Chicago

bigdataeverywhere

Leading a Healthcare Company to the Big Data Promised Land:

A Case Study of Hadoop in Healthcare

Mohammad Quraishi (IT Senior Principal - Cigna)
atif71@gmail.com

About me

- BS in Computer Science and Engineering from University of Connecticut
- In the Healthcare Industry for over 19 years
  - Programmer most of my career - Architect, Designer
  - Worked in the SOA space for a number of years
  - Lead engineer in the mobile application space
  - Now Lead engineer in the Big Data Analytics Space - Hadoop

In my spare time

- Love to travel with the family
- Video games, music, movies
- Community relations work
- Fan of College basketball

Breakdown of the Hadoop Journey

1

Making the case
Vision
Architecture

2

The blowback
What we accomplished

3

Roadmap to the future
Lessons Learned
Questions?

The Elephant in the room

Image Credit: Guian Bolisay/Flickr

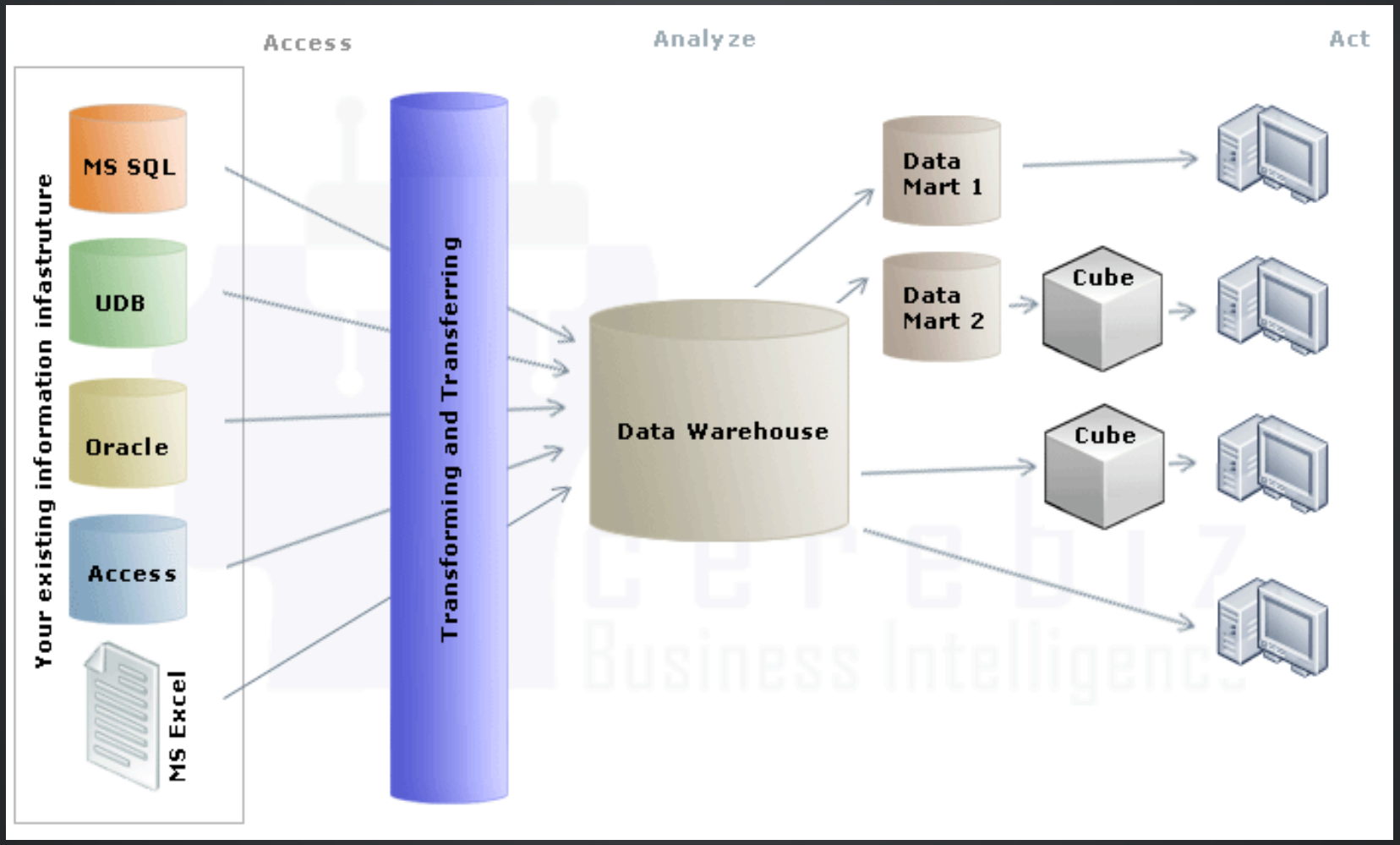
What's the problem?

We already have a mature data analysis infrastructure

And it looks something like this...

What we already do

- We have independent data marts
- We have the Hub-and-spoke architecture, the centralized warehouse



What is the vision?

The ability to perform

- Descriptive, Predictive *and* Prescriptive Analytics

Remove the traditional IT barriers separating the business users from insights

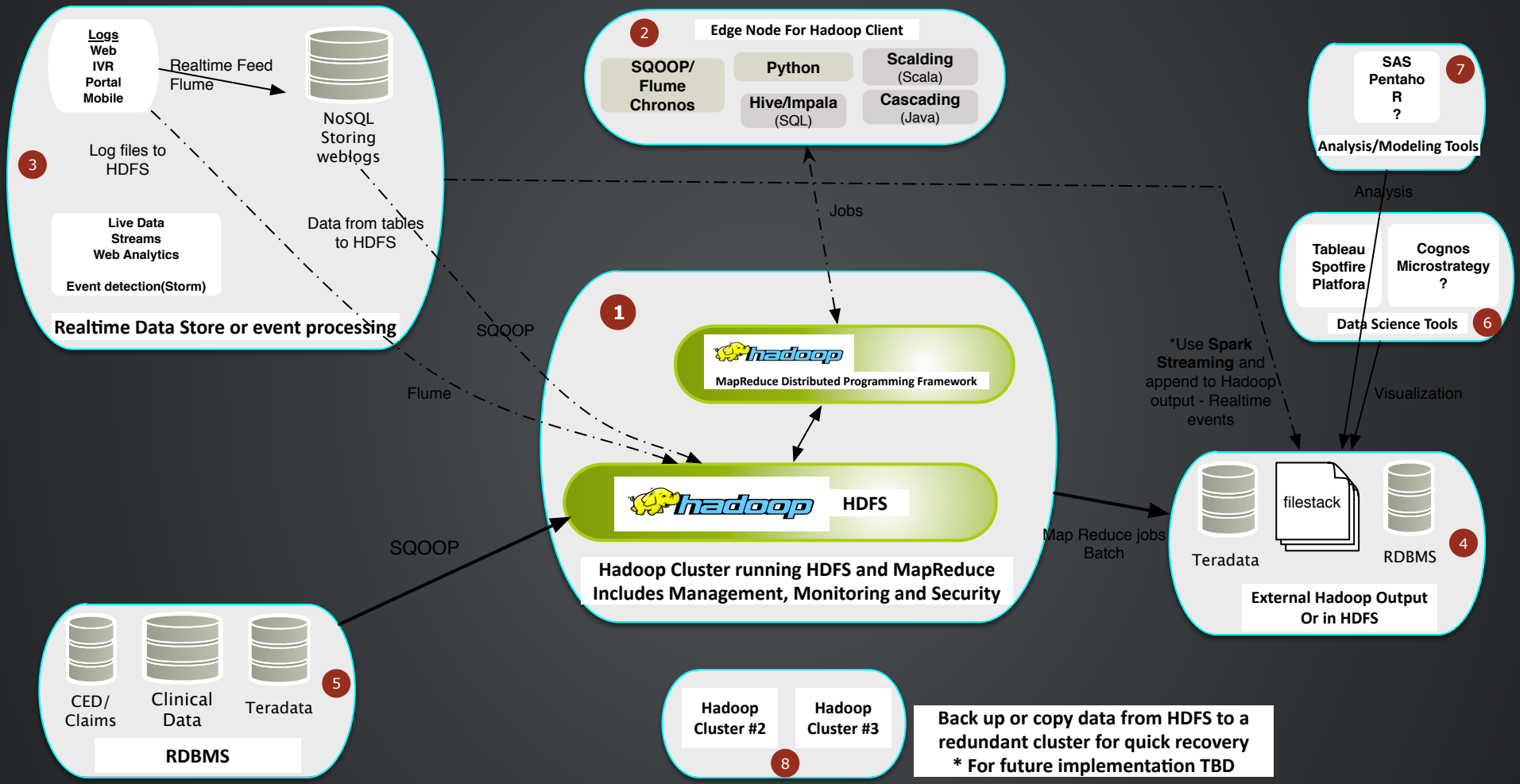
Benefits of Big Data

- Hadoop has the lowest cost per TB ratio of any data technology available
- Getting started with Hadoop is fairly inexpensive
  - “Entry-level” clusters relatively inexpensive
  - Grow in small steps

Benefits of Big Data

You don’t have to throw away data anymore!

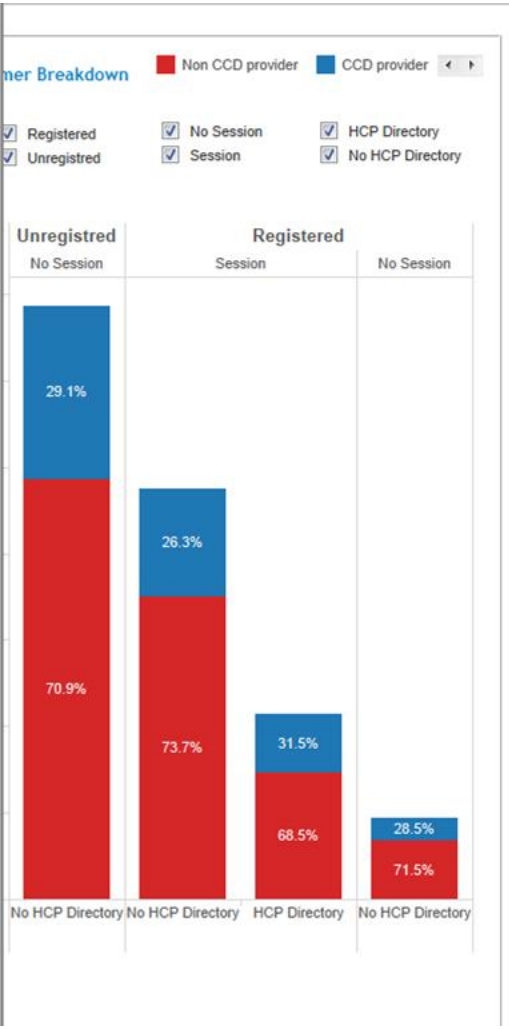
Vision - Reference Architecture

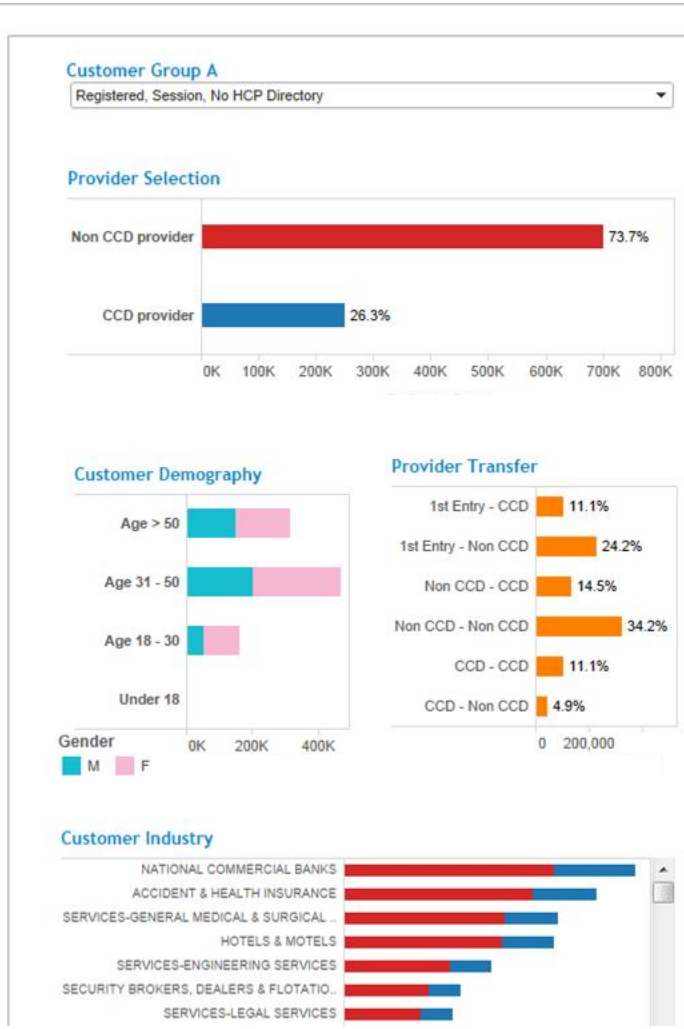


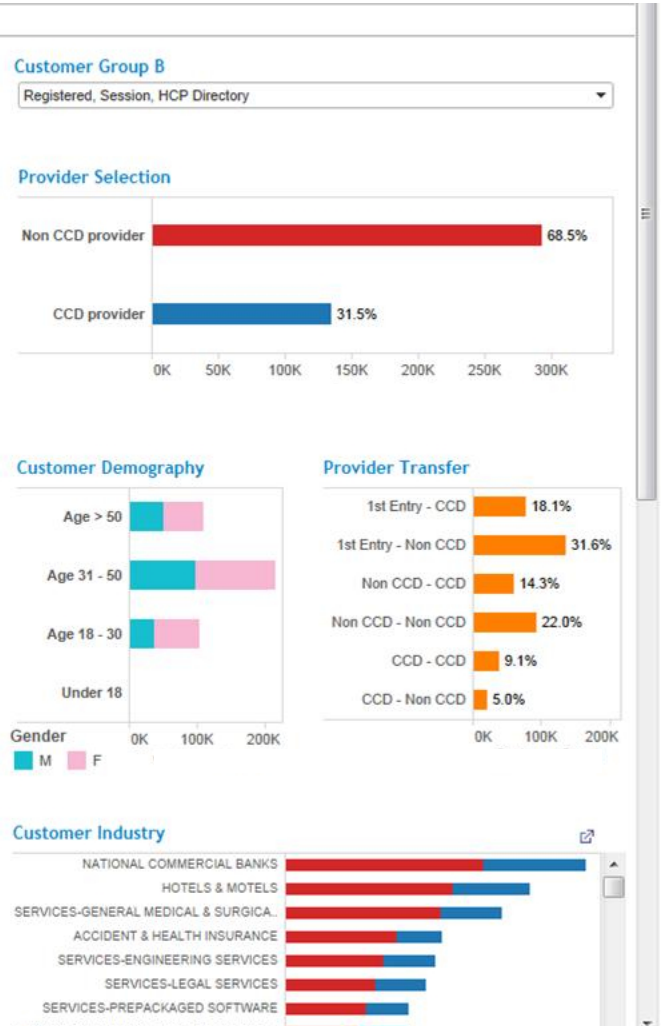
The Initial Evaluation

- Vendor Evaluation: Which relationship best fits our needs without lock-in?
- Selection of use cases for demonstration
- Visualization of those use cases

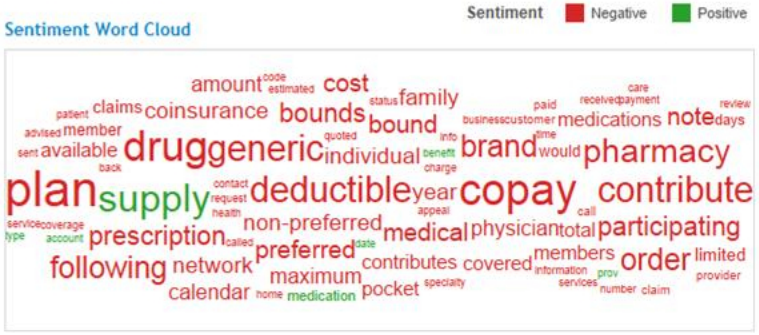
Use Case 1

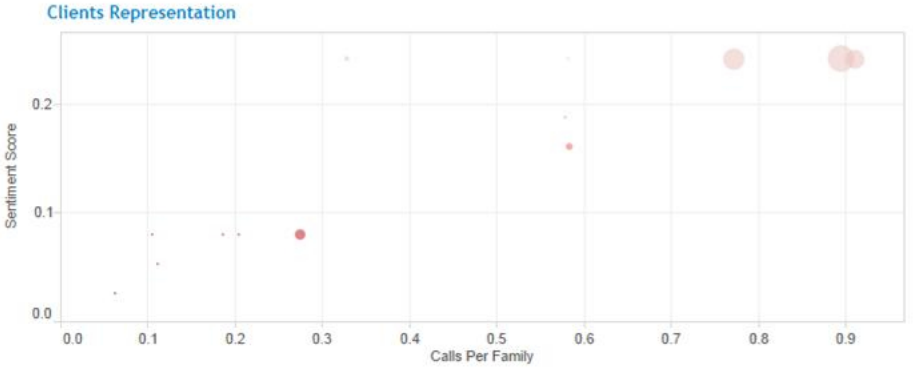

mer Breakdown
Non CCD provider
CCD provider
Registered
Unregistred
No Session
Session
HCP Directory
No HCP Directory
Unregistred
Registered
No Session
Session
No Session
29.1%
70.9%
26.3%
73.7%
31.5%
68.5%
28.5%
71.5%
No HCP Directory
No HCP Directory
HCP Directory
No HCP Directory

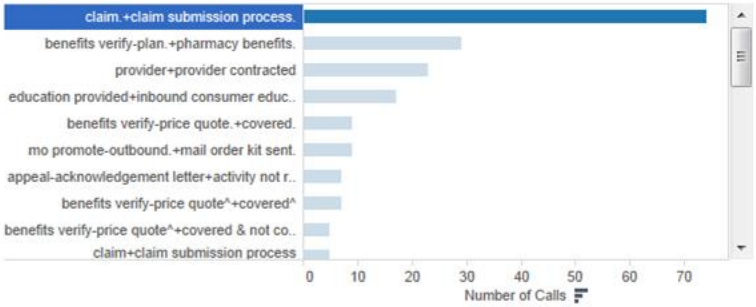

Customer Group A
Registered, Session, No HCP Directory
Provider Selection
Non CCD provider
73.7%
CCD provider
26.3%
0K 100K 200K 300K 400K 500K 600K 700K 800K
Customer Demography
Age > 50
Age 31 - 50
Age 18 - 30
Under 18
Gender
M
F
0K 200K 400K
Provider Transfer
1st Entry - CCD
11.1%
1st Entry - Non CCD
24.2%
Non CCD - CCD
14.5%
Non CCD - Non CCD
34.2%
CCD - CCD
11.1%
CCD - Non CCD
4.9%
0 200,000
Customer Industry
NATIONAL COMMERCIAL BANKS
ACCIDENT & HEALTH INSURANCE
SERVICES-GENERAL MEDICAL & SURGICAL ..
HOTELS & MOTELS
SERVICES-ENGINEERING SERVICES
SECURITY BROKERS, DEALERS & FLOTATIO..
SERVICES-LEGAL SERVICES

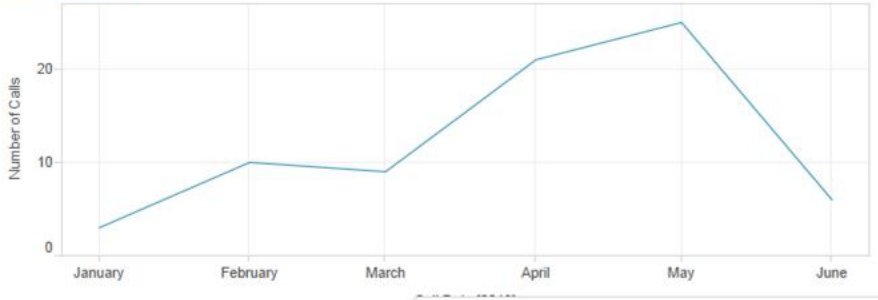

Customer Group B
Registered, Session, HCP Directory
Provider Selection
Non CCD provider
68.5%
CCD provider
31.5%
0K 50K 100K 150K 200K 250K 300K
Customer Demography
Age > 50
Age 31 - 50
Age 18 - 30
Under 18
Gender
M
F
0K 100K 200K
Provider Transfer
1st Entry - CCD
18.1%
1st Entry - Non CCD
31.6%
Non CCD - CCD
14.3%
Non CCD - Non CCD
22.0%
CCD - CCD
9.1%
CCD - Non CCD
5.0%
0K 100K 200K
Customer Industry
NATIONAL COMMERCIAL BANKS
HOTELS & MOTELS
SERVICES-GENERAL MEDICAL & SURGICA..
ACCIDENT & HEALTH INSURANCE
SERVICES-ENGINEERING SERVICES
SERVICES-LEGAL SERVICES
SERVICES-PREPACKAGED SOFTWARE

Use Case 2

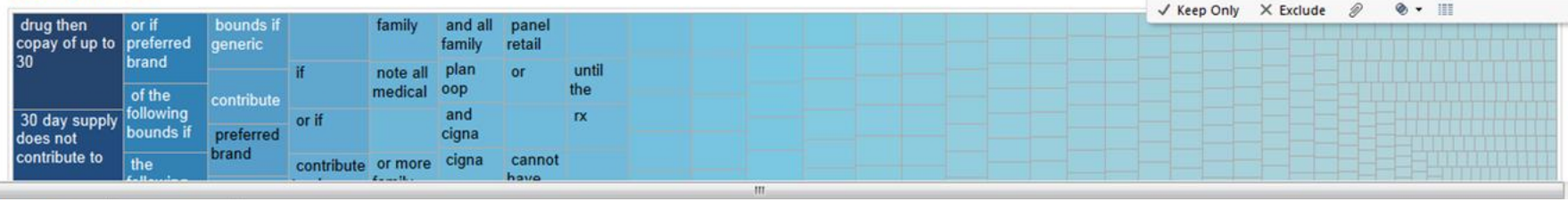








Associated Phrases



Success!

- Ready to tackle tougher more complicated problems
- Went out looking for more use cases

Ran into misconceptions

“Let’s use Hadoop as ETL!”

“Help us move data.”

“Can we back up data for archiving?”

… & Challenges

But Why?

- Overuse of the words “Big” & “Data”
- There was an overlap with other tools and platforms
- Hadoop looked like a swiss army knife
- Will it take over the world and replace other platforms?

Broader impact - Business Benefits

- Building a Customer Persona
- Service Ops efficiency
- Being Customer Centric
- Product Efficiency
- Brand Impact

Broader impact - IT Benefits

- Predictive threat modeling
- Data Archival
- Network Efficiency

Hadoop and Big Data

- Big Data = Hadoop + Relational + other suitable task related technologies
- Hadoop is complementary

Hadoop is Complementary

- Hadoop excels at processing and analyzing large volumes of distributed, unstructured, structured and semi-structured data in batch or near real-time fashion for analysis
- NoSQL databases are adept at storing and serving up multi-structured data in near-real time for web-based applications
- Massively parallel OLAP databases are best at providing analysis of large volumes of mainly structured data - Teradata
- SAS/R - Modeling and Business Intelligence
- Tableau - Visualization

Embrace the Most Important Change: *Culture*

Democratize your data and reap the benefits!

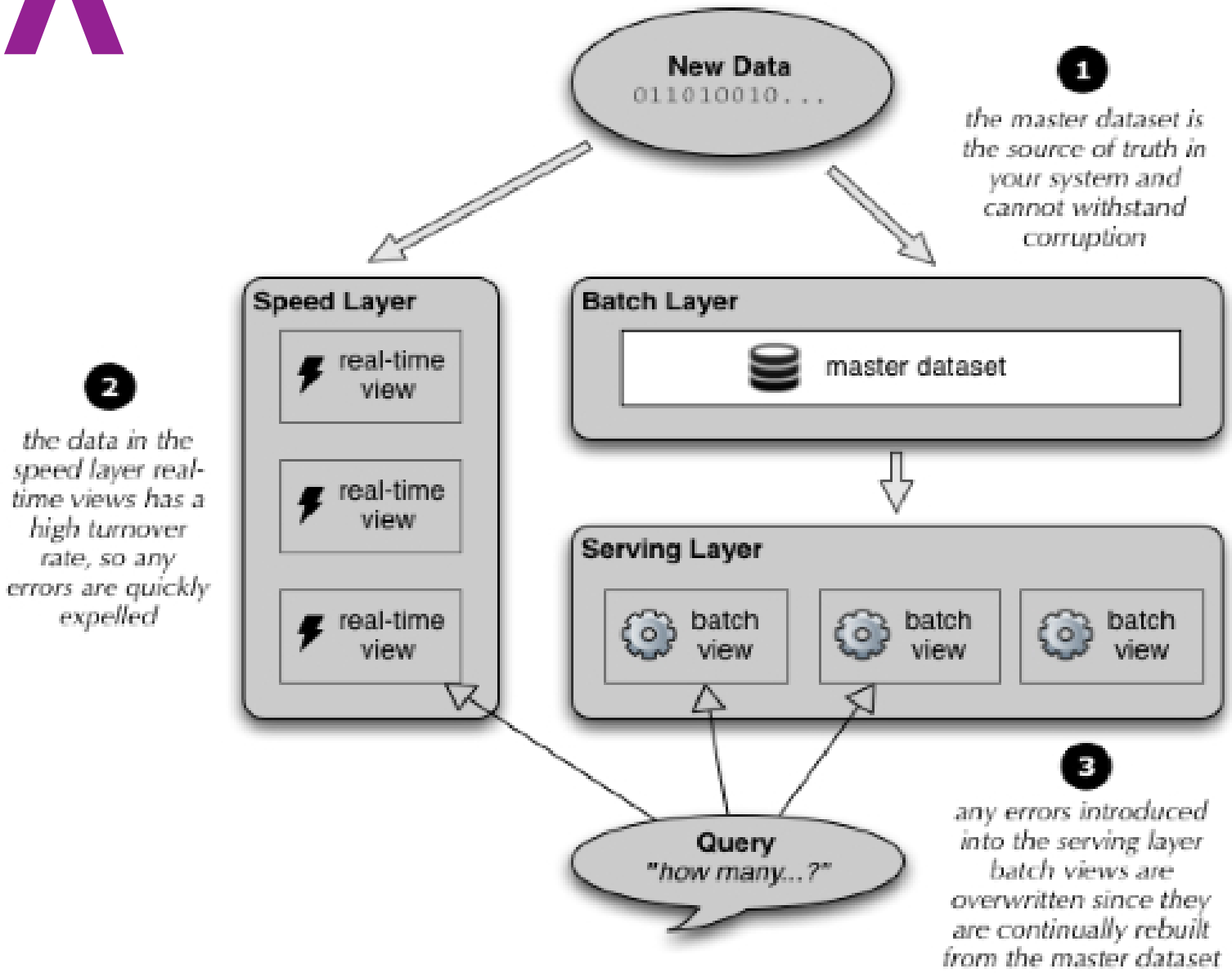
Why is Hadoop Complementary?

	Hadoop	Relational
Analysis type	Exploratory analysis to uncover value in data	Operational analysis of what was uncovered
Data granularity	Store High Volumes of Highly Granular data – lowest level; disk is cheap	Store transformed, aggregated data – conserve processing and storage costs
Time frame	Volumes and Varieties of data that is analyzed is streamed directly into Hadoop	Long term trending analysis from data that is provided by utilizing Hadoop

	Hadoop	Teradata
Maturity	Rapid evolution. Documentation and tooling are rough around the edges.	Stable, mature system.
Cost	Lowest $/GB available.	~10-100x the cost of Hadoop.
Data Model	Full spectrum from relational to unstructured, i.e., suitable for queries to machine learning problems.	Relational only.

What we accomplished?

- Evangelized Hadoop
- Linked Hadoop to BI Tools
- R on Hadoop
- A fail fast iterative analytics approach

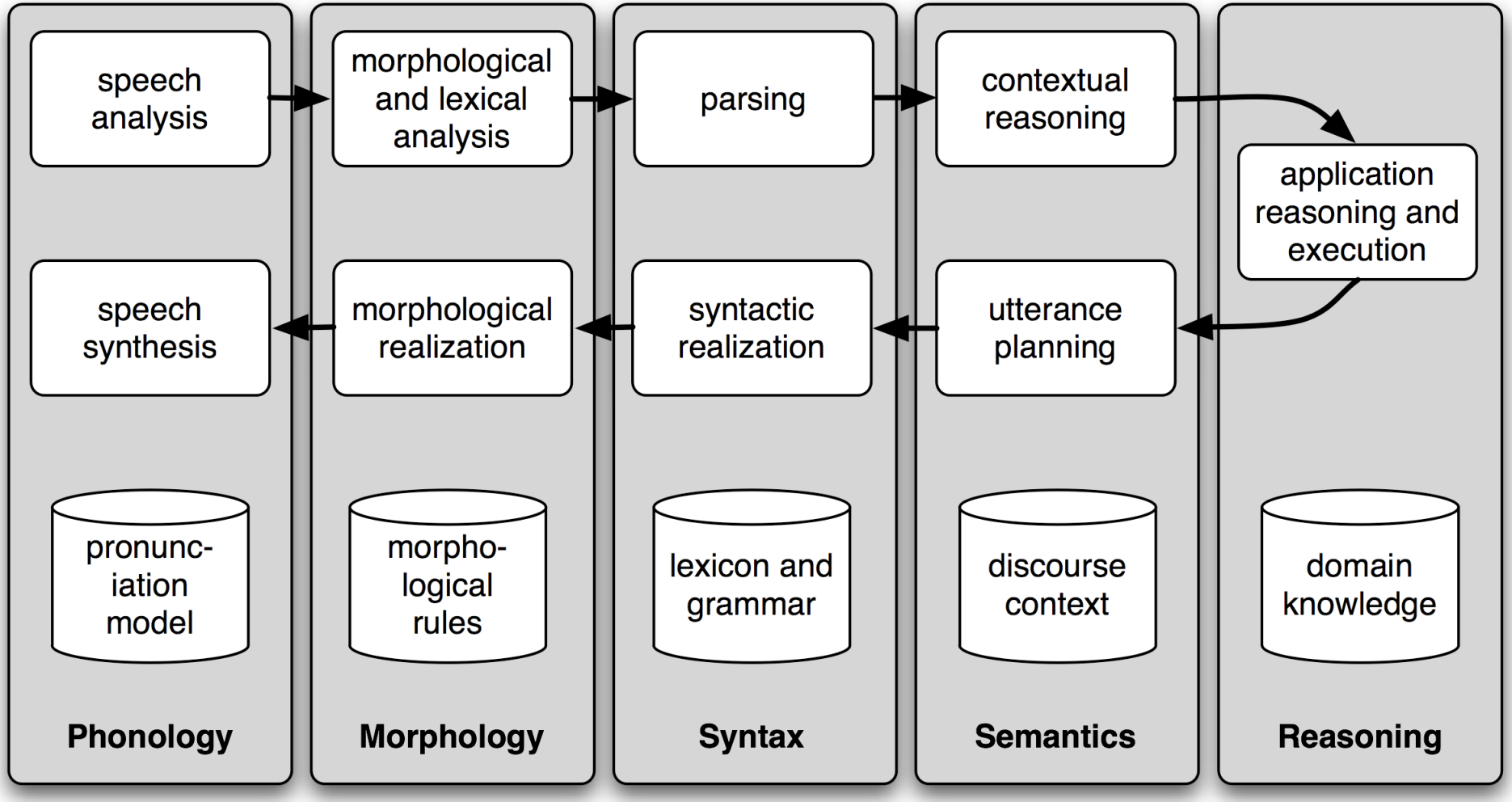


Credit Nathan Marz - Big Data

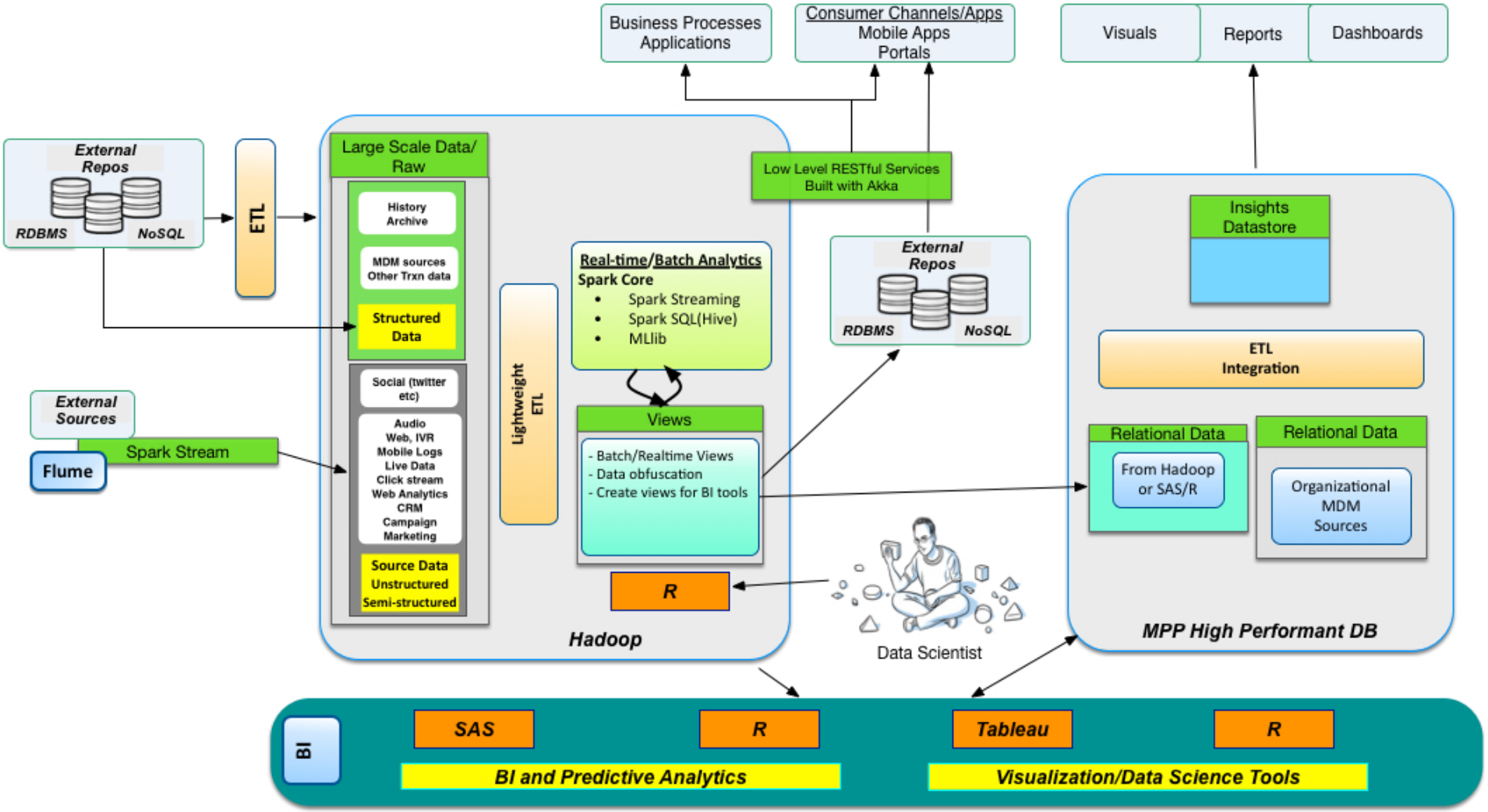
What we accomplished?

- ETL - Ingest, Transform and Move patterns
- Logs generated from consumer channels were ingested with Flume
- Standardized on Parquet (Storage) and Snappy (Compression)
- Lifecycle and organization of Data on HDFS
- LUKS - dm-crypt — for data at rest encryption
- Sentry and LDAP for Role Based Access Control

A Custom NLP Framework



A Roadmap to the Future

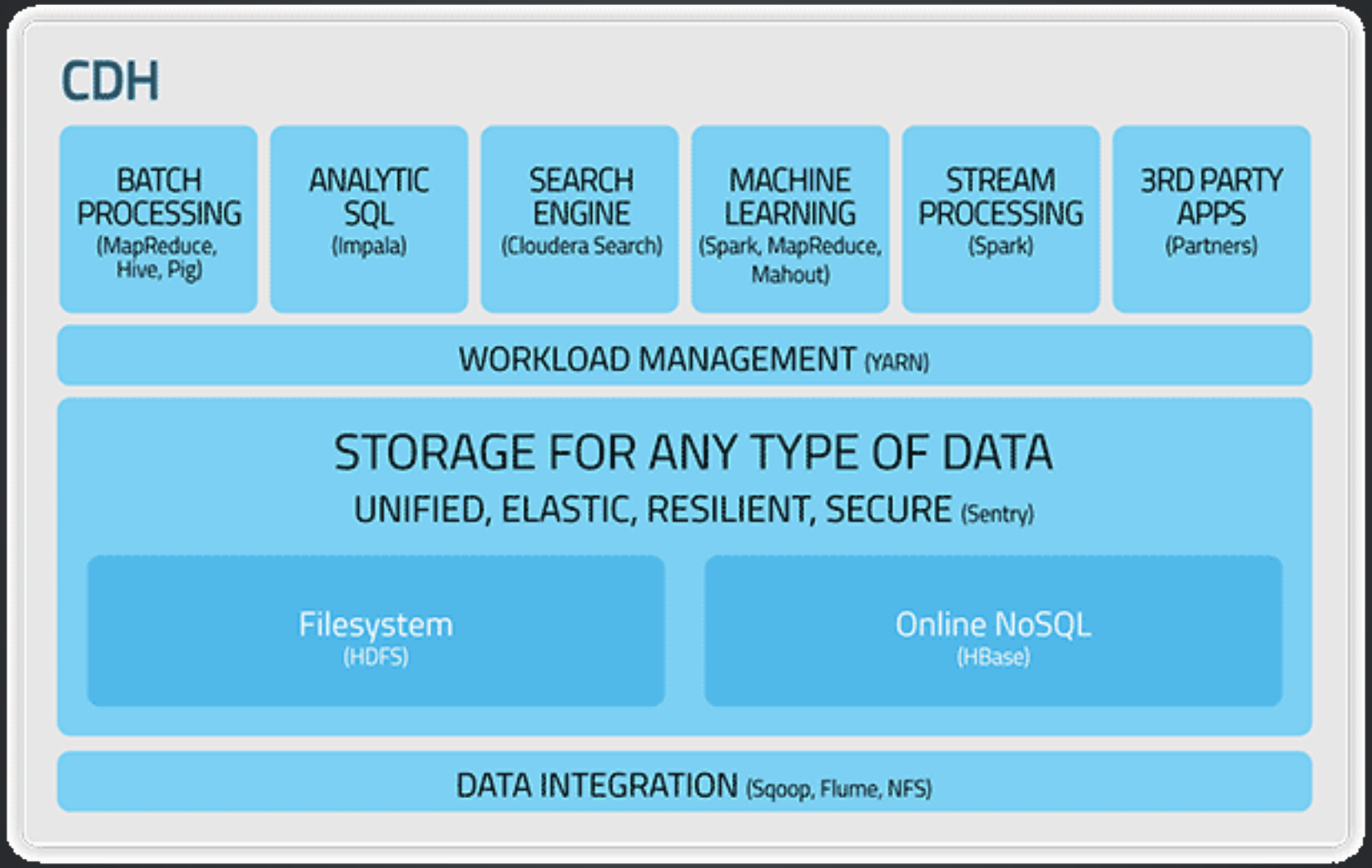


A Roadmap to the Future

Data Driven Solutions + FP

“Functional Programming: I came for the concurrency, but I stayed for the Data Science”
Dean Wampler

The Hadoop Stack – Advanced View



There's also Workflow Management with Oozie.

Lessons Learned

- Overuse of the words “Big” & “Data”
- The overlap
- Everyone found a use for Hadoop
- Big Change/Baby Steps
- Agility + Process = Cognitive Dissonance

Healthcare company needs

- Security
- Vendors
- Vendor Partnerships

WWYS

"Difficult to see. Always in motion
is the future…"
Yoda

"Many of the truths that we cling to depend on
our point of view."
Yoda

The Journey of a thousand miles begins with one cluster…

Questions?

Mohammad Quraishi (IT Senior Principal - Cigna)
atif71@gmail.com