Introduction to CAT4 Development Project

Andrew Holster
ATASA Research
Aotearoa New Zealand
September 2021

All material in this presentation ©Andrew Holster 2021
Introduction to CAT4

CAT4 refers to two things:

- A logical method for representing information – called the CAT4 fact network, and:
- An information system based on this – called the CAT4 database.

CAT4 changes info system design at a fundamental level:

- It provides a logically complete way to represent all information.
- It puts all information in a common form, and makes all logical relationships between elements of information accessible in a simple and transparent way.
Introduction to CAT4

Aim of this presentation:

• Here we look at CAT4 as a software development project, and focus on these two questions:
• What does it do? How does it replace current software? How does it solve problems?
• How hard is it to build? What does a development project look like? How much time and cost is involved?
• We only give a brief explanation of how it works here – but proof of concepts are well established in technical documentation - the method is explained in more detail in separate presentations.
Introduction to CAT4

Proof of claims.

- There is a mathematical theory behind CAT4, which we refer to technical documentation:
  - Introduction to CAT4. Part 2. CAT2

- Here we explain key features for a software development.

- The aim is to build a CAT4 database application.
Introduction to CAT4

CAT4 is a complete way to represent complex information.

• Current information systems can store masses of data but are poor at dealing with complex information.
• And systems lack critical functionality because of this.
• The solution lies in a better logical form of representation - it does not lie in more complex algorithms for processing – it is the data representation that is flawed.
• We can see historical examples in the change from Hierarchical DBs to Relational DBs and to Graph DBs → the form of representation is fundamental.
• CAT4 provides a much more general method: it gives a natural and complete representation of all information.
Introduction to CAT4

In technology terms, CAT4 can be seen as the true generalisation of relational and graph databases.

• It developed out of relational and graph concepts – but it is a generalisation of these – and the generalisation makes it simpler and more powerful!

• It enables high-level automation methods for machine learning, automation and AI.

• It will provide much better tools for analysts, researchers, managers – anyone dealing with complex information.

• It will make more robust systems and save billions of hours of manual programming.

• **Special feature:** it can also capture natural language and allow people to work in their own natural languages.
Introduction to CAT4

As a software development we want to know:

• First, what does it do? What problem does it solve? How does it replace present software?
• Second, how hard is it to build? How long will it take to develop as a system or a product? And how does it integrate with existing software systems?
• We are not trying to replace RDBs as large data silos –
• We start using CAT4 for purposes that it is uniquely suited to – and RDBs and Graph DB and OO-programming, etc, are poor at – managing large amounts of open-ended complex data.
Introduction to CAT4

What useful things does it do?

• CAT4 allows all information normally scattered in multiple applications, documents, databases, etc to be represented in a single fully scalable system.

• Computerised information is also stratified into multiple levels – “soft data”, tables and programmed objects, meta-data, functions, software code …

• These layers all contain real information, but only a fragment of information is accessible as soft data …

• It is very difficult to integrate information across different applications and at different levels -
Introduction to CAT4

- CAT4 allows integration of all types of information into a single hub – represented in a single table.
- This makes it much easier to deal with, and much more transparent to querying and analysis.

Multiple data apps ➔ Small CAT4 networks ➔ Big CAT4 network

- Tables ➔ Table imports
- DBs ➔ DB imports
- HTML ➔ HTML imports
- Spread-sheets, text docs, etc ➔ etc imports
- CAT4 expands seamlessly to include new content.
Introduction to CAT4

The data hub function.

• Pulling information together in one place takes up a lot of the time of analysts and researchers – core of “BI”.
• CAT4 can integrate information from conventional sources in one system → universal data hub.
• And it integrates very easily with present DBs – good for storing large data sets but only simple relations.
• It also acts as a generic database – you could manage all data entirely within a CAT4 system – so it is a powerful tool for managing complex information.
Introduction to CAT4

How does it do this?

• The functionality we have been describing is very powerful – *but how is it possible to achieve?*

• How does it work? How is it possible to solve this problem of fully automated data integration?

• Thousands of programmers have tried to create such functionality – but methods so far never achieve the *desired level of generalisation* – they just become limited tools-kits or methods for special purposes.

• So it is important to explain how the CAT4 system works in principle.
Introduction to CAT4

How it works.
It is based on a discovery that there is a single pattern underlying all symbolic information:

• This pattern is a simple type of graphical network.
• When we see information in this pattern, it is very intuitive – because the network reflects semantic relations that are natural to our thought.
• And the network of information can be represented by a single table – CAT4 only requires one data table!
• This single table design is what makes programming the system very simple!
Introduction to CAT4

How it works: a single fact table.
- In programming terms, all information is stored in a single table – we call it a fact table.
- Every table row represents an individual fact.
  - Rows taken alone only represent entities.
  - Information consists of relations between facts.
- CAT4 only has four relations between facts. These are called the CAT4 joins.
  - The four CAT4 joins play generic roles (they are really semantic relations).
- This is illustrated for CAT2 (the first two joins of CAT4) next.
Introduction to CAT4

**CAT2 Fact Table.** Every row is indexed with an *ID number*, and *ID1* and *ID2* refer back to this.

Fact Table for CAT2 is a self-joining table. ID is the primary key for ID1 and ID2.

The data illustrated here represents just one tiny conventional “table” – but can be extended to any number of tables, graphs, trees, etc, all at once.

<table>
<thead>
<tr>
<th>ID</th>
<th>ID1</th>
<th>ID2</th>
<th>CONTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>WORLD POINT</td>
</tr>
<tr>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>ZERO POINT</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Client Objects</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>Clients</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>Client Properties</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>Mr. A</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>Mrs. B</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>Ms. C</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>Name</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>Age</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3</td>
<td>Balance</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>7</td>
<td>“A”</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>9</td>
<td>$1</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>7</td>
<td>“B”</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>8</td>
<td>35</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>9</td>
<td>$5</td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>7</td>
<td>“C”</td>
</tr>
<tr>
<td>17</td>
<td>6</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
<td>9</td>
<td>$7</td>
</tr>
</tbody>
</table>
Introduction to CAT4

**CAT2 Table** ↔ **CAT2 Network**

The CAT2 table represents a structured network.
Introduction to CAT4

The CAT2 network is used to represent data from all kinds of structures, e.g. here from a normal “data table”.

<table>
<thead>
<tr>
<th>R</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>r_{11}</td>
<td>r_{12}</td>
<td>r_{13}</td>
<td>r_{14}</td>
</tr>
<tr>
<td>b</td>
<td>r_{21}</td>
<td>r_{22}</td>
<td>r_{23}</td>
<td>r_{24}</td>
</tr>
<tr>
<td>c</td>
<td>r_{31}</td>
<td>r_{32}</td>
<td>r_{33}</td>
<td>r_{34}</td>
</tr>
</tbody>
</table>
Introduction to CAT4

Fact Table ↔ Network ↔ Conventional Forms

The fact table is the form for data storage, the network is what functions operate on, the conventional forms are “shapes” we project information into – screens, reports, etc.
Introduction to CAT4

CAT2 network lattice is general.
• Extends down as far as you like.
• Captures any relational data.
• Each level down corresponds to a *joined table* in a RDB.
• Note there are generally *multiple facts* in any position (with same parents).
• Every point has a single *flat lattice above* – shown here for the “$12” point – but the lattice is multi-layered.
Introduction to CAT4

- The lattice can be extended downwards, corresponding to sub-tables in RDBs, e.g. here we add more details of an account balance.
- It always collects to a single point above, and at the very top is always a single point, called the world point.
- There can be multiple lattices embedded within each other.

Can expand down to more details…
Introduction to CAT4

- All kinds of networks can be represented in CAT3.
- The third join (red) represents object identity.
- Left is a recursive lattice for a file-folder system.
- Right is a Cartesian product $\rightarrow$ dimensional DB (cubes)
Introduction to CAT4

• Many CAT4 lattices can be embedded in one.
• We can combine or separate networks about different subjects →
• The CAT4 data engine is designed to manage a library of CAT4 tables – all with the same table design and the same functions.
Introduction to CAT4

Representing complex information.

• This lets us make an automated data hub → solves a major problem for structured data systems.

• And also works for unstructured data → natural language contains more information than anything else – our brains can decode NL into meaningful information –

• But there is no known way to decode NL into a structured form that makes its information explicit –

• It is stored in text strings, but their meaning or their information is opaque to the computer.

• But CAT4 can turn NL into structured information, integrated with structured data.
Introduction to CAT4

This data integration problem has never been solved.

- Programmed systems for data integration are complex and very ad hoc – but they remain very inflexible, with high manual input and maintenance costs.
- Methods range from enterprise data warehousing, to linking tables across databases, to hyperlinking text, down to copying and pasting data between spreadsheets and tables and documents –
- This involves mapping data from one program into another program, which represents it in a different form – but methods to do this are extremely ad hoc.
- CAT4 does this in a simple highly automated way.
Introduction to CAT4

CAT4 is an extendable database.

• To do this integration, the CAT4 database is fully extendable: it extends to capture any information - just by entering “soft data” in standard forms.

• There is no programming for the user. You do not build new objects or program new code – you just enter information in a natural way.

• This is only possible because CAT4 identifies a general pattern behind all information – this pattern is captured by the special CAT4 network.
Introduction to CAT4

CAT4 is a general network structure for information.

- The discovery that there is a general network that can capture all information is the secret.
- This network is a type of structure with simple properties, and a recursive fractal composition.
- It means bundles of information relating to any subjects can be added or removed from a CAT4 network, by highly automated processes.
- It is much more comprehensive than tables or graphs or object hierarchies, etc – but also:
- The generalised structure makes it simpler to program!
Introduction to CAT4

How difficult is it to build?

• The software is surprisingly simple to build – especially considering the scope of what it does.

• An initial CAT4 DB is only about as complex as a mid-level business database –

• In 2007-09 I built a prototype with powerful features – with only 1,000 programmed functions – running dozens of screens and complex processes – tiny!

• This was to test higher-level functions to make sure the logic works → it led to the present solution.

• A good commercial version can now be built.
Introduction to CAT4

How difficult is it to build?

• Also a version can be built using any RDMS as a platform, and this is a great advantage:

• It can be provided as a functional add-on to a standard database ➔ very portable!

• The long-term goal is to make a generic application, optimised for its own processing functions –

• But we can build versions that plug into standard RDB systems, used as platforms –

• This also shows its power to easily integrate with other software platforms.
Introduction to CAT4

How difficult is it to build?

- We stage the development – the critical step is to build a core data engine.
- This is equipped with the essential file handling and core functions and the CAT4 functional language.
- Extended functions are built on top of this.

<table>
<thead>
<tr>
<th>1. Core Data Engine:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• data tables</td>
</tr>
<tr>
<td>• core functions</td>
</tr>
<tr>
<td>• screens</td>
</tr>
</tbody>
</table>

- Math & Stats
- Data Integration
- Special Graphics
- Natural Language - AI
Introduction to CAT4

How difficult is it to build?

• The critical first step is to build a core data engine.
• This creates the essential file structures and core functions and primary screens.
• A team of 3 full-time programmers could build a high quality CAT4 data engine in six months → to be commercially useable.
• This supports a first set of targeted user applications.
• And is the foundation for extended functions – which can also be developed by users within CAT4.
Introduction to CAT4

What specific application can be targeted first?

• Want to aim for specific real-world uses from the start → for real-world demonstrations and support.

• As a start-up we can stage some products along the way → then go to complete a mass-market product.

• It can be used for many purposes, but its unique strength lies in its ability to integrate complex data – which is scattered across databases, spreadsheets, text files, html or xml, meta-data, coded functions, etc – so we start with this data hub function.
Introduction to CAT4

Target a data hub for analysts/researchers first.

- This is an application for analysts and researchers in complex projects → to pull together data and references and writing and graphics etc → it aims to be the single hub for complex projects.
- No one has made a viable system for this.
- It replaces a variety of ad hoc tools.
- It works equally well for any business or academic projects → anyone who needs to relate lots of details between documents, data, text, publications, etc.
Introduction to CAT4

What specific application can be targeted first?

- It is powerful for sharing information across networks of people → enables distributed information systems.
- Allows parts of information to be easily extracted and shared → ideal for website back-ends.
- Special interest for researchers in some areas → like network theorists, system architects, semantic logicians – and in computational linguistics → systems research.
- So this is what we can target a user product to first:
- Larger orgs with extensive business analysis systems or complex research projects.
Introduction to CAT4

What general applications can be targeted?

• Once it is established it has open-ended potential, to provide generic DB-apps for all kinds of businesses.

• It makes complex flexible databases → you make a “database” just by typing in examples → removes the development process!

• It removes the bulk of manual database management and programming work → setting up systems up and maintaining them.

• A full database version is a mass product market → requires a lot more product refinement.

• We can aim for some targeted business uses first.
Introduction to CAT4

To recap the general concept.

• We can map any kind of relational or graph structures into a CAT4 network →
• And there are algorithms to do these mappings →
• These algorithms are represented within CAT4 itself →
• It provides its own machine-learning capability →
• This leads into a full AI-machine learning system →
• It extends to a natural language interpreter.
Introduction to CAT4

Representational Power.

• The key aspect of the theory is *representational power*.
• The *representational method* has a quite different focus to the *computational focus* of traditional theory.
• CAT4 gets its power by representing information as *meaning* – or equally, *meaning as information*.
• This brings key insights from *advanced semantic logics* → *intensional, modal, counterfactual logics*.
• And this area is about natural language semantics and NL translation and computational linguistics.
Introduction to CAT4

Representational power: natural language translator.

• A powerful use of CAT4 is for linguistics → it lets us make a real natural language translator.

• This is a famous unsolved problem for over 50 years – since Chomsky’s theory of “universal grammar”.

• Logicians have tried to solve this by decoding NL “grammar” from syntax – but this doesn’t work.

• CAT4 solves it because it can map language statements to their meanings – and logically process meaning as facts – and map facts back into language.

• And its generic machine learning capability is perfect for this → CAT4 can learn languages!
Introduction to CAT4

Representational power: natural language translator.

- CAT4 lets us make a real natural language translator.
- Example of a CAT4 network for a phrase \( \rightarrow \) the meaning is represented in CAT4.
Introduction to CAT4

Representational power: natural language translator.

• This natural language translator shows the revolutionary power of the system most clearly.
• But we cannot program this in the first stage –
• We have to develop the core data engine first, and then enhance secondary functions – and build the natural language translator as a third-level function.
• Then it suddenly becomes a full AI system!
• But there is also a much simpler method for NL we can introduce immediately.
Introduction to CAT4

Special Feature: Screen Language Translator.

• The full *natural language interpreter-translator* is a higher-order function –

• But there is a practical language function that can be provided immediately – this is a simple method for *translating the screen language for users*. 

• You can change the language that appears on screens at the click of a button – to English or Maori or Samoan...

• This is superficial translation compared to the meaning translator - but it is very effective – and this is also uniquely enabled by the CAT4 method!
Introduction to CAT4

Changing the screen language.

• It means you can switch *all text that appears on any screen into an alternative language.*

• People can work in their preferred language – and switch exactly the same information to another language for another user.

• This can be a powerful feature for people who want to support indigenous language environments in computer systems.
Introduction to CAT4

Next step: Business Development.

• That introduces what the CAT4 project is about. It has real power and unique potential - and it is very doable!
• The key step is to get it into a business development.
• It can be staged with intermediate goals → cash-flow
• But because it has power to solve critical problems, it should be strategized with a high future value.
• It only requires a small team to build the primary data engine → then expand to commercial uses.
• It needs support from partnership with a larger org → with a long–term business strategy.
Introduction to CAT4

Partnerships can include:

• Capital investor funding start-ups for part-ownership.
• Software entrepreneur with developer & product expertise → fully profit-oriented enterprise.
• Academic/research group wanting a software solution & research opportunities (… academic talent).
• NGO or org wanting a software solution and future stakeholder benefits → social-good aspect.
• A large business prepared to fund a CAT4 system for their own business needs → without ownership.
• We need a partner with strong values, who will ideally become the long-term owner → manage a large-scale enterprise.
Introduction to CAT4

To conclude: the project has special values.

• To me, there is a strong *public good aspect*: a purely commercial motive will not work.

• It has a special public value because of its *language power* – it can support *minority languages* ➔ *central to supporting opportunities for indigenous cultures* ➔ can be done by an indigenous NGO.

• It has potential to become a valuable mass product.

• It challenges the IT corps ➔ *software models* and the *cloud paradigm of centralised control systems*.

• It initiates a new type of AI ➔ large social impact!
Introduction to CAT4

Thank you for your time, and please feel welcome to email me below.

CAT4 is an ATASA Research project.
https://atasanz.com

Dr. Andrew Holster. atasa030@gmail.com