



# Promoting Data-Centric Architectures

SATURN 2010 Conference Day 1  
Session Architecture Design and Analysis

May 19th 2010  
Minneapolis, MN, U.S.A.

Michael C. Jaeger, Uwe Hohenstein,  
Gerald Kaefer, Ravi K. Madipadaga

Siemens AG, Corporate Research and Technologies  
Global Technology Field "System Architecture and Platforms"  
Munich, Germany

© 2010 Siemens AG. All Rights Reserved.

## Overview

- Stepping out of software architecture
  - *Example from an almost different domain: processors*
- Data Centric Architecture
  - *Where do we find them? - or do we have them already?*
- The essence
  - *It is not about knowledge, but it moves towards about the ,how'*
- Conclusions

## And now for Something Completely Different Microprocessors

SIEMENS

- Basically, processors perform data manipulation and math calculation [Smith 1997]
- General purpose processors ('GP CPU', e.g. Motorola 68K, x86) try to perform well at both disciplines with focus on the first purpose
- However, when it comes to repeating sequential calculations (well known application: signal processing), GP CPUs lack the efficiency
  - They are just not optimal for math calculations
  - Specific architecture design is necessary
- One result are Digital Signal Processors (DSPs) which are optimized for high bandwidth I/O in conjunction with efficient calculation units



Source: Wikipedia



Source: Wikipedia



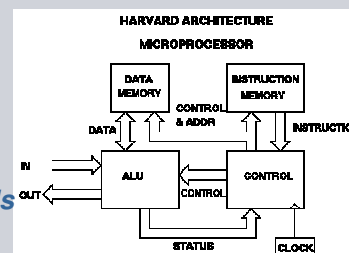
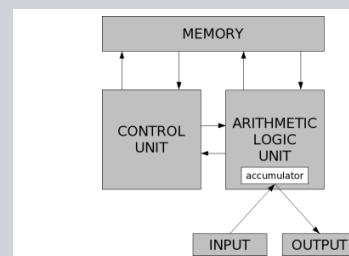
Source: Wikipedia

## Microprocessors and DSPs

SIEMENS

To achieve optimal solution for data intensive processing:

- DSPs are aligned to Harvard architecture
  - GP CPUs have modified Harvard but are conceptually von Neumann
- DSPs are optimized for sequence of data
  - Modern CPUs: SIMD / vector units
  - Modern DSPs: MIMD with VLIW
- DSPs lack the performance for branched application flow.
  - GP CPU also good for multiple threads
- **DSPs are designed to serve particular applications more efficiently than GP CPUs**

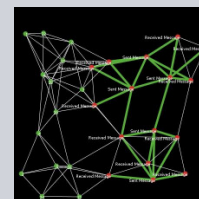


... that is very data centric. However, ...

**Searching the Web and you will find:  
Data Centric Architecture in Sensor Networks**

When it comes to DCA, much is in sensor networks, ad-hoc networks, etc. [cf. Estrin et al.]

- General problem is to transfer data from A to B
- Problem constraints are low power consumption, low processing power, low bandwidth
- Approaches went from address-centric to data-centric routing
  - Aggregation of different data flows can bring benefits in global efficiency
- The problem is moving the data from A to B.



Source: Wikipedia, 'Sensor Network'

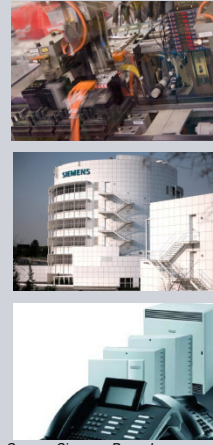
- ***It is about the efficiency which is the driver to such architecture!***

## Discovering Data Centric Challenges Siemens Cases

SIEMENS

*Data Centric Architecture in sensor networks is OK,  
but we have cases also at Siemens:*

- **Industry Automation:** An engineering tool for distributed object trees that require transactions and versioning
- **Building Technology:** Management of engineering data with varying data structures, in particular different versions must be managed by tooling
- **Communications Technology:**  
A telecommunications middleware must support multiple tenants and faces customization of the data model



Source: Siemens Press Images

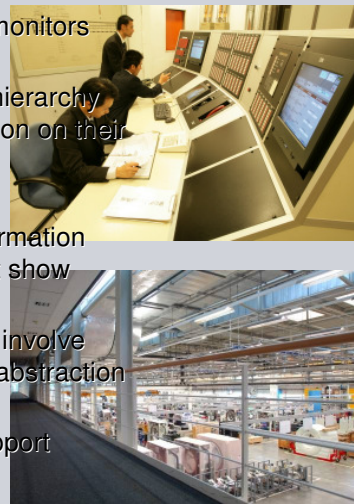
## Industry Automation Case

SIEMENS

- Automation with software that controls and monitors physical processes and involved hardware
- The structure of the installation also shows hierarchy where sub-controlling nodes show aggregation on their involved hardware nodes as well

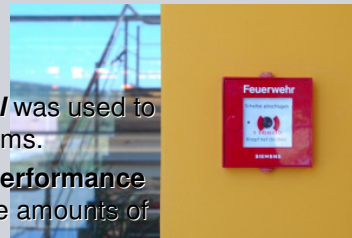
### Challenges:

- **Distribution of the data:** partitioning of information results in the fact that a central unit does not show the actual state
- **Compatibility with hardware updates** that involve updates in the data model. This leads to an abstraction layer at efficiency loss
- **Versioning of data model** required that support different versions at run time



## Building Technology Case

- Engineering and configuration tool
  - 1000s of devices, installed in office building
- Tool must handle configuration data from 3rd party vendors
  - Reselling, integrated into the Siemens infrastructure
- Product lines, functionality, configuration data are constantly evolving



### Challenges:

- An approach where an **abstract data model** was used to maintain all possible kind of process-able items.
- The additional abstraction layer came with **performance penalties** when retrieving and handling huge amounts of configuration data

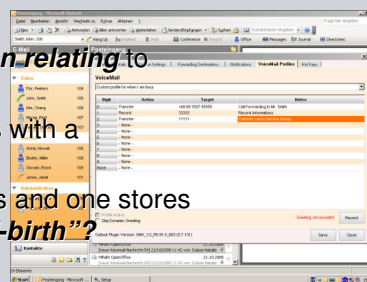
## Communication Infrastructure Case

- Communications middleware with user management
- Different clients or applications use middleware
  - Several extensions necessary to basic user data model
- An additional abstraction (like key-value pairs) to user related information is feasible.



### Challenges:

- Key-value-pairs shows **disadvantage when relating to** such key-value-stored data items
- Querying involving such information comes with a **performance penalty**
- What if the platform serves two applications and one stores **"birthday"** while the other stored **"date-of-birth"**?



## Summary on the Siemens Cases

### Main observations

- *Data centric architecture is a design view on solving a data intensive problem optimally*
- We have several problems in Siemens related IT where data handling represents core functionality
- Software developers are influenced by OOD, by java and .net as leading technology in the industry
  - OOD is not wrong, but it leads to abstraction which could be different to the data encapsulated by objects.
- *Naturally, one starts to identify* real world objects with characteristics
- *Naturally, one starts to identify* functional building blocks
  - Perfect match for services in a SOA
- *But how to achieve a data centric solution for such problems?*

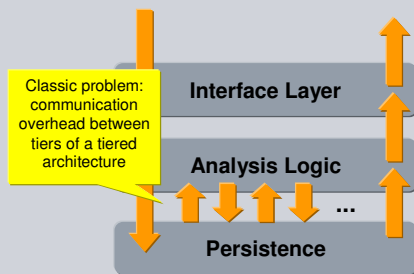
## It is not About Knowledge

- Software architects and developers have the knowledge
  - Modeling languages exists for defining data views
    - Data flow diagrams, entity-relationship diagrams
  - Patterns exist that cover specific data-oriented problems
    - DAO, DTO ...
  - Architectural styles exist that are proven
    - Publish/subscribe to name the most prominent
- New technology is on the way for industry IT to cope with data centric applications
  - *Cloud computing*, example cases:
    - 99 designs, a platform for graphic design templates serving 3.1 million designs
    - PhotoWall, a photo sharing application serving community portals
    - and, of course, Amazon and Google services itself ...

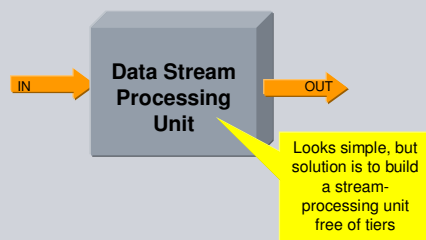
## What would be an example?

Think of an video analysis application that must process large streams of data – provided by building security systems

*classic Enterprise IT approach  
maybe set by technology choice,  
for example, application server*



*however, the wish is  
data stream processing unit that  
takes input stream and computes  
output stream with less interactions*



## Cloud Computing

- Cloud Computing delivers technology and resources for data intensive problems
- Cloud computing enables the engineering of novel applications and business models
  - Opportunity for new ideas
  - Threat to existing businesses
- New products implies new designs
  - New storage products: **Blobs and Tables**
  - **New application frameworks** with new API and programming models
  - **New communication styles** in a distributed environment (on-premises and cloud space)
- **Data centric architectures become even more important!**



Source: Press Image Hewlett-Packard



Source: Press Image Microsoft Europe

## Technology for Data Centric Systems

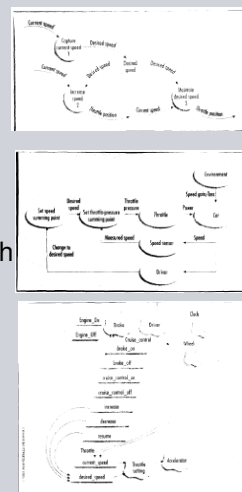
- Cloud computing offers also new opportunities with regard to a data-centric design.
  - Cloud computing := known resources in a novel business model
  - Technically: blobs, tables, RDBMS and file storage can be obtained very efficiently
  - Cloud computing enables new business models and applications
  - Photo-storage services, backup-services, handling large amount of genetic data [cf. Amazon use cases]
- But, how does this affect the design of a data centric architecture?**

photoWall  
image community platform  
fotopedia  
image database  
PicTranslator  
translates words on photos  
elephantdrive  
backup  
Jungle Disk  
backup  
Zmanda  
backup  
Washington Post  
TIFF to PDFconversion  
search box  
search solution

Trademarks of the respective owners. Please see <https://aws.amazon.com/solutions/case-studies/> for more information  
© 2010 Siemens AG. All Rights Reserved.

## Conclusions

- DCA is about data in the centre of engineering
  - Another view: **data view** for designing architecture, for example in addition to [cf. Kruchten 1995]:
    - Logical** and **development** view, **process** view and a **physical** view, and ... the scenarios
- Data centric problems require strong focus on data**
- Assumption:** DCA achieved by data centric approach
  - A design style similar to OOD, Jackson System Development approach, Ward Mellor approach, [cf. Shaw 1995], i.e.:
    - OOD puts the objects as real world entities in the centre, then continues with the data and interaction
    - JSD: identifying messages and events, then OOD
    - Ward/Mellor: Start with data flow diagrams and events, then model state-transitions ...



images: Mary Shaw 1995, Comparing Architectural Design Styles



## Summary

- Modeling languages, patterns, arch. styles and technology exist
  - Clearly, data centric architecture is not the solution to every problem
  - As the examples show:  
*much is already out there!*
  - But as the cases show: *how does 'much out there' do for more efficient systems?*
  - And ... *with emerging cloud computing technology the need for such architecture rises*
  - *Facts seem to confirm our assumption:*
- ***Do we think about a design approach to create this architecture?***

## Thank you for your attention!

**Michael C. Jaeger**  
michael.c.jaeger@siemens.com  
**Uwe Hohenstein**  
uwe.hohenstein@siemens.com  
**Gerald Kaefer**  
gerald.kaefer@siemens.com

Siemens AG, CT T DE IT1  
Corporate Technology,  
Global Technology Field System  
Architecture and Platforms  
Otto-Hahn-Ring 6  
81730 München

**Ravi K. Madipadaga**  
Ravi.Krishna@siemens.com

Siemens Corporate Technology-India  
No. 84 Keonics, Electronics city,  
Hosur road  
Bangalore - 560010, India

*Within Corporate Technology the Global Technology Field System Architecture and Platforms focuses on software architectures for a wide range of software-types. This includes embedded systems, distributed applications, and enterprise software.*

*In the recent field of cloud computing the focus is cloud computing architecture for cloud platform stacks and applications. Cloud computing architecture is key for scalability, cost efficiency, and meeting of legal and business requirements. These activities are completed by the industry focused evaluation of strategic cloud computing platforms in order to support customers on their way to cloud computing.*

**References**

[Shaw 1995] Mary Shaw: Comparing Architectural Design Styles. IEEE Software 12(6): 27-41 (1995)

[Kruchten 1995] Philippe Kruchten: Architecture Blueprints - the "4 1" View Model of Software Architecture. TRI-Ada Tutorials 1995:540-555

[Estrin et al. 2000] Chalermek Intanagonwiwat, Ramesh Govindan, Deborah Estrin: Directed diffusion: a scalable and robust communication paradigm for sensor networks. MOBICOM 2000:56-67

Amazon Case Studies available at:  
<http://aws.amazon.com/solutions/case-studies/>