SYBASE IQ PROVIDES ADVANCED ANALYTICS TO UNLOCK CRITICAL BUSINESS INSIGHTS

DIAL NUMBERS:

1-866-803-2143

1-210-795-1098

PASSCODE: SYBASE

YOUR HOSTS FOR TODAY

Your Host…

David Jonker
Product Marketing
Sybase, an SAP Company

Guest Speaker…

Philip Howard
Research Director,
Bloor Research

SYBASE | An SAP Company

HOUSEKEEPING

Questions?

Submit via the ‘Questions’ tool on your Live Meeting console, or call **1-866-803-2143** United States, **1-210-795-1098** Other
Password SYBASE
Press *1 during the Q&A segment

Presentation copies?

Select the printer icon on the Live Meeting console

SYBASE | An SAP Company

Big data and the warehouse

Philip Howard

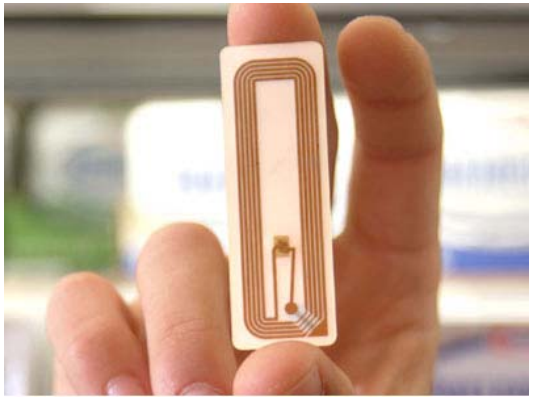
Research Director – Bloor Research

Agenda

- What is different about big data?
- Why would you want to take a big data-based approach?
- How would you deploy it using Sybase IQ?

How is big data different?

But …



But ...

DIGITAL MOBILE

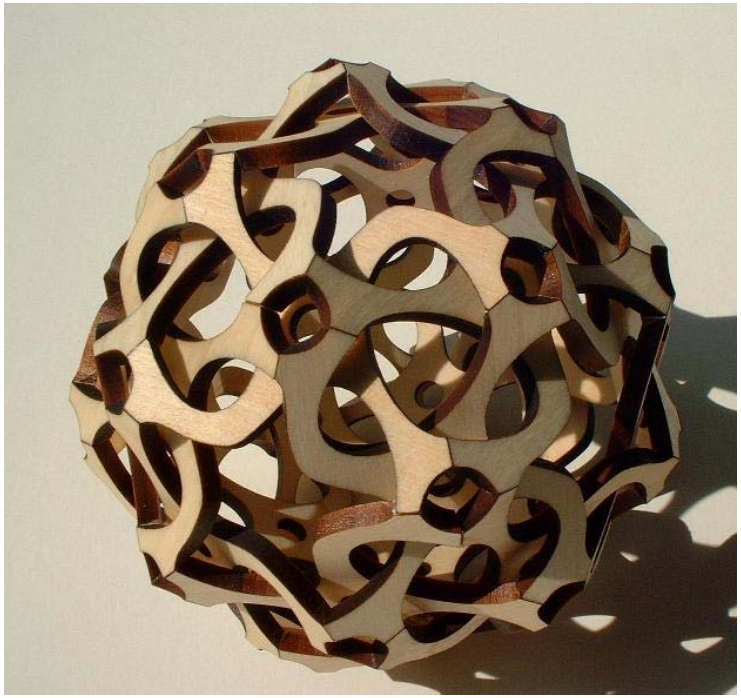
Example of a typical Call Detail Record

MOBILE TERMINATED RECORD
- RECORD LENGTH: xxxxxx
- RECORD TYPE: xxx
- RECORD NUMBER: xxxxxxxxxxxxxx
- RECORD STATUS: xxxxxxxx
- CHECK SUM: xxxxxxxxx
- CALL REFERENCE: xxxxxxxx
- EXCHANGE ID: xxxxxxxxx
- INTERMEDIATE RECORD NUMBER: xxxxxxxxxxx
- INTERMEDIATE CHARGING INDICATOR: xxxxxx
- A SUBSCRIBER NUMBER: xxxxxxxxxxx
- B SUBSCRIBER IMSI: xxxxxxxxxxxxxxx
- B SUBSCRIBER IMEI: xxxxxxxxxxxxxxxx
- B SUBSCRIBER NUMBER: xxxxxxxxxxx
- B CATEGORY: xxxxxxxxxxxxx
- MS CLASS MARK B: xxxxxxxxxx
- IN CIRCUIT GROUP NUMBER: xxxxxxxxxxx
- B SUBSCRIBER FIRST LOCATION: xxxxxxxx
  - LOCATION ID: xxxxxxxxxxxxxxx
  - CELL ID: xxxxxxxxxxxxxxxxxxxx
- B SUBSCR. LAST LOCATION EX ID: xxxxxxxx
- B SUBSCRIBER LAST LOCATION: xxxxxxxxxx
  - LOCATION ID: xxxxxxxxxxxxxxxx
  - CELL ID: xxxxxxxxxxxxxxxxxxxx
- CHARGEABLE SERVICE TYPE: xxxxxxxxxx
- CHARGEABLE SERVICE CODE: xxxxxxxxxxx
- SECONDARY SERVICE TYPE: xxxxxxxxxxx
- SECONDARY SERVICE CODE: xxxxxxxxxx
- NON TRANSPARENCY INDICATOR: xxxxxxxx
- HALF RATE INDICATOR: xxxxxxxxxxx
- SET UP START TIME:
  - DATE: xxxxxxxxxxxxxxxx
  - TIME: xxxxxxxxxxxxxxxx
- CHANNEL ALLOCATED TIME:
  - DATE: xxxxxxxxxxxxx
  - TIME: xxxxxxxxxxxxx
- CHARGING END TIME:
  - DATE: xxxxxxxxxxxxx
  - TIME: xxxxxxxxxxxxx
- CHARGEABLE DURATION: xxxxxxxxxx
- SUCCESS INDICATOR: xxxxxxxxxxxxx
- DATA VOLUME: xxxxxxxxxxxxxxxx
- CALL TYPE: xxxxxxxxxxxxxxxxxxxx
- TARIFF CLASS: xxxxxxxxxx
- DTMF SENDER INDICATOR: xxxxxxxxxx
- ADVICE OF CHARGE INDICATOR: xxxxxx

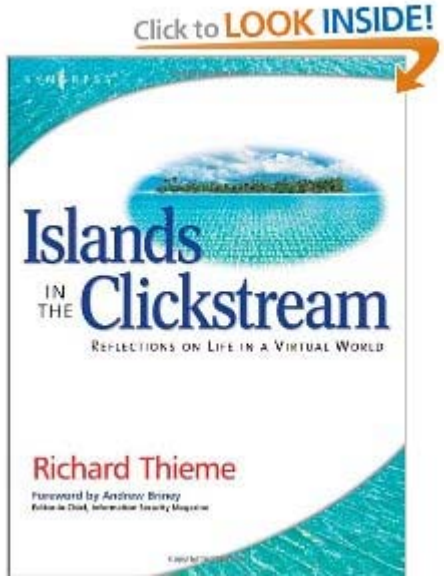
Illustration A: Sample of information to be found in a Call Detail Record (CDR)

But …

2 types of big data

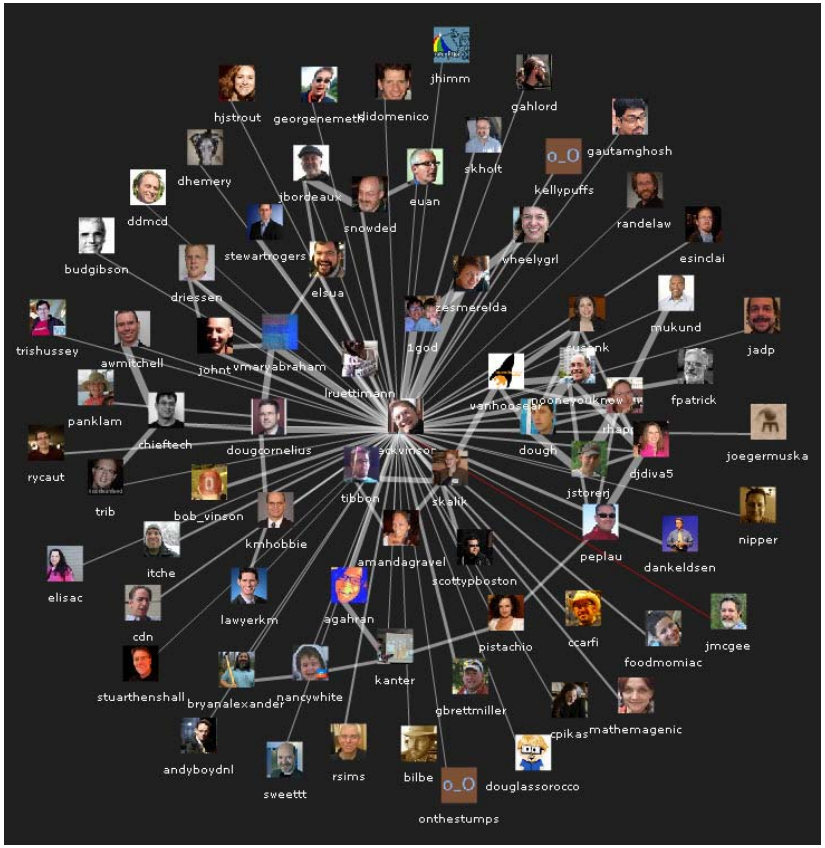


Structured data

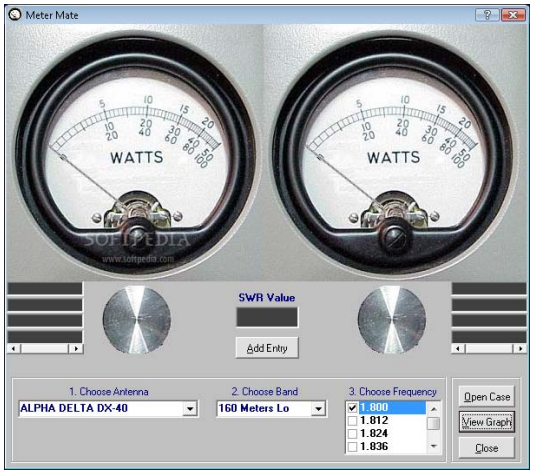


Unstructured data

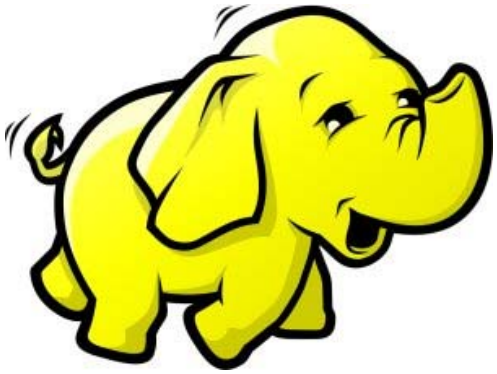
Why - unstructured



Why - structured



How - Hadoop & NoSQL



Key-value model: warehousing

Column-family model: write once

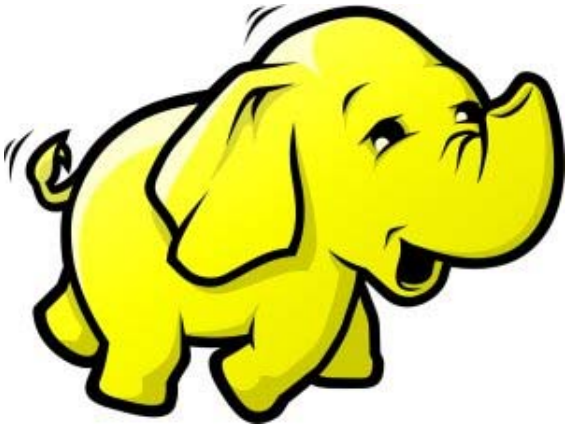
Document model: operational

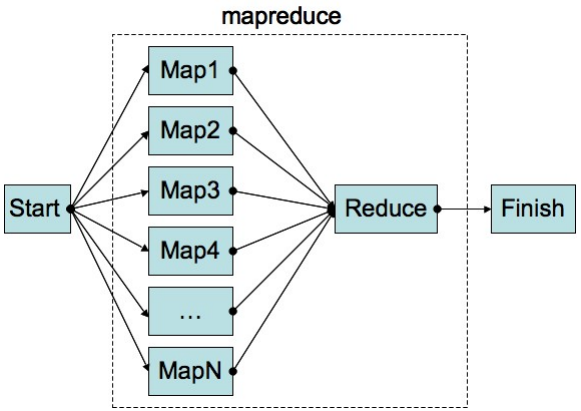
NoSQL vs Triple Stores

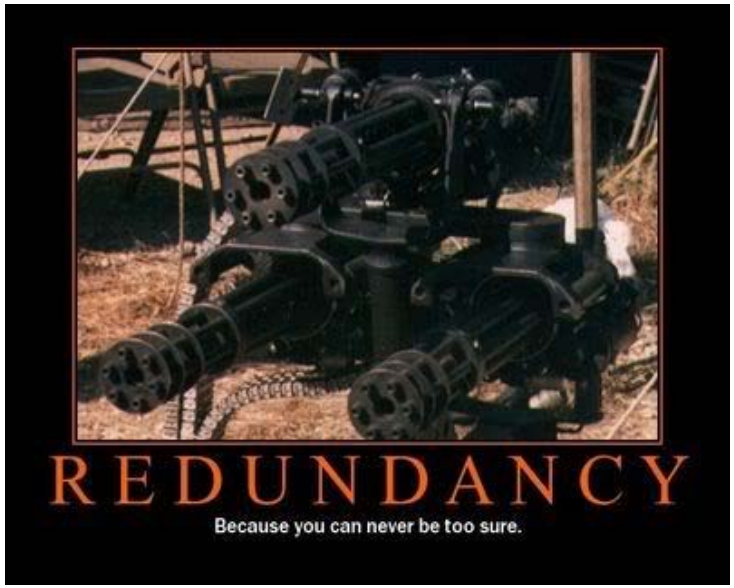
Stardog

Confidential © Bloor Research 2012

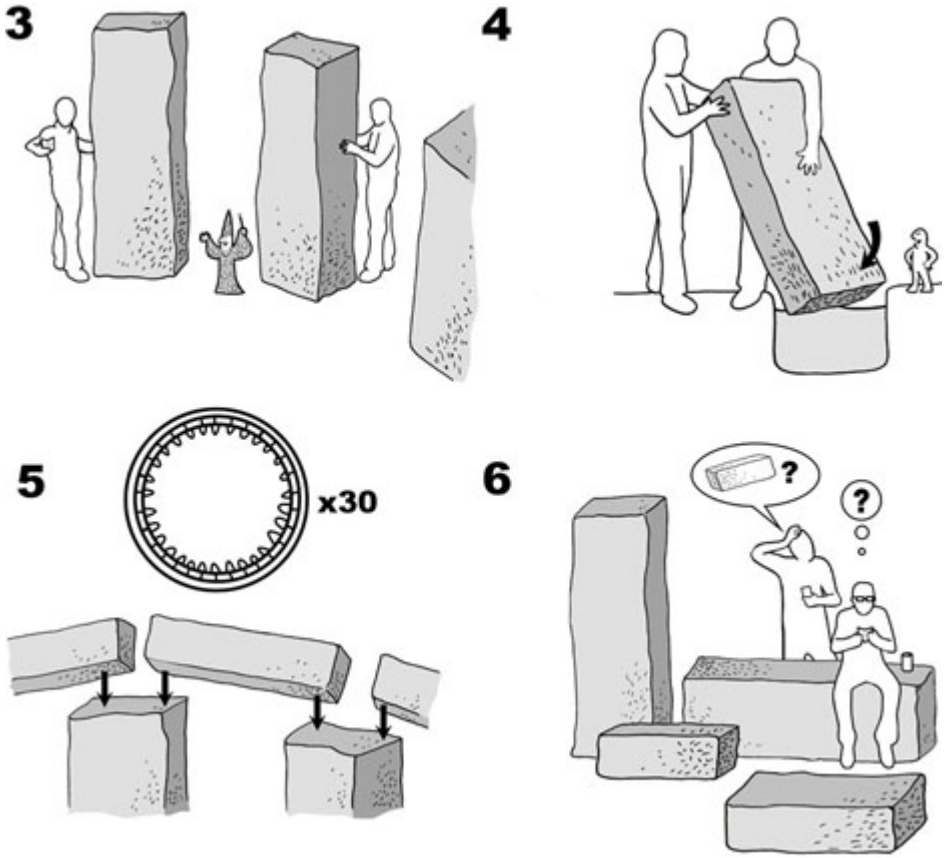
How – Hadoop basics



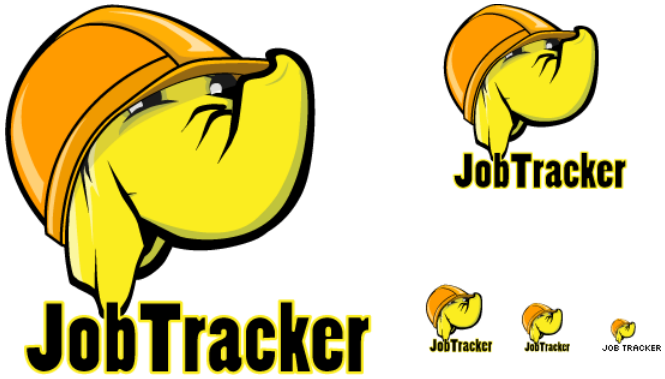


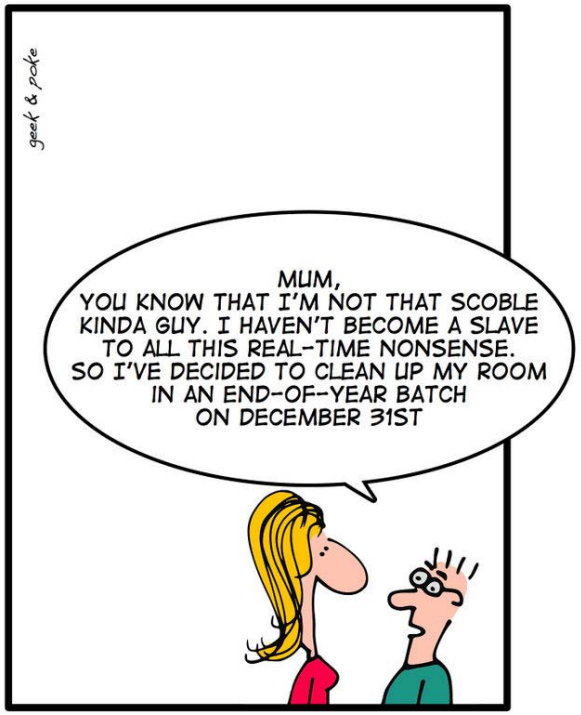


MapReduce



But …



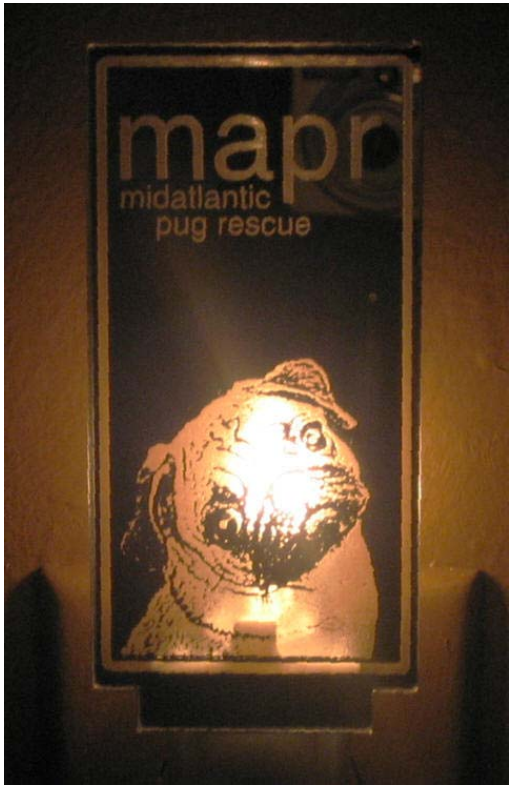


BATCH VS. REAL-TIME

Alternatives/extensions to Hadoop

Cassandra

HADAPT

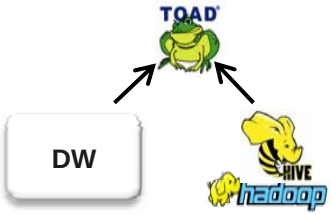


Confidential © Bloor Research 2012

Hadoop and the warehouse

Technology

1. **Client side federation**



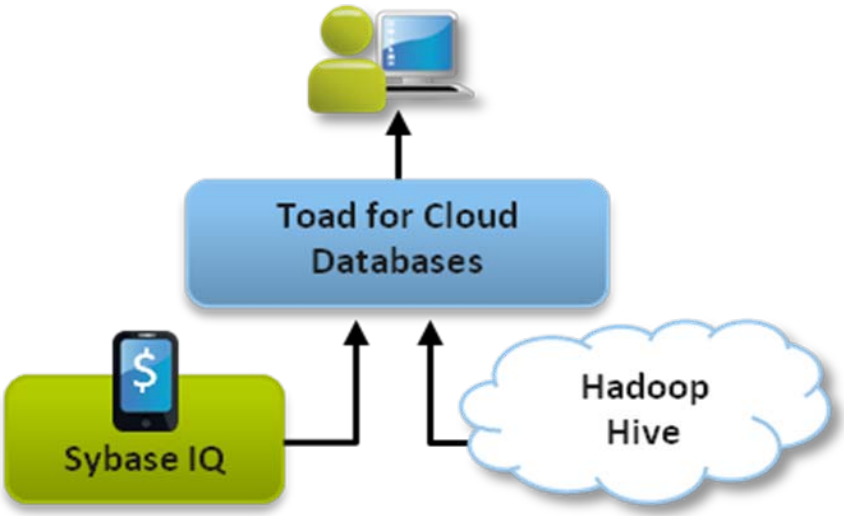
Results and Applications

✓ **Join pre-computed DW and Hadoop results**

For example, in telecoms you might use a data warehouse to provide aggregated customer loyalty data while Hadoop holds aggregated network utilisation data; and then a 3rd party product such as Quest Toad for Cloud can bring the data together from both sources, linking customer loyalty to network utilisation or network faults such as dropped calls.

Confidential © Bloor Research 2012

Hadoop and Sybase IQ



For example, in telecoms you might use a data warehouse to provide aggregated customer loyalty data while Hadoop holds aggregated network utilisation data; and then a 3rd party product such as Quest Toad for Cloud can bring the data together from both sources, linking customer loyalty to network utilisation or network faults such as dropped calls.

Confidential © Bloor Research 2012

Hadoop and the warehouse

Technology

2

ETL Hadoop data into DW

DW

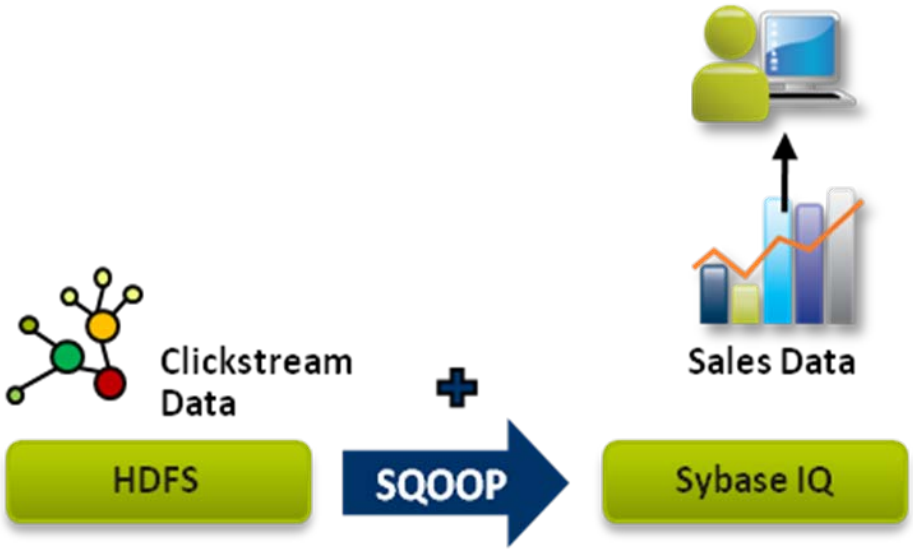
Results and Applications

✓ Fetch HDFS subsets into warehouse
✓ Process entirely in warehouse

An example use case is in eCommerce, where clickstream data from weblogs are stored in HDFS, and outputs of MapReduce jobs on that data (to study browsing behaviour) are loaded into the data warehouse via an ETL process. The transactional sales data in the warehouse is joined with the clickstream data to understand and predict customer browsing and buying behaviour.

Confidential © Bloor Research 2012

Hadoop and Sybase IQ



An example use case is in eCommerce, where clickstream data from weblogs are stored in HDFS, and outputs of MapReduce jobs on that data (to study browsing behaviour) are loaded into the data warehouse via an ETL process. The transactional sales data in the warehouse is joined with the clickstream data to understand and predict customer browsing and buying behaviour.

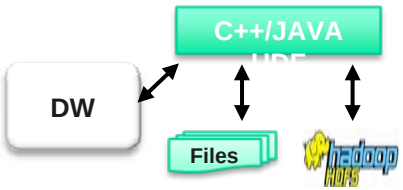
Confidential © Bloor Research 2012

Hadoop and the warehouse

Technology

3

Federate Hadoop data into DW



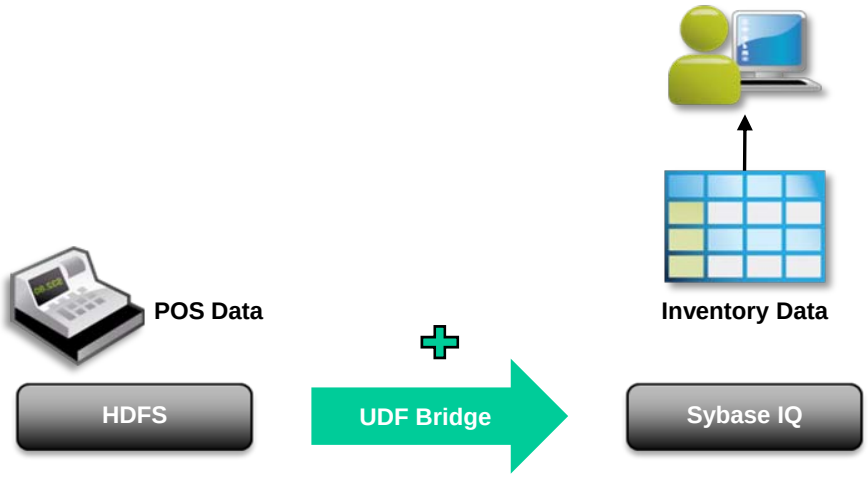
Results and Applications

- ✓ Fetch HDFS subsets by materializing inside DW as in-memory virtual tables
- ✓ Triggered on the fly as part of query

An example here would be in retail, where point of sale (POS) detailed data is stored in Hadoop. The warehouse fetches the POS data at fixed intervals from HDFS for specific hot selling SKUs, combined with inventory data from the warehouse to predict and prevent inventory "stockouts".

Confidential © Bloor Research 2012

Hadoop and Sybase IQ



An example here would be in retail, where point of sale (POS) detailed data is stored in Hadoop. The warehouse fetches the POS data at fixed intervals from HDFS for specific hot selling SKUs, combined with inventory data from the warehouse to predict and prevent inventory "stockouts".

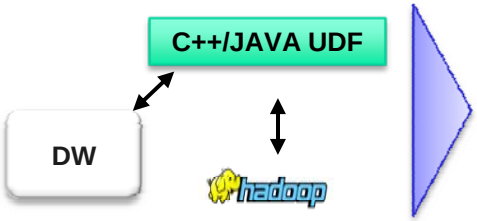
Confidential © Bloor Research 2012

Hadoop and the warehouse

Technology

4

Federate Hadoop jobs into DW



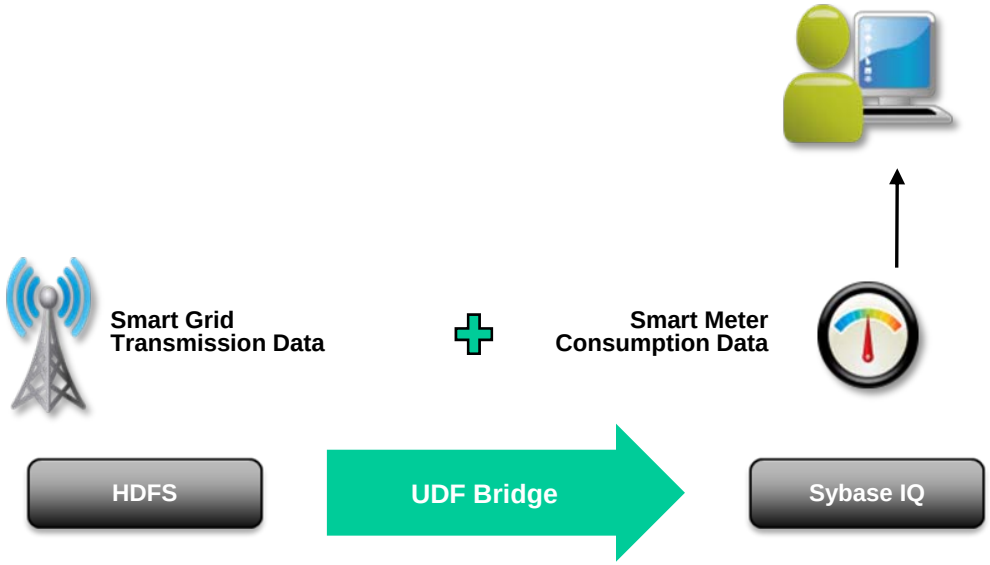
Results and Applications

✓Trigger MapReduce job in both DW and Hadoop and join the results on the fly

You might use this approach in the utility sector, with smart meter and smart grid data combined for load monitoring and demand forecasting. Smart grid transmission quality data (multi-attribute time series data) stored in HDFS can be computed via Hadoop MapReduce jobs triggered from the data warehouse and combined with smart meter data stored in the warehouse, to analyse demand and workload.

Confidential © Bloor Research 2012

Hadoop and Sybase IQ



You might use this approach in the utility sector, with smart meter and smart grid data combined for load monitoring and demand forecasting. Smart grid transmission quality data (multi-attribute time series data) stored in HDFS can be computed via Hadoop MapReduce jobs triggered from the data warehouse and combined with smart meter data stored in the warehouse, to analyse demand and workload.

Confidential © Bloor Research 2012

Hadoop and Sybase IQ

- NoSQL scale out is easier and cheaper than scale up
- Warehouse is faster, more robust and is not limited to batch processing
- SQL programmers are less expensive and easier to get
- Time stamps may be an issue

Confidential © Bloor Research 2012

Conclusion

- We can now analyse data in ways that were not previously possible that can provide new insights into customer (and other) behaviour
- We can do this cost-effectively
- But there are limitations that will typically require a combined warehouse/Hadoop environment
- There are various ways in which this can be achieved.

THANK YOU
FOR MORE INFORMATION
WWW.SYBASE.COM/SYBASEIQBIGDATA

SYBASE | An SAP Company


SYBASE®
An SAP Company