Oxford Economics

Technical Interview Scenarios

Anything is possible!

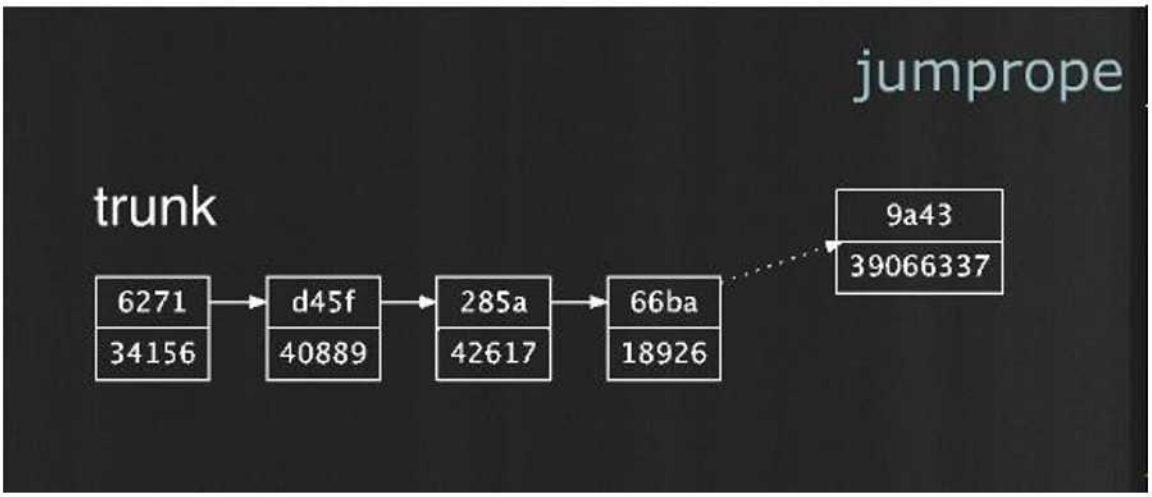
Authored by	Software Engineering Team
Abstract	This document contains a technical scenario which software engineer job candidates are required to analyse, problem solve, and replay in a Q&A session with the interviewing team. • Please do not send any code or written documents. We will discuss your ideas in person at the interview. • A flip chart/whiteboard and pens will be provided. The aim of this interview is to appraise your technical problem-solving skills and to see how you manage the interaction with us. For example, drawing pictures would be a good way to explain your thinking and demonstrate your group communication skills. We expect thorough preparation for this interview, so please plan this into your schedule.
Version	Version 11.0
Status	Final
Date of issue	20/11/2019

Scenario 1

The objective of our first meeting with candidates for the Software Developer role at Oxford Economics to assess their conceptual thinking, and ability to take technical concepts and ideas, and apply them in different problem / solution scenarios.

To prepare for this, candidates should view and study the following InfoQ video on data structures.

https://www.infoq.com/presentations/Data-Structures

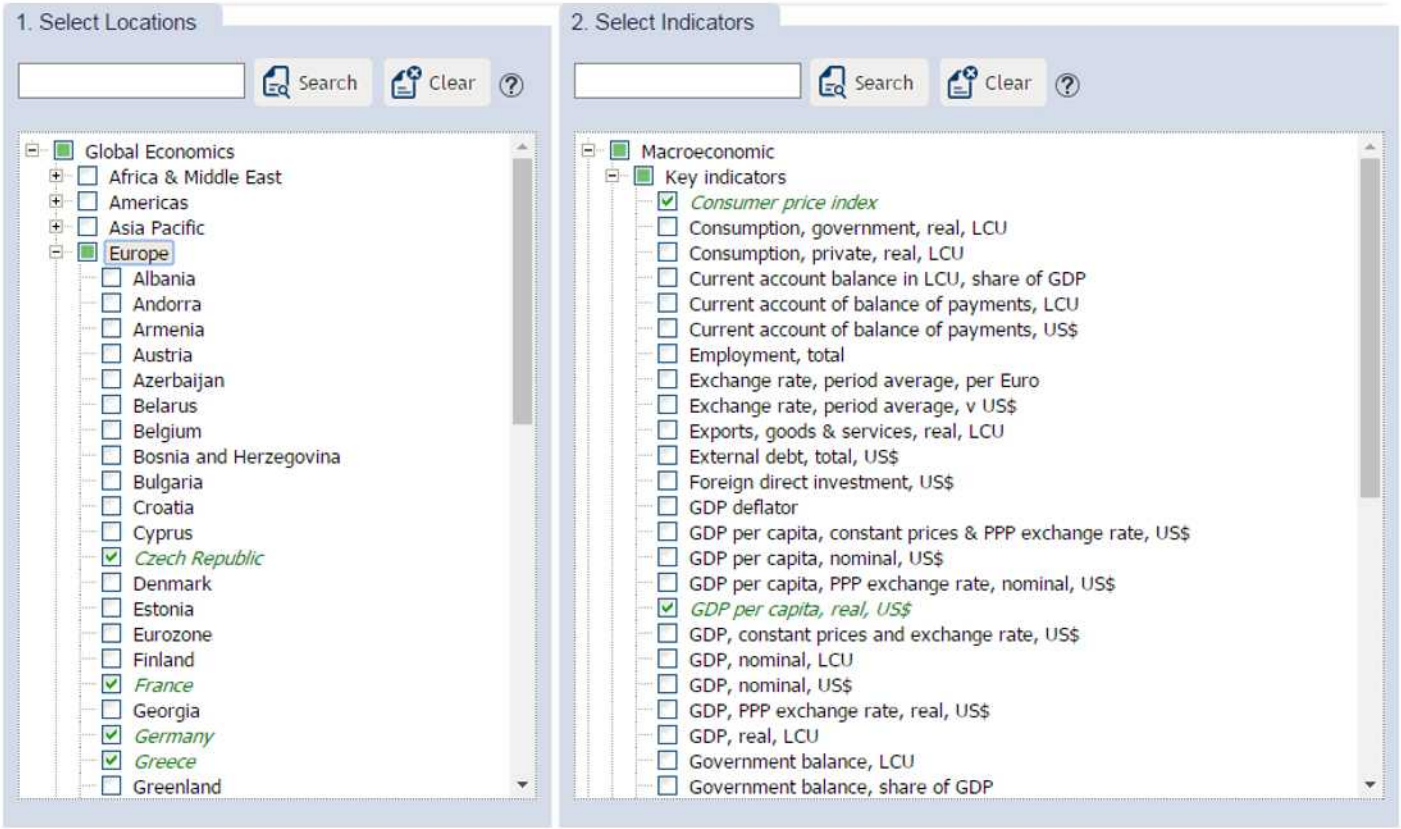


During the interview we would like you to present one or two of the structures and explain how they could be useful in practice.

Scenario 2

Overview

Our online databank allows data to be selected by choosing a set of indicators (e.g. GDP, CPI, etc.) and a set of locations (e.g. UK, Germany, France) from the available pool. This set of selections resolves to the underlying data series for each permutation of indicator and location.



In an ideal world, there is a data series for every selection permutation but, due to the lack of availability and varying quality of data sources for countries in the world, this is generally not possible.

The result is our data has varying indicator coverage from country to country which leads to a number of challenges that need to be addressed.

Challenges

User feedback [15 minutes]

Design a user experience for real-time feedback on the coverage of results as the user is making indicator and country selections in the user interface. The feedback could show expected number of results and some measure of the 'quality' of the selection.

Ensure you address the following concerns in your discussion: the user experience should be intuitive; screen real estate is limited and maintaining the user's context is very important; potentially large datasets need to be processed on the back end depending on the number of selections made; server performance can be adversely impacted depending on the technical/design choices you make and you should demonstrate how you would evaluate and/or justify option selection; results can be sparse despite making many selections; results can even resolve to an empty set.

Data pivoting [30 minutes]

Design a data structure and processing system to efficiently pivot ("cross tabulate") a multi-dimensional data set.

We would like to tabulate our economic data series in a variety of ways. "Cross tabbing", "table reshaping" are common terms used to describe this. Typical tabular operations are transpose, pivot, unpivot, merge, join, append, aggregate, transform.

For example, the unpivoting operation can reshape several "YYYY" columns in a table of data series for different location and indicator combinations, into two columns, namely "Year" and "Data". The "YYYY" measures have effectively been reshaped into their proper dimensions. The impact on the table shape is that it goes from being a "short, wide" table to a "tall, skinny" one.

	Location	Indicator	2015	2016	2017	2018	2019	2020
1	Africa	Consumer price index	148.4037	158.2262	166.8495	176.3823	186.1483	196.3048
2	Argentina	Consumer price index	122.4448	153.9463	180.3576	200.7494	214.9622	225.615
3	Asia Pacific	Consumer price index	120.01	122.4193	124.6211	127.3973	129.7948	131.6364
4	Austria	Consumer price index	110.7644	111.564	112.7208	113.4321	113.0592	112.0471
5	Belgium	Consumer price index	100.9025	101.0356	100.8121	100.7639	100.4455	100.1205
6	Brazil	Consumer price index	166.9455	180.7531	193.5347	204.3246	212.8017	221.57
7	BRICs	Consumer price index	128.3669	132.4056	135.654	139.3869	142.4556	144.7684
8	Bulgaria	Consumer price index	6553.042	6530.23	6684.556	6930.256	7121.899	7302.448
9	Canada	Consumer price index	126.479	127.6394	129.4678	130.2238	130.0111	129.4892
10	Chile	Consumer price index	108.9325	113.2185	116.9544	119.7645	121.283	122.9598
11	China	Consumer price index	132.6501	134.0105	135.427	137.3882	138.6542	138.861
12	Croatia	Consumer price index	107.4083	107.3487	108.5752	111.2471	113.4124	115.6628
13	Czech Republic	Consumer price index	123.6045	123.9182	123.1601	125.2476	127.6102	129.1143

▼

	Location	Indicator	Year	Data
1	Africa	Consumer price index	2015	148.4037
2	Africa	Consumer price index	2016	158.2262
3	Africa	Consumer price index	2017	166.8495
4	Africa	Consumer price index	2018	176.3823
5	Africa	Consumer price index	2019	186.1483
6	Africa	Consumer price index	2020	196.3048
7	Argentina	Consumer price index	2015	122.4448
8	Argentina	Consumer price index	2016	153.9463
9	Argentina	Consumer price index	2017	180.3576
10	Argentina	Consumer price index	2018	200.7494
11	Argentina	Consumer price index	2019	214.9622
12	Argentina	Consumer price index	2020	225.615
13	Asia Pacific	Consumer price index	2015	120.01
14	Asia Pacific	Consumer price index	2016	122.4193
15	Asia Pacific	Consumer price index	2017	124.6211
16	Asia Pacific	Consumer price index	2018	127.3973
17	Asia Pacific	Consumer price index	2019	129.7948
18	Asia Pacific	Consumer price index	2020	131.6364
19	Austria	Consumer price index	2015	110.7644
20	Austria	Consumer price index	2016	111.564
21	Austria	Consumer price index	2017	112.7208

It should be easy to combine dimensions and assign them to the x and y axes of a table or chart to display the resulting data set and its aggregates. In this example, an unpivoted table of Employment data (similar to the earlier of CPI data) has been aggregated over the Location dimension computing average employment over the period 2015 to 2020.

	Location	Average Employment 2015 -2020
1	Argentina	18267.835
2	Australia	12212.06
3	Austria	4301.478833
4	Belgium	4639.187667
5	Brazil	91198.04333
6	Bulgaria	2994.814167
7	Canada	18005.985
8	Chile	8113.270333
9	China	764549.2
10	Croatia	1329.088833
11	Czech Republic	4994.617167
12	Denmark	2684.426
13	European Union	230975.4333
14	Eurozone	152887.9833
15	Finland	2503.553
16	France	27969.87167
17	Germany	43137.77667
18	Greece	3583.231167
19	Hong Kong, China	3733.135833
20	Hungary	4206.107
21	India	443270.5
22	Indonesia	122467.6667

Be prepared to answer the following questions:

If your solution had to be implemented in-memory

- What kind of data structures would you use to efficiently store the data series measures and dimensions?
- What algorithm, calculation or procedure would you apply to fetch data, (A) matching a boolean combination, or (B) to aggregate, over any given set of measures or dimensions?
- What issues do you see affecting performance of such queries?

If you had to use a database instead of an in-memory solution

- What database technology would you use, and why?
    - For example, a relational database (SQL) vs. a document/key-value store (NoSQL) solution vs. map-reduce (Hadoop) vs. multi-dimensional database (cube)?
- Discuss the pros and cons of your choice in terms of the data representation scheme?
- Discuss the business considerations of your database technology choice?

What and why are they important?