APICS

(American Production and Inventory Control Society)

A Presentation by

SEARS HOLDINGS CORP

on

Real-Time Inventory Tracking with Big Data Tools

- attended by Socrates

on Mar 25, 2015

Overview

- Fourth largest on-line retailer in United States
  - Next to Amazon, Wal-Mart, and Target
- Only one Datacenter – Troy, Michigan
- Went to online shopping in early 2000
- Purchase can be done
  - Online and Home Delivery
  - Online and Store Delivery
  - From Brick and Mortar store

Data Warehouse tools

- Terradata
  - Market leader
  - Provides both hardware and software together
  - Costly to scale (spent $11 MM)
- Greenplum
  - Now EMC / Pivotal company
  - Offered Map/Reduce
  - S/W only. But do have appliances too
- Netezza
  - From IBM

- Now all got consolidated to Terradata
  - Capacity: 1 PB

Challenges

- Due to non-Real-time (delayed) inventory
  - Product inventory sync-up took 24 hours till 2010, due to batch
  - Online orders were placed, but not delivered for store pick-up
  - Impacted shipment of items to online customers from warehouse
  - Impacted delivery of inventory from Warehouse to Stores
  - Revenues declined from $50B to $42B between 2008 and 2011
  - Lead to poor Customer satisfaction
- Data stored in different formats in different Legacy Systems (Mainframe and Terradata)
- No single version of Truth
- Volume: Average 100,000 Message /day
  Peak 77,000 Message /hour on Black-Friday

Big Data Setup

- Installed 100 nodes clusters from Cloudera in 2010
  - With Cloudera Manager and Cloudera Enterprise
- Scaled up with 50 more nodes
  - Happen to relinquish Cloudera
- Currently 638 Data Nodes
  - Each node has 4 TB x 2 disks
- Total Capacity (4 x 2 x 638) : ~ 5.1 PB
  - With Repl Factor 3, actual storage capacity: ~ 1.7 PB

Big Data Facts

- Initial slogan was, “We will keep all the data forever”
- Currently, almost 80% of 1.7 PB is filled
- Planning to add another 150 clusters in phased manner
- Do NOT have plan to
  - get rid off collected Hadoop data
  - migrate Terradata (DWH) into Hadoop
- Inventory period reduced from 24 hours to few seconds/minutes – REALTIME

Leveraging Big Data setup

- Web Harvesting
  - Getting prices from competitors' websites
  - Adjusting prices automatically
- Product Analysis
- Customer behavior analysis
- Analyzing Store Wi-Fi Data

Big Data Results

Value for the Organization

- Increased Sales by improving product availability
- Increased Customer Satisfaction
- Accurate and Real-time Inventory, Pricing, and Sales data

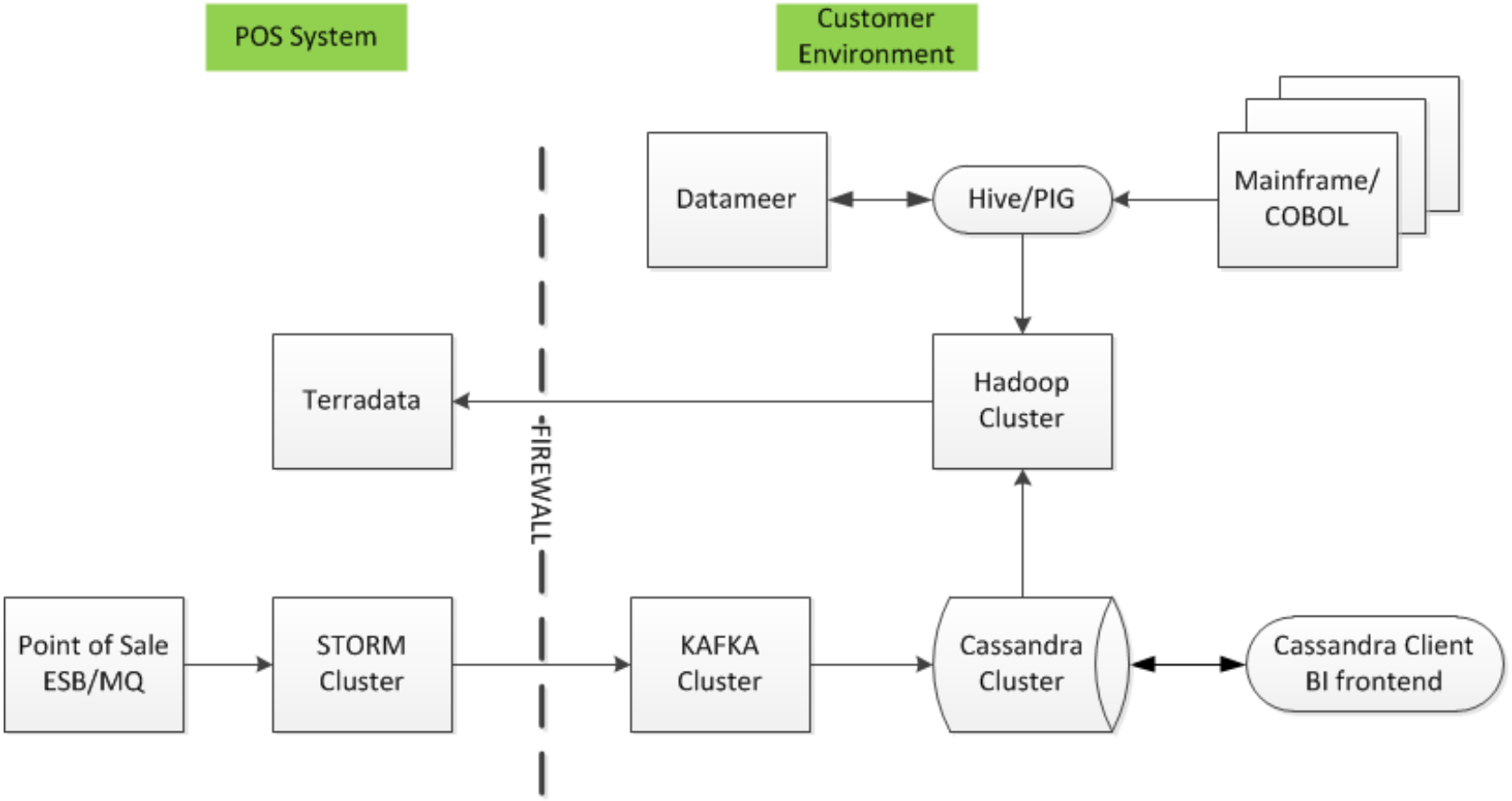
Cost savings from

- reduced truck times
- reduced Customer Service Center Calls

Big Data technology

- Apache STORM and KAFKA: To transfer PoS data
- Cassandra: Column-Family DB to consolidate data
- Hive and PIG: To transfer data from other system to Hadoop
- PIG: to migrate Mainframe/ COBOL programs
- Datameer: To integrate, visualize, and analyze data
- Infographics: To visualize the output in various forms (pie-chart, graph, histogram, linear model, etc)

High-level Architecture



MetaScale – Subsidiary of Sears

Thank you!