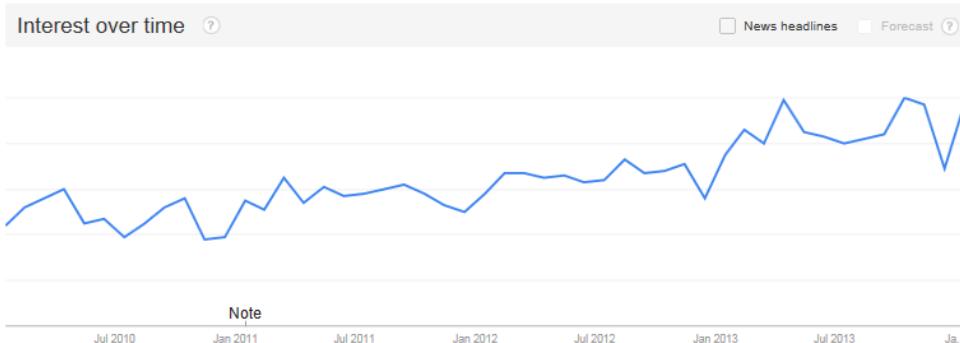
Business Intelligence



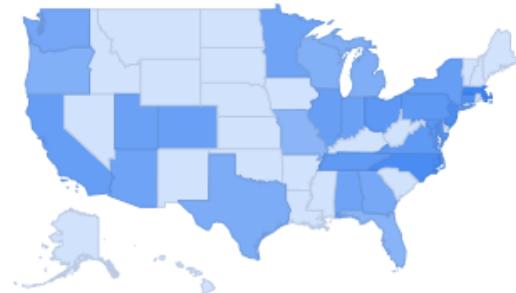
Worldwide > United States



Business Analytics

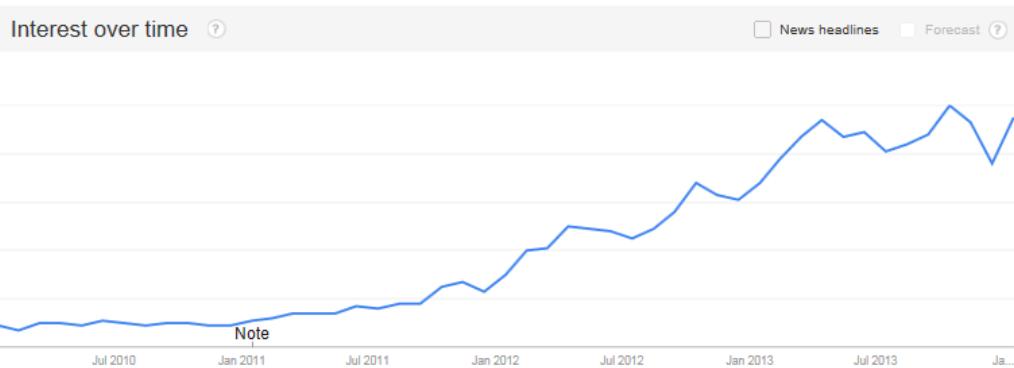


Worldwide > United States

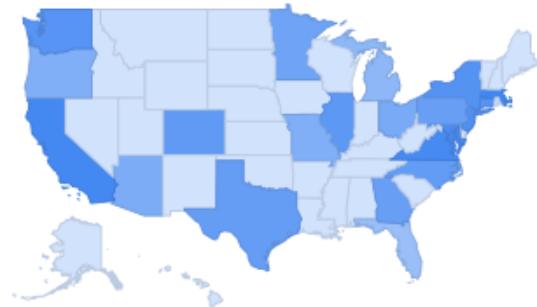


Google Trends Analysis

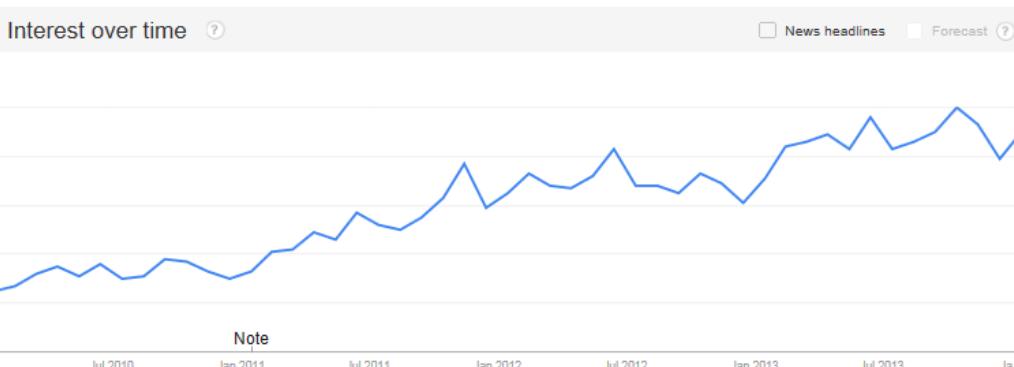
Big Data



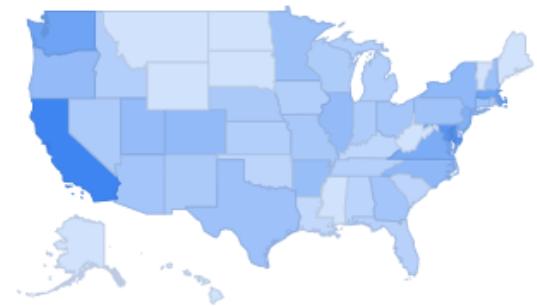
Worldwide > United States



Hadoop



Worldwide > United States

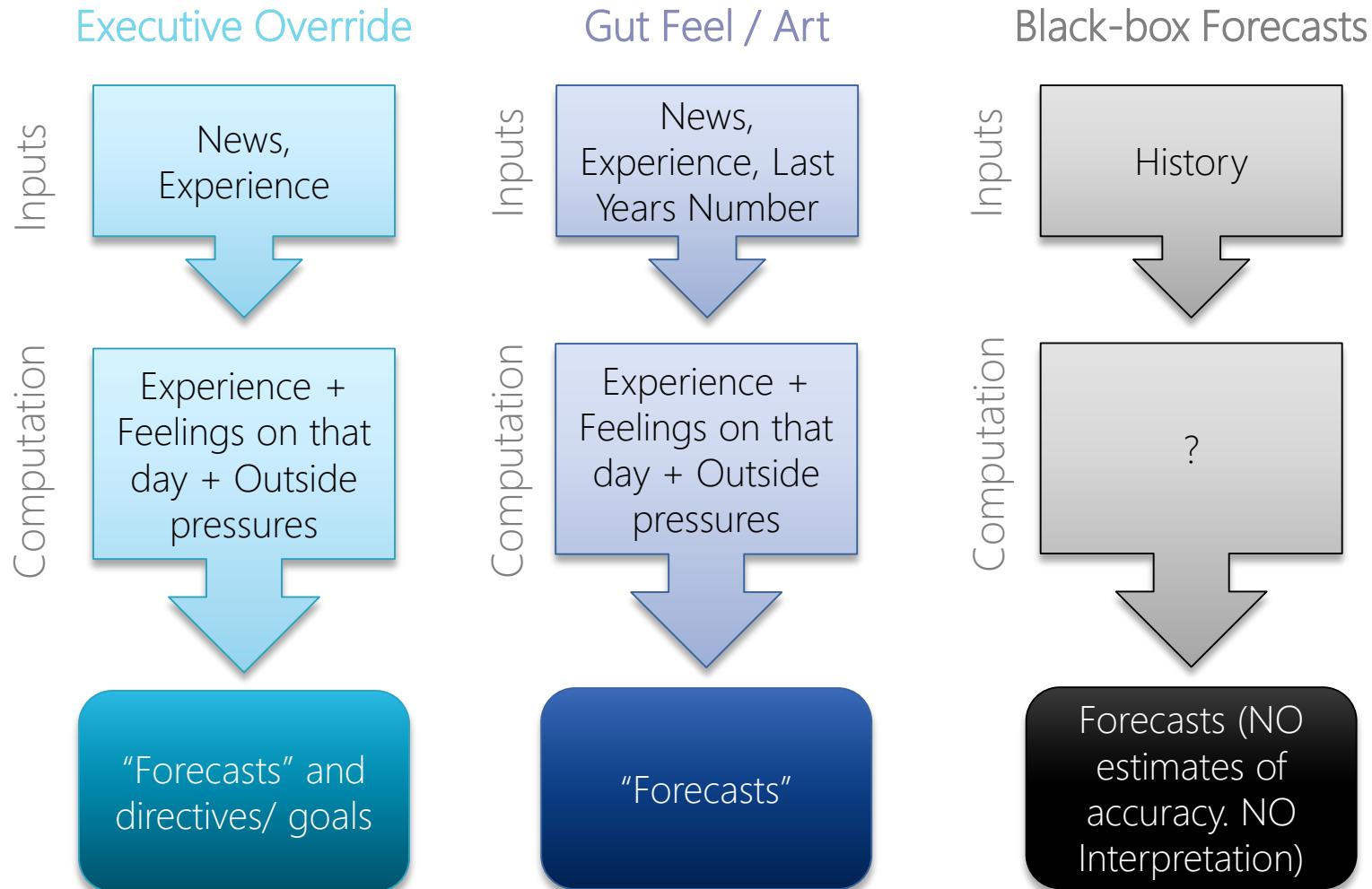


Applied Predictive Analytics: Forecasting Example

The Importance of Forecasting

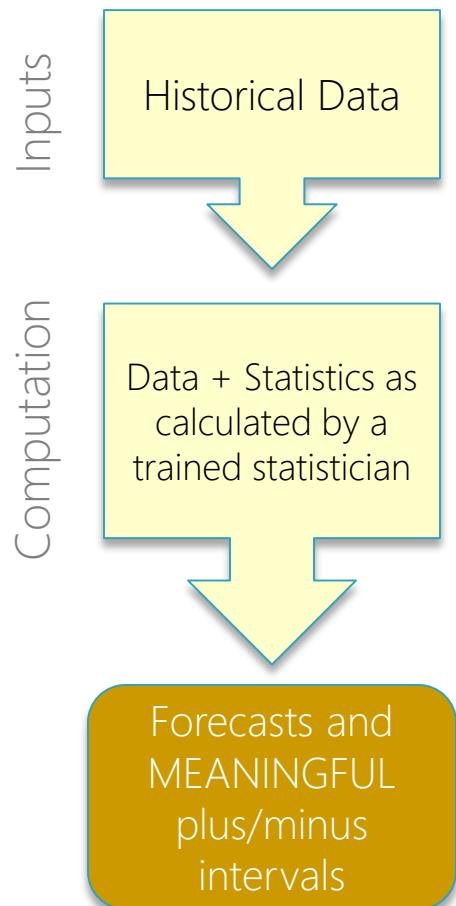
- ❖ Business owners can end up with many problems without planning and forecasting the business.
- ❖ Predictive modeling allows you to know when to take more risks and when it is time to scale back resource spending and allocation.
- ❖ Managers will make better decisions as a result if they analyze historical performance and try to predict future performance.
- ❖ Sustained success can be achieved when organizations become proficient at making business decisions after careful deliberation and modeling, instead of reacting to crises or circumstances.
- ❖ **Question:** How do you develop your forecast?

Poor Forecasting Philosophies

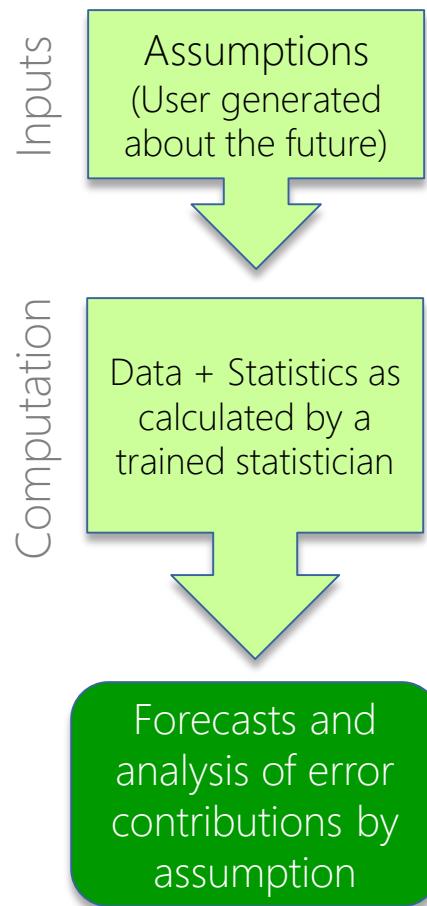


Better Forecasting Philosophies

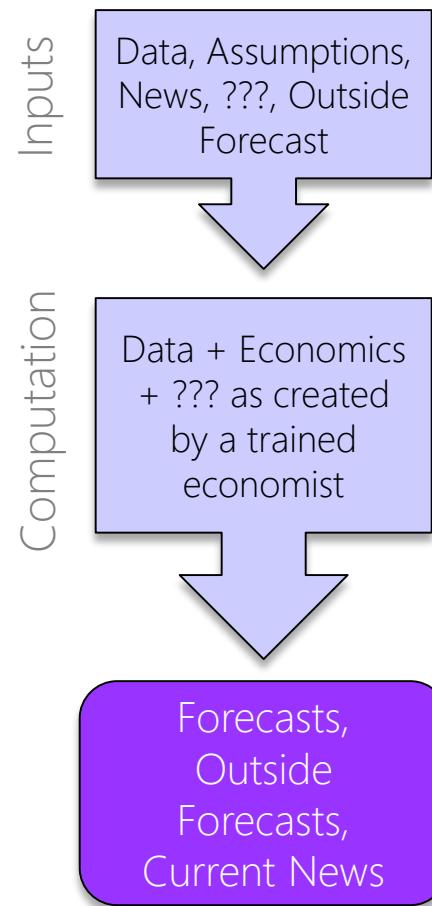
Statistical Models



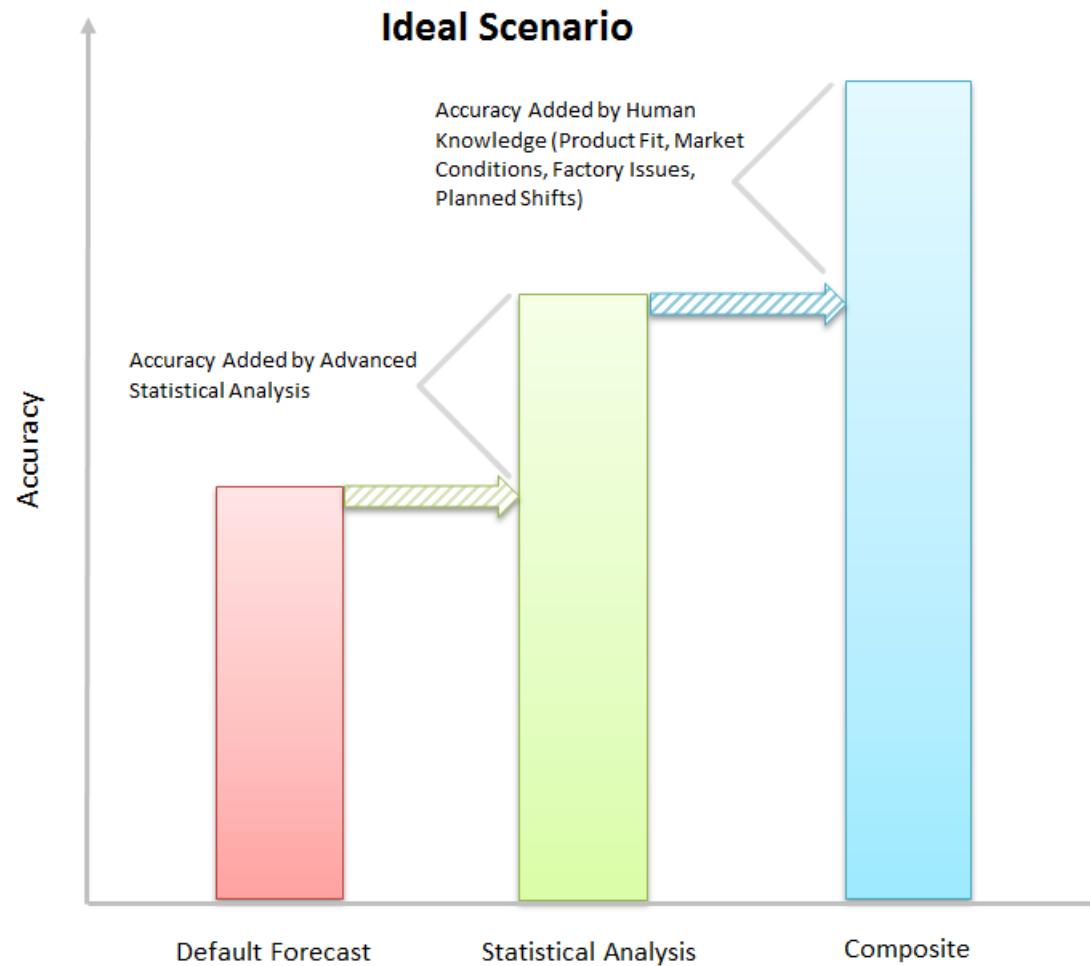
Assumption Models



Economic Models



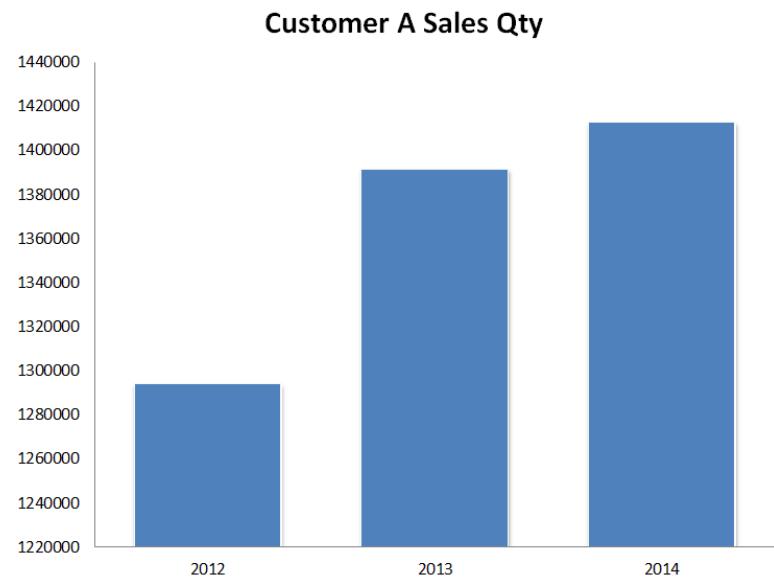
Strengthening the Forecast



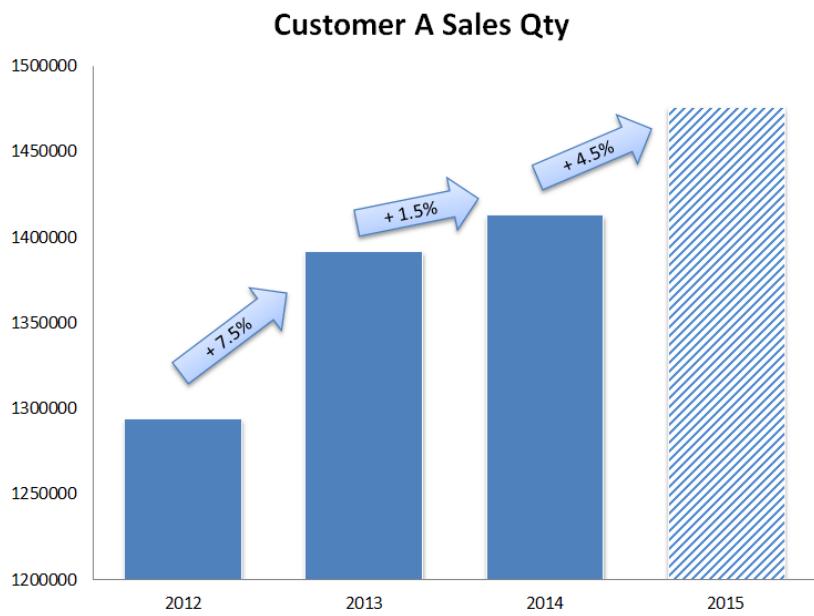
Predictive Analytics in Action

A VP of Sales is tasked with completing sales objectives for the sales staff and financial planning for the CFO.

- ❖ How will they develop an appropriate forecast target for this customer?
- ❖ Will they show an overall sales decline, stagnant sales, or sales growth?
- ❖ How do they reach this conclusion?
 - ❖ Some sales professionals will conservatively estimate ("sandbag") the overall sales growth so that they can readily obtain their performance objectives.
 - ❖ Another approach that sales professionals will take involves simple averaging of the sales trends over a period of time.



Predictive Analytics in Action



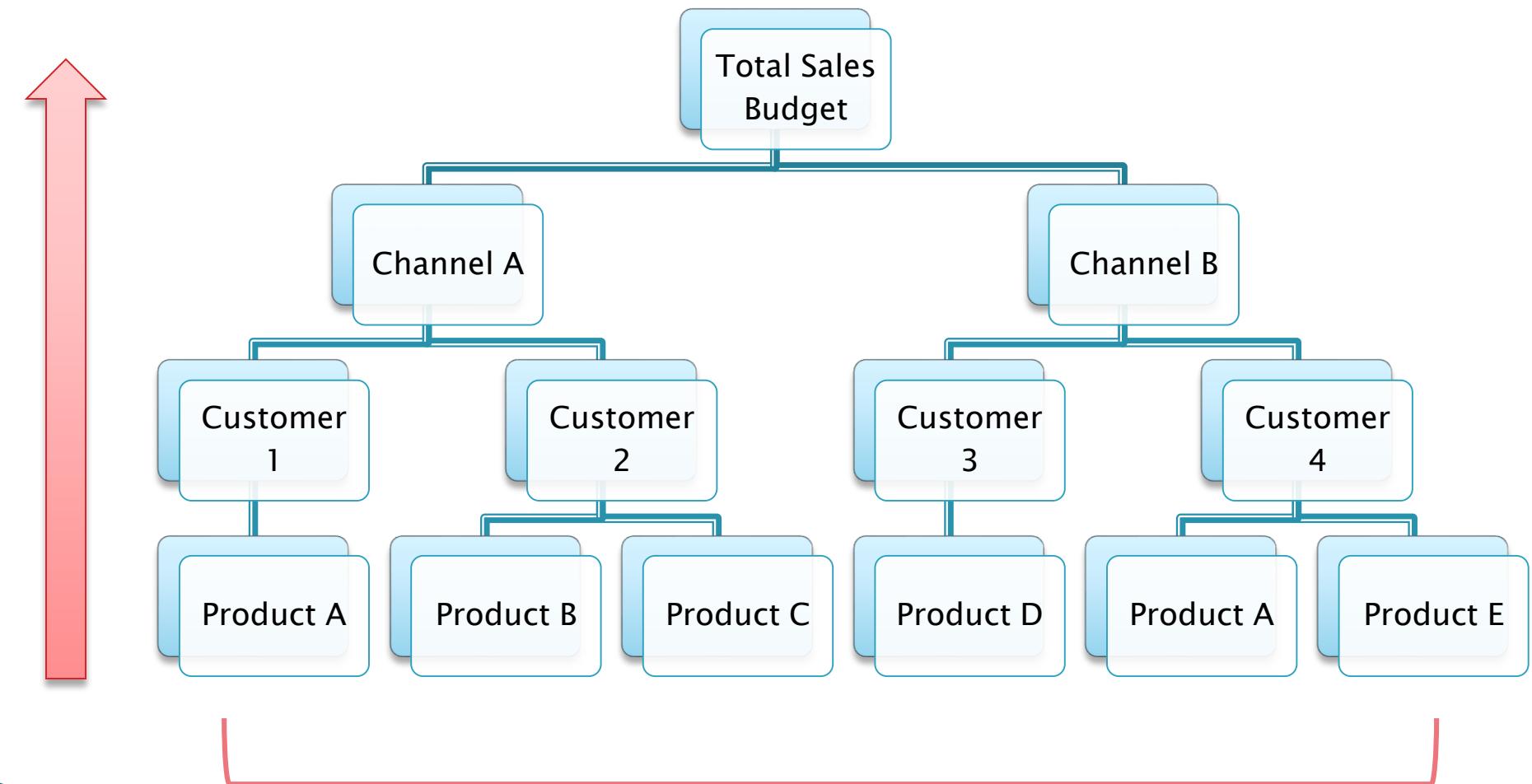
- ❖ Our VP of Sales has decided to take the last 2 years growth rates and average them to come up with the sales projection ($7.5\% + 1.5\% = 4.5\%$)
- ❖ This seems like a reasonable approximation (verified by GDP growth of 4%) and creates an optimistic picture (no sandbagging) for 2015.
- ❖ Furthermore, this 4.5% sales growth helps to align with the CEO's strategic initiative of improving EBITDA by 3%, which looks great on their part.
- ❖ This approach is a top-down approach and is not robust enough to strategically align all of the departments to achieve their respective performance objectives.
 - ❖ Ex. Working Capital Reductions, Margin Development, CAPEX Planning, Order Fulfillments, etc...)

Predictive Analytics in Action

- ❖ The following example will showcase how we can easily transform this forecasting process into a robust methodology using advanced statistical methods.
- ❖ Additionally, we will showcase how this approach can be leveraged to help optimize the supply chain/manufacturing process and how it can be combined with the finance/cost accounting to help fortify an operating budget / forecast.



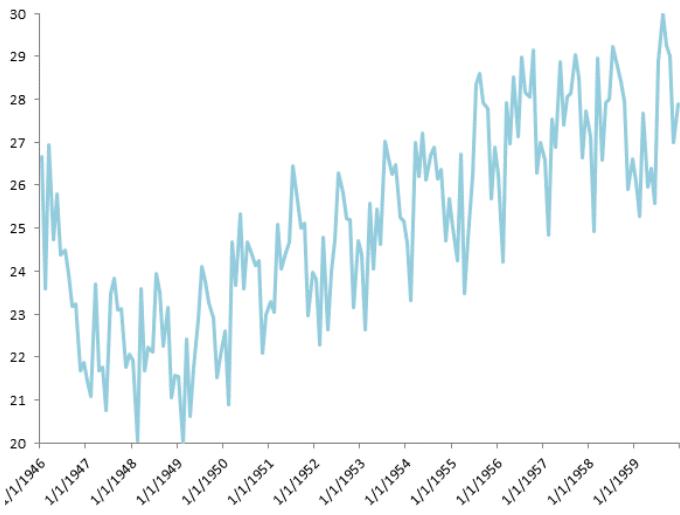
Predictive Analytics in Action



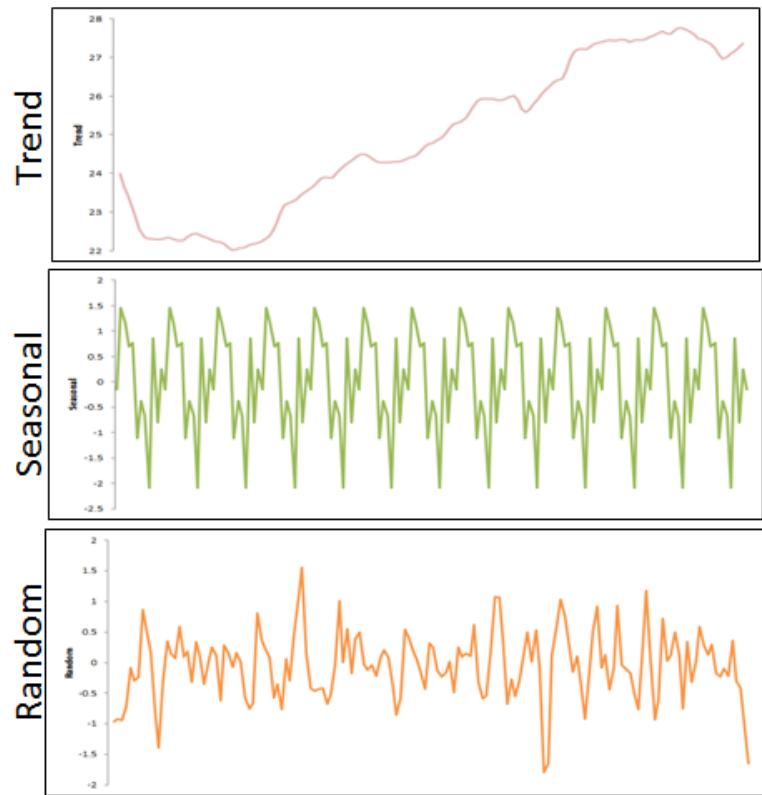
Deep granular level is the focus
for our Forecast

Predictive Analytics in Action

- ❖ The ARIMA time series statistical model breaks down the product sales quantities into three fundamental components (Trend, Seasonality, and Random Fluctuations).
- ❖ These components are then projected out for the forecast period and then aggregated to create a powerful forecast approximation.

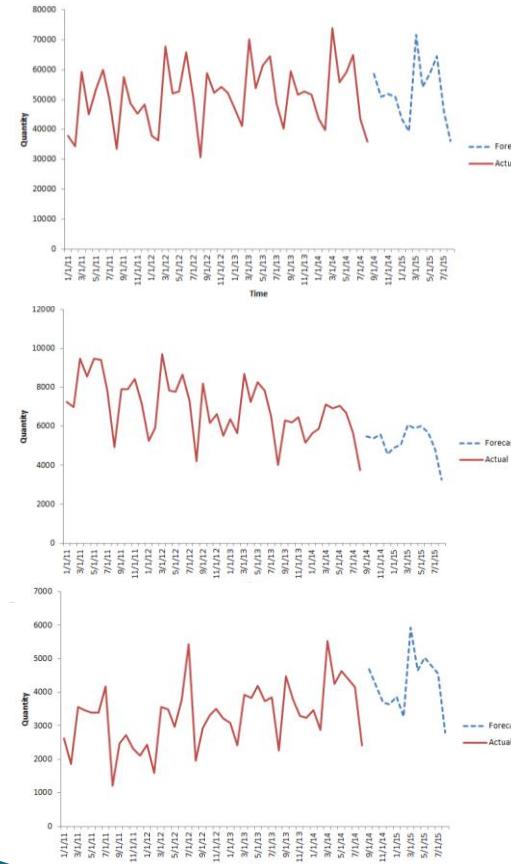


Decomposition of additive time series



Predictive Analytics in Action

Customer Level & Product Specific Forecast

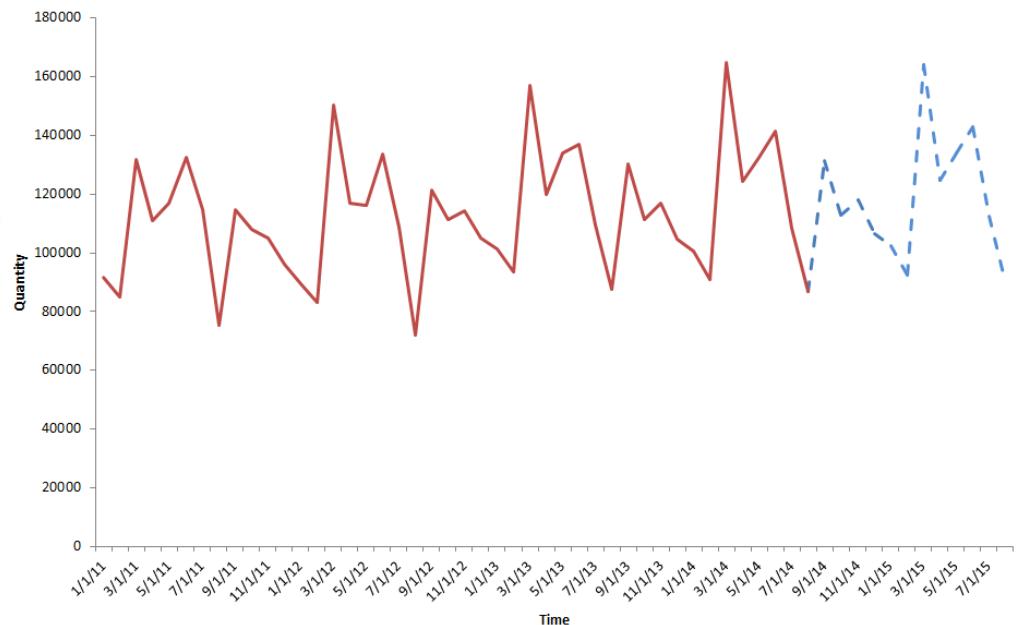


Aggregate

Aggregate

Aggregate

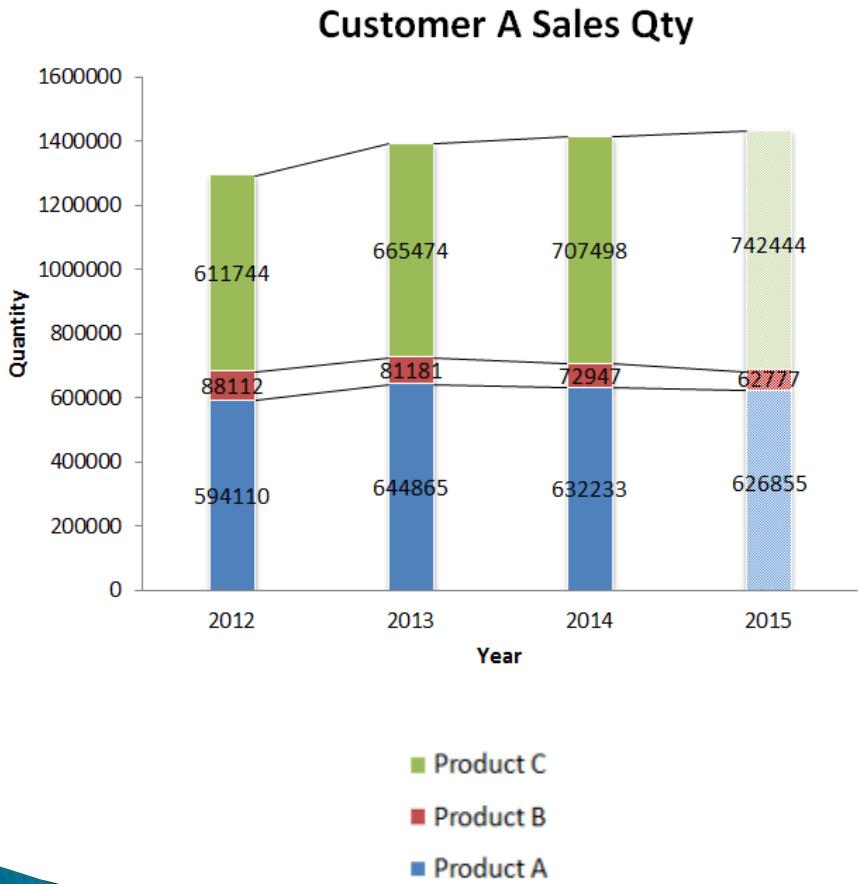
Customer Level Forecast Summary



Forecast Approach Includes:

- ❖ Seasonality
- ❖ Trending
- ❖ Purchasing Irregularities

Predictive Analytics in Action



- ❖ The composite forecast now is built up from a granular level of detail.
- ❖ The ARIMA forecast shows YOY growth of 1.37% compared to the 4.5% that was originally presented by our VP of Sales.
- ❖ This is showing a contraction of the growth rate of 1.5% from 2013-14.
- ❖ This approach can now help to form a narrative to close the gap between the statistical model and the VP's projection.
- ❖ While this projection should not be taken without proper vetting, this is now a data driven approach which can be further scrutinized and improved upon.

Predictive Analytics in Action



- ❖ Operations and Supply Chain can aggregate the sales detail by product rather than Customer to form the basis for manufacturing capacity planning.
- ❖ This forecast not only shows the total sales volume but also the development over time.
- ❖ When we combine this forecast with order and fulfillment lead times, we can effectively drive down working capital and improve customer satisfaction.
- ❖ This forecast can also be integrated into the product cost and pricing/margin projections from the finance department.

Predictive Analytics in Action

- ❖ When we aggregate all the products/customers/channels forecasts from our model, we can budget down to the gross margin level (EBITDA) within the P & L.
- ❖ More importantly, we can tie all of the functional operational areas within the organization together from this forecast and leverage strategic initiatives with greater precision and measurability.
- ❖ With a proper forecast, we are now able to coordinate improvements to Working Capital, Customer Satisfaction, Strengthen our Sales Projection, and better account for the Impact of Product Cost and Pricing Shifts at a macro/micro Level.
- ❖ These pieces combine to create a stronger operational budget and become the catalyst for further budgeting/ forecasting process improvements. (Ex. 12 Month Rolling Forecasts)

Fine