

Indian Institute of Technology Madras
ZANZIBAR CAMPUS

Z3003 · DATA MINING: SCALING TO LARGE DATA

OLAP Operations: Moving Through the Cube

Lecture 11 · Thursday 8 October 2026

Dr. Innocent Nyalala | BS Semester 5 (DSAI) | IIT Madras Zanzibar

Session Outline

IIT Madras Zanzibar

1. (0–5 min) Recap: cube, cuboids, star schema
2. (5–12 min) What OLAP is — the Google Maps analogy
3. (12–30 min) Roll-up and drill-down, on real Zanzibar numbers
4. (30–48 min) Slice, dice and pivot — the 2017 drought story
5. (48–58 min) OLAP in SQL (GROUP BY ROLLUP / CUBE) and pandas
6. (58–66 min) Bitmap indexes: answering in milliseconds
7. (66–75 min) Exercise, summary, questions

Learning Outcomes

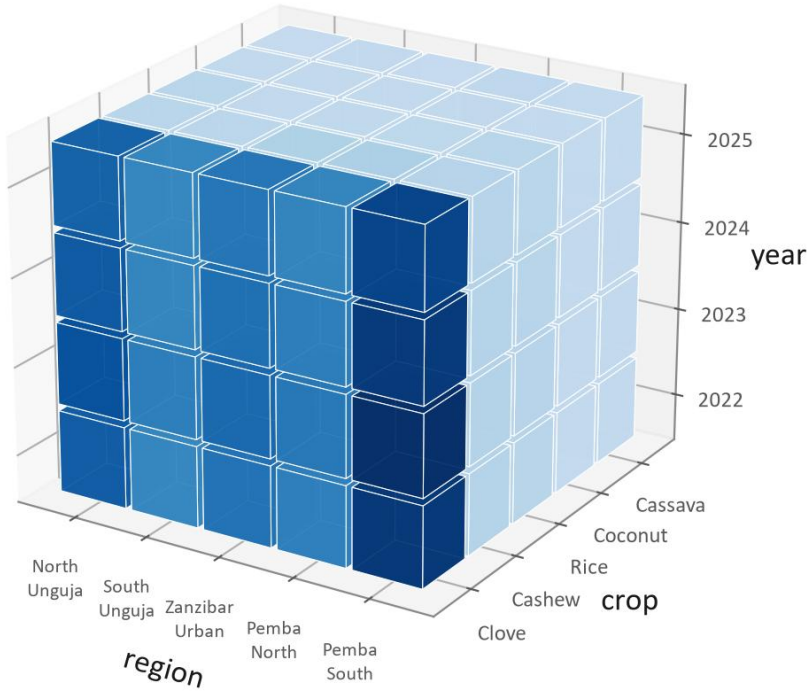
IIT Madras Zanzibar

By the end of this session, you will be able to:

Define	roll-up, drill-down, slice, dice and pivot, each with a picture
Apply	each operation to our 200,000-row cube and read the answer
Write	each operation as SQL (GROUP BY, WHERE, ROLLUP, CUBE) and as pandas
Explain	how a bitmap index answers 'crop = Clove AND region = Pemba South' fast
Compare	ROLAP, MOLAP and HOLAP servers

Recap: The Cube We Built Yesterday

Each small cube = total revenue for one (region, crop, year)



1. **Cube**
   revenue for every (region, crop, year)

2. **Cuboids**
   2^n group-bys; $\prod(L_i + 1)$ with hierarchies

3. **Hierarchies**
   district → region; season → year

4. **Star schema**
   fact_harvest + dim tables

Today's question: how does an analyst move around this cube to find a story?

OLAP Operations: Moving Through the Cube

IIT Madras Zanzibar

→ **Foundations: What OLAP Is**

Roll-Up and Drill-Down

Slice, Dice and Pivot

OLAP in SQL and pandas

Indexing for Fast OLAP

In-Class Exercise

What Is OLAP? Think Google Maps

IIT Madras Zanzibar

R	**Roll-up**	**Maps: zoom out**	fewer details, bigger totals
D	**Drill-down**	**Maps: zoom in**	more detail
S	**Slice**	**Maps: one layer**	fix ONE dimension to ONE value
D	**Dice**	**Maps: select an area**	keep ranges on SEVERAL dimensions
P	**Pivot**	**Maps: rotate the map**	swap rows and columns

OLAP = OnLine Analytical Processing: interactive moves through a cube, each answered in seconds.

OLAP Operations: Moving Through the Cube

IIT Madras Zanzibar

Foundations: What OLAP Is

→ **Roll-Up and Drill-Down**

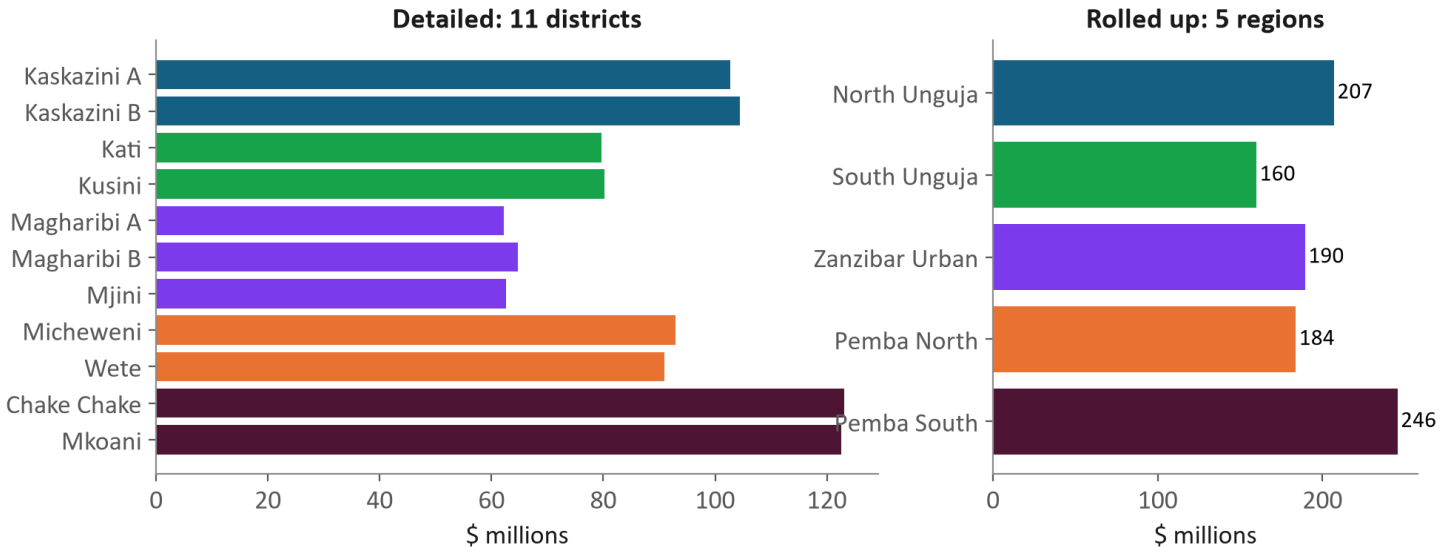
Slice, Dice and Pivot

OLAP in SQL and pandas

Indexing for Fast OLAP

In-Class Exercise

Roll-Up: Climb the Hierarchy



1 **Definition**
summarize by moving UP a hierarchy (district → region)

2 **Check**
Chake Chake 123.0 + Mkoani 122.5 = 245.6 $M

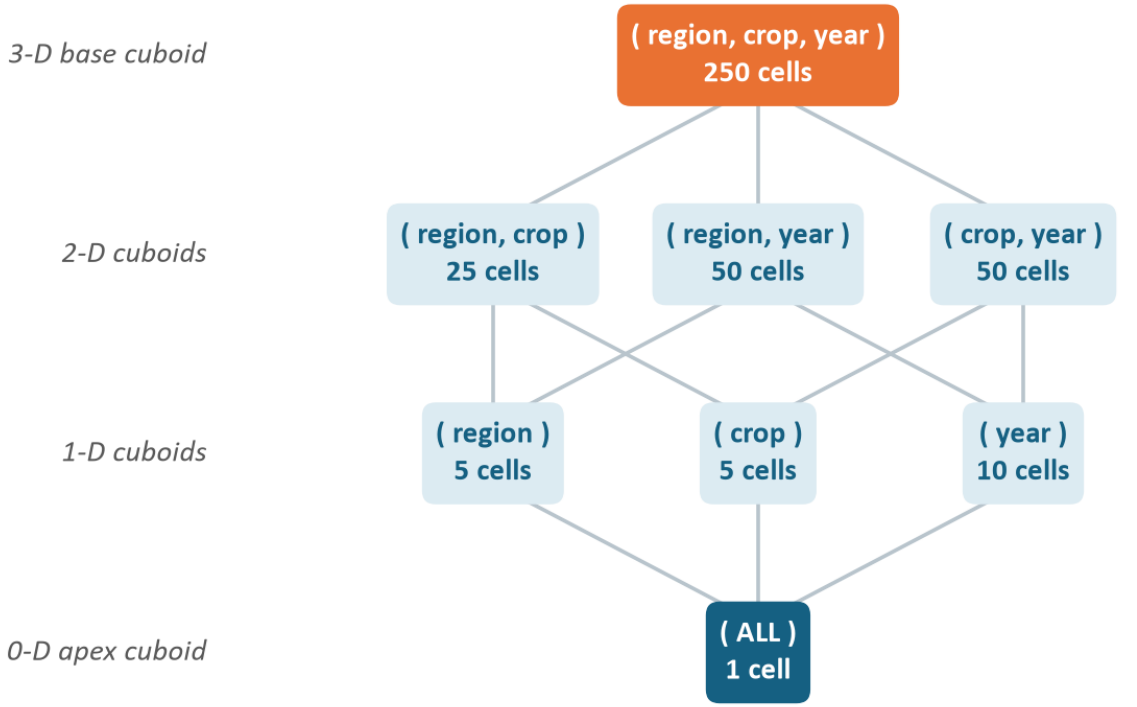
```
SELECT region, SUM(revenue_usd)
FROM farm_seasons
GROUP BY region;
```

Everyday version: your M-Pesa app lists each transaction; the monthly statement ROLLS UP to one total per month and type.

Roll-Up by Dropping a Dimension

IIT Madras Zanzibar

3 dimensions → 2^3 = 8 cuboids (group-by combinations)

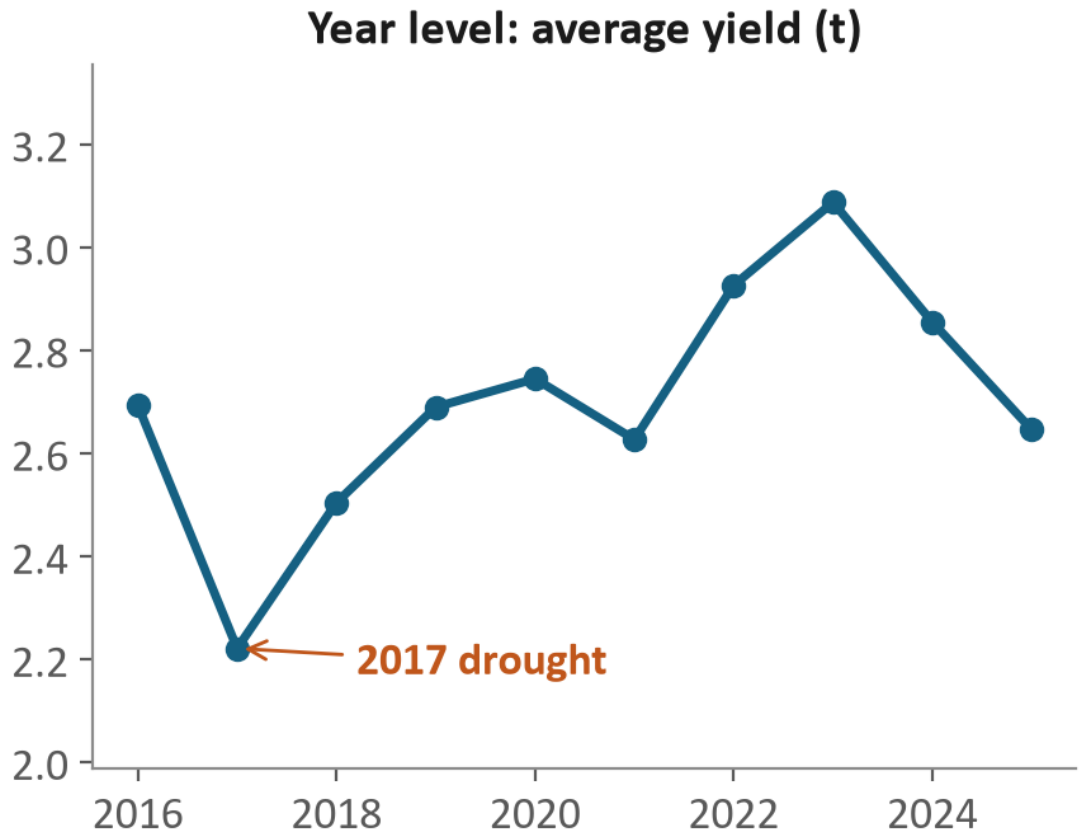


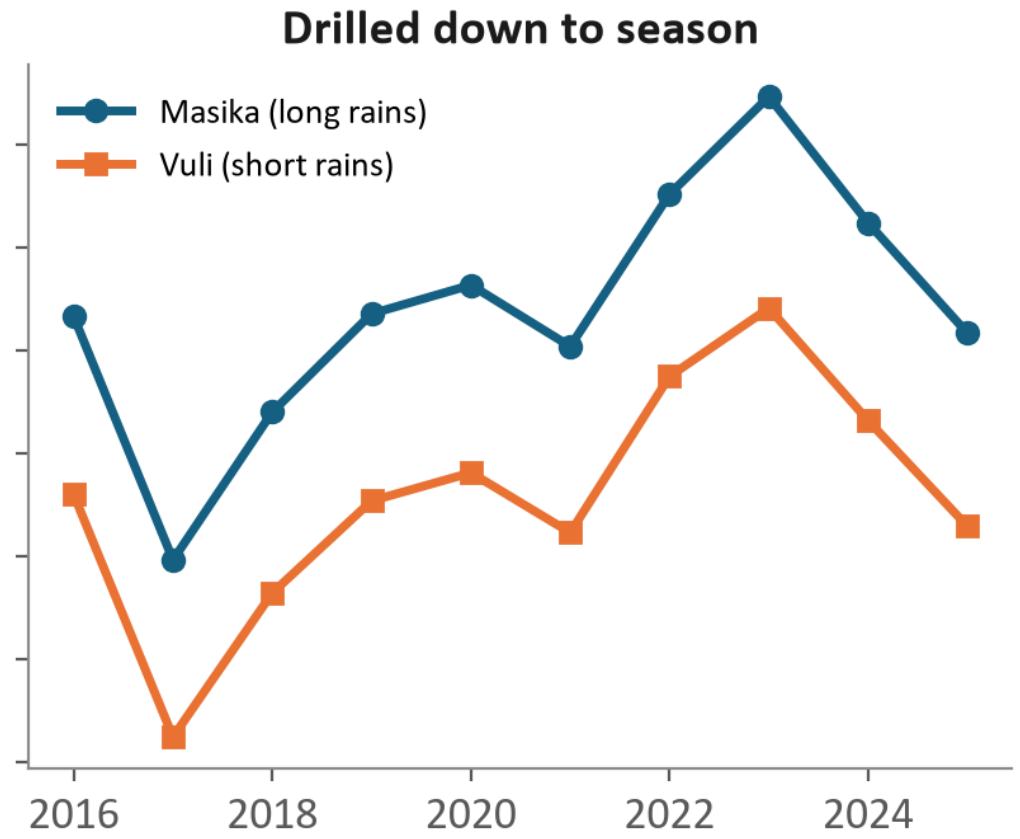
1. **region × crop**
   25 cells
2. **drop crop → region**
   5 cells, e.g. Pemba South 245.6 $M
3. **drop region → ALL**
   1 cell: 986 $M in total

Two kinds of roll-up: climb a hierarchy (district → region) or drop a dimension (region × crop → region).

Drill-Down: Find the Story Behind a Number

IIT Madras Zanzibar





1 **Spot**
2017 average yield 2.22 t vs ≈ 2.7 t normally

2 **Drill down**
year → season: BOTH seasons fell

3 **Explain**
island-wide drought, not a one-season accident

OLAP Operations: Moving Through the Cube

IIT Madras Zanzibar

Foundations: What OLAP Is

Roll-Up and Drill-Down

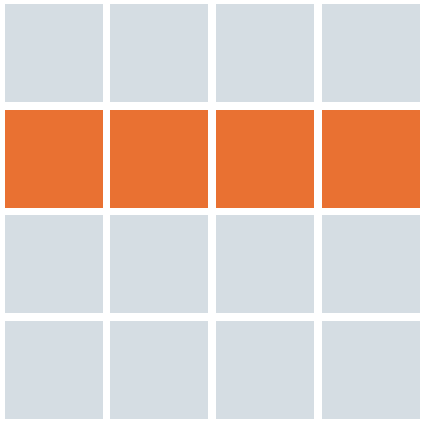
→ **Slice, Dice and Pivot**

OLAP in SQL and pandas

Indexing for Fast OLAP

In-Class Exercise

Slice: Fix One Dimension to One Value



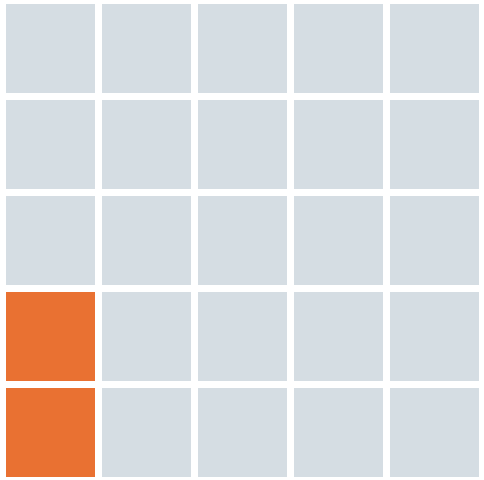
one layer:
year = 2017

2017 avg yield (t)	Clove	Cashew	Rice	Coconut	Cassava
North Unguja	1.58	1.03	2.26	2.82	5.09
South Unguja	1.56	1.04	2.19	2.81	4.98
Zanzibar Urban	1.56	1.01	2.21	2.81	5.15
Pemba North	1.53	1.03	2.15	2.75	5.01
Pemba South	1.56	1.01	2.25	2.71	5.23

```
SELECT region, crop, AVG(yield_tonnes) FROM farm_seasons WHERE year = 2017 GROUP BY region, crop;
```

Read it: every crop is low in EVERY region (e.g. clove ≈ 1.56 t vs 1.85 t normally) — the drought was island-wide.

Dice: Cut Out a Sub-Cube



regions: Pemba N & S
crop: Clove · years 2021–25

Clove revenue ($M)	2021	2022	2023	2024	2025
Pemba North	12.3	14.0	14.8	14.1	13.3
Pemba South	19.5	22.0	23.2	21.8	20.8

1 **Rows selected**
11,862 of 200,000

2 **Revenue**
175.8 $M

```
WHERE region IN ('Pemba North','Pemba South')
  AND crop = 'Clove' AND year BETWEEN 2021 AND 2025
```

Slice vs dice: slice fixes ONE dimension to ONE value; dice restricts SEVERAL dimensions to sets or ranges.

Pivot: Rotate the View

IIT Madras Zanzibar

Crops as rows — compare crops

crop \ year	2021	2022	2023	2024	2025
Clove	73.4	82.4	87.9	83.0	79.6
Cashew	9.9	11.3	11.8	11.4	10.8
Rice	6.2	7.2	7.7	7.2	6.8
Coconut	4.0	4.5	4.9	4.6	4.4
Cassava	3.6	4.1	4.4	4.2	3.9

Years as rows — read the trend

year \ crop	Clove	Cashew	Rice	Coconut	Cassava
2021	73.4	9.9	6.2	4.0	3.6
2022	82.4	11.3	7.2	4.5	4.1
2023	87.9	11.8	7.7	4.9	4.4
2024	83.0	11.4	7.2	4.6	4.2
2025	79.6	10.8	6.8	4.4	3.9

Pivot: swap rows and columns. Same numbers ($ millions of revenue, 2021–25), nothing recomputed — only the reading changes. Excel's PivotTable does exactly this.

The Five Operations summary

IIT Madras Zanzibar

Operation	What it does	Our example	SQL idea
Roll-up	up a hierarchy / drop a dimension	district → region	GROUP BY the higher level
Drill-down	down a hierarchy / add a dimension	year → season	GROUP BY the lower level
Slice	one dimension = one value	year = 2017	WHERE on one dimension
Dice	several dimensions restricted	Pemba × Clove × 2021–25	WHERE on several dimensions
Pivot	swap rows and columns	crop × year ↔ year × crop	change the display only

Workflow: roll up to spot something odd → drill down to explain it → slice or dice to focus → pivot to present.

OLAP Operations: Moving Through the Cube

IIT Madras Zanzibar

Foundations: What OLAP Is

Roll-Up and Drill-Down

Slice, Dice and Pivot

→ **OLAP in SQL and pandas**

Indexing for Fast OLAP

In-Class Exercise

OLAP in SQL: GROUP BY ROLLUP and CUBE

IIT Madras Zanzibar

```
SELECT region, district, SUM(revenue_usd)
FROM farm_seasons
GROUP BY ROLLUP(region, district);
```

region	district	SUM ($M)
Pemba South	Chake Chake	123.0
Pemba South	Mkoani	122.5
Pemba South	**ALL**	**245.6**
…	…	…
ALL	**ALL**	**986**

```
SELECT region, crop, SUM(revenue_usd)
FROM farm_seasons
GROUP BY CUBE(region, crop);
```

1. **region × crop**
   25 rows
2. **region, crop = ALL**
   5 rows
3. **crop, region = ALL**
   5 rows
4. **ALL, ALL**
   1 row: grand total

Difference: ROLLUP follows a hierarchy (n + 1 groupings); CUBE computes every combination (2^n groupings).

OLAP in pandas (Tutorial 4 Preview)

IIT Madras Zanzibar

```
# roll-up and drill-down
df.groupby('region')['revenue_usd'].sum()
df.groupby(['year', 'season'])['yield_tonnes'].mean()

# slice and dice
df[df.year == 2017]
df[df.region.isin(['Pemba North', 'Pemba South'])
   & (df.crop == 'Clove') & df.year.between(2021, 2025)]

# pivot, with ALL totals like ROLLUP
df.pivot_table(index='crop', columns='year', values='revenue_usd',
               aggfunc='sum', margins=True, margins_name='ALL')
```

This afternoon: you will run every one of these on all 200,000 rows in Tutorial 4.

OLAP in the Real World

IIT Madras Zanzibar

1 Mobile money / bank apps

monthly totals → expand to each transaction

2 Google Analytics

visitors by country → drill to city; slice by date

3 Power BI / Tableau

every click on a dashboard is an OLAP operation

4 DHIS2 health reports

national cases → drill down to district and facility

5 Supermarkets

slice one branch; dice a category × a week

6 University portal

results pivoted by semester or by course

Common thread: zoom out to see the pattern, zoom in to explain it.

OLAP Operations: Moving Through the Cube

IIT Madras Zanzibar

Foundations: What OLAP Is

Roll-Up and Drill-Down

Slice, Dice and Pivot

OLAP in SQL and pandas

→ **Indexing for Fast OLAP**

In-Class Exercise

Why Indexes? Scanning 200,000 Rows Every Click

IIT Madras Zanzibar

Without an index

read all 200,000 rows for every click

a real warehouse: billions of rows → minutes per click

With an index

jump straight to the matching rows

like the index at the back of a textbook

OLAP columns are low-cardinality: only 5 regions, 5 crops, 2 seasons, 10 years. For such columns the best shortcut is a BITMAP index.

Bitmap Index: One Row of Bits per Value

IIT Madras Zanzibar

row	crop	region	Clove bit	Pemba S. bit	AND
1	Clove	Pemba South	1	1	**1**
2	Rice	Pemba South	0	1	**0**
3	Clove	North Unguja	1	0	**0**
4	Clove	Pemba South	1	1	**1**
5	Cassava	South Unguja	0	0	**0**
6	Cashew	Pemba North	0	0	**0**
7	Clove	Pemba South	1	1	**1**
8	Coconut	Pemba North	0	0	**0**

1 **Clove bitmap**
1 0 1 1 0 0 1 0

2 **Pemba South bitmap**
1 1 0 1 0 0 1 0

3 **Bitwise AND**
1 0 0 1 0 0 1 0 → rows 1, 4, 7

Size: 200,000 rows = 200,000 bits = 25 kB per value; all 5 crops = 125 kB. No row is read — only bits are ANDed, 64 at a time.

ROLAP, MOLAP, HOLAP: Where the Cube Lives

IIT Madras Zanzibar

Type	How the cube is stored	Strength	Weakness
ROLAP	star-schema tables; cube answered by SQL	scales to huge data	slower per query
MOLAP	pre-computed cells in an array engine	very fast answers	large; slow to refresh
HOLAP	summaries in MOLAP, details in ROLAP	best of both	more complex

Tomorrow: MOLAP's speed comes from PRE-COMPUTING cuboids — but which ones? That is cube computation at scale.

OLAP Operations: Moving Through the Cube

IIT Madras Zanzibar

Foundations: What OLAP Is

Roll-Up and Drill-Down

Slice, Dice and Pivot

OLAP in SQL and pandas

Indexing for Fast OLAP

→ **In-Class Exercise**

In-Class Exercise (10 minutes, pairs)

IIT Madras Zanzibar

IN-CLASS TASK

Cube: farm_seasons with dimensions location (district → region), time (season → year), crop; measure SUM(revenue_usd).

1. Name the operation: (a) region totals → district totals; (b) only 2023; (c) swap crops and years; (d) Pemba × Clove × 2021–25.

2. Write the pandas line for: average yield per region in the Vuli season of 2017.

3. Rows: crops Rice, Clove, Rice, Cashew. Write the Rice bitmap and the bitmap for NOT Rice.

4. A dashboard's monthly total is wrong only in March. Which two operations do you use to find the cause, in which order?

Solution

IIT Madras Zanzibar

1 **Q1**
(a) drill-down (b) slice (c) pivot (d) dice

2 **Q2**
df[(df.year == 2017) & (df.season == 'Vuli')].groupby('region')['yield_tonnes'].mean()

3 **Q3**
Rice = 1 0 1 0; NOT Rice = 0 1 0 1 (flip every bit)

4 **Q4**
slice to March, then drill down (month → day → branch/transaction) until the faulty rows appear

Solution Q3: Rice Bitmap and NOT Rice

IIT Madras Zanzibar

row	crop	Rice bit	Clove bit	Cashew bit	NOT Rice
1	Rice	**1**	0	0	**0**
2	Clove	**0**	1	0	**1**
3	Rice	**1**	0	0	**0**
4	Cashew	**0**	0	1	**1**

1 **Rice bitmap**
1 0 1 0

2 **NOT Rice bitmap (flip every bit)**
0 1 0 1

3 **Check: Rice OR NOT Rice**
1 1 1 1 → every row is covered once

Rule: One bit per row per value. NOT Rice flips every bit, and each row has exactly one 1 across the crop bitmaps.

Summary & Exam Checkpoints

IIT Madras Zanzibar

Roll-up / drill-down

Up or down a hierarchy (or drop / add a dimension). Roll up to spot, drill down to explain.

Slice / dice / pivot

Slice: one dimension = one value. Dice: several restricted. Pivot: rotate the view.

ROLLUP and CUBE

ROLLUP follows a hierarchy (n + 1 groupings); CUBE gives all 2^n. pandas: groupby, pivot_table(margins=True).

Bitmap indexes

One bit per row per value; answer AND / OR queries with bit operations. Ideal for low-cardinality columns.

Common Questions

IIT Madras Zanzibar

Q **Is a slice just a filter?**

A Yes — a filter that fixes one dimension to one value. The name comes from cutting one layer off the cube.

Q **Does pivot change any numbers?**

A No. It only swaps rows and columns; the cells are identical.

Q **Why not index every column with a normal B-tree index?**

A B-trees suit many distinct values (IDs). For 5 crops or 2 seasons, bitmaps are smaller and combine faster.

Q **Can you drill down below the fact table?**

A No — the fact table (one farm in one season) is the finest detail the cube stores.

What's Next

IIT Madras Zanzibar

Tut **Tutorial 4 (today, 4:45)**
star schema and all five OLAP operations on 200,000 rows in pandas

Fri **Cube computation at scale**
which cuboids to pre-compute; iceberg cubes; the bridge to pattern mining

Wk 5 **Frequent pattern mining**
Apriori and FP-Growth on 120,000 farm-supply baskets

References & Further Reading

- Han, J., Pei, J. & Tong, H. (2023). Data Mining: Concepts and Techniques, 4th ed. — Chapter 3 (OLAP operations, indexing, OLAP servers).
- Chaudhuri, S. & Dayal, U. (1997). An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26(1), 65–74.
- O'Neil, P. & Quass, D. (1997). Improved query performance with variant indexes. Proc. ACM SIGMOD, 38–49. (Bitmap indexes.)