LDBC SPB:
Semantic Publishing Benchmark

LDBC

Outline

- Introduction to LDBC SPB
- Workloads
- Software
- Sample results
- Next steps

BBC's DSP Approach

- The SPB was inspired by the Dynamic Semantic Publishing
  - First applied at BBC's FIFA Worldcup 2010 website
  - Next at website of BBC and the official one for the 2012 Olympics
  - Now continuously used at BBC Sport and by many others

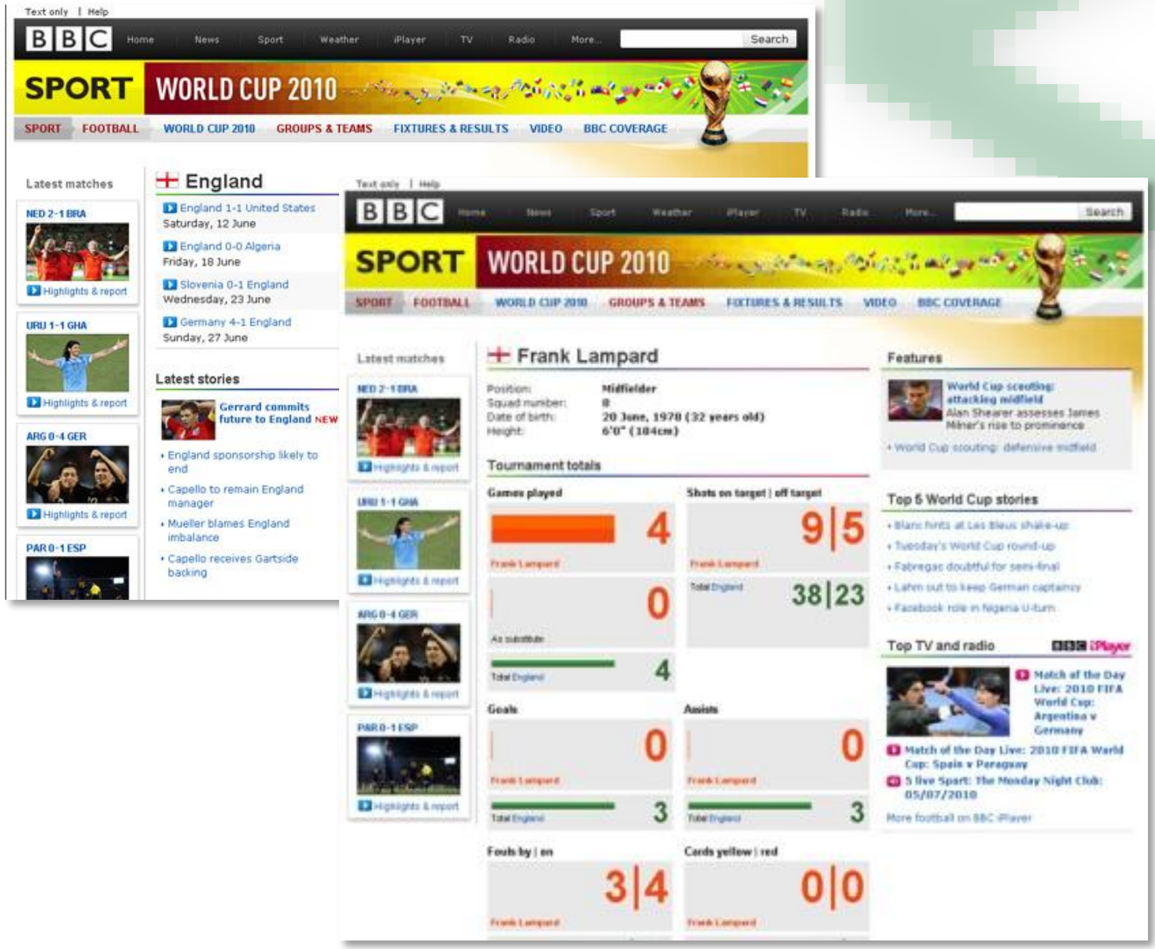
- The BBC's primary use case: rich, deep, dynamic websites
  - Text-mining automatically annotates articles with entities
  - Editor curates the metadata before storing it in a triplestore
  - Thematic web pages are generated on-the-fly through SPARQL

BBC's DSP Approach

"The goal is to be able to more easily and accurately aggregate content, find it and share it across many sources. From these simple relationships and building blocks you can dynamically build up incredibly rich sites and navigation on any platform."

John O'Donovan

Chief Technical Architect, BBC



England

England 1-1 United States
Saturday, 12 June

England 0-0 Algeria
Friday, 18 June

Slovenia 0-1 England
Wednesday, 23 June

Germany 4-1 England
Sunday, 27 June

Latest stories

Gerrard commits future to England NEW

- England sponsorship likely to end
- Capello to remain England manager
- Mueller blames England imbalance
- Capello receives Gartside backing

Frank Lampard

Position:	Midfielder
Squad number:	8
Date of birth:	20 June, 1978 (32 years old)
Height:	6'0" (184cm)

Tournament totals

Games played: 4 (Frank Lampard); 0 (As substitute); 4 (Total England)

Shots on target | off target: 9|5 (Frank Lampard); 38|23 (Total England)

Goals: 0 (Frank Lampard); 3 (Total England)

Assists: 0 (Frank Lampard); 3 (Total England)

Fouls by | on: 3|4 (Frank Lampard)

Cards yellow | red: 0|0 (Frank Lampard)

Features

World Cup scouting: attacking midfield
Alan Shearer assesses James Milner's rise to prominence

- World Cup scouting: defensive midfield

Top 5 World Cup stories

- Blanc hints at Les Bleus shake-up
- Tuesday's World Cup round-up
- Fabregas doubtful for semi-final
- Lahm out to keep German captaincy
- Facebook role in Nigeria U-turn

Top TV and radio BBC iPlayer

Match of the Day Live: 2010 FIFA World Cup: Argentina v Germany

Match of the Day Live: 2010 FIFA World Cup: Spain v Paraguay

5 live Sport: The Monday Night Club: 05/07/2010

More football on BBC iPlayer

Publishing & Media Domain

- Publishing & Media Domain
  - Constantly generating new content
  - Constantly updating existing content
  - Constantly consuming content

- Semantic technologies in the publication pipeline
  - Annotation of content; often through text mining
    - Also know as “metadata enrichment” or just “tagging”
  - Content multi-purposing
    - Multiple media (press/TV, web, mobile), multiple products (specific channels, data feeds, etc.)

LDBC SPB

- LDBC-Semantic Publishing Benchmark [1]
  - A benchmark for *RDF Databases*
  - Simulates a media organization which maintains a catalogue of meta-data for its assets
  - Simulates operations performed by real users:
    - consumption of meta-data
    - management of meta-data
  - Measures the performance of both types of operations

[1] https://github.com/ldbc/ldbc_spb_bm/tree/master/doc

LDBC SPB Requirements

- Support for quadruples; triples plus context/named graphs
  - Support for TRIG or NQ RDF serialization syntax
- SPARQL Query 1.1 support
- SPARQL Update 1.1 support
  - READ COMMITTED transaction isolation
  - Consistent handling of updates
- SPARQL Protocol 1.1 (known as “SPARQL End-point”)
- Inference support – RDFS semantics is sufficient

Outline

- Introduction to LDBC-SPB
- Workloads
- Software
- Sample results
- Next steps

WORKLOADS

- Editorial agents
  - Simulate the work performed by journalists or editors with the system
    - E.g. enriching journalistic assets with meta-data: description, creation date, location etc.
  - Run simultaneously
  - Provide a constant stream of update operations
  - Editorial operations:
    - INSERT, UPDATE, DELETE

WORKLOADS

- Aggregation agents
  - Simulate the interactions of end-users or semi-automated tools with the system
    - E.g. queries generated by an application that dynamically generates web pages for wide range of topics (teams, players, events, etc.) at BBC Sport website
  - Run simultaneously
  - Provide a constant workload of queries from two available query mixes:
    - Basic query mix of 9 queries
    - Advanced query mix of 25 queries

WORKLOADS

- Basic (interactive) query mix (9 Queries) :
  - Search queries
  - Full-text search queries
  - Aggregation queries
  - Geospatial queries

- Advanced (interactive + analytical) query mixes (25 Queries):
  - Analytical queries
  - Faceted search queries
  - Drill-down queries

Outline

- Introduction to LDBC SPB
- Workloads
- Software
- Sample results
- Next steps

DATA GENERATOR

- Real ontologies provided by the BBC
- Real reference datasets provided by the BBC, DBpedia and Geonames
- Parallel data generation
- Deterministic

DATA GENERATOR

- Scalable
  - Scales to billions of triples

- Generated datasets simulate the activity of a publishing organization for a period of time
  - Media assets are enriched by metadata called *‘Creative Works’*

LDBC-SPB Test Driver

- The SPB test driver:
  - Open Source
  - Available on GitHub: https://github.com/ldbc/ldbc_spb_bm
- Runs Editorial and Aggregation agents simultaneously
  - Parallel execution
  - Provide a steady update stream during query workload
- Validates Results
  - Query results
  - Update operations; batching is not allowed
- Gathers and reports performance metrics

Outline

- Introduction to LDBC SPB
- Workloads
- Software
- Sample results
- Next steps

SPB Sample Results

Engine	Variant	Scale	R/W Agents	Hardware	Load time (sec)	Reads /sec	Updates /sec
GraphDB SE 6.0	Inter. Basic	50M	14/2	2xXeon, 256GB, SSD	2 045	32	12
GraphDB SE 6.0	Inter. Basic	50M	14/2	AWS c3.4xlarge (16GB, SSD)	2 045	27	11
GraphDB SE 6.1	Inter. Basic	50M	14/2	2xXeon, 256GB, SSD	2 045	34	17
GraphDB SE 6.0	Inter. Basic	1B	14/2	2xXeon, 256GB, SSD	41 400	10	2
GraphDB SE 6.1	Inter. Basic	1B	14/2	2xXeon, 256GB, SSD	41 400	10	6

Outline

- Introduction to LDBC SPB
- Workloads
- Software
- Sample results
- Next steps

Next Steps

- Generate rich set of public results
  - Ranging across different scales, variants, engines, hardware, etc.
- Richer reference knowledge; more links between entities
- Queries that explore links between entities
- Experiment with simple OWL inference

Next Steps

- Extension/ variant for testing enterprise features
  - Online backup
  - Various **failover scenarios**, e.g. seamless failover, hardware or network failures
  - Various **stress tests**, e.g. reaction of the cluster to a “Disk Full” event
  - Various **workload tests**, i.e. failover and stress tests passing under ongoing LDBC-SPB interactive workload

Questions?

Thank You