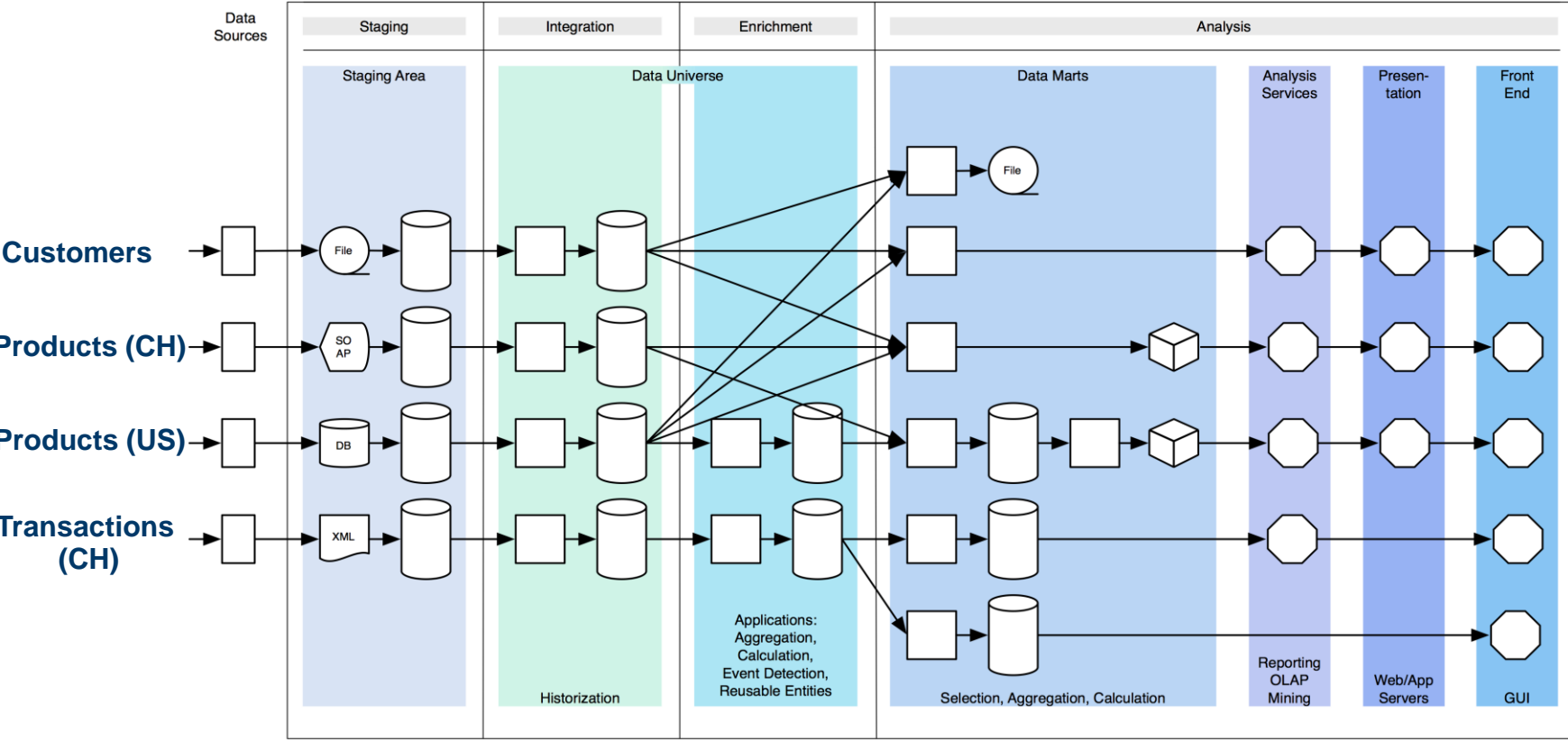


The Parallel Universe Effect between Data-Intensive Physics and Banking

Dr. Kurt Stockinger
Data Warehouse & Business Intelligence Architect
Credit Suisse, Zurich
May 28, 2013

Typical Data Warehouse Reference Architecture



Main Characteristics of Major Industries in DWH



**Source: M. Ehrenmann, R. Pieringer, K. Stockinger. Is there a Cure-All for Business Analytics?
Business Intelligence Journal, Vol. 17, No. 3, Sept. 2012**

Commonalities and Differences Between Banking and Physics in Data Management

	Banking	Physics
Data volume	Large (PB range)	Very large (multi PB)
Data sources	Large (thousands)	Small (<10)
Data types	Numerical + text	Numerical + images
Data Storage	Relational databases (DB2, Oracle, Teradata) + Cubes (Microsoft)	Open-source (Root, HDF5, Flat files + FastBit)
Data Integration	ETL (PowerCenter,...)	Typically not the main challenge
Data Analysis	Business intelligence tools (BusinessObjects, Microstrategy, SAS,...) + data mining tools (Clementine)	Root, parallel R, ...
Data Visualization	See above	Root, VisIt
Main challenge	Data modeling & integration	Managing the volume

The Grand Unification? Big Data Technology

- Physics:
 - Most of the software has been home-grown due to lack of scaling of commercial alternatives (bleeding edge + cost effective)
- Banking:
 - Most of the software has been bought (early followers, stable solution)
- Big data technology (scalable + cost effective):
 - Hadoop (HDFS), Pig, Hive, Mahout (Machine Learning)
- Both physics and banking can benefit from new technology
- However, both areas seem to be skeptical and adoption rate seems modest
 - Physics:
 - CHEP 2012 showed “promising” results with Hadoop (experiment log analysis, HDFS as new storage)
 - Banking:
 - Early success stories, e.g. DWH query log analysis: <http://www.ecc.ethz.ch/research/xadoop>
 - Still struggling with integration of DWH+BI solutions with Big Data technology