

Production ready hadoop

By Deepak Rao
National Head – Datawarehousing
Bajaj Finserv

Agenda

- Data in today's BFSI world
- Modern Data Lake
- Use cases & prototyping
- Big data impact in BFSI
- Thank you!!

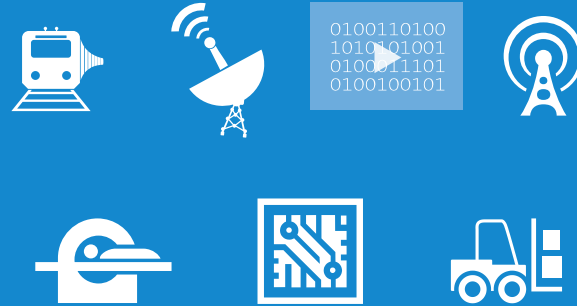
Definition of data

Today *all* types of data should be available for analysis

Structured



Semi structured



Unstructured



Ability to analyze transactional data with semi & unstructured data is key to modern analytics

Big data can help bring customer & market data from various formats to one place, **data lake**

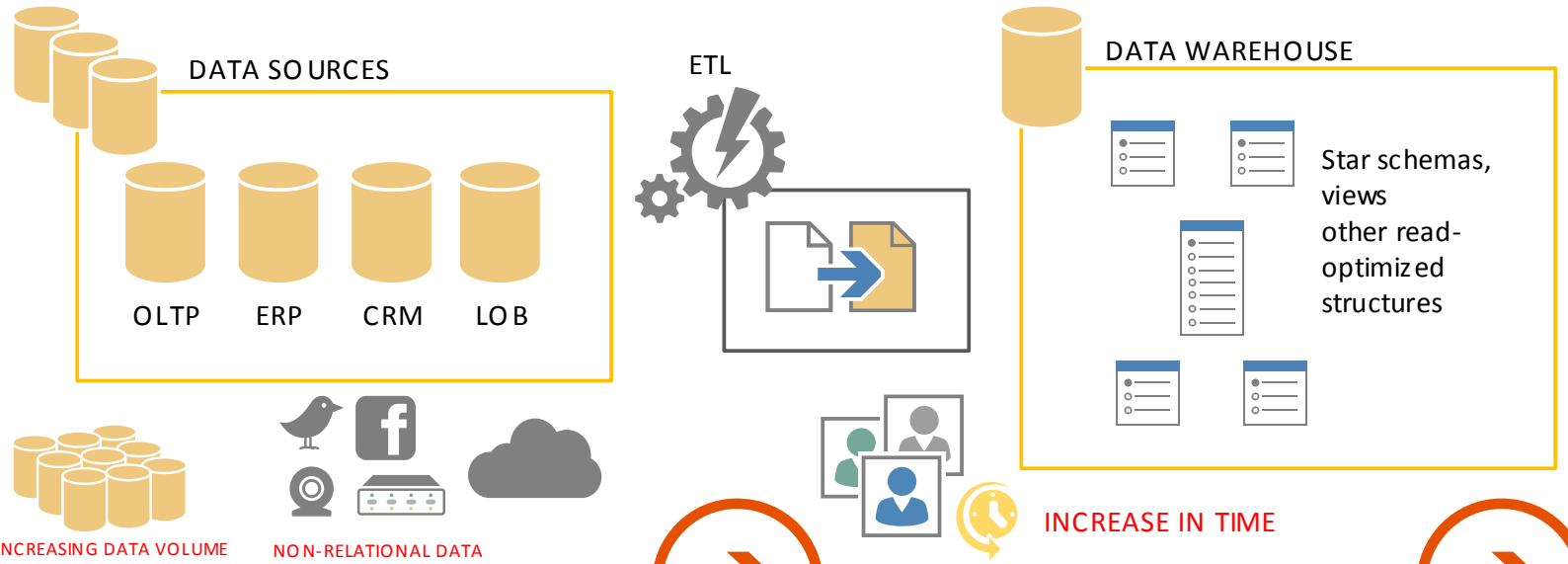
What is a data lake?

Storage repository that holds a vast amount of data in its native form.

- Most commonly, Hadoop based solution
- Large data pool of “schema on read” or “schema less” data
- A place to store unlimited amounts of data in any format inexpensively
- Allows collection of data needed by multiple departments in one place

Traditional data warehouse

Current state of a data warehouse



Increase in variety of data sources

Increase in data volume

Increase in types of data



Rigid transformations can no longer keep pace with changing data

Ensuring availability with costly appliances is a myth

Inability to transform large volumes

Always on EDW redesign mode

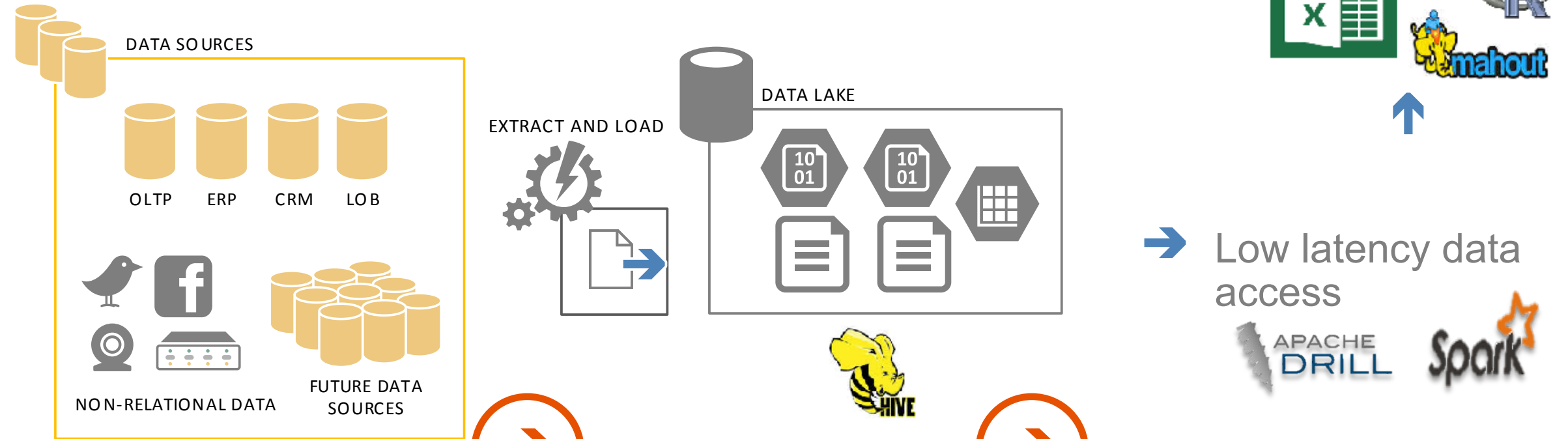


Reports become invalid or unusable

Delay in preserved reports increases

Users begin to “innovate” to relieve starvation

Modern Data lake



All data sources (internal & external) are valid source for data lake



Extract and load data

No need to archive for performance!!

Data always stored in native format & always online

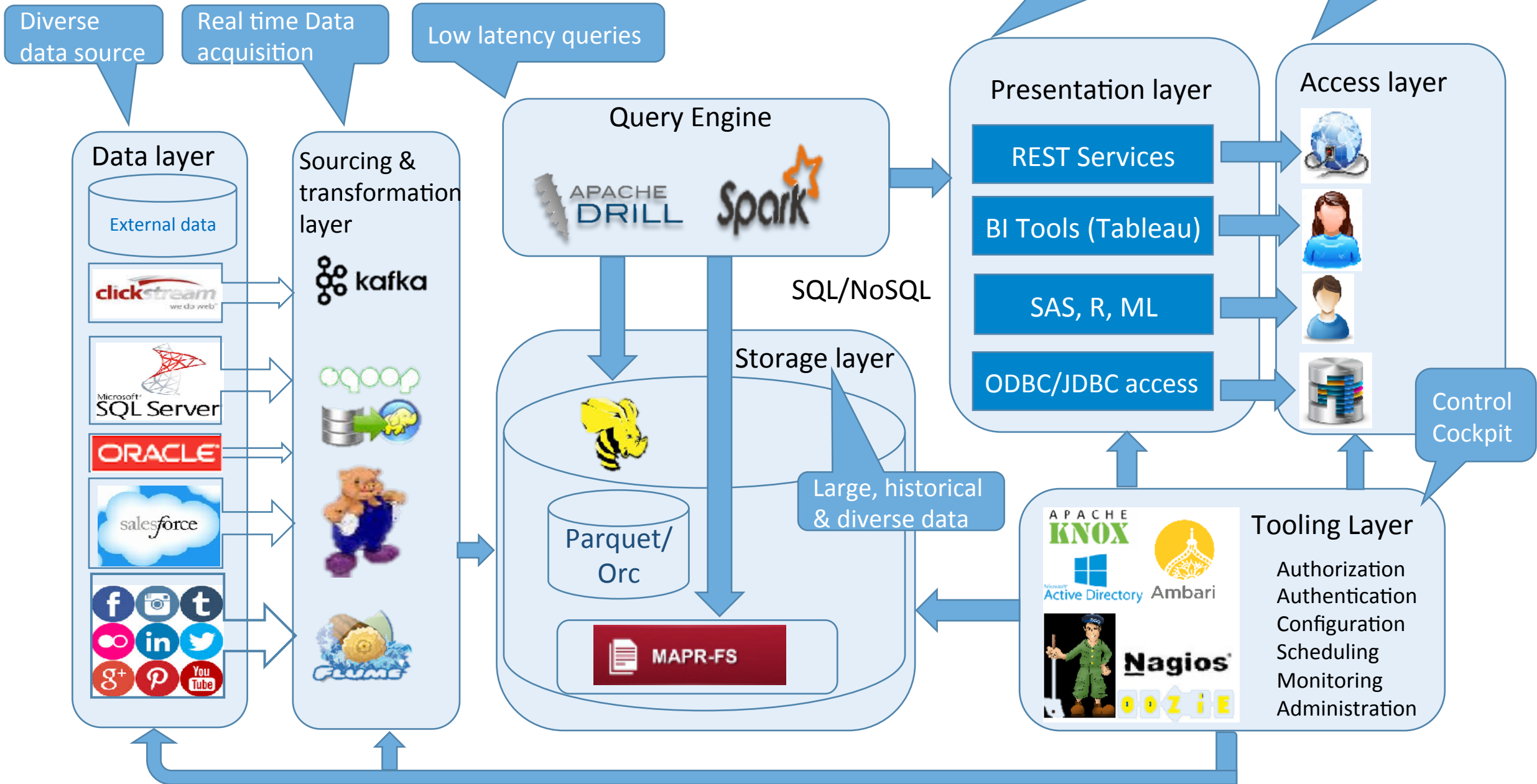
Performance & availability based on Large scale commodity infrastructure



Schema on Read or Schema less systems are the future on large scale analysis

Users discover published data sets/ services using familiar tools

BFL data lake framework



Future of Analytics is not just SQL queries

Increasing complexity & value of analytics

Analytics - Goals



Maturity model of Analytics

Technology: + Algorithms +Machine Learning

- How can we make it happen?
- Simulations, Recommendations

**Prescriptive
Analytics**

Technology: + Statistics

- What can happen?
- Predict Sales, demands, sentiments & patterns

**Predictive
Analytics**

Technology: + hadoop

- Why did it happen?
- Historical trends, Causality, Mining

**Diagnostic
Analytics**

Technology: Traditional EDW Database & Reports

- What happened?
- Ad-hoc & Cube reports

**Descriptive
Analytics**

How about some performance numbers?

- During SQL test, Apache drill was ~2x faster than commercial RDBMS.
- During ETL test, we were able to get a network throughput of 600GB per hour consistently with MapR NFS.
- 20X Compression (Raw data vs ORC) in storage.

Note: Above numbers are from our internal tests & the same is not in Production yet.

Cluster config tested: MapR M5, 6 node cluster, each with (12 cores, 64 GB RAM, 4TB SATA, 2*10GBPS network)

Statutory warning

”fools rush in where angels fear to tread”

1709

- From “An essay on criticism” by Alexander Pope,

Big data is exciting, but (as always) its also a minefield of failures if not thought well.

To avoid big data adoption failure:

- 1) Start small, if you fail, fail quick & get back on track wiser
- 2) Choose partners carefully
- 3) Be agile in building projects, no single projects beyond 2 months delivery
- 4) Build roadmap to add use case incrementally

Prototype based Big data adoption

- Clearly defined use cases with expected outcome & timeline
- Prototype before going on a journey
- Build small teams to solve problems

Example use case:

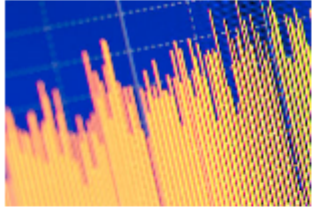
Problem statement: Reduce computation time for customer segmentation

Approach: Build distributed in-memory based solution on hadoop.

Results: Customer segmentation execution time was reduced from 8 hours to 15 sec

What's happening globally in BFSI?

Trends:



For financial institutions, the potential of big data is huge, and it will bring disruptive change to the industry.

Big data, a buzzword in recent years, is an all-encompassing term for any collection of data sets that is so large and complex that quick processing with hands-on data management tools or traditional applications becomes nearly impossible. Big data has

four striking features: 1) a vast amount of information—generally more than 10 terabytes of data; 2) fast speed, consistent with the one-second rule in data processing; 3) diversification, which includes both text that is easy to store and unstructured data such as images and video; and 4) realization of value—high-value output, but relatively low average value per data piece.

Big data is extensively used across industries touching various points of daily life. Amazon uses its powerful big data analysis to recommend a number of goods targeted to consumers based on their perceptions and interests. Netflix, before launching the American hit TV show *House of Cards*, used big data to evaluate the audience's fast-forward and replay time and then to determine the perfect combination of director, leading actor, and plot.

Topics: [Analytics and Visualization](#) | [Business Intelligence](#) | [M&A and Investment](#)

Big data startups set to disrupt credit reporting industry

ZestCash brings Big Data to consumer lending

Company launches improved credit assessment method that it says will let it lend to 25% more people

— MORE LIKE THIS

AllTuition is the star pupil at Launch conference

Are Online Lending Platforms Beating Banks at Big Data?

by COLIN WILHELM

How Big Data Is Revolutionizing the Credit Scoring Industry

Like it?

Posted February 9, 2015

0

comments

The Online Lenders That Could Break the Payday Loan Racket

Same-day loans provide quick cash to many Americans—but with crippling fees. New online lenders may provide a path to good credit.

BY SOPHIE QUINTON

[Follow on Twitter](#)

4 startups reinventing finance from L.A., not NYC

We're witnessing "big data everywhere", literally

What   did to commercial emails & what  did to proprietary Unix

Today, we are at the tipping point where  is challenging the domination of RDBMS in Analytics & Warehouse solutions

Today: There are no major big data driven Financial firms in India.

Tomorrow: Big data driven Financial services is an eventuality in India, (It's "when", not "if")

Thank you!!!

Change, before you have to.

- Jack Welch