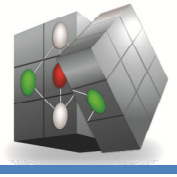
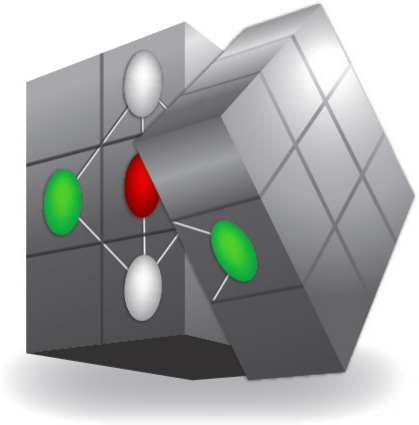
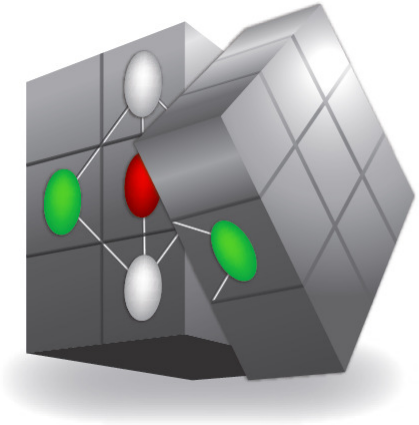




cubist

Your Business Intelligence

Project Overview

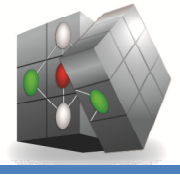


cubist

Your Business Intelligence

CUBIST in a nutshell: Developing an approach for semantic and user-friendly Business Intelligence by

- **augmenting Semantic Technologies with BI capabilities, and**
- **providing conceptually relevant and user friendly visual analytics.**



Motivation

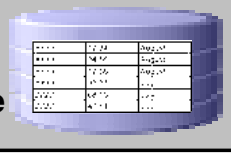
- Increased proportion of unstructured data (>80%)
  - Not accessible for classical BI solutions
  - Can be better leveraged by means of Semantic Technologies (ST)
- Insufficient user interfaces for Business Intelligence (BI)
  - Improved visual analytics, based on Formal Concept Analysis (FCA), for qualitative Data Analysis
  - Complementing to existing approaches for quantitative Data Analysis

CUBIST Main Idea
From classical to semantic BI

Classical Business Intelligence

restricted queries / analytics

Data Warehouse



ETL

databases

Output

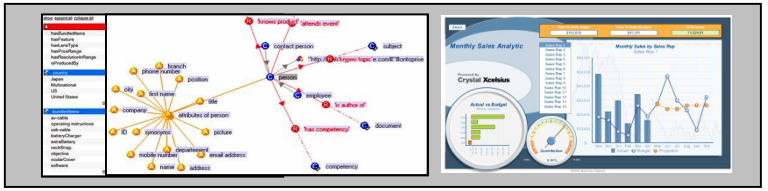
User Interaction

Store

Gathering Information

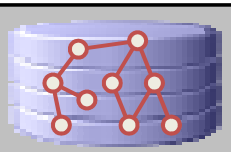
Data sources

Semantic Business Intelligence



flexible and visual queries / analytics

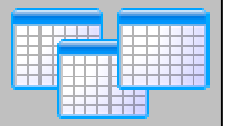
Triple Store



Semantic ETL

databases

Forums, blogs



Office docs

CUBIST Main Idea
From classical to semantic BI

CUBIST: Developing an approach for semantic and user-friendly BI

Providing conceptually relevant and user friendly visual analytics.

- Formal Concept Analysis / Galois Lattices
- Faceted navigation
- Graph-based navigation

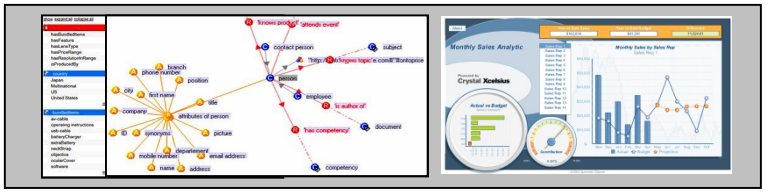
Augmenting Semantic Technologies with BI capabilities

- Triple store as persistency layer
- Flexible Data Warehouse design
- Extending SPARQL with OLAP functionalities
- Reasoning / Deriving implicit facts

Federating data from both unstructured and structured sources

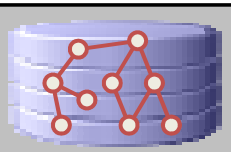
- Enhanced ETL
- Text Mining
- Information Extraction

Semantic Business Intelligence

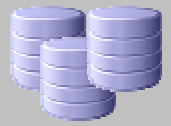


flexible and visual queries / analytics

Triple Store

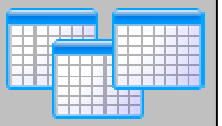


Semantic ETL

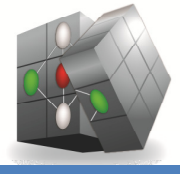


databases

Forums, blogs



Office docs



Comparing classical BI with ST

Feature	Classical BI	Semantic Technologies
Federating data from unstructured and structured sources	bad	good
Explicit meaning of data / explicit relations	bad	good
Querying/accessing huge amounts of data	good	average
OLAP functionalities	good	bad
Agile DW schema	bad	good
Inferring implicit facts	bad	good
ETL flexibility	bad	average
Data reuse	bad	good
Use of external datasources / data location independence	bad/average	good

Research Challenges

UI challenges for FCA-based VA

- Look like keyword but semantic search front-ends
- Faceted navigation
- Graphical UI to build graph pattern queries

Technical challenges for FCA-based VA

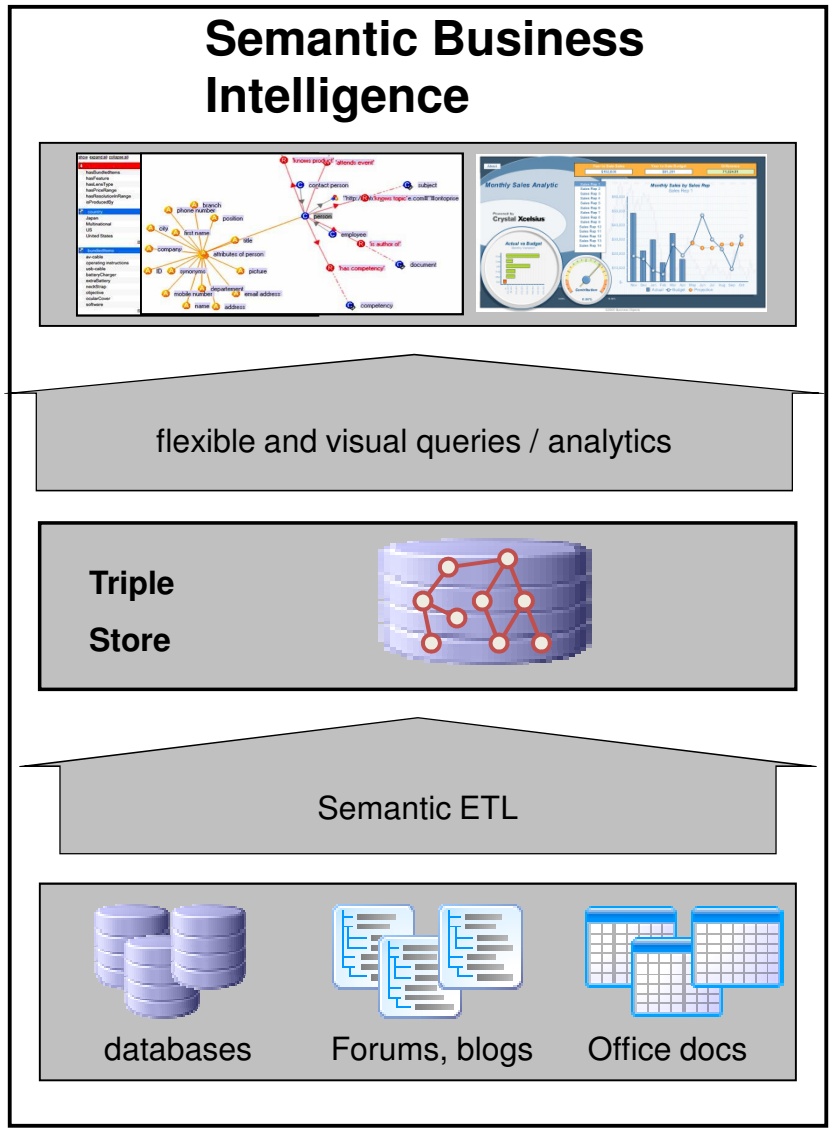
- Parallel processing algorithms
- Algorithms for many-core architectures
- Making use of Single Instruction-Multiple Data (SIMD) architecture

BI enabled Triple Store

- No existing query language: extending SPARQL with OLAP functionalities
- Performance and scalability of triple store

Semantic ETL

- Quality of source data: particularly “noisy” unstructured data
- Quality of the extraction process: uncertainty of data has to be addressed
- Identity resolution
- Federating unstructured and structured sources

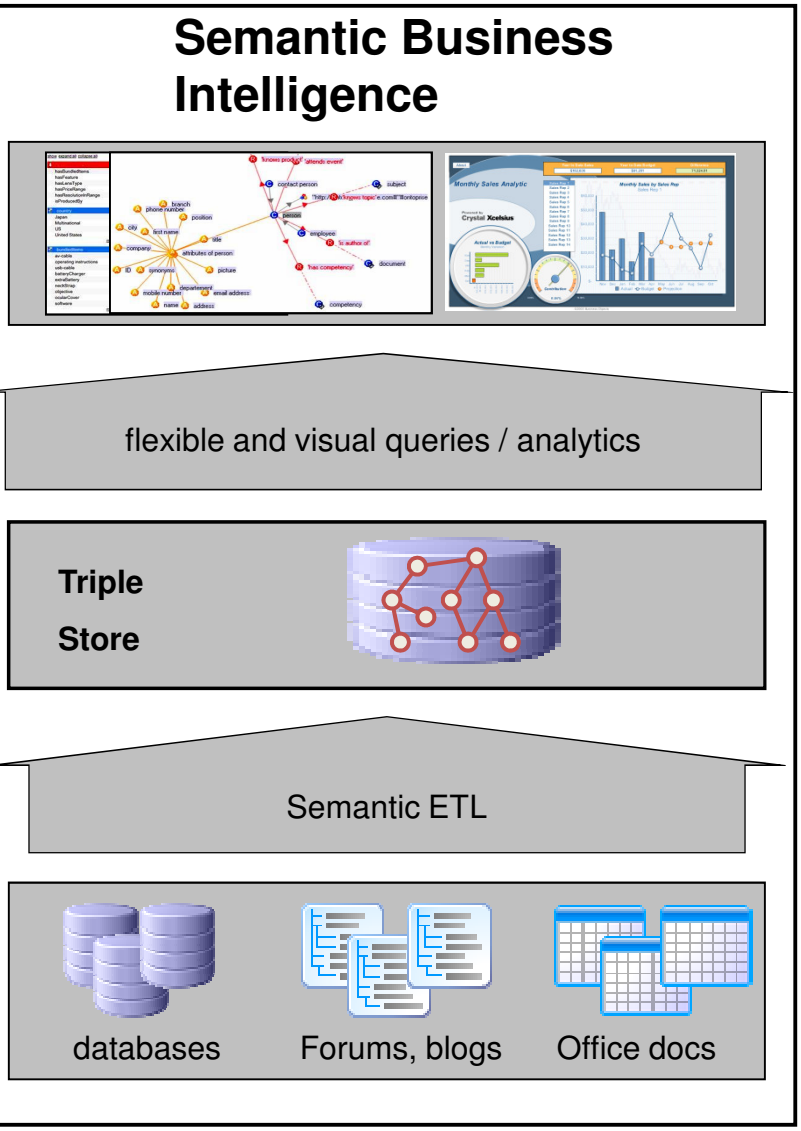


Expected Results

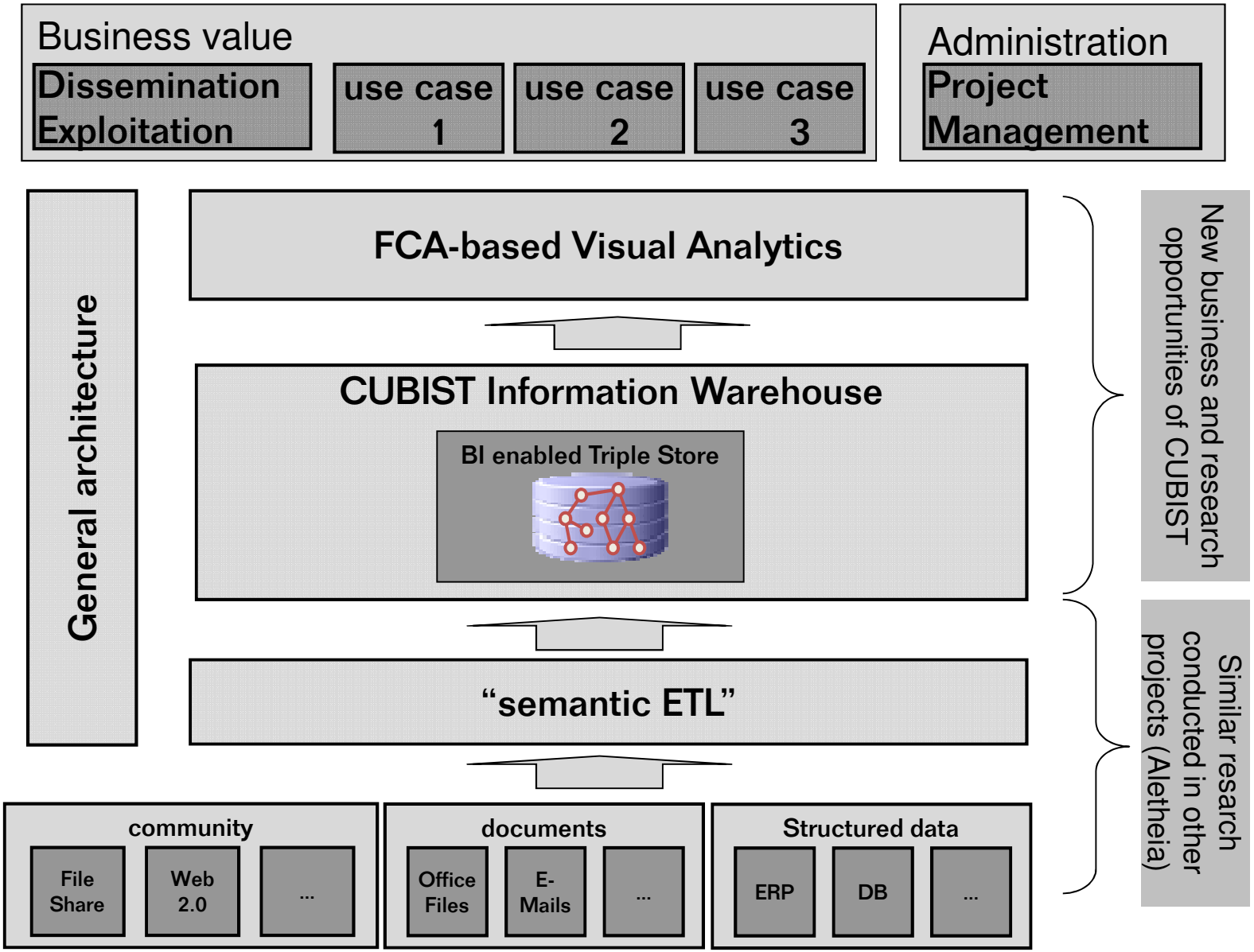
FCA-based VA
- Performance formal concept miner
- Large scale FCA capability
- FCA visualization of BI

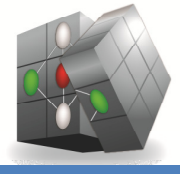
BI enabled Triple Store
- OLAP extensions for SPARQL
- Discovery mechanism for finding implicit information
- Triple store based information warehouse
- Integration of triple store with FCA

Semantic ETL
- Semantically enriched lineage information
- Error detection within extracted data and CUBIST information warehouse
- Semantic ETL component that provides SPARQL endpoints for various data sources



CUBIST Architecture





CUBIST – Project Details

Instrument

- Instrument: STREP
- Theme: ICT-2009-4.3
- Call: FP7 Call 5
- Lead: SAP Research
- Duration: 36 Months
- Start: 2010/10

Consortium

Technological Partners

- SAP (Germany)
  - Coordinator and technological partner
- Ontotext (Bulgaria)
  - Expertise in Semantic Technologies
- Sheffield Hallam University (UK)
  - Expertise in FCA
- Centrale Rechereche S.A. (France)
  - Expertise in FCA and Visual Analytics

Use Case Partners

- Heriot-Watt University (UK)
- Space Applications Services (Belgium)
- Innovantage (UK)

Thank You!

- **Dr. Frithjof Dau**
  Researcher, Business Intelligence Practice I SAP Research Dresden
- **SAP AG** I Chemnitzer Str. 48 I 01187 Dresden I Germany
-
- T +49 351 4811-6152 I F +49 6227 78-51425 I mailto: frithjof.dau@sap.com
- www.sap.com

© 2011 SAP AG. All rights reserved

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. The information contained herein may be changed without prior notice.

Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors.

Microsoft, Windows, Excel, Outlook, and PowerPoint are registered trademarks of Microsoft Corporation.

IBM, DB2, DB2 Universal Database, System i, System i5, System p, System p5, System x, System z, System z10, System z9, z10, z9, iSeries, pSeries, xSeries, zSeries, eServer, z/VM, z/OS, i5/OS, S/390, OS/390, OS/400, AS/400, S/390 Parallel Enterprise Server, PowerVM, Power Architecture, POWER6+, POWER6, POWER5+, POWER5, POWER, OpenPower, PowerPC, BatchPipes, BladeCenter, System Storage, GPFS, HACMP, RETAIN, DB2 Connect, RACF, Redbooks, OS/2, Parallel Sysplex, MVS/ESA, AIX, Intelligent Miner, WebSphere, Netfinity, Tivoli and Informix are trademarks or registered trademarks of IBM Corporation.

Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

Adobe, the Adobe logo, Acrobat, PostScript, and Reader are either trademarks or registered trademarks of Adobe Systems Incorporated in the United States and/or other countries.

Oracle is a registered trademark of Oracle Corporation.

UNIX, X/Open, OSF/1, and Motif are registered trademarks of the Open Group.

Citrix, ICA, Program Neighborhood, MetaFrame, WinFrame, VideoFrame, and MultiWin are trademarks or registered trademarks of Citrix Systems, Inc.

HTML, XML, XHTML and W3C are trademarks or registered trademarks of W3C®, World Wide Web Consortium, Massachusetts Institute of Technology.

Java is a registered trademark of Sun Microsystems, Inc.

JavaScript is a registered trademark of Sun Microsystems, Inc., used under license for technology invented and implemented by Netscape.

SAP, R/3, SAP NetWeaver, Duet, PartnerEdge, ByDesign, SAP BusinessObjects Explorer, StreamWork, and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and other countries.

Business Objects and the Business Objects logo, BusinessObjects, Crystal Reports, Crystal Decisions, Web Intelligence, Xcelsius, and other Business Objects products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Business Objects Software Ltd. Business Objects is an SAP company.

Sybase and Adaptive Server, iAnywhere, Sybase 365, SQL Anywhere, and other Sybase products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Sybase, Inc. Sybase is an SAP company.

All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

The information in this document is proprietary to SAP. No part of this document may be reproduced, copied, or transmitted in any form or for any purpose without the express prior written permission of SAP AG.

This document is a preliminary version and not subject to your license agreement or any other agreement with SAP. This document contains only intended strategies, developments, and functionalities of the SAP® product and is not intended to be binding upon SAP to any particular course of business, product strategy, and/or development. Please note that this document is subject to change and may be changed by SAP at any time without notice.

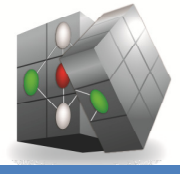
SAP assumes no responsibility for errors or omissions in this document. SAP does not warrant the accuracy or completeness of the information, text, graphics, links, or other items contained within this material. This document is provided without a warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

SAP shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials. This limitation shall not apply in cases of intent or gross negligence.

The statutory liability for personal injury and defective products is not affected. SAP has no control over the information that you may access through the use of hot links contained in these materials and does not endorse your use of third-party Web pages nor provide any warranty whatsoever relating to third-party Web pages.

Appendix

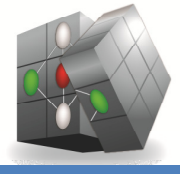
- FCA in three Minutes



What is Formal Concept Analysis?

Formal Concept Analysis is:

- a mathematization of the philosophical understanding of concepts
- a human-centered method for conceptually clustering and structuring data
- a method to visualize data and its inherent structures, implications and dependencies



FCA in three Minutes (i)

How can we describe the concept "BI products from SAP"?

- Extensionally by enumerating all ***objects***:
  - BO Xcelsius, BO Crystral Reports, …
- Intensionally through ***attributes***:
  - "is an SAP product", "is a BI tool", …

Generally, a ***concept*** is divided into two mutually dependent parts:

- Its ***extension*** are all objects that share all the attributes of the concept,
- Its ***intension*** are the attributes which precisely describe the objects of the concept.

The concepts form a hierarchy: A concept C1 is a ***subconcept*** of C2, iff

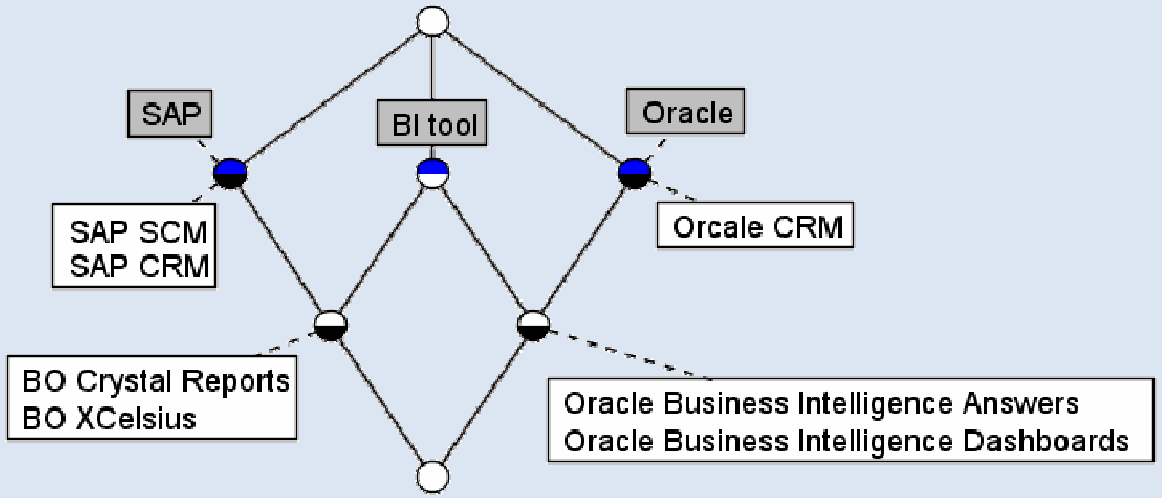
- the extension of C1 is a subset of the extension of C2
- the intension of C2 is a subset of the extension of C1

} equivalent

Theorem: For a given universe of discourse, the concept hierarchy is a complete lattice

FCA in three Minutes (ii)

A toy **formal context**

	SAP	Oracle	BI tool
BO XCelsius	X		X
BO Crystal Reports	X		X
SAP CRM	X		
SAP SCM	X		
Oracle Business Intelligence Dashboards		X	X
Oracle Business Intelligence Answers		X	X
Orcale CRM		X	
Oracle Fusion GRC		X	

Its derived **concept lattice**