



# Big data

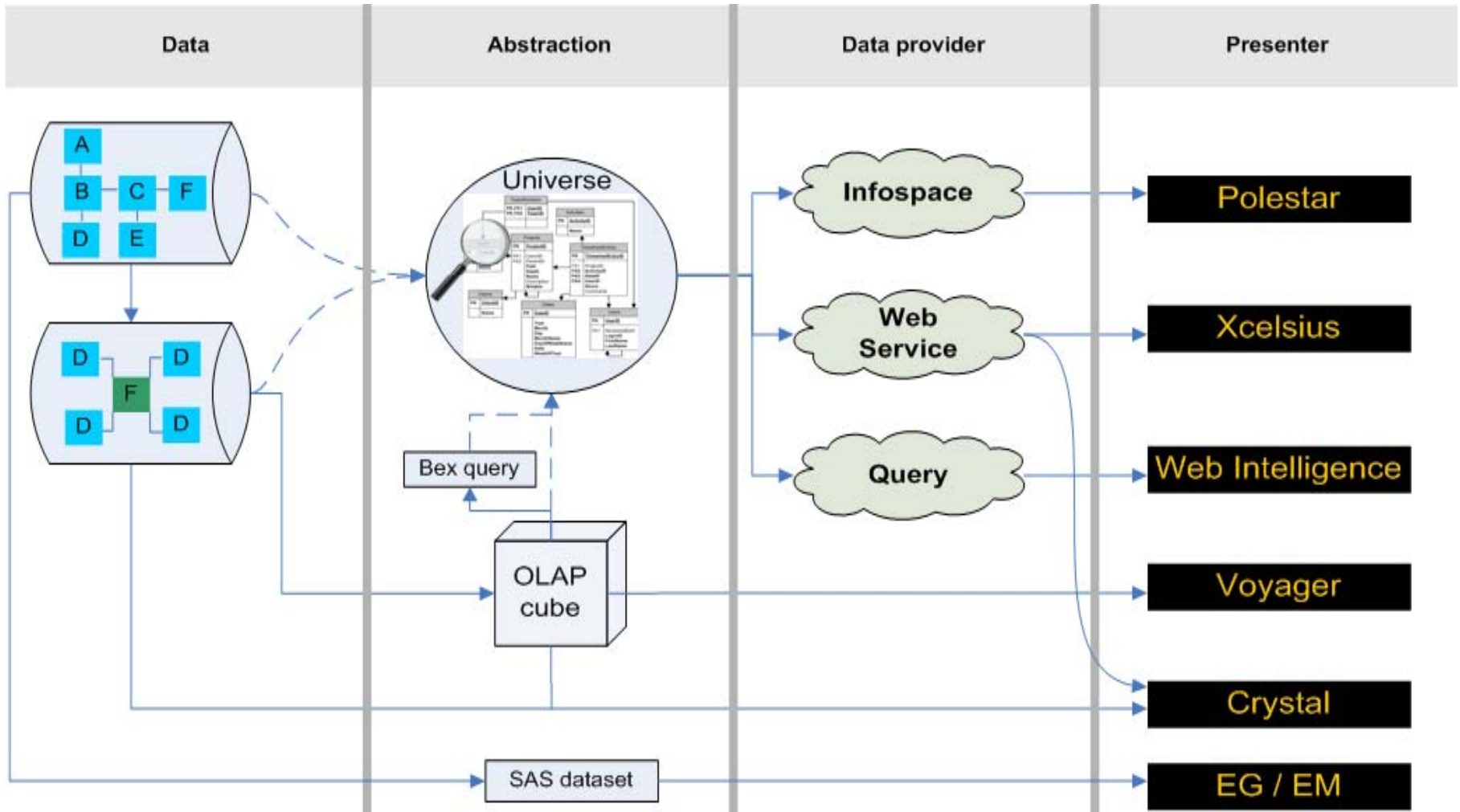
Improving performance through optimal data structures

**Presenter: Jeffrey Strauss – BI Architect**

**March 12, 2010**

[Youtube video](#)

# Where to focus?





## The problem:

**I want to do analytics. But this Business Objects query takes so long...**



## Temporary solutions:

**Offline mode**

**Schedule the report**



## More permanent solutions:

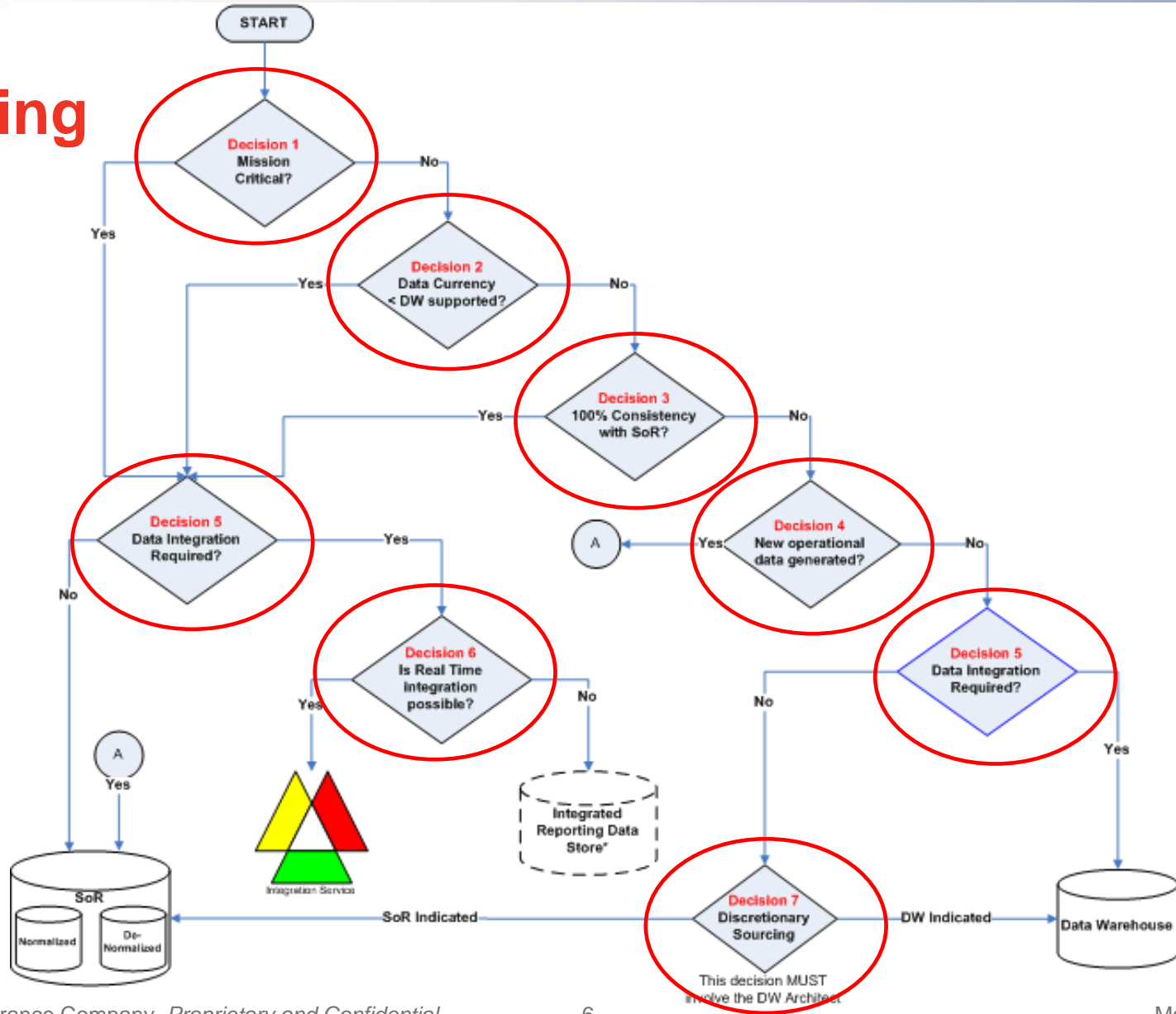
**Data appliances**

**Indexing**

**Aggregate awareness**

**Having a data structure strategy in place**

# Data sourcing



# Data structures



Structure	Short description	Use case
Logical view	Logical RDBMS representation of one or more tables joined together	<ul style="list-style-type: none"> <li>Encapsulation of a common shared reusable query</li> <li>Access security for a subset of DB columns</li> </ul>
Materialized view	Physical RDBMS representation of one or more tables joined together	<ul style="list-style-type: none"> <li>Same as logical view</li> <li>Added benefit of report performance without having too much ETL complexity within the refresh logic</li> </ul>
Denormalized table	A stand-alone DB table with all columns necessary to support a given report, or series of reports	<ul style="list-style-type: none"> <li>Good for standard reporting / dashboards when the underlying 3NF structure does not provide for effective query performance</li> </ul>
Star schema (Dimensional model)	Multi-table DB structure that supports a wide array of static and ad-hoc business requirements.	<ul style="list-style-type: none"> <li>Performance ideal for majority of agile reporting needs</li> <li>Requires presentation modeling exercise</li> </ul>
OLAP cube	Multi-dimensional DB structure that supports hierarchies and cross-level analysis by high indexed intersections	<ul style="list-style-type: none"> <li>Structure for "OLAP aware" interfaces</li> <li>Good for slice and dice against summary low cardinality data</li> </ul>
Derived table (Nested inline query)	Business Objects universe construct that is a logical formulated view of one to many relational tables.	<ul style="list-style-type: none"> <li>Embedded non-reusable focused business logic</li> <li>Ability to have an in-line reporting prompt</li> </ul>
Infospace	Business Objects indexed result set construct that is focused on a specific area of analysis	<ul style="list-style-type: none"> <li>Structure for "Explorer" interface</li> <li>Certified shareable result set</li> </ul>



# Tune, Tune, Tune...

- Indexing
- Aggregate awareness
- Hints
- Shortcut joins
- Verify volume of data coming back to the report





## Questions

### Contact info:

**Jeffrey Strauss**

**Allstate Insurance – BI / Data Architect**

**847-402-3027**