**“Three Your Data” (Three – Free – Get it?)**

|  |  |  |
| --- | --- | --- |
| **Intro :** | **Ken :** Team Kojak Presents  **Ken :** Three Your Data  **Ken :** It’s an ordinary Monday Morning |  |
| **Scene 1 – Introduction**  What is the business scenario and problem? | **Jochen :** Hi Ken in Legacy Systems!  **Ken :** Hi Jochen, CEO of top financial institution National Kensington Bank. What can I do for you?  **Jochen** : I’ve just acquired New Insurance Co, a major insurance company and our share price will double in the next 24 hours. That’s because we’ll focus on customers who make us money. Customers who keep having accidents or who have no cash will be seeing some pretty sharp increases in costs, or ( menacingly ) maybe they’ll leave us all together.  **Ken :** Brilliant idea Sir!  **Jochen :** So I just need a list of those customers. \****pause\**** I can wait…  **Ken :** Errr... There's going to be a problem, Sir  **Jochen :** Problems! You legacy systems people. What is it this time? I'm an important CEO and I need that information!  **Ken :** Well that information is in two \***completely\*** different databases and organised in different schema. I just can’t run a query that will compare the data.  **Jochen :** But I need it by lunchtime! How long will it take?  **Ken :** Well data unifications exercises usually get estimated at three months, and take two years if they finish at all  **Jochen :** Not exactly lunchtime, Ken from Legacy Systems? Think man ,think!  **Ken :** What if I ask Adam from the Future?  **Jochen :** You mean mean Adam who’s just installed MarkLogic 7?  **Ken :** Exactamundo | Critical to the success of business in the 21st century is being responsive to changing business needs and costs. The quicker merged companies can save costs, the more profitable and successful an acquisition will be.  Legacy integration is very costly. You need to be able to pull in information from disparate data sources, without spending months on schema redesign. |
| **Scene 2 – Adam Helps**  Show how to import and query data.  Also show inferencing – finding joint customers. | **Ken :** Adam, I’ve got this terrible problem. We’ve just acquired New Insurance Co. Jochen from the boardroom wants a list of all our joint customers, so he can get rid of the unprofitable ones.  **Adam :** So?  **Ken :** So that’s going to mean unifying heterogenous schema from disparate data sources. Here's the two schema. They're nothing alike! How can I get this done!  **Adam :** Well Ken, it so happens I've been importing claims Documents from New Insurance Company and I've set up connections to those two databases and amazingly someone from the MarkLogic Community has developed this easily re-usable rdb2rdf thing that should do exactly what you need!  **Ken**: **\*sounding relieved\*** Fantastic!  **Adam**: Even better Ken, it uses the W3C Open Standard for direct import of RDBMS data to a triple store and we're all over Open Standards. Lets see …  *Show the rdb2rdf import screen and list the databases*  **Adam**: Which database?  **Ken**: NewInsuranceCo and NationalKensingtonBank  **Adam**: OK, NewInsuranceCo first  **Ken**: Yeah right. Like you can just import data as-is!  **Adam**: Watch me.  *Import of all tables of NewInsuranceCo runs*  **Ken**: Velocity-tastic!  **Adam**: You betcha! Now the other one...  *Import of NationalKensingtonBank*  **Adam**: Ha! Done! Lets take a quick look  *Show sparql widget of query for all 'NIC Client' entities, and click on one to view their info, and their relationships to other tables*  **Adam**: All the data, ready to explore!  **Ken**: Nice! … but what about finding joint customers. | Creating a merged schema and moving systems over takes months, if not years. Often companies just replace systems rather than make use of existing, valuable data.  Open Standards means there are people with common, existing skills you can use, reducing vendor lock in and increasing productivity of your IT team.  Importing data as-is, to either the document database or triple store, allows you to take decisions and implement new systems quickly, reducing your bottom line and giving you competitive advantage. |
| **Scene 3 – Inferencing**  Discovering new facts from existing information. | **Adam**: We’ll run a query that creates Joint Customer entities for people with the same DoB and names. This adds more triples to MarkLogic. Our future queries will be much simpler – just search for 'Joint Customer' entities!  **Ken**: Just like SQL!  **Adam**: More of a Subject-Predicate-Object man myself. As you see, we're linking the insurance company clients to bank customers. I'm also adding triples that link claims documents to Joint Customers, so we can search by documents too.  **Ken**: Easier than creating a new schema! That takes months!  **Adam**: And the beauty is we're not altering data – just adding new facts.  *Inferencing query executing in the background*  **Adam**: Done. Let’s find join customers.  *Executes SPARQL UI widget query for all Joint Customer entities. Clicks on one and shows links to both original customer objects*  **Adam**: From here Ken, we search for all related information. What else did you need? | Rather than write spaghethi SQL you can infer simple facts from complex queries, then use those simple facts in the future. This is called Inferencing and is the basis for a lot of semantic solutions.  Storing these inferred facts rather than carrying them out on the fly increases query performance.  Being able to browse all related data in a user friendly interface is key for applications aimed at business people. |
| **Scene 4 – Dodgy customers**  Combining information from multiple data sources and formats to drive business insights. | **Ken**: I need dodgy customers – those with bank balances below GBP 100 and with expensive claims, the most expensive being claims involving pedestrians. Oh but, they're in documents, not a database! That's totally different data! We're stuck again.  **Adam**: Calm down dear, we can totally do that!  *Edit SPARQL query so that you search for JointCustomers with NKB Accounts with Balances < 50. Executing this also shows related documents to those Customers*  **Adam**: OK so here are your poor joint customers. Now combine this with a content search to find all customers with claims documents with the word 'pedestrian' in them.  *Adds 'pedestrian' as a word query*  **Adam**: Click and you get a summary for the CEO's viewing pleasure! Job Done Ken!  **Ken**: Job done indeed! I've never seen software so flexible and powerful!  **Adam**: No Ken, it's MarkLogic 7 – Time for us to get in the fast lane! | Merging facts (triples) and documents in a single search environment is unique to MarkLogic's Enterprise NoSQL database. This enables you to drive decision making with ALL the available data, no matter the format or source, and to do so quickly as required when new data sources become available. This reduces IT landscape complexity, and cost of creating new business insights. |
| Scene 5 – credits, smily faces, etc. | Ken : I’ve got that data sir.  Jochen : Good man. I knew I could count on you.  Ken : At first I thought it was really hard, but Adam helped me realize …  Adam : It was just a matter of semantics  Jochen : You guys … |  |