Analytics-R'-Us

Schema Integration and Justification Team

Demand Prediction Analysis

DSE 203 Presentation #4

11/17/2017

Team:

Josh Wilson

Amisha Bhanage

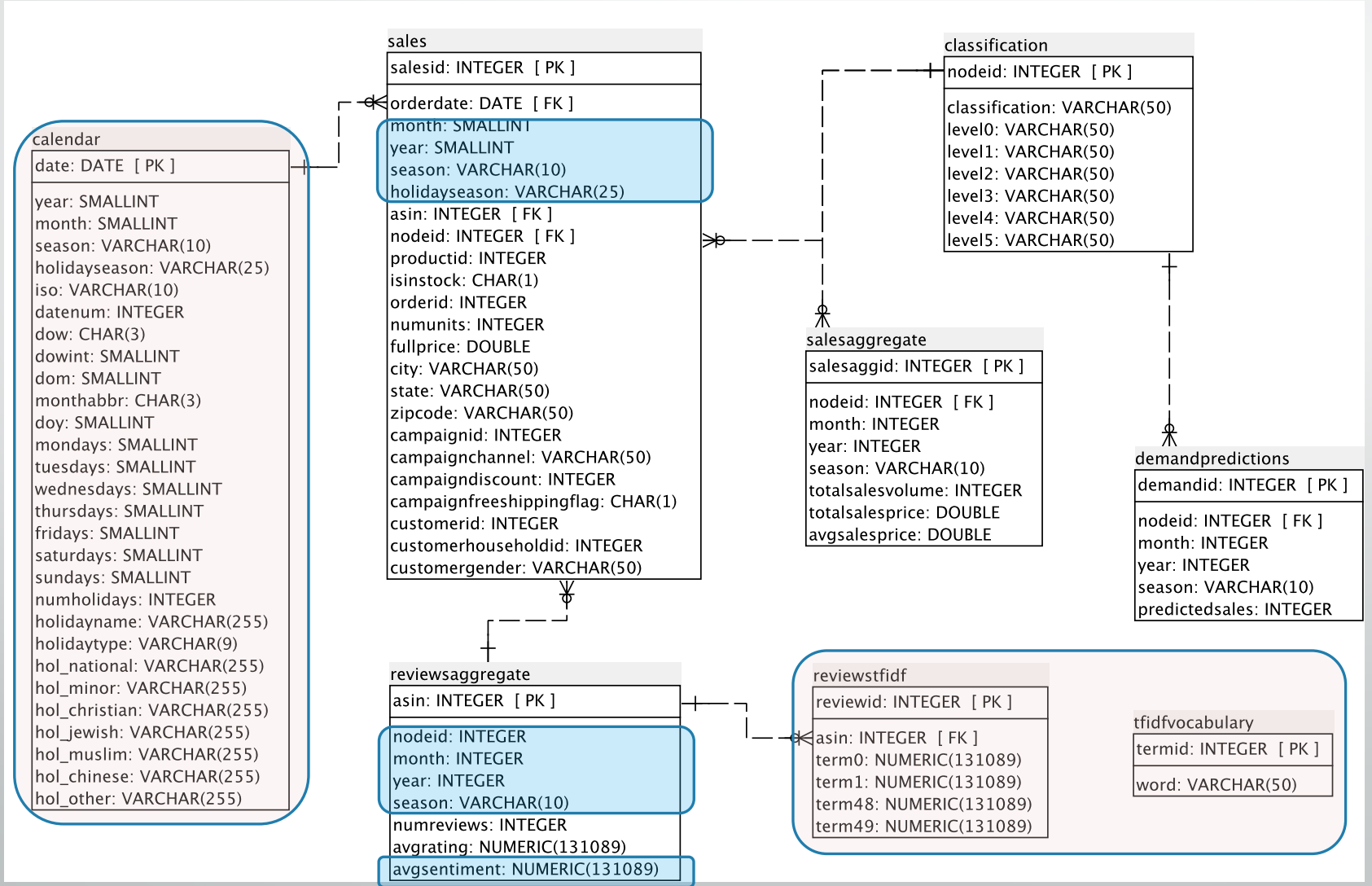
Ken Kroel

Mai Huynh

Agenda

- Schema update and views
- Handling of non-SQL cases
- Query unfolding status

Schema Update and Views – Previous ERD



Schema Update and Views

- Changes to existing mediated tables
  - reviewsAggregate – Add nodeID, month, year, season, average sentiment polarity score
- Potential additions
  - Views for various types of aggregations – yearly, quarterly, previous period(s)
  - View listing TopN categories
- Proposed removals
  - Eliminate calendar
    - Include month, year, season, holidayseason info in sales table
  - Eliminate tf-idf info
    - ML team only needs sentiment analysis from review text

Handling Non-SQL Cases

- Present view of non-SQL data that can be queried with datalog
- Mapping of datalog query to source data will occur in query engine
  - Information needed to perform mapping will be provided to query team
- Example: Nested category data from AsterixDB

Datalog query against mediated schema:

```
Ans(nodeid, classification, level0, level1) :-
    classification(nodeid, classification, level0, level1),
    nodeid = 1
```

classification

nodeid: BIGINT
classification: VARCHAR(100)
level0: VARCHAR(100)
level1: VARCHAR(100)
level2: VARCHAR(100)
level3: VARCHAR(100)
level4: VARCHAR(100)
level5: VARCHAR(100)

SQL++ query against flattened data:

```
SELECT c.nodeID, c.classification, c.level0, c.level1
FROM classification c
WHERE c.nodeID = 1;
```

```
SELECT c.nodeID, c.classification, c.level0, c.level1
FROM
    (SELECT
    c_nested.nodeID, c_nested.classification,
    c_nested.category.level_0 AS level0,
    c_nested.category.nested.level_1 AS level1,
    c_nested.category.nested.nested.level_2 AS level2,
    c_nested.category.nested.nested.nested.level_3 AS level3,
    c_nested.category.nested.nested.nested.nested.level_4 AS level4,
    c_nested.category.nested.nested.nested.nested.nested.level_5 AS level5
    FROM ClassificationInfo c_nested) AS c
WHERE c.nodeID = 1;
```

Result:

```
{ "nodeID": 1,
  "classification": "Arts & Photography",
  "level0": "Books",
  "level1": "Arts & Photography" }
```

Query Unfolding Status

- Have written functions to decompose datalog query text into head, body, conditions, group_by, order_by, and topn clauses
- Working on generic process to unfold datalog query against mediated schema into equivalent query on data sources
  - Uses dictionaries to store information on attribute position, source, matching source attribute, source attribute position
  - Remaps mediated schema body relations and attributes to source relations and attributes
- Adhering to datalog standards per
  - http://logic.stanford.edu/reports/LG-2012-01.pdf
  - https://pages.iai.uni-bonn.de/manthey_rainer/IIS_1718/manualDES4.1.pdf

Q & A

Thank you!