Advanced Data Modelling & Modelling for Big Data

Training Course

Overview

A 3-day[1] course exploring Data Modelling for Big Data together with the techniques and uses for data models beyond Relational DBMS development.

Objectives

In the modern era, the volume of data we deal with has grown significantly. As the volume, variety, velocity and veracity of data keeps growing, the types of data generated by applications become richer than before. As a result, traditional relational databases are challenged to capture, store, search, share, analyse, and visualize data. Many companies attempt to manage big data challenges using a NoSQL ("Not only SQL") database and may employ a distributed computing system such as Hadoop. NoSQL databases are typically key-value stores that are non-relational, distributed, horizontally scalable, and schema-free.

Many organisations ask, *"do we still need data modelling today?"* Traditional data modelling focuses on resolving the complexity of relationships among schema-enabled data. However, these considerations do not apply to non-relational, schema-less databases. As a result, old ways of data modelling no longer apply.

This course will show Data modelling approaches that apply to not only Relational, but also to Big Data, NoSQL, XML, and other formats. In addition, the uses of data models beyond simply development of databases will be explored.

What you will learn

At the end of the course, delegates would have gained the following:

Level set understanding & terminology:

- Learn about the need for and application of Data Models in Big Data and NoSQL environments
- See the areas where Data modelling adds value to Data Management activities beyond Relational Database design
- Understand the critical role of Data models in other Data Management disciplines particularly Master Data Management and Data Governance.

Pragmatic Learning

- Learn the best practices for developing Data models for Big Data and NoSQL environment

[1] 2-day if previously have attended Data Modelling Fundamentals (or similar)

- Understand how to create data models that can be easily read by humans
- Recognise the difference between Enterprise, Conceptual, Logical, Physical and Dimensional Data models
- Through practical examples, learn how to apply different Data modelling techniques

Course Outline:

- **Data modelling recap**:[2] Data modelling basics, major constructs, identifying entities, model levels and the linkage between them.
- **Data Modelling – Back to the Future?** Data Modelling didn't start with relational! This may be a surprise to many people, but the first uses of data models were well before Relational data bases became the norm. The techniques are applicable to many of the modern non-relational formats we see today.
- **Data Modelling for Big Data & NoSQL:**
  - What must change when we are developing data models for a Hadoop or other Big Data environments?
  - Do modelling tools support Big Data technologies, what are the restrictions and considerations?
  - What data modelling techniques are applicable when targeting a Big Data platform?
  - Does normalisation still have a place in the Big Data world?
  - Where's our metadata in the model now?
  - Massively denormalised files: Is modelling needed? How do we create data models?
- **Modelling for hierarchic systems & XML:** What must change when developing data models for XML & Hierarchic systems?
- **Services Oriented Architecture (SOA):** Why data models are essential for success.
- **Dimensional Data Models:**
  - How (and why) do we create a dimensional model?
  - Converting an ER model to a Dimensional model.
  - Slowly changing dimensions, what types and when are they applicable.
  - Beyond the basics with aggregates, conformed dimensions, bridges, junk dimensions & fact less facts.
- **Application Packages & Data Models:** Do we need to develop data models when implementing a COTS package? Uses and benefits.
- **Data Models & Data Governance:** The essential role that Data Models play in Data Governance. The use of rich metadata in our data models. Using a data modelling repository as the basis of Business Glossary.
- **Using Data Models for Data Integration & Lineage:** How to exploit data models for design of data integration approaches and in data lineage.

[2] Omitted if taken as a 2-day course

- **Requirements capture for data models:**
  - Top down requirements capture: When is it appropriate, what are the limitations.
  - Bottom up requirements synthesis: When this works, where is it appropriate. How do we cope with existing DBMS's and systems.
  - The 3rd way – hybrid.
- How to capture **requirements** for both Data and Process needs.
- **Checking** the Data vs the MetaData; why does it matter?
- Use of **standard data model constructs**, and **pattern** models:
- Understanding the **Bill of materials** (BOM) construct. Where can it be applied, why it's one of the most powerful modelling constructs.
- **Party**; **Role**; **Relationship**: Why mastering this construct can provide phenomenal flexibility.
- **Mastering Hierarchies**: Different approaches for modelling hierarchies.
- **Alternative Data Modelling Notations** and tooling
- **Normalisation**: Progressing beyond 3NF. 4NF, 5NF Boyce-Codd, and why, and when to use them.

Audience

Practitioners who will need to read, consume or create data models, particularly for Big Data and non-RDBMS environments. Users who wish to gain a better understanding of data during Information Management initiatives including:

- Business Intelligence & Data Warehouse developers & architects
- Data Modellers
- Developers
- Data Architects
- Data Analysts
- Enterprise Architects
- Solution Architects
- Application Architects
- Information Architects
- Business Analysts
- Database Administrators
- Project / Programme Managers
- IT Consultants
- Data Governance Managers
- Data Quality Managers
- Information Quality Practitioners

Tutor Biography

Christopher Bradley has spent 38 years in the forefront of the Information Management field, working for International organisations in Information Management Strategy, Data Governance, Data Quality, Information Assurance, Master Data Management, Metadata Management, Data Warehouse and Business Intelligence.

Chris is an independent information strategist & recognised thought leader. He advises clients including, British Gas, Alinma Bank, American Express, ANZ, Bank of England, BP, Celgene, Cigna Insurance, Enterprise Oil, Emirates NBD, GSK, HSBC, NAB, National Grid, Riyad Bank, SABB, Saudi Aramco, Shell, Statoil, and TOTAL.

He is the inaugural Fellow of DAMA CDMP, Vice President of Professional Development for DAMA International, member of several standards committees, an author of DMBoK 2 and author & examiner for professional certifications.

In 2016 Chris received the lifetime achievement award from DAMA International for exceptional services to furthering Data Management education & to the International Data Management community.

Recently he has delivered a comprehensive appraisal of Information Management practices at an Oil & Gas super major, Data Governance strategy for a Life Sciences Company, and Information Management training for a Government Organisation.

Chris guides Global organizations on Information Strategy, Data Governance, Information Management best practice and how organisations can genuinely manage Information as a critical corporate asset. Frequently he is engaged to evangelise the Information Management and Data Governance message to Executive management, introduce data governance and new business processes for Information Management and to deliver training and mentoring.

Chris is Director of the E&P standards committee “DMBoard”, is an officer of DAMA International, an author of DMBoK 2.0, a member of the Meta Data Professionals Organisation (MPO) and a holder at “Fellow” level and examiner for the various professional certifications.

Chris is an acknowledged thought leader in Data Governance, author of several papers and books including ***“Data Modelling for the Business”*** and is an expert judge on the annual Data Governance best practice awards.

Email: chris.bradley@dmadvisors.co.uk
Blog: http://infomanagementlifeandpetrol.blogspot.com/
Twitter @Inforacer
Tel: +44 (0)1225 923000 **Mobile** +44 (0)7973 184475