Big Data: Impact and applications in Grocery Retail
Data growing exponentially

Data is growing at a 40 percent compound annual rate, reaching nearly 45 ZB by 2020

Data in zettabytes (ZB)

Source: AT Kearney
Sources of Big Data of increasing interest to the Retail business

Source: HortonWorks/Terradata
At a macro level
- THOUSANDS of stores
- HUNDREDS OF THOUSANDS of product sku’s to be sourced, shipped, sold
- TENS OF MILLIONS of customers
- BILLIONS of transactions
- 500 Petabytes of Data stored at US retailers alone

And a micro level – each loyalty customer is a walking data generator
- Dozens of transactions a year, online and offline
- Multiple communication channels: Web, Mobile, Email, Kiosk, Phone, Social etc..
- Massive amounts of (unstructured) data generated by mobile and social in particular
- No way to capture, process, analyse these growing datasets in most organizations today

Taming these Big Datasets brings competitive advantage: speed, efficiency, intelligence
80% of retailers are aware of Big Data concept

47% understand impact of Big Data to their business

30% have executed a Big Data project

5% have or are creating a Big Data strategy

Source: "State of the Industry Research Series: Big Data in Retail" from Edgell Knowledge Network (EKN)
How do we boil this vast data world down to make it relevant for us?

Big Data Analytics brings value to our entire value chain.
How do we boil this vast data world down to make it relevant for us?

Big Data Analytics brings value to our entire value chain
Macy’s

Dynamic Price Optimization:

- 876 stores in 45 states, 10,000 sku’s
- Analyses more than two terrabytes of price information to determine optimal pricing.
- Moved to cloud-based solution and speed of analysis increased more than 60-fold
- Now takes just 20 minutes to analyse 270 million different factors across 73 million products on a daily basis to auto-adjust pricing to drive higher sales

Time required for analysis 22 hours → 20 minutes
How do we boil this vast data world down to make it relevant for us?

Big Data Analytics brings value to our entire value chain
Reducing waste in stores:

- Tesco sells half a billion reduced price products every year as they approach sell-by dates
- Some products more price elastic than others and can have prices reduced later than others
- Tesco runs models to see how much they need to drop prices by and when in order to shift product and reduce waste
- Project delivers £30 million in savings a year

Stock Management in Distribution Centers:

- Tesco holds 12,000 items from 700 suppliers in average DC
- Different suppliers deliver with different lead times and with different rates/volumes
- Traditionally the more you order the higher the discount
- But Tesco deals with individual pieces and orchestrates using Big Data analytics in MatLab running full simulations of their stores.
- Simulations take 3 full days to run, but free up £50 million a year in working capital
How do we boil this vast data world down to make it relevant for us?

Big Data Analytics brings value to our entire value chain
Long-term commitment to Big Data analytics:

- David Dillon says Big Data is his secret weapon in the grocery wars
- Dunnhumby has 120 analysts working full time on Kroger business
- Teams sift through 300 terrabytes of data from 40 billion purchases made from 4 billion shopping trips over two years made by 42 million card-carrying customers
- Personalization of communications/promotions has been driving incremental sales and profitability for over a decade.
- Big Data analytics alone drives annual revenues in excess of $100 Million
- Pilot store in Cincinnati leverages data to serve up real time coupons to consumers in different aisles in the store

Mailing Program:

- Widely recognized in the industry as having the most sophisticated and influential retailer databases in the US
- 11 million households get regularly mailed coupons, 97% of those are personalized
- Focus is on growing basket size rather than acquiring new customers

Challenges: Privacy & Data Security

Grocery retail is exposed in many different ways

- **Customer personal data** exposed through Loyalty platforms
- **Financial data** exposed through payment/credit card process
- Liabilities different in different regions (US vs Europe for example)
- Cloud solutions add risk, encryption often essential
- Firewalls not designed for distributed computing architectures

Target

- Malware infection of POS system – Dec 2013
- 40 million customer credit/debit cards exposed to fraud
- 70 million customers had personal data exposed
- Separation of internal company data systems, especially those related to payment/credit cards essential

Adobe

- Malware infection of POS system – Dec 2013
- 40 million customer credit/debit cards exposed to fraud
- 70 million customers had personal data exposed
For the enterprise two key outcomes from a Big Data approach

**HUGE (ELASTIC) STORAGE**

**FAST/EFFICIENT PROCESSING & ANALYTIC CAPABILITY**

**Key challenges**

- Big Data expensive and difficult to mine with traditional means
- Low knowledge and agility in the organization
- Extending the past to grow into the future should be avoided
- Strategize what the framework of the future should look like, then start filling in the pieces
Big Data at Delhaize
Challenges

Legacy System Landscape

- Data stored in different systems
- Different teams in the organization need cross-system access to datasets
- Difficult to have automated and self-service reporting

Diverse datasets and data structures, accessed through different tools for a variety of purpose
Legacy System Landscape

- Data stored in different systems
- Different teams in the organization need cross-system access to datasets
- Difficult to have automated and self-service reporting

Diverse datasets and data structures, accessed through different tools for a variety of purpose
Example: Digital marketing environment tomorrow

- TB of item-level transaction data
- Pricing data
- Promotions
- Web analytics
- Weather Data
- Mobile data
- Social media Data

Digital Tools

- Web, Mobile, Social, Print, etc...
- Content Publishing Tools (Adobe etc.)

Loyalty Platform

- CRM Tools (Unica etc.)
- Analytics (SAS etc.)

Data Systems

- Enterprise data warehouse

BIG DATA FRAMEWORK

DB
DB
DB
DB
DB
Objective:
• Explore the power and potential of Big Data technologies
• Explore Cloud and Visualization technology
• Lay the foundation for a Big Data strategy

Goals:
• Partner with Google to explore capabilities and ease of use of Big Query
• Understand ETL processes associated with transferring large datasets from internal systems to Google’s Compute Engine
• Run benchmark queries to determine how much faster BQ runs than current internal systems.
• Identify gaps in expertise within our organization to transition to eventual real-time data operations (BI, Marketing, Digital, etc..)
Extract and Load of Data – no easy process

- Mainframe download to Zip (and then encrypted)
  56 hours (est. manually by Delhaize)

- Zipped Data
  56 hours (manually by Crystalloids Innovations)

- Cloud Storage
  28 hours (manually)

- BigQuery
  5 minutes for one years data

- BigQuery (Preprocessed Data)
Step 2: Run benchmark analysis to determine step change in speed

**Benchmark test to baseline BQ vs own internal systems:**
- Run substitutability analysis of pairs of product within a given product category

- Delhaize mainframe download to Zip (and then encrypted)
  
  60 hours to download + 56 hours to zip

- Import into Big Query
  
  5 minutes

Result:
- Internal systems 2 hours per category to analyse single pairs of product

- Big Query 2 mins per category: **60 times faster**
Step 3: Run cross-category substitutability and complementarity (impossible on current internal systems)

- BigQuery (Preprocessed Data)

  1167 queries executed in total
  15h 40m total query execution time
  100 €cloud costs sofar

- 2010 Cross Category Substitutability / Complementarity

Results: query successfully executed 15 hours 40 mins
Big Data in our US business
Example: Weather Impact on Location Sales
Weather impacts many parts of our business: Overall store sales, basket mix, replenishment, forecasting and others. Volumes are high so small variations translate into millions of dollars.
**Observation** – Both high temperature and rain impact beer sales. High temperature positively & rain negatively.
Coast vs Inland – Impacts of Temperature

**Observation** – In temperatures over 30°C seafood sales at the coast soar, whereas inland we see that there’s a preference for beef, veal & lamb.
Danger of being suffocated by growing data volumes

Need right tools, right infrastructure and the right skill sets to sift through

This drives efficiency, speed, better decision making and so a better value proposition for our customers
THANK YOU!