An Overview of Data Science and Business Analytics

Sunil Subrahmanyam Yadavalli

Senior Consultant, KPMG

About the Speaker

Senior Consultant in KPMG’ s Digital Advisory practice

8 years of experience in Telecommunications & IT industry with key focus on Revenue Assurance, Enterprise Risk, Business Intelligence & ICT transformation.

TM Forum certified Revenue Assurance Practitioner

Previously worked with Subex, Cognizant and TCS.

A guest speaker and visiting faculty at VIT University, Vellore having trained 800+ graduates in project management & digital transformation.

Served as a resource person at various forums for AICTE, APSSDC, TASK and member at various panel discussions

Published 5 research papers and paper presentation in international conference at IISc (Indian Institute of Science), Bangalore

Types of Data

DEST	DISTANCE	FL_DATE	FL_NUM	ORIGIN	Weather
JFK	213	37987	4760	DCA	0
JFK	213	37988	4760	DCA	0
JFK	213	37989	4760	DCA	0
JFK	213	37990	4760	DCA	0
JFK	213	37991	4760	DCA	0
JFK	213	37992	4760	DCA	0
JFK	213	37993	4760	DCA	0
JFK	213	37994	4760	DCA	0

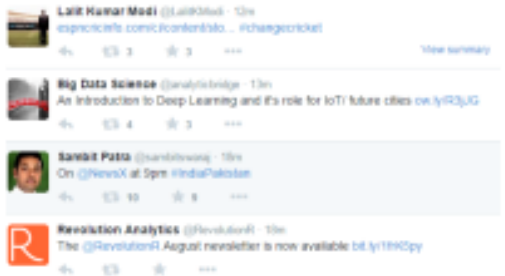
Structured Data

http://timesofindia.indiatimes.com/topic/Amitabh-Bachchan

https://mail.google.com/mail/u/0/#inbox/14f4b017c67de062

Quasi-Structured Data

Semi-Structured Data



Unstructured Data

Structured:
Data containing data type, format etc. The data sets which are easy to understand

Semi-Structured:
Textual Data files with a distinct pattern e.g. XML files

Quasi-Structured:
Textual data with irregular data formats e.g. Web click stream data

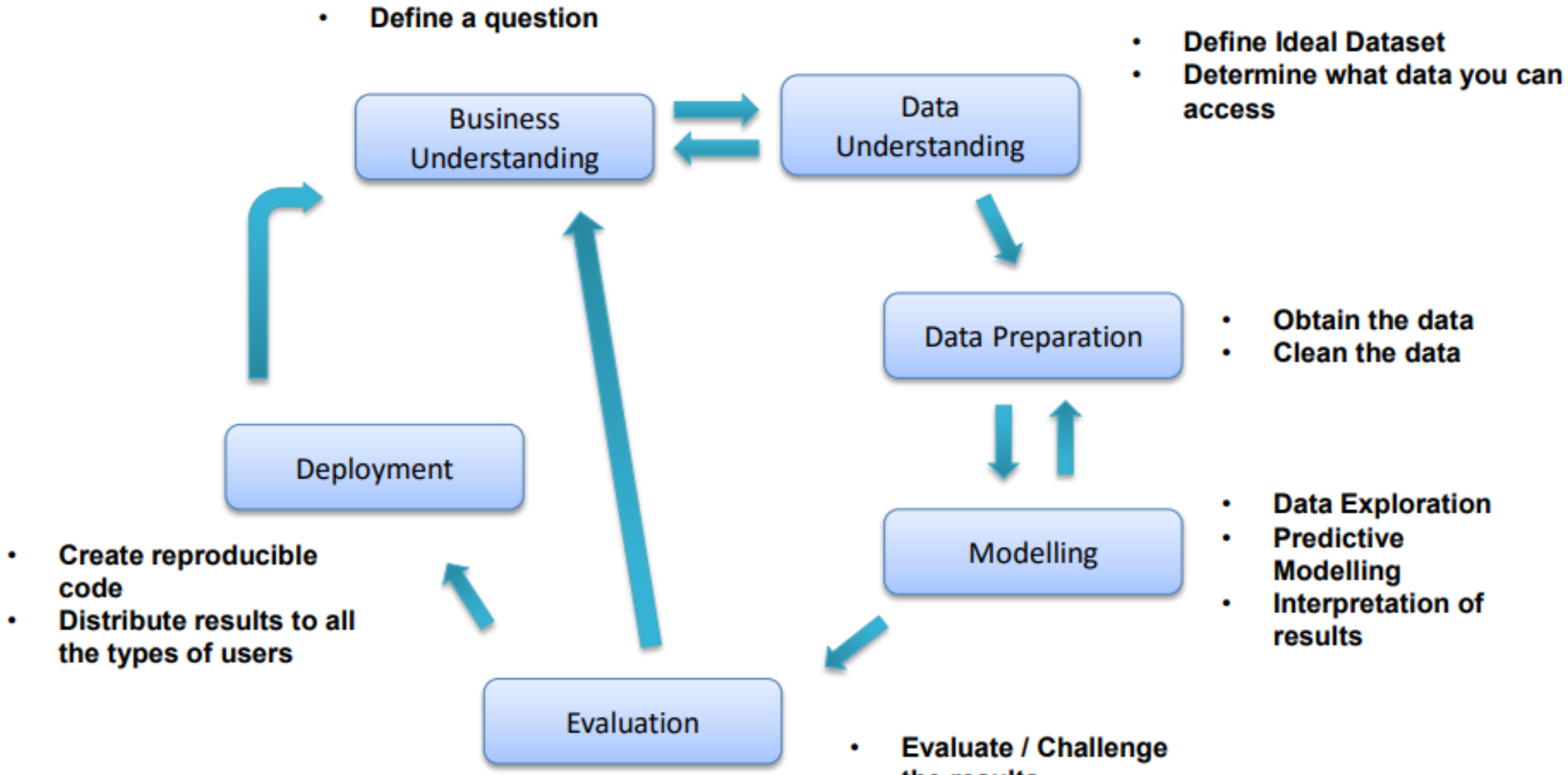
Unstructured data:
No structure at all e.g. Twitter tweets, FB posts, WhatsApp posts etc.

Big Data and its Key Characteristics

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate

Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, and information privacy

Analytics Life Cycle



- Define a question

- Define Ideal Dataset
- Determine what data you can access

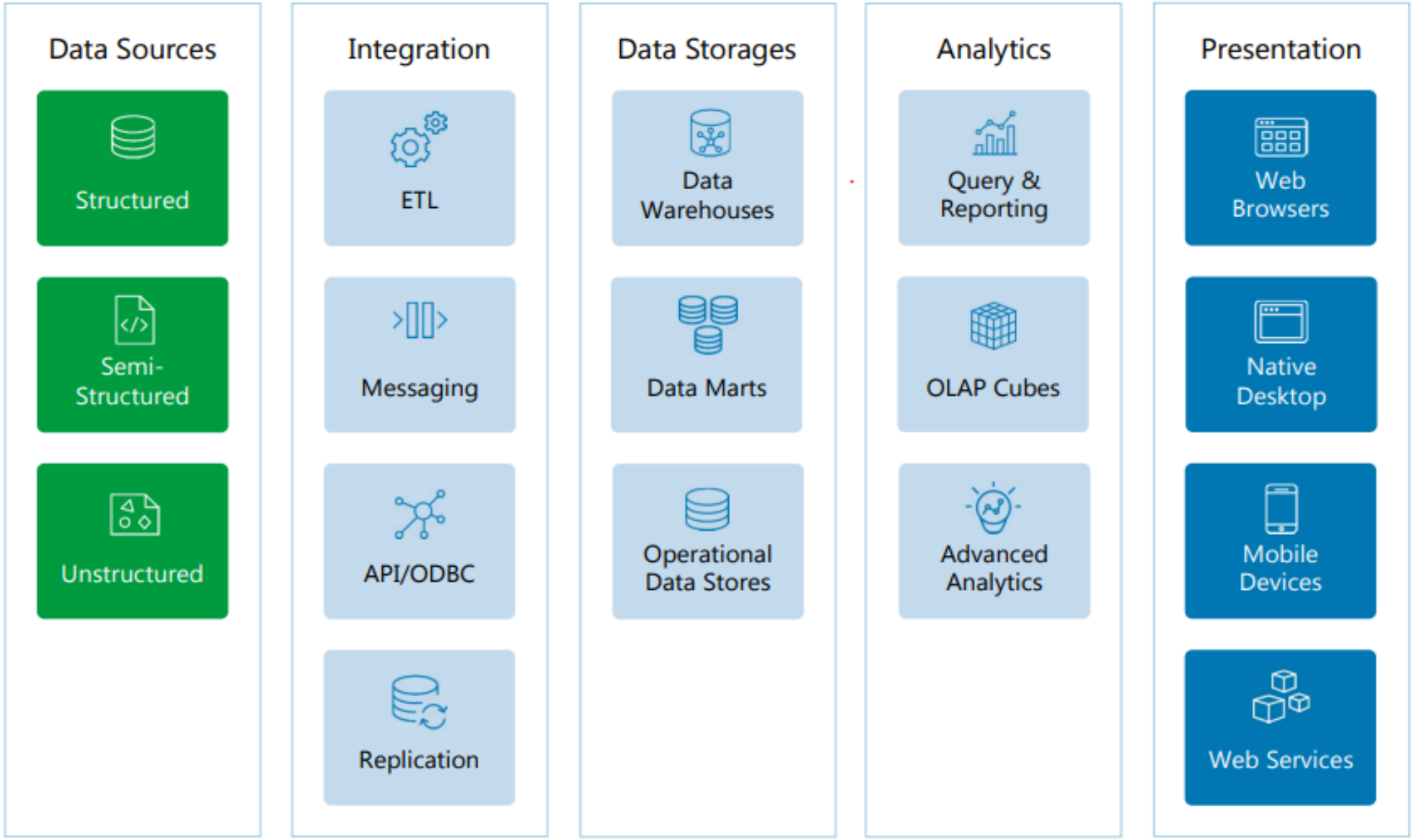
- Obtain the data
- Clean the data

- Data Exploration
- Predictive Modelling
- Interpretation of results

- Evaluate / Challenge the results
- Synthesize / write-up results

- Create reproducible code
- Distribute results to all the types of users

Data Architecture



Source: SoftServe

Qualitative vs. Quantitative

Qualitative
- Any Nominal and Ordinal data
- It can be Numeric or non-Numeric
- Various graphs and charts for Qualitative data are available
  - Frequency distribution
  - Relative & Percent frequency distribution
  - Bar graphs and Pie charts

Quantitative
- Any Interval and Ratio data
- It is Numeric data
- Various graphs and charts for Quantitative data are available
  - Frequency distribution
  - Relative & Percent frequency distribution
  - Cumulative distribution
  - Histograms & Polygons, Ogive curves, Dot plots

Original Data about service quality of a telecom operator (Sample size = 200)
[Excellent, Good, Average, Poor]

E, A, A, G, A, E, G, P, P, E, A, ..., G, P, E, A

Frequency Distribution for Rating

Rating	Frequency
Excellent	40
Good	70
Average	60
Poor	30
Total	200

Rating	Frequency	Relative Frequency
Excellent	40	= 40/200 = 0.2
Good	70	= 70/200 = 0.35
Average	60	= 60/200 = 0.3
Poor	30	= 30/200 = 0.15
Total	200	1

Data Mining Techniques

Classification	classify a data item into one of several predefined classes
Regression	map a data item to a real-value prediction variable
Clustering	identify a finite set of categories or clusters to describe the data
Summarization	find a compact description for a set (or subset) of data
Dependency Modeling	Describe significant dependencies between variables or between the values of a feature
Change and Deviation Detection	Discover the most significant changes

Market Basket Analysis

Finding patterns or sequences in the way that people purchase products and services.

Walmart Analytics

- Obvious: People who buy Gin also buy tonic.
- Non-obvious: Men who bought diapers would also purchase beer.

Cluster Analysis

Grouping data into like clusters based on specific attributes.

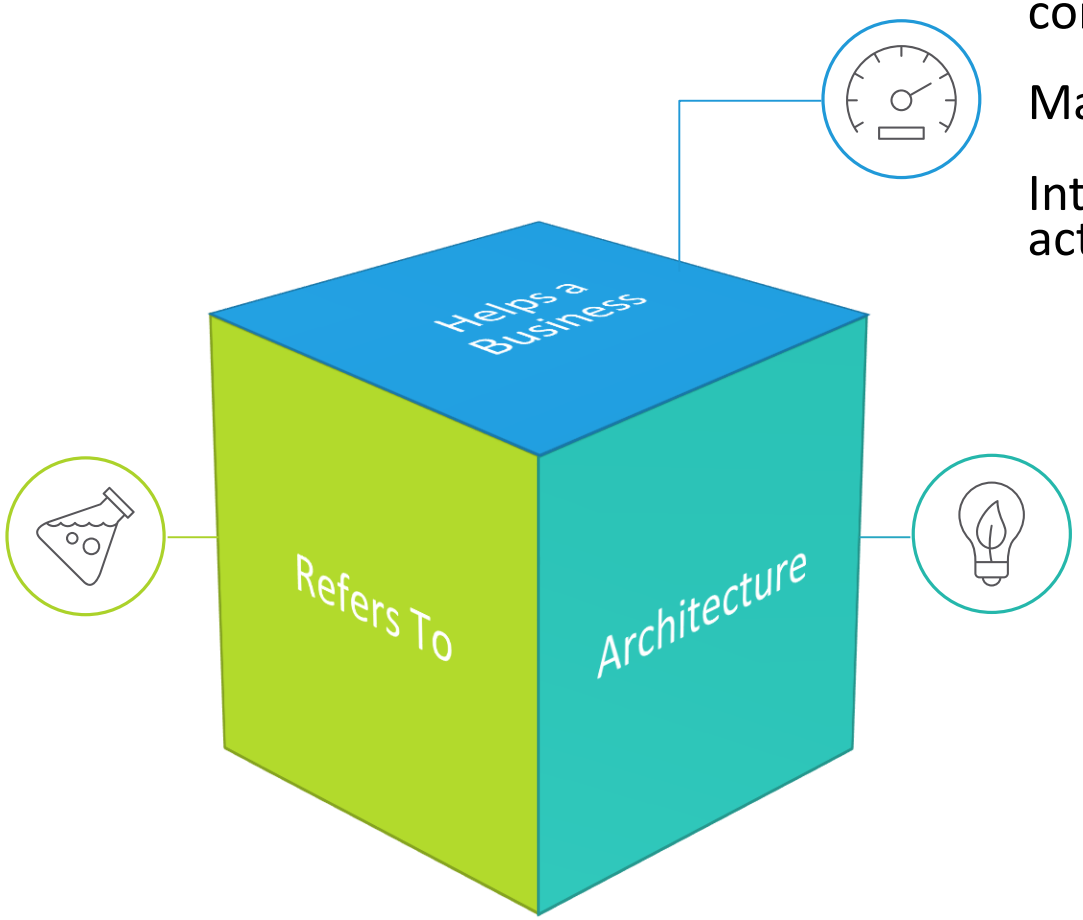
Examples

- Crime map clusters to better deploy police.
- Where to build a cellular tower.
- Outbreaks of virus.

What is Business Intelligence

Business intelligence combines a complex set of analysis and reporting tools to help you streamline company operations.

Skills
Technologies
Applications
Practices



Acquire a better understanding of its commercial context

Make better decisions

Interpret the presented facts and guide actions towards desired goals

Insight to business users about their business so that they can make effective and timely decisions to get ahead of their competition

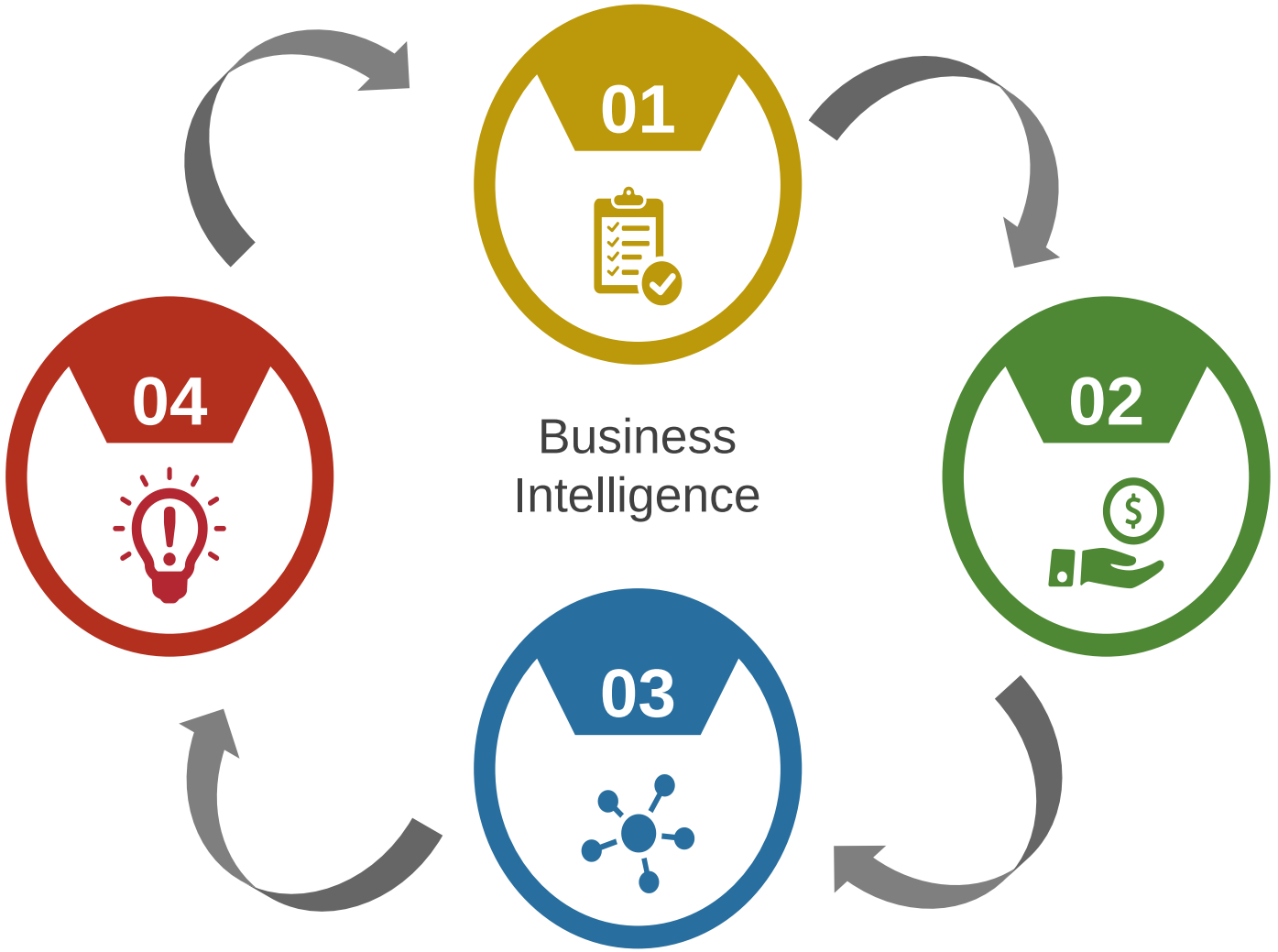
Historical, current and predictive view of the business

Key Requirements for Business Intelligence

Technology: Hardware, database platforms, data management software, delivery tools

People:

Roles/responsibilities, empowering knowledge workers to use information



Data:

Definitions, cleansing process, consistency, trustworthiness

Business Practices: Governance, decision, operational processes

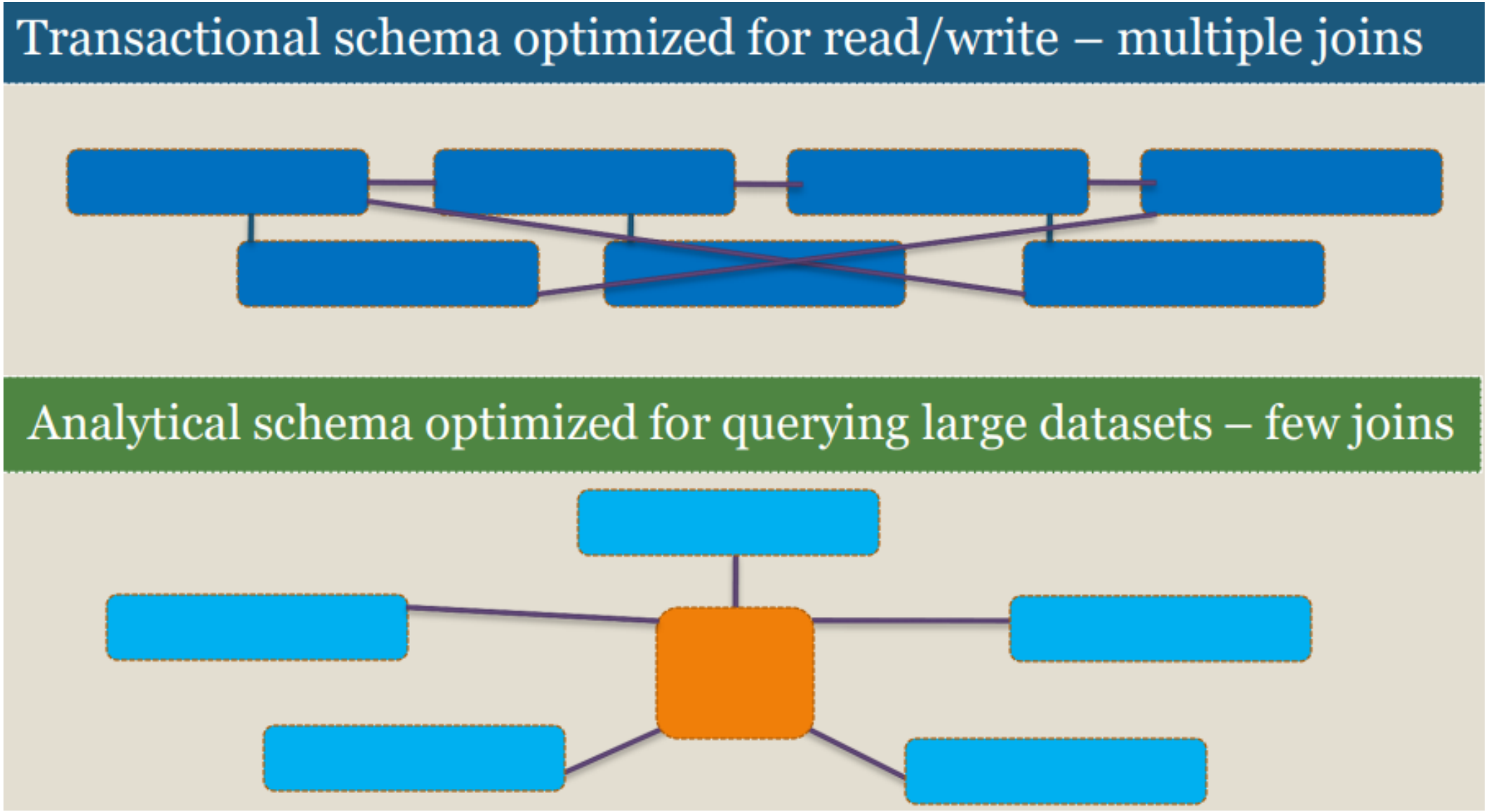
Examples

- Show me the most effective promotions
- Show me customers most likely to switch
- Show me products that are not profitable
- Compare sales with this quarter with sales a year ago
- Show me the bottom five products this year
- Show me lowest/highest ranked salesperson by region
- Show me sales for each district by month
- Show me average sales amount for this quarter
- Show me the top ten customers this year

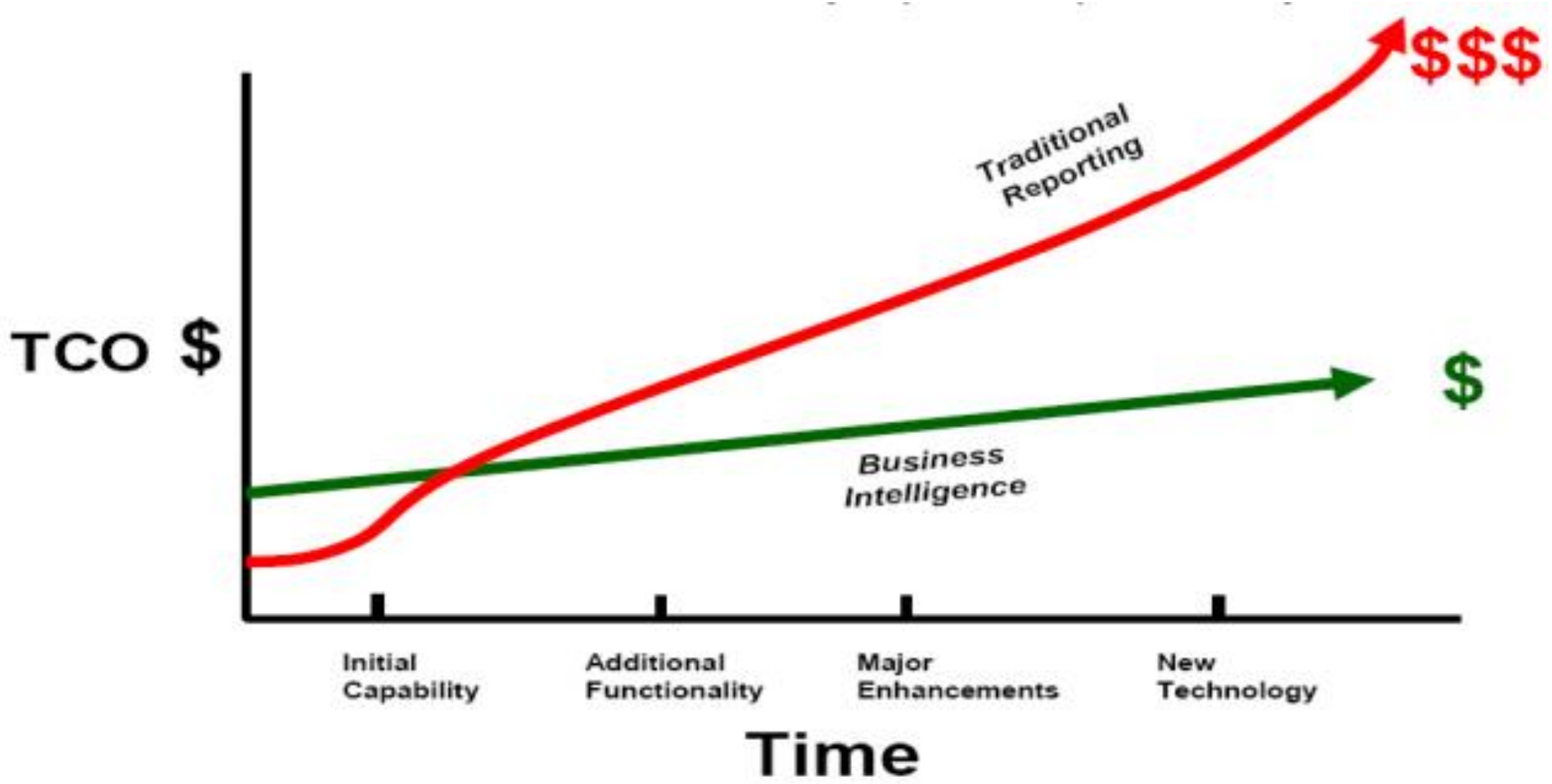
OLTP vs OLAP

Areas	OLTP	OLAP
Users	Data Entry, IT Professional	Analyst, Manager, Director, CEO, CFO
Function	Day-to-day, data optimized for storage, read/write performance	Decision support, data optimized for query performance (denormalized, minimal joins)
DB Design	Application oriented	Subject oriented
Data	Current, up-to-date, detailed, flat file, relational, isolated	Historical, summarized, multi-dimensional, integrated, consolidated
Calculations	Little data aggregations, reports require calculation	Data pre-calculated for processing
Usage	Repetitive, Data entry, data retrieval	Ad-hoc, reports, charts & pivot tables

OLTP vs OLAP



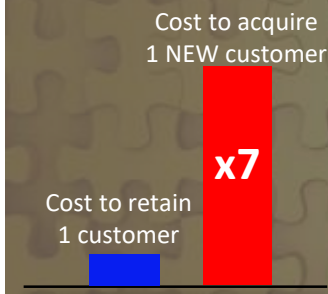
Total Cost of Ownership (TCO) Comparison



Churn Analytics

Customer, Network Experience and Dealers

Challenges



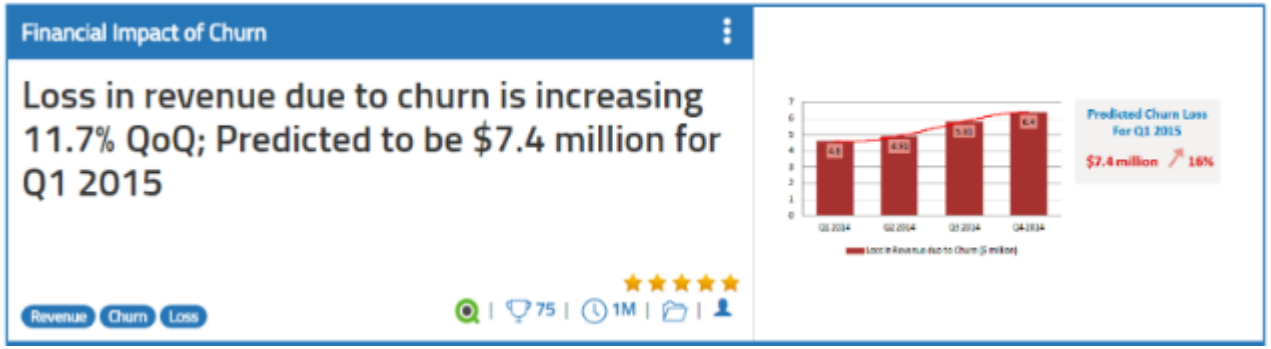
- Price Sensitive
- Service Quality
- Technology - Handset Support
- Competition
- Technology - Network

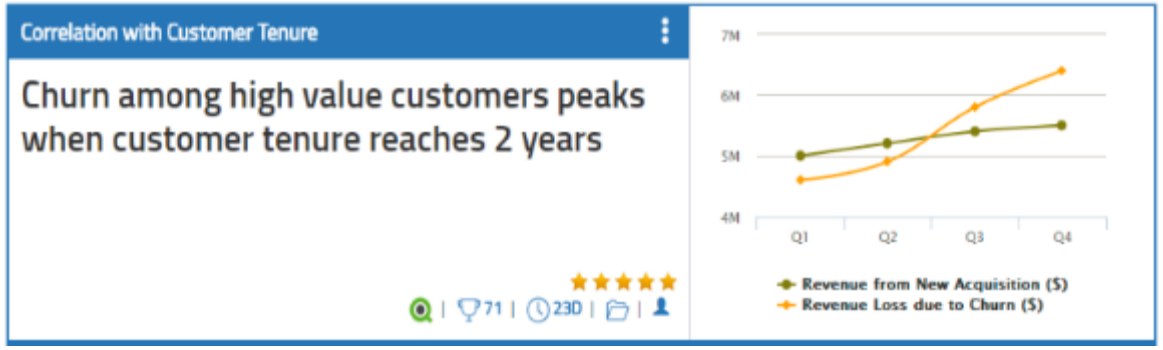
Adopted Approach

- Churn diffusion analysis on churn behaviour
- Geospatial hotspot of churn behavior
- Behavioral analysis of churning customers
- Product affinity analysis with churn
- Handset and service based association analysis
- Association of demographics attributes to customer attrition

Outcome

- Predict customers likely to churn in next quarter
- Impact on revenue from churn experience
- Comparative analysis between churn and acquisition
- Bad experience to churn conversion analysis
- Behavior and DNA of high value churners
- Identify potential risk with new customers





Product Optimization Analytics

Rate-plans, Network Capacity, Customer Experience, Revenue & Margin

Challenges

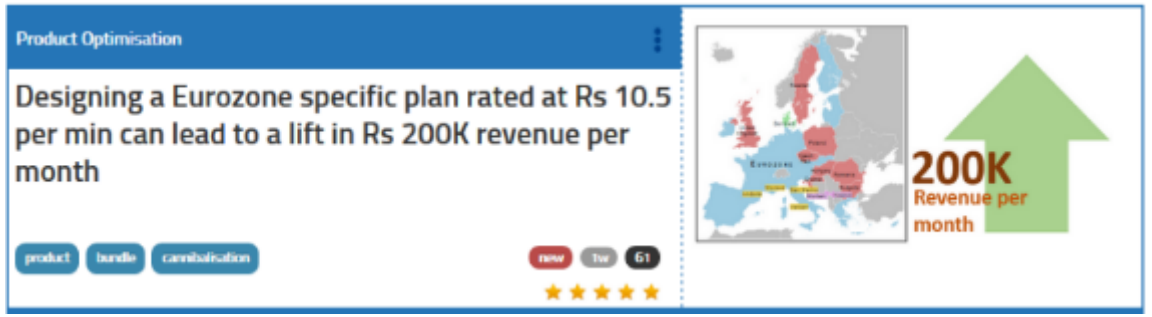
- Declining ARPUs
- Customer Acquisition Cost
- Revenue & Margin Enhancement
- Internal Cannibalisation
- Call Center Cost

Adopted Approach

- Price elasticity analysis to optimize the product
- Simplex optimization to estimate the components of the products
- Predict ARPU per customers for products
- Estimation of cost in network maintenance and eco-system
- What-if analysis to optimize margin/revenue

Outcome

- Optimize product bundle to enhance revenue and margin
- Behavior matching between usage behavior and product used
- Translating customer usage to product catalogue
- Identifying the target customers for a particular product



Call Center Analytics

Interaction Experience

Challenges

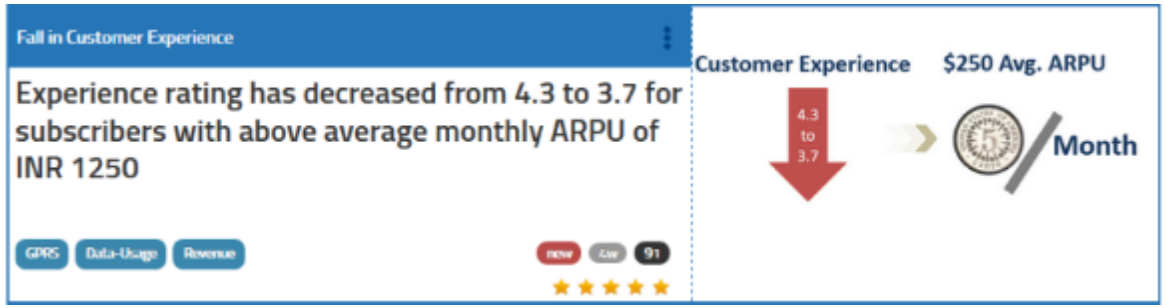
- **High Cost of servicing calls**
- **Long hold times**
- **Complaints address inadequately**
- **Service Quality**
- **Complaints not getting resolved quickly; too many repeat calls**

Adopted Approach

- Top factor of repeat calls to call center
- Pattern analysis to understand what drives propensity to call
- Estimation of prediction and drivers for calls to call-center
- Temporal analysis of call center volumes

Outcome

- Prediction of who, when and why will call call-center
- Behavior and DNA of repeat callers
- Insights into product/offers/demographic attributes strongly associated with call center complaints
- Geo-spatial analysis of high density of calls to call center



Customer Experience Analytics

Challenges

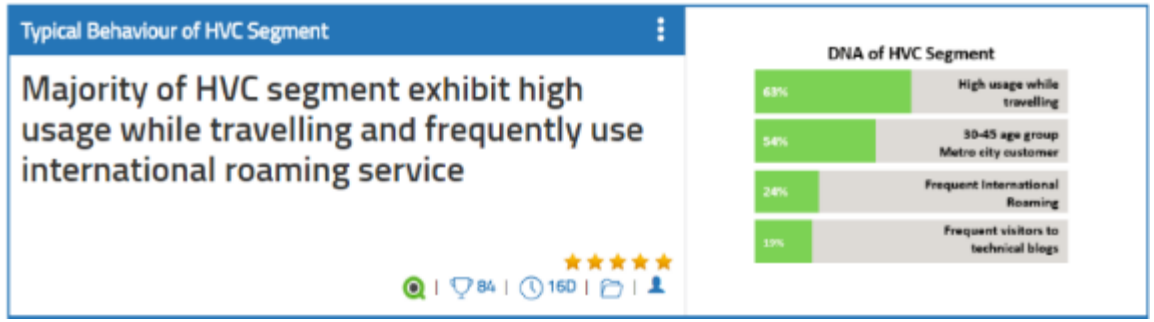
- Low Onboarding experience
- High Complaint Volume
- -ve sentiment on social media
- Long On-boarding Process
- Poor Interaction Experience
- Too many touch points in order to first bill

Adopted Approach

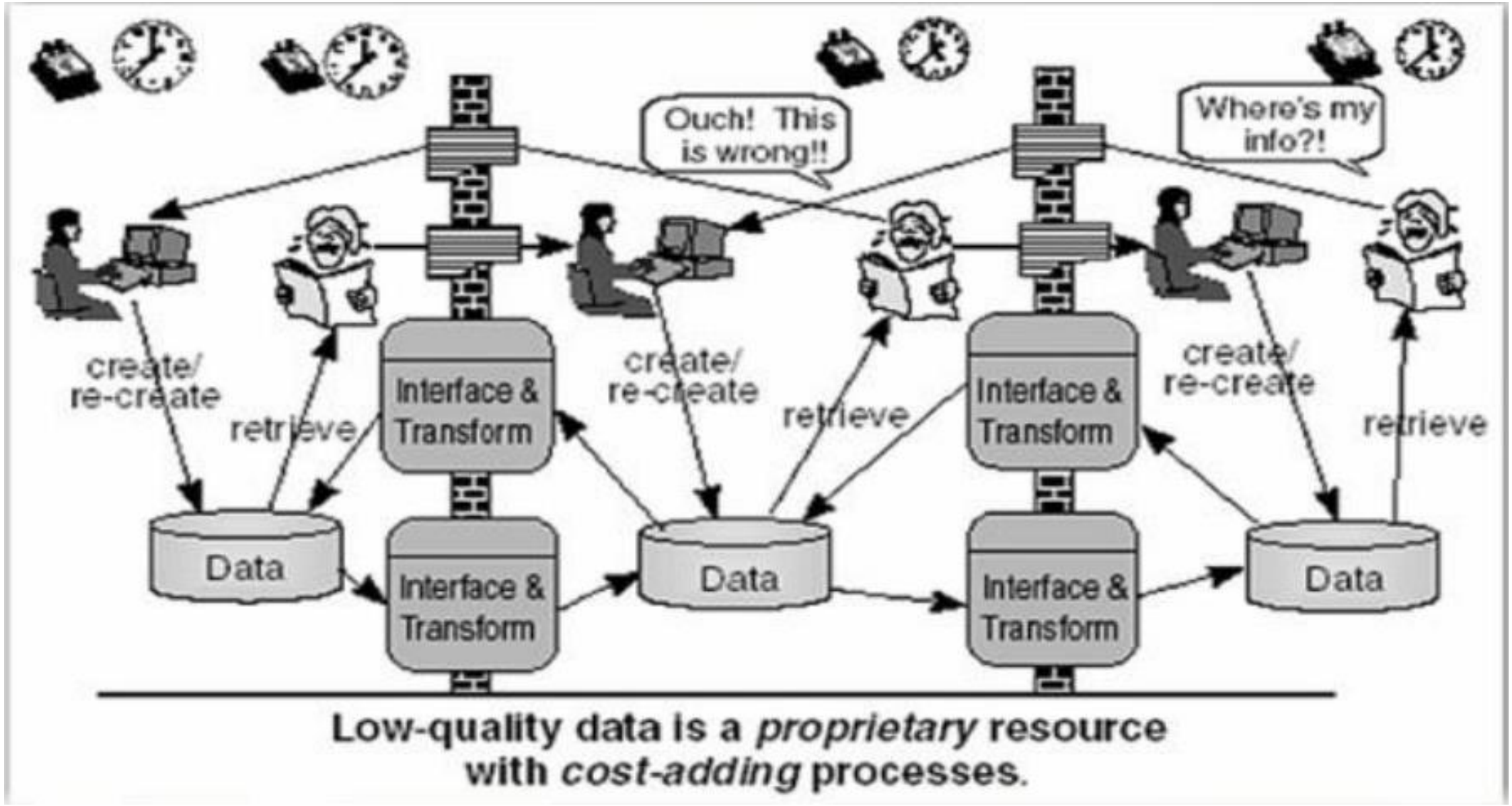
- Diffusion behavior analysis on call network
- Geospatial spread of analysis over temporal dimension
- Call network analysis to identify the advocates
- Behavioral analysis of unhappy customers
- Top factor of first time resolution

Outcome

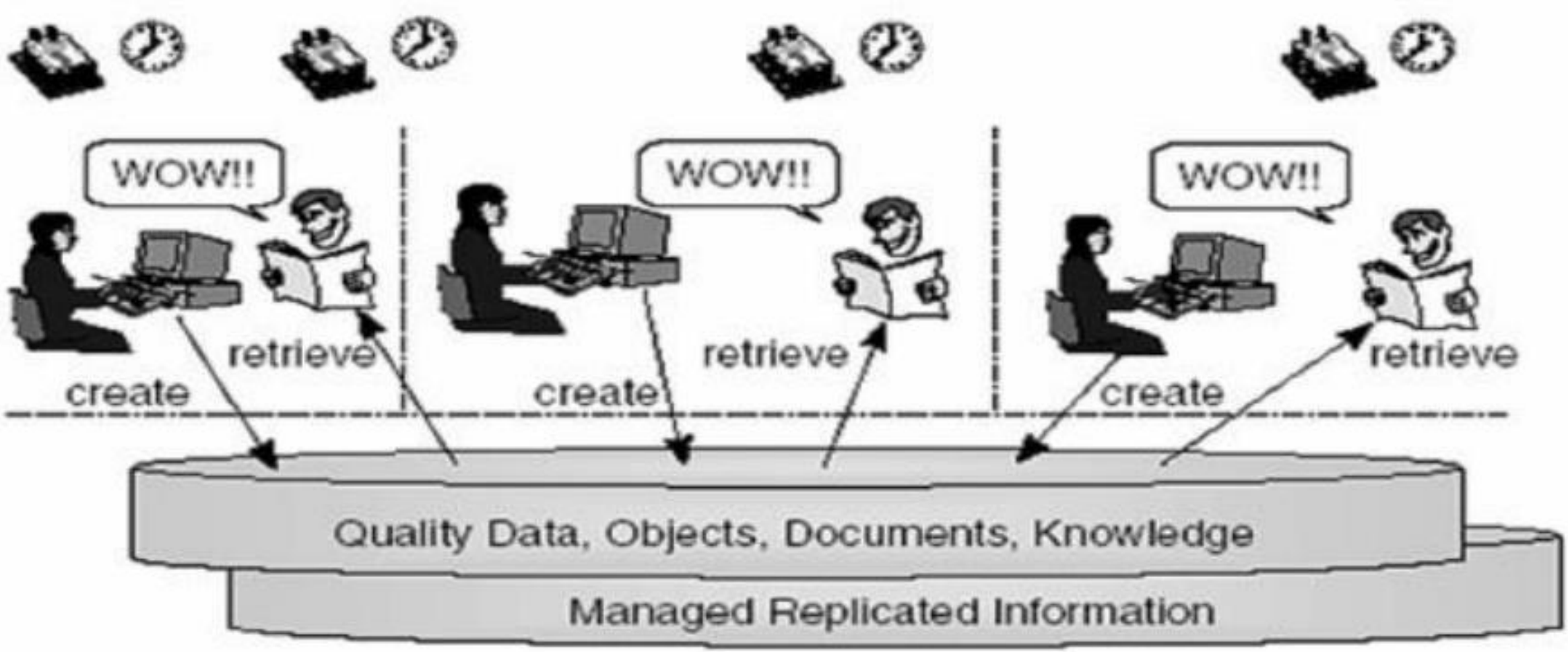
- Predict customers likely to be experiencing bad experience
- Impact on revenue from bad experience
- Impact on high value customers from bad experience
- Bad experience to churn conversion analysis
- Bad experience to bad debt conversion analysis



Total Cost of Ownership (TCO) Comparison – ERP system



Total Cost of Ownership (TCO) Comparison – ERP and BI



pwc

COURSE CERTIFICATE

08/14/2019

Sunil Subrahmanyam Yadavalli

has successfully completed

Data-driven Decision Making

an online non-credit course authorized by PwC and offered through Coursera

Alex Mannella
Principal
Data and Analytics Consulting

Verify at coursera.org/verify/DQGJDNSJAYLJ

Coursera has confirmed the identity of this individual and their participation in the course.

This certificate is issued by PricewaterhouseCoopers LLP with an address at 300 Madison Avenue, New York, New York, 10017.

Stay Strong

Thank you