Complex Event processing for Time evolving graphs

Charith Wickramaarachchi
cwickram@usc.edu

Image source: http://whatis.techtarget.com/definition/3Vs
Volume

• Batch processing
  • Map-Reduce
  • Pregl
  • Impala
  • GoFFish/Gopher

Image source: [http://wikibon.org/wiki/v/Real-time_IO_Centric_Processing_for_Big_Data](http://wikibon.org/wiki/v/Real-time_IO_Centric_Processing_for_Big_Data)
Downsides of Batch processing

- Jobs must be run through to completion
- Data changes might cause to re-run the operations
- Will not work for some BI applications – acting in minutes/seconds is significant
- Value of Data degrades over time
• Complex Event processing
  ◦ combines data from multiple sources
  ◦ detect patterns
  ◦ attempt to identify either opportunities or threats

• Operational Intelligence (OI)
  ◦ real time data processing + CEP
  ◦ insight into operations by running query analysis against live feeds and event data

http://srinathsvi...
Current usage

- Detect customer buying patterns
  - Historical data + real time data
- Assembly lines
  - Reduce time, Detect Errors
- Social media threat monitoring
  - React to negative posts, tweets

<table>
<thead>
<tr>
<th>Solution</th>
<th>Organization</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storm</td>
<td>Twitter</td>
<td>Streaming</td>
<td>Twitters streaming big-data analytics solution</td>
</tr>
<tr>
<td>S4</td>
<td>Yahoo</td>
<td>Streaming</td>
<td>Stream computing platform Yahoo</td>
</tr>
<tr>
<td>Hadoop</td>
<td>Apache</td>
<td>Batch</td>
<td>Map-Reduce Impl</td>
</tr>
<tr>
<td>Spark</td>
<td>AMPLab USB</td>
<td>Batch</td>
<td>In Memory data analytics</td>
</tr>
<tr>
<td>Disco</td>
<td>Nokia</td>
<td>Batch</td>
<td>Map reduce</td>
</tr>
<tr>
<td>HPCC</td>
<td>LexisNexis</td>
<td>Batch</td>
<td>HPC cluster for Big-data</td>
</tr>
</tbody>
</table>
• Twitter's distributed and fault tolerant stream processing engine.

• Key Concepts
  ◦ Tuples: Ordered List of Elements
  ◦ Streams: Unbounded $seq^n$ of Tuples
  ◦ Spouts: Stream source
  ◦ Bolts: Processing units which consumes streams and create new ones (Pellets in Floe)
• Topology: Directed Graph of Spouts and Bolts

Word count

Source:
https://github.com/nathanmarz/storm/wiki/Tutorial
From Last week...

- Early work on Big data processing
  - More focus on volume and verity dimensions
- “Big Data in Real time” getting lot of traction.
- State of the art
  - OI, CEP, Steam processing
  - Twitter storm
Overview for this week..

- Complex event processing systems
  - Overview
  - Operation abstractions
- Motivational problems /use cases
- Hypothesis
- Open problems
- Related work
- Directions ? Feedback ?
Complex Event Processing

- Processing that combines data from multiple sources and infer events or patterns that suggest more complex scenarios

- Queries
  - Filters
  - Windows
  - Joins
  - Patterns and Sequences
  - Event Tables
  - Partitions

[Image: http://srinathsview.blogspot.com/2013/08/understanding-complex-event-processing.html]
Motivation

- Early pattern identification in dynamic graphs
  - False rumor detection on social networks
  - Early fault detection and prevention in sensor networks/data centers
  - D2R use cases in Smart grid?
Hypothesis: Online graph event detection system, that consider data streams as streams of updates, to a changing dynamic graph, enable “better” online graph analytics and event detection.
Big Picture

Dynamic graph streams

System X

DSL - X

Events
Example Queries

- Find a Formation of a Cluster with diameter 8
- Find a formation of a path of length 5
- Find a Cluster transformation
- Find a Cluster destruction
• Applications
• Language model for online graph pattern detection
• Processing models which compiles this language and perform pattern detection
  ◦ Distribution of Graph
  ◦ Performance optimizations
  ◦ Pattern Matching algorithms.

Open problems
Related work

- Towards Efficient Query Processing on Massive Time-Evolving Graphs, Arash Fard at al... (Survey paper)
  - Partitioning time evolving graphs. (When to do re partitioning, Node replication to minimize communication)
  - Sub graph pattern marching on TEG.
    - Using diffs in different graph snapshots (Using Sparse bit matrices)
• Kineograph: Raymond at al..
  ◦ In memory graph storage
  ◦ Create time snapshots of dynamic graph
  ◦ Supports iterative propagation based graph mining.
• An Event-based Framework for Characterizing the Evolutionary Behavior of Interaction Graphs, Sitaram Asur et al...
  ◦ Defines set of critical events in time evolving graphs
    • K-Merge – (k% merge)
    • K-Split
    • Form
    • Dissolve etc..
• Graph pattern matching, dynamic graph analytics getting some traction.
• The growth of “network” big data require a new class of real time data processing mechanisms
• Fast online analytics on dynamic graphs is a still a wide open research area.
Questions ? Directions ? Feedback ?