

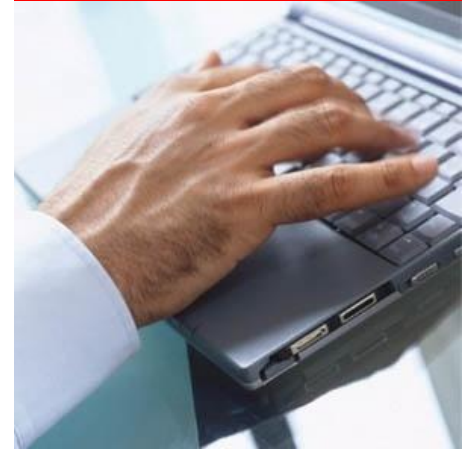
Oracle Enterprise Architecture

People. Process. Portfolio.

Acquiring Big Data to Realize Business Value

Agenda

- What is Big Data?
- Common Big Data technologies
- Use Case Examples
- Oracle Products in the Big Data space
- In Summary: Big Data Takeaways

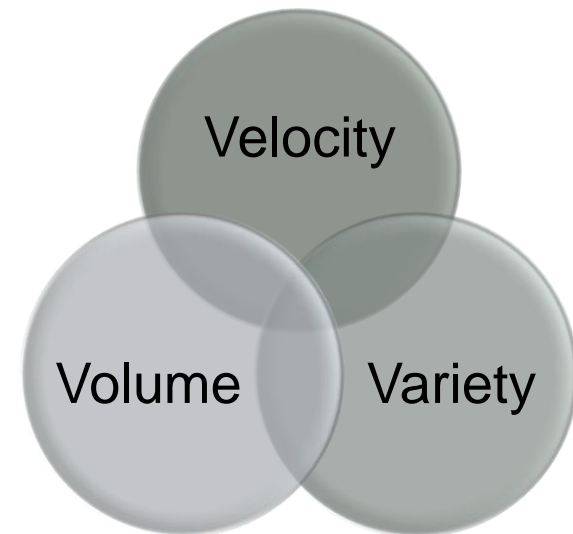


What is Big Data?

Big Data is the *convergence* of vast quantities of both structured and unstructured information sourced internally and externally to the business. New technologies are used along with traditional RDBMS systems to drive revenue or lower costs for an organization.

- Big Data is the *Convergence* of the 3 “V’s”¹
 - **Velocity**
 - Moves at very high rates
 - *Valuable* in its temporal, high velocity state
 - **Volume**
 - Fast-moving data creates massive historical archives
 - Valuable for mining patterns, trends and relationships
 - **Variety**
 - Structured (logs, business transactions)
 - Semi-structured and unstructured
- Companies are swimming in data
 - TBs/day of data arrival is here to stay

3 V's of Big Data



The Data Universe



Structured Data

Row and column structure
Related

Unstructured Data

Enormous transactional data
Enormous unstructured information
Too big for relational databases alone
New tools are needed

Velocity – Every 60 Seconds on the Net



Tapping into Diverse Data Sets

Information Architectures Today:

- Decisions based on classical transactional data



Video and Images



Big Data:

- Decisions based on all your data

Documents



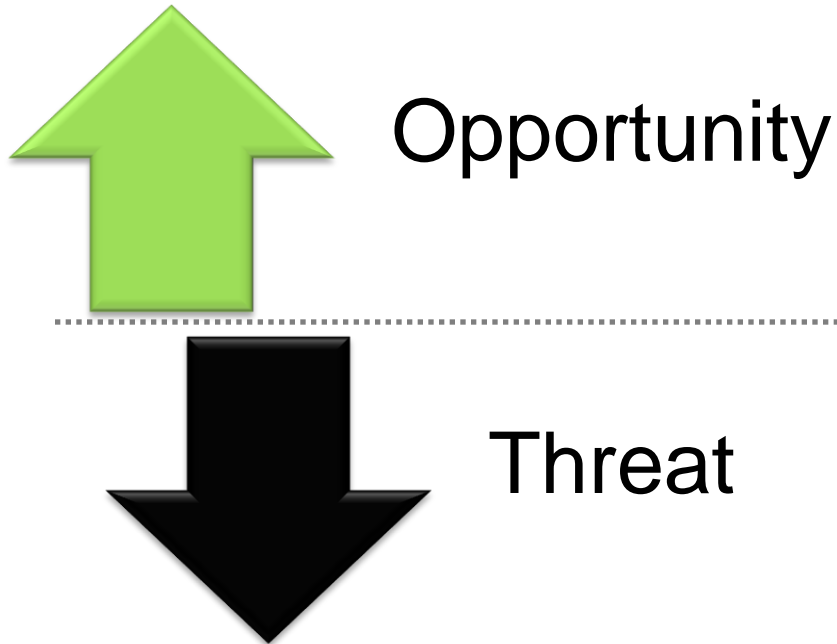
Social Data



Machine-Generated Data



Why Do Big Data?



- Big data can improve your top line today!
 - Big data can make you much more agile
 - Provides an edge over your competitors
-
- Big data is here – now
 - Your competitors will not miss out on the opportunity
 - Act now! Start building a big data platform for your organization

Big Claims of Big Data¹

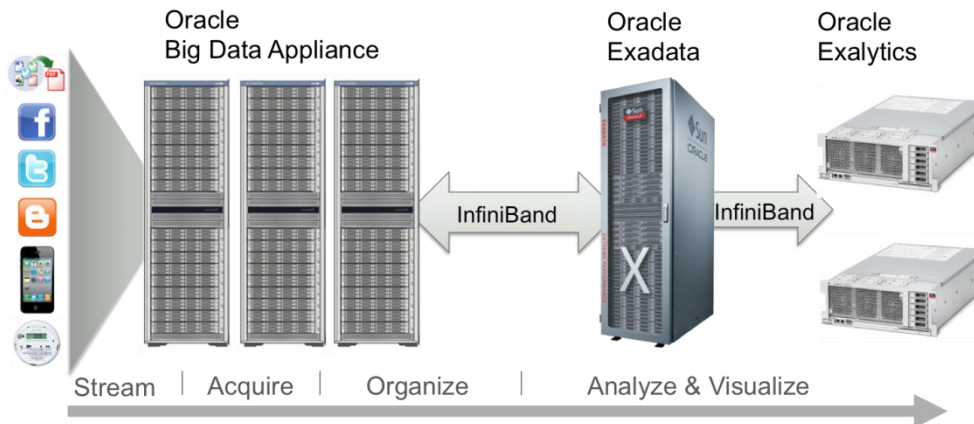
Industry	New Data	What's Possible	Why?
Communication Product Innovation, Operational Efficiency, Strategic Positioning	Customer Usage Data, Mobile Social Media Streams	New product innovation, pricing mix, cross marketing oppty	Product Differentiation, New revenue potential
Retail One size fits all marketing	Weblog, click streams	Micro-segmentation, recommendations	Increase net margin by 60%
Banking Fraud detection; risk analysis, Predictive Market Analysis	Weblogs, transaction systems, fraud reports	Semantic discovery; pattern detection	Billions of Dollars lost in bank fraud annually
Location-Based Services Based on home zip code	Personal location data	Geo-advertising, traffic, local search, more.	Increase revenue for providers by \$100B+
Utilities Resilient and adaptable grid	Smart meter reading, call center data	Realtime and predictive utilization analysis	Energy use expected to grow by 22 percent by 2030
Healthcare Improve Quality and Efficiency	Practitioner's notes; machine statistics	Best practices, reduced hospitalization	Increase industry value by \$300 B per year

Why Should the Business Care?

- There is a major industry shift underway – Big Data is Disruptive
 - The inclusion of vast amounts and variety of data is giving firms competitive advantages
 - New technologies allow us to leverage this convergence
 - Firms can act in the short term for value
- Oracle is releasing core products around Big Data

Oracle Big Data Appliance

Usage Model



- The business is looking for direction – ***Big Data is here now!***

Common Big Data Technologies

NoSQL DB

- Not Only SQL - Transactional data that does not have a row and column structure, is inconsistent, or sparse
- Dynamic and rapidly changing schema
- Predictable and bounded low latency store

MapReduce

- Break problem up into smaller sub-problems
- Distribute across thousands of nodes
- Can be exposed via SQL and in SQL-based BI tools

Hadoop

- Leading MapReduce implementation (from Apache)
- Highly scalable batch processing
- Highly customizable infrastructure

HDFS

- Hadoop Distributed File System
- Low cost, distributed, highly scalable storage
- Write once, read many times

Common Big Data Technologies (cont'd)

Apache Hive

- A query processor built on top of Hadoop
- Much more convenient than writing MapReduce logic
- Support limited SQL constructs and UDFs

HBASE

- Non-relational database modeled after Google's BigTable on top of HDFS
- Provides a fault-tolerant way of storing large quantities of sparse data

Open Source R

- Programming language for statistical analysis
- Language support for generalized linear models, cluster analysis, and other advanced statistical techniques

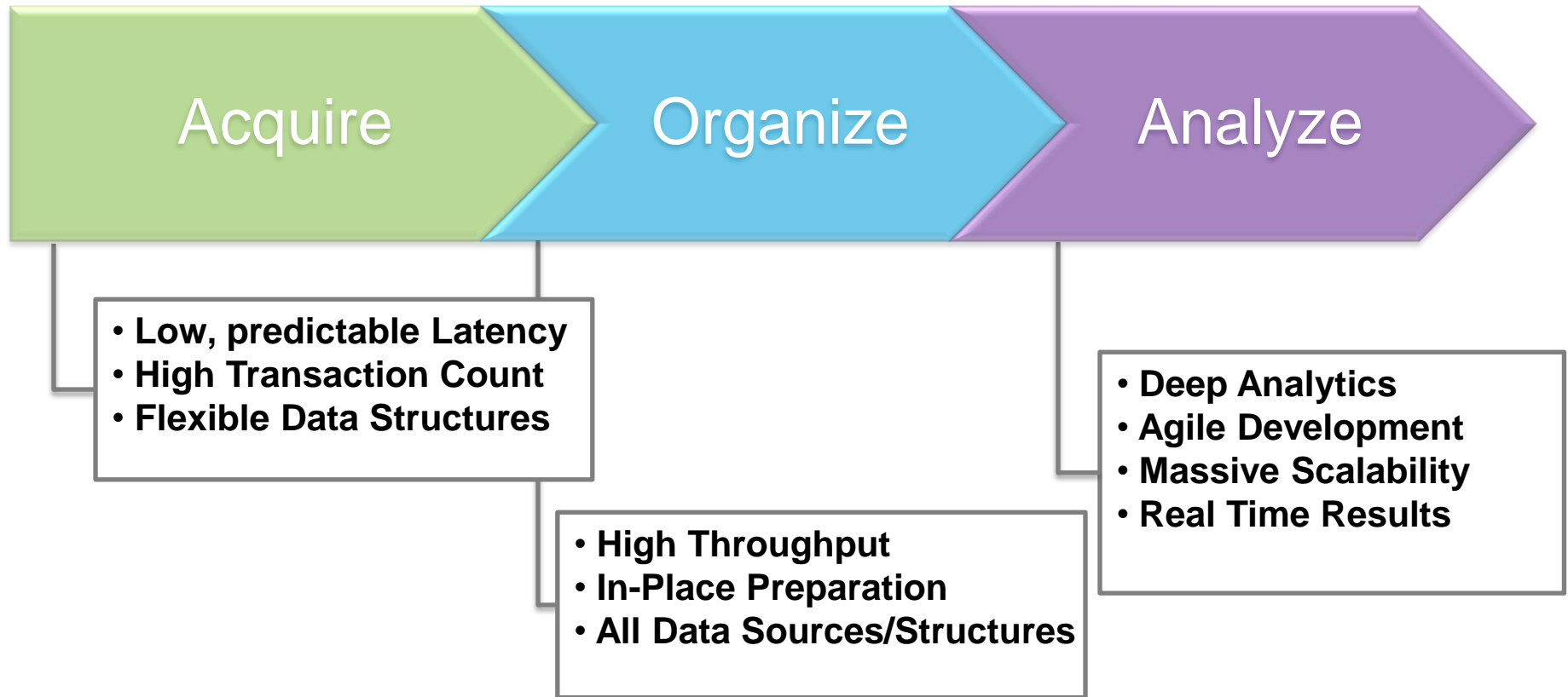
PIG

- A new programming paradigm to simplify MapReduce programming
- Developed at Yahoo – runs 30% of their jobs

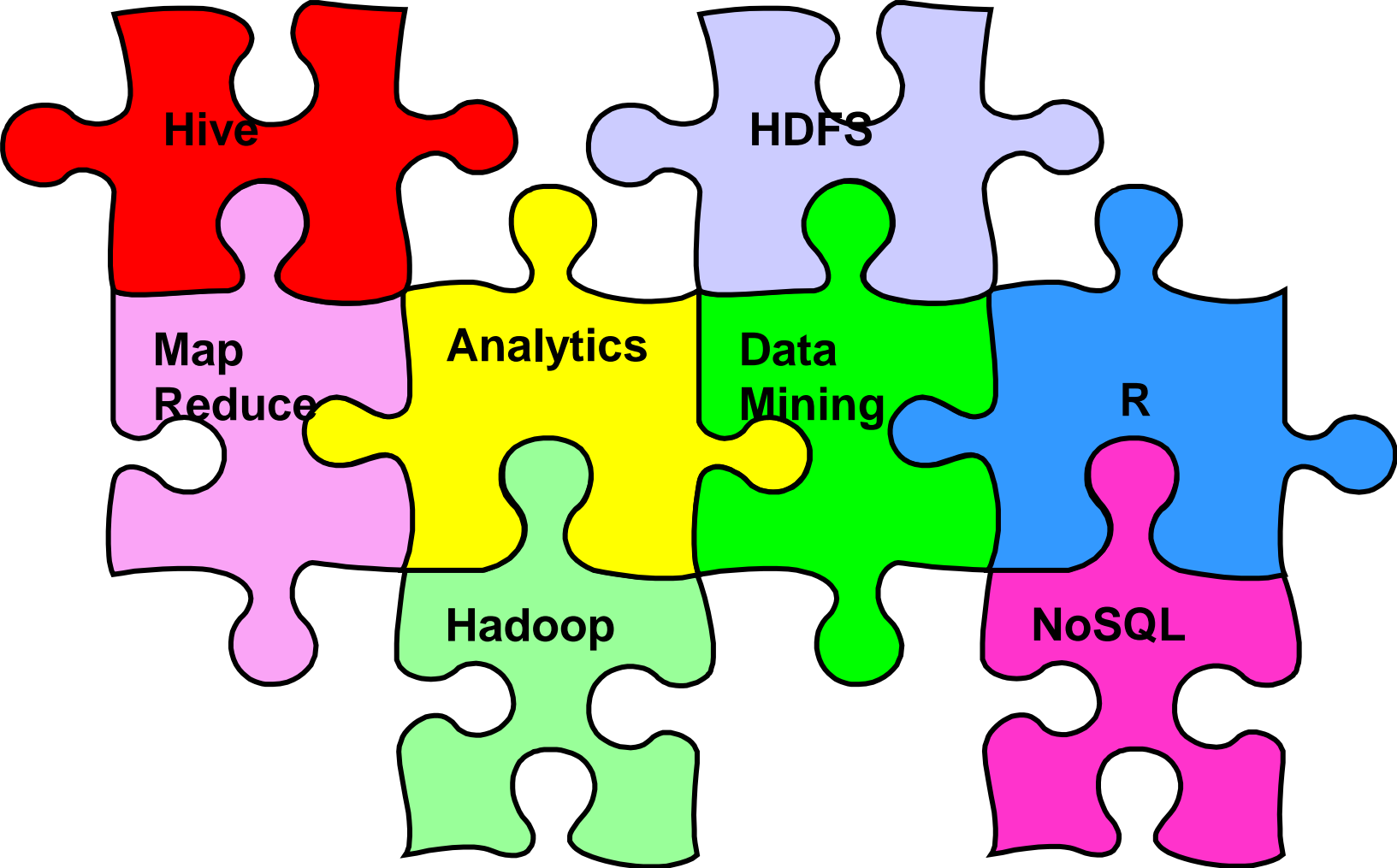
Sqoop

- SQL-to-Hadoop
- Parallel data import/export tool from Cloudera
- Tools to connect Hadoop to Relational Databases
- Export MapReduce results back to RDBMS end users

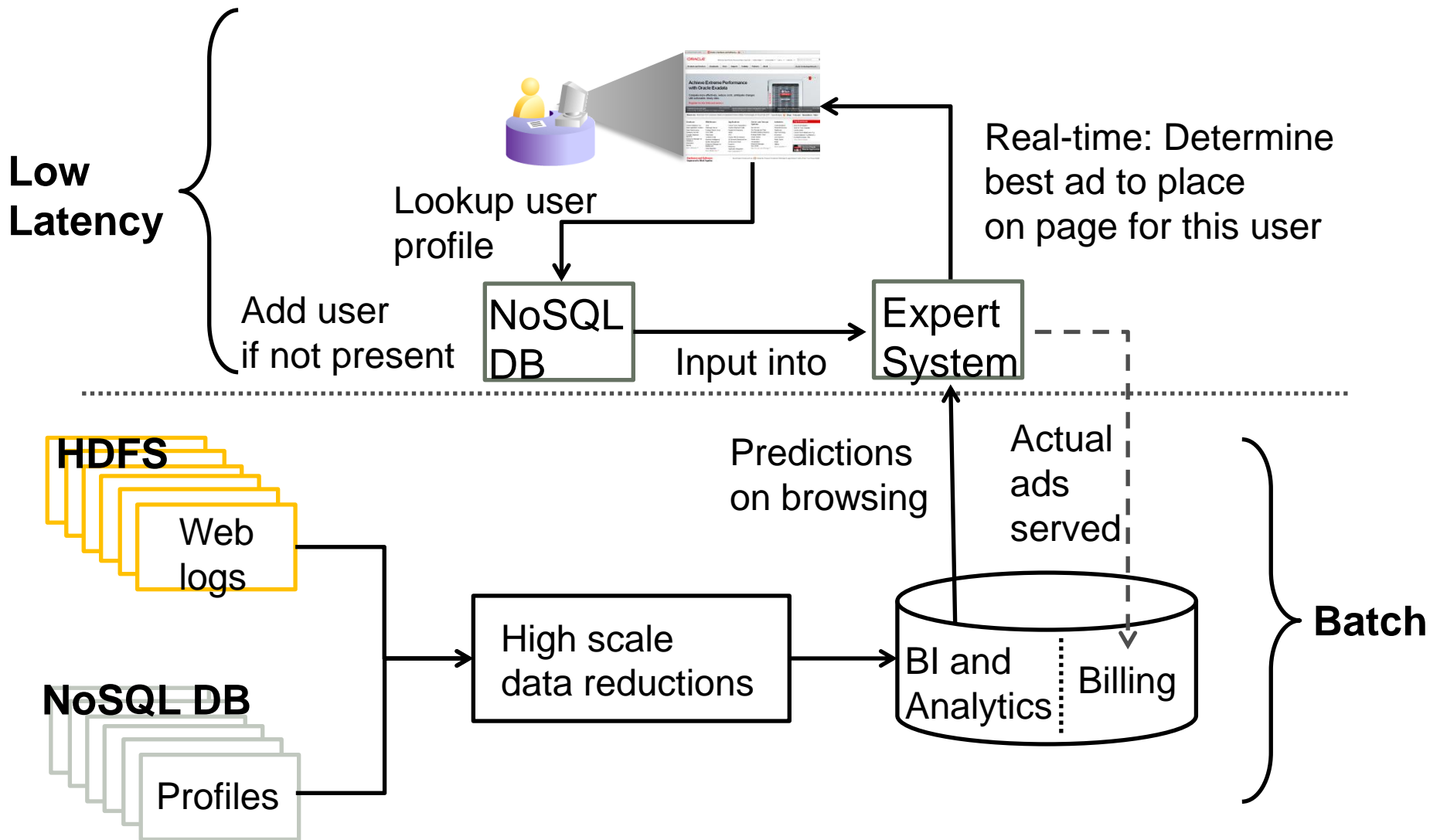
Process Flow of Big Data



Use Cases: Putting the Pieces Together



A Common Use Case: Predictive Ad and Content Generation

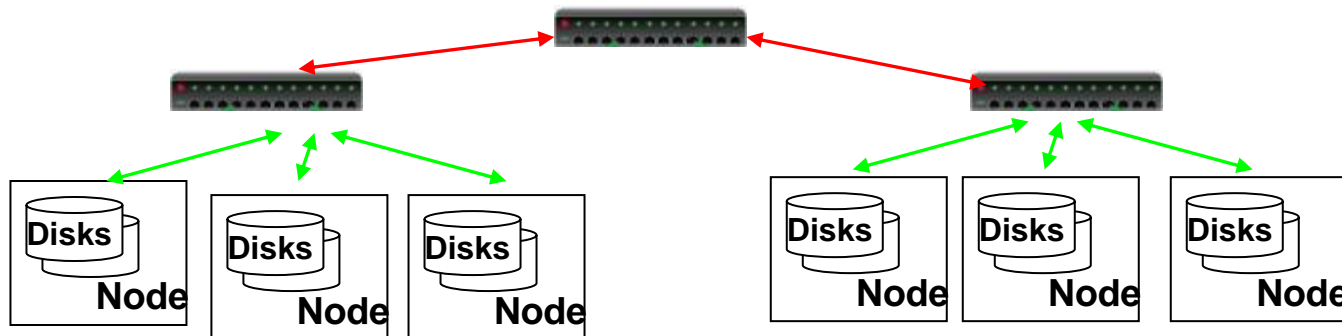




Data Warehousing @ Facebook using Hive & Hadoop

Looks like this ...

4800 cores, Storage capacity of 5.5 PetaBytes



Hadoop & Hive Cluster @ Facebook

- **Hadoop/Hive Warehouse – the new generation**
 - 4800 cores
 - 12 TB per node, Storage capacity of 5.5 PetaBytes
 - Two level network topology
 - 1 Gbit/sec from node to rack switch
 - 4 Gbit/sec to top level rack switch
- **Statistics per day:**
 - 4 TB of compressed data added per day
 - 135TB of compressed data scanned per day
 - 7500+ Hive jobs per day

Hadoop & Hive Usage @ Facebook

Types of Applications:

➤ **Reporting**

Eg: Daily/Weekly aggregations of impression/click counts
Measures of user engagement
Microstrategy dashboards

➤ **Ad hoc Analysis**

Eg: how many group admins broken down by state/country

➤ **Machine Learning** (Assembling training data)

Predictive Optimization for Advertising
Eg: User Engagement as a function of user attributes

➤ **Many others**

Large Bank “A” - Big Data Analytics Strategy

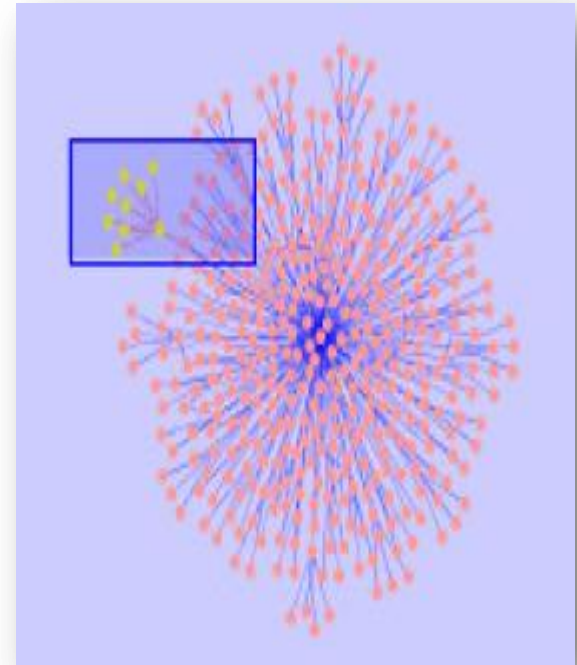
Offering Hadoop as a fully engineered shared service

5 of 7 lines of business are actively using Hadoop

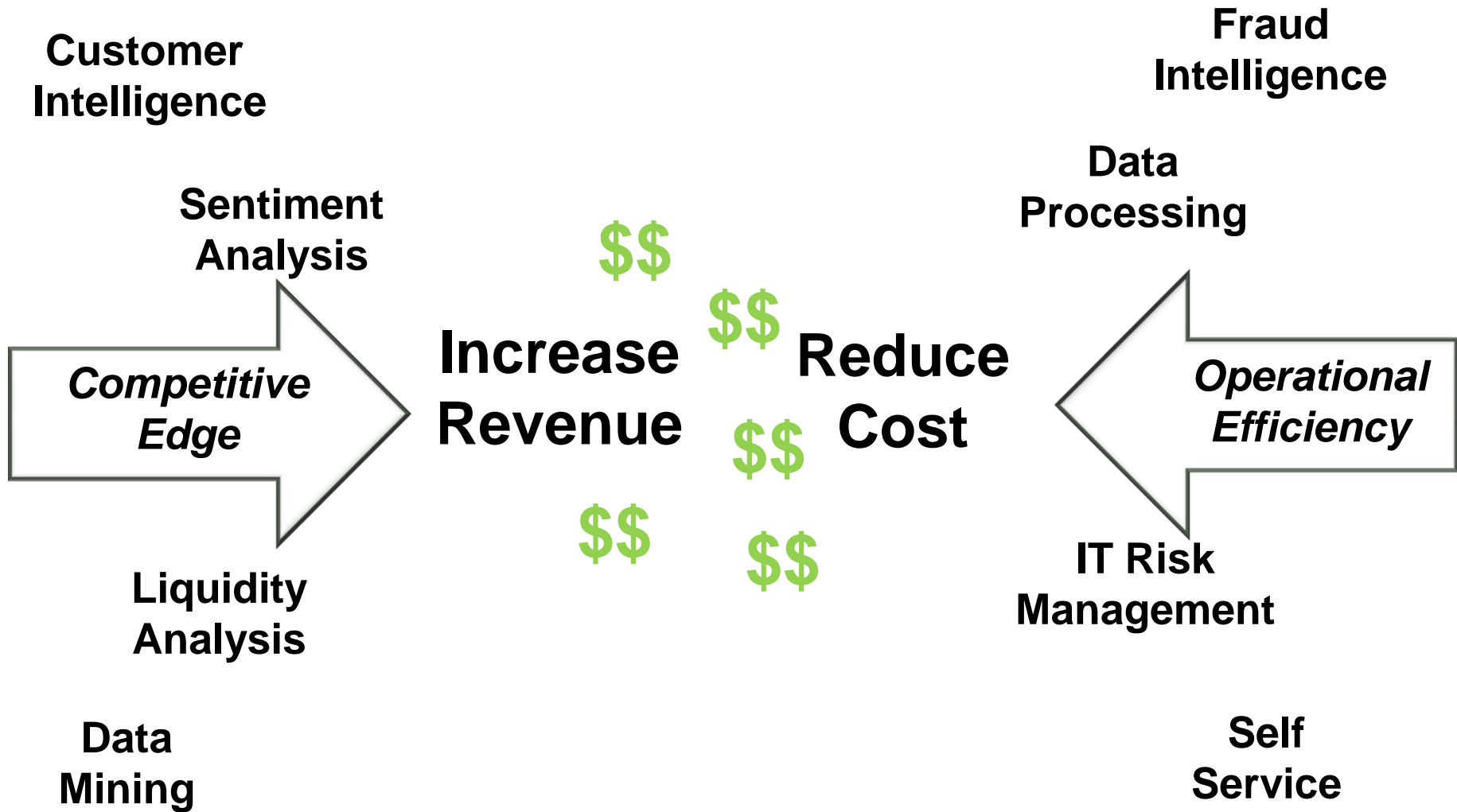
Building data scientist team

- A new role in most companies
- Has about 50, hiring & training more

Trying to influence the evolution of Big Data tools



How Bank "A" Uses Big Data



Bank “A” Use Case: Data Mining

Analysis of large datasets
(web logs, transactions, social
media)

Empowering Data Scientists to
bypass data modeling
process

Exploration of data across
business lines to identify useful
intelligence

Analysis across structured and
unstructured data

Retail Bank & IT Infrastructure

Fraud Prevention

Investment Management

Trade quality analysis

Recorded Futures

The Web is Loaded with Predictive Signals



Silicon Valley executives head to Vail, Colorado next week for the annual Pacific Crest Technology Leadership Forum

Drought and malnutrition hinder next year's development plans in Yemen...

The carrier may select partners to set up a new carrier as early as next month



"2010 is the year when Iran will kick out Islam. Ya Ahura we will."

"...opposition organizers plan to meet on Thursday to protest..."

"... Dr Sarkar says the new facility will be operational by March 2014.."



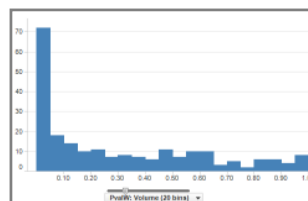
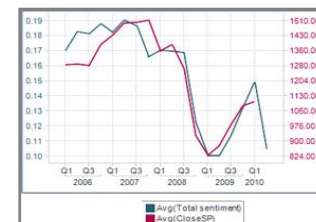
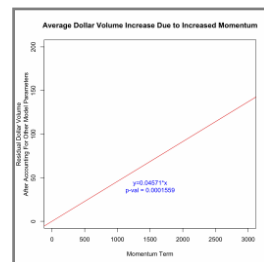
"Excited to see Mubarak speak this weekend!"

"Strange new Russian worm set to unleash botnet on 4/1/2012.."

"According to TechCrunch China's new 4G network will be deployed by mid-2010"

Recorded Futures: Selling Big Data as a Service

- Predicting trading volumes
- Predicting stock returns from sentiment
- Predict volatility by future events
- Single blog impact analysis



Banking

Analyze complex semantic graphs

- Data coming in from multiple input streams
- Unstructured text
- Schema-based data

- Consolidate data from multiple streams
- Extract entities
- Generate RDF/OWL capturing relationships between entities

- Identify relationships between data items
- Translate bits and bytes into higher level facts
- Rationalize data across silos
- Query large graphs

Identify fraud: Users with multiple identities

Utilities

- Data feeds coming in from multiple types of smart meters
- Data in multiple vendor-specific formats
- Semi-structured feeds

- Extract relevant location, time and consumption values from raw data feeds
- Transform data into a standard spatial format

Spatial analysis

- Aggregate results based on geographic service areas
- Perform what-if analysis on the distribution network

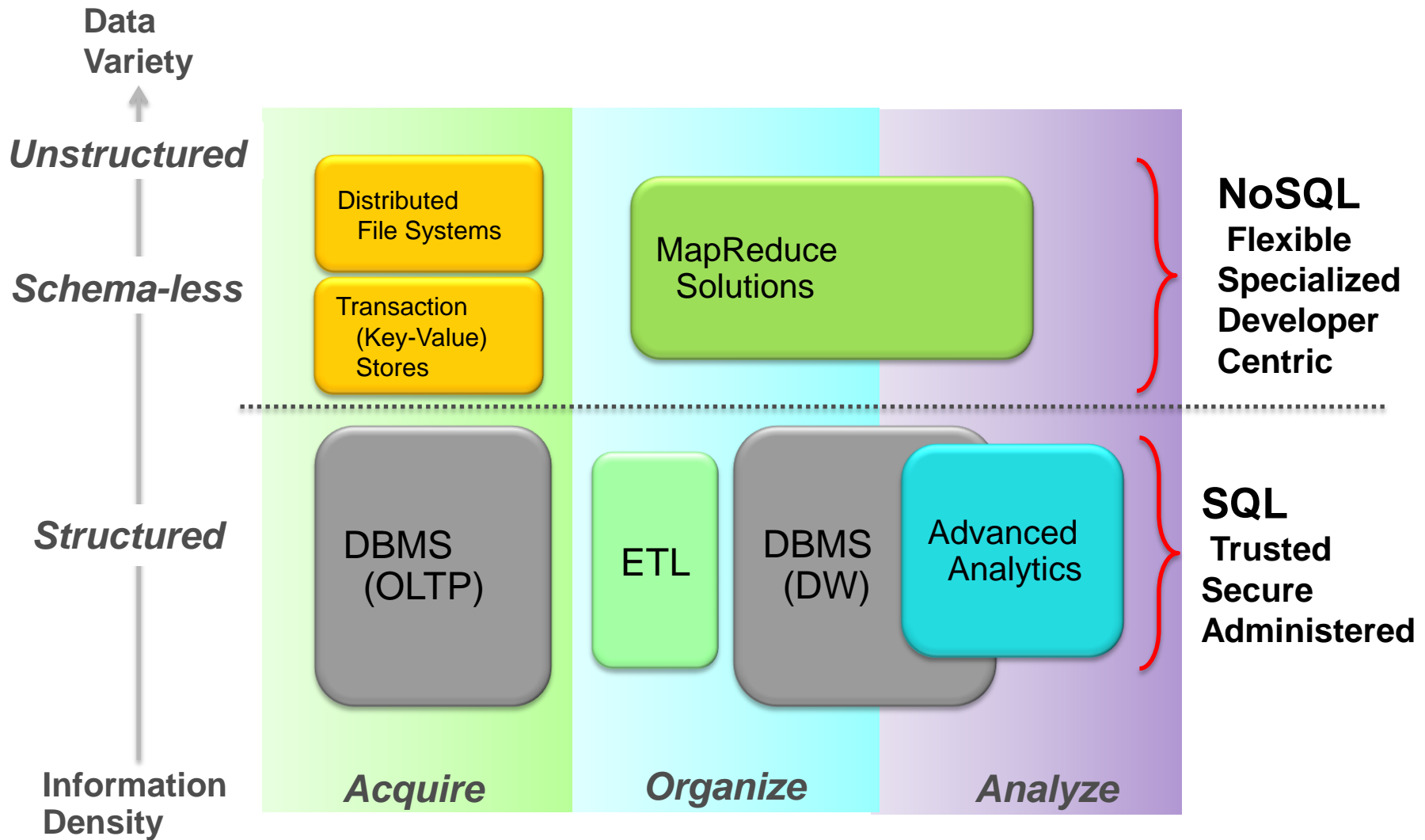
Analyze and display service areas with high or low utilization based on smart meter data

Oracle Products in the Big Data Space

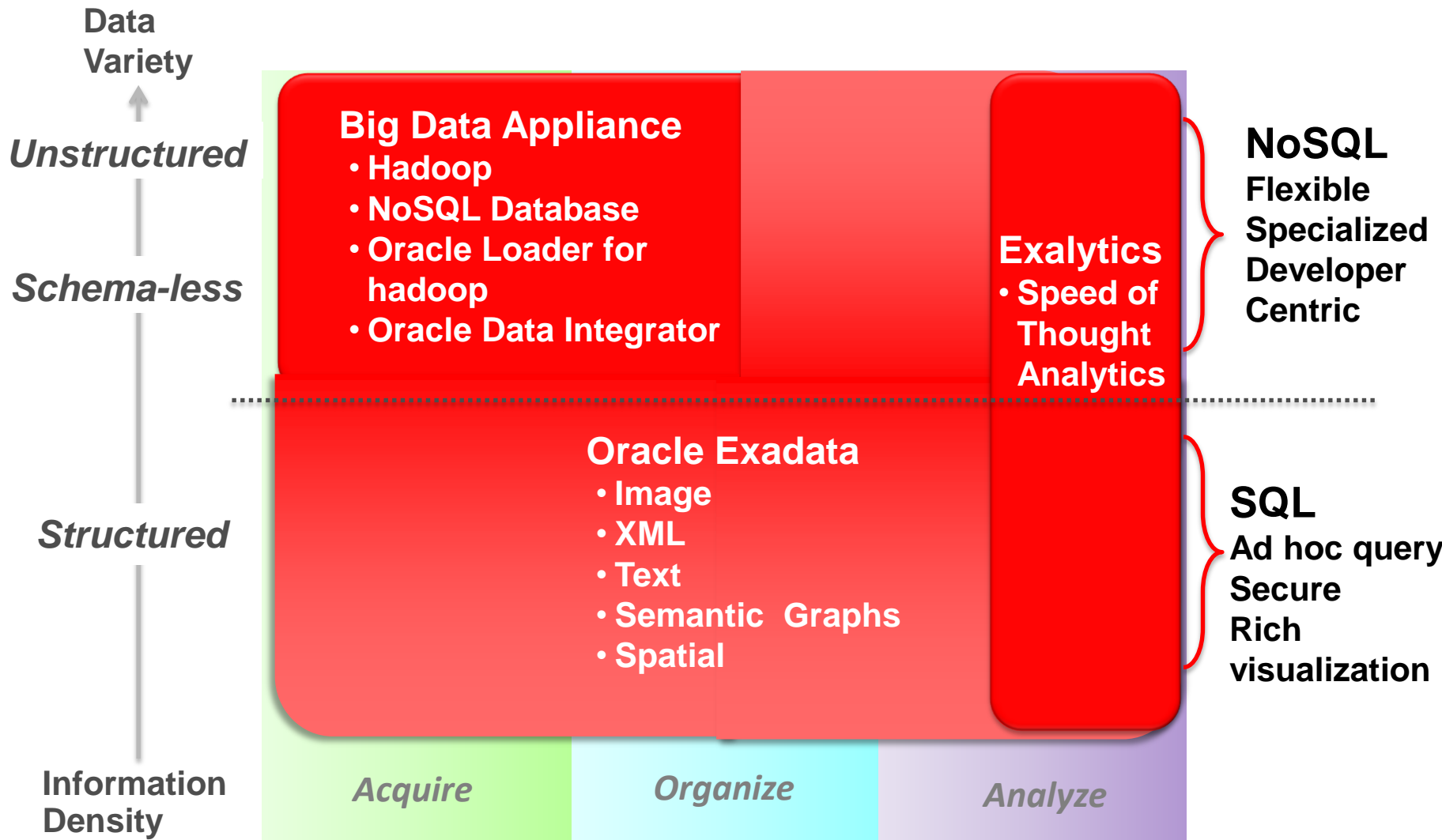


Bridging the Chasm Between Structured/Unstructured Data

Total Data Solution Spectrum

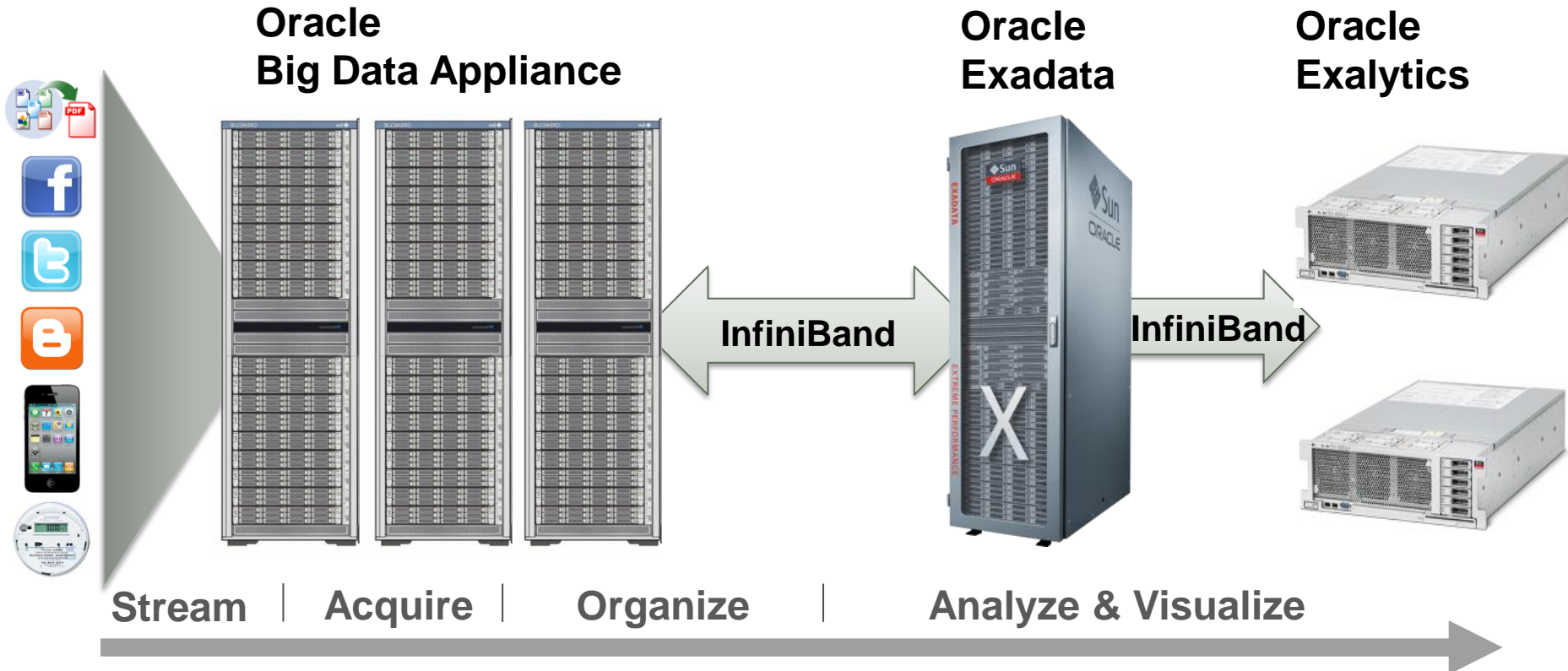


Oracle Engineered Solutions



Oracle Big Data Appliance

Usage Model



Oracle Big Data Appliance

- Optimized for acquiring, organizing, and loading unstructured data into Oracle Database
 - 18 Intel CPU Servers
 - 216 cores
 - 864GB memory
 - 432 TB storage
 - 40 Gb p/sec InfiniBand
- Oracle Big Data Appliance includes:
 - Oracle NoSQL to capture and manage data
 - Open source Hadoop to transform data
 - Application Adapter for Hadoop to load an Oracle Database
 - Open source distribution of R
 - Oracle Enterprise Linux



In Summary: Big Data Takeaways

- Its here!
- Can't be ignored
- The technology is maturing
- We need to be prepared to advise our clients
- The 3Vs: Volume, Velocity, and Variety
- Adds value to business decision making process
- Will lead to a competitive advantage
- We need to be prepared to advise our customers, both internal and external



Q&A