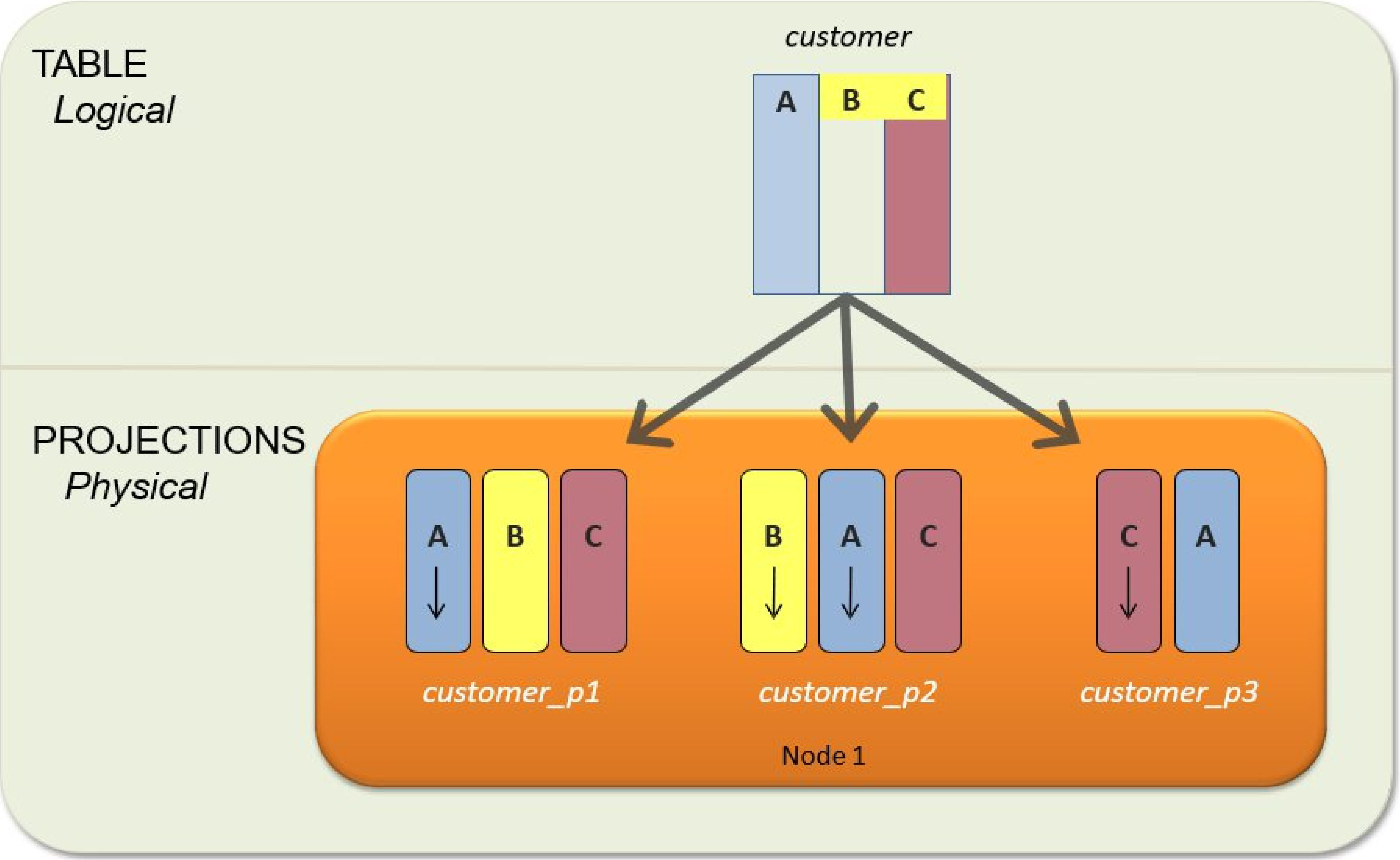
VERTICA
by opentext™

Vertica

Logical and Physical Representations

Logical and Physical Representations



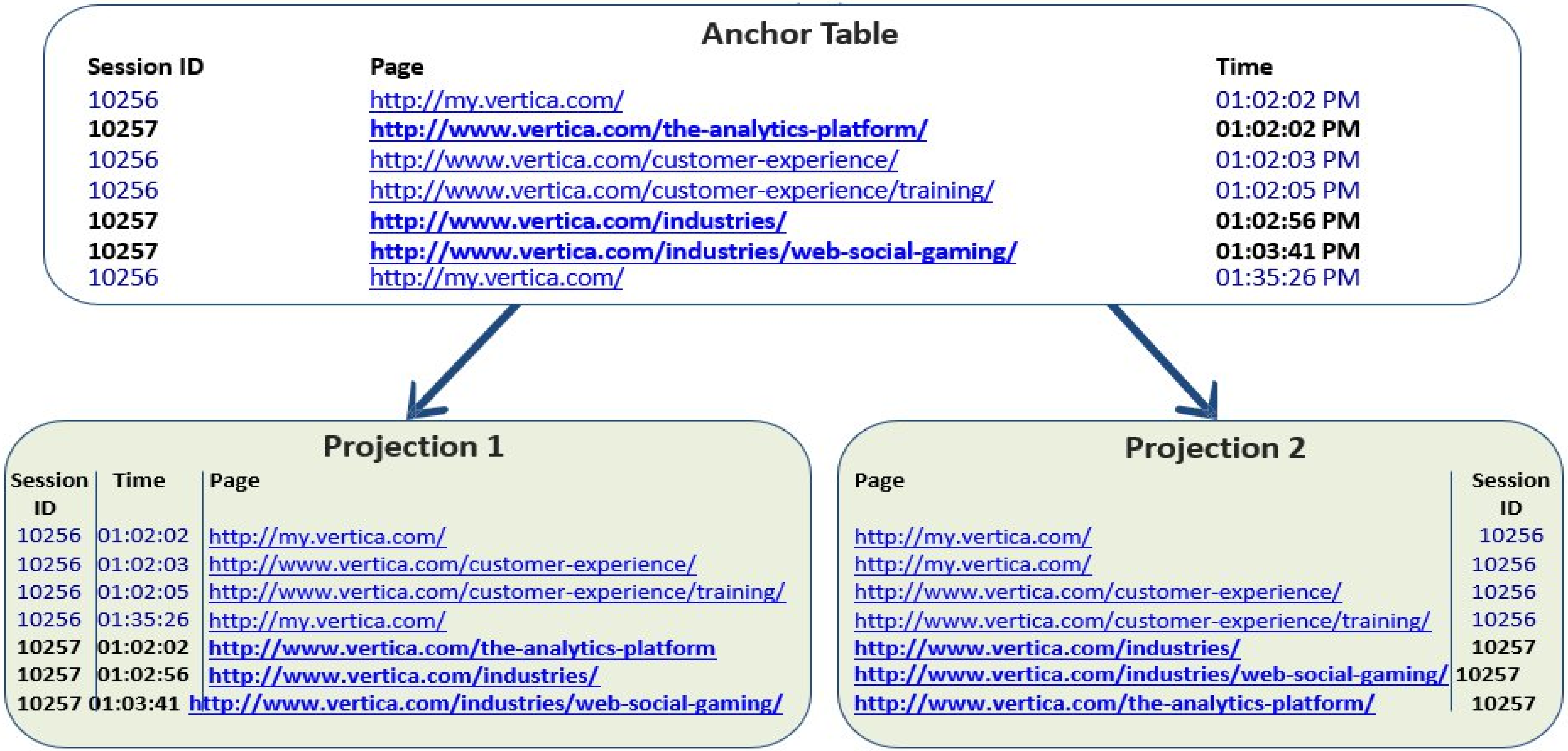
Tables

- Tables are logical objects in Vertica
- Queries reference tables
- Optimizer picks best projection(s) to answer queries
- Tables are owned by a user and created in a schema
- \d to see all tables

Types of Temporary Tables

- GLOBAL – default, DDL visible in all sessions
- LOCAL – DDL visible only in the session created

Tables versus Projections



Anchor Table

Session ID	Page	Time
10256	http://my.vertica.com/	01:02:02 PM
10257	**http://www.vertica.com/the-analytics-platform/**	**01:02:02 PM**
10256	http://www.vertica.com/customer-experience/	01:02:03 PM
10256	http://www.vertica.com/customer-experience/training/	01:02:05 PM
10257	**http://www.vertica.com/industries/**	**01:02:56 PM**
10257	**http://www.vertica.com/industries/web-social-gaming/**	**01:03:41 PM**
10256	http://my.vertica.com/	01:35:26 PM

Projection 1

Session ID	Time	Page
10256	01:02:02	http://my.vertica.com/
10256	01:02:03	http://www.vertica.com/customer-experience/
10256	01:02:05	http://www.vertica.com/customer-experience/training/
10256	01:35:26	http://my.vertica.com/
10257	**01:02:02**	**http://www.vertica.com/the-analytics-platform**
10257	**01:02:56**	**http://www.vertica.com/industries/**
10257	**01:03:41**	**http://www.vertica.com/industries/web-social-gaming/**

Projection 2

Page	Session ID
http://my.vertica.com/	10256
http://my.vertica.com/	10256
http://www.vertica.com/customer-experience/	10256
http://www.vertica.com/customer-experience/training/	10256
http://www.vertica.com/industries/	**10257**
http://www.vertica.com/industries/web-social-gaming/	**10257**
http://www.vertica.com/the-analytics-platform/	**10257**

Projections are optimized for common query patterns

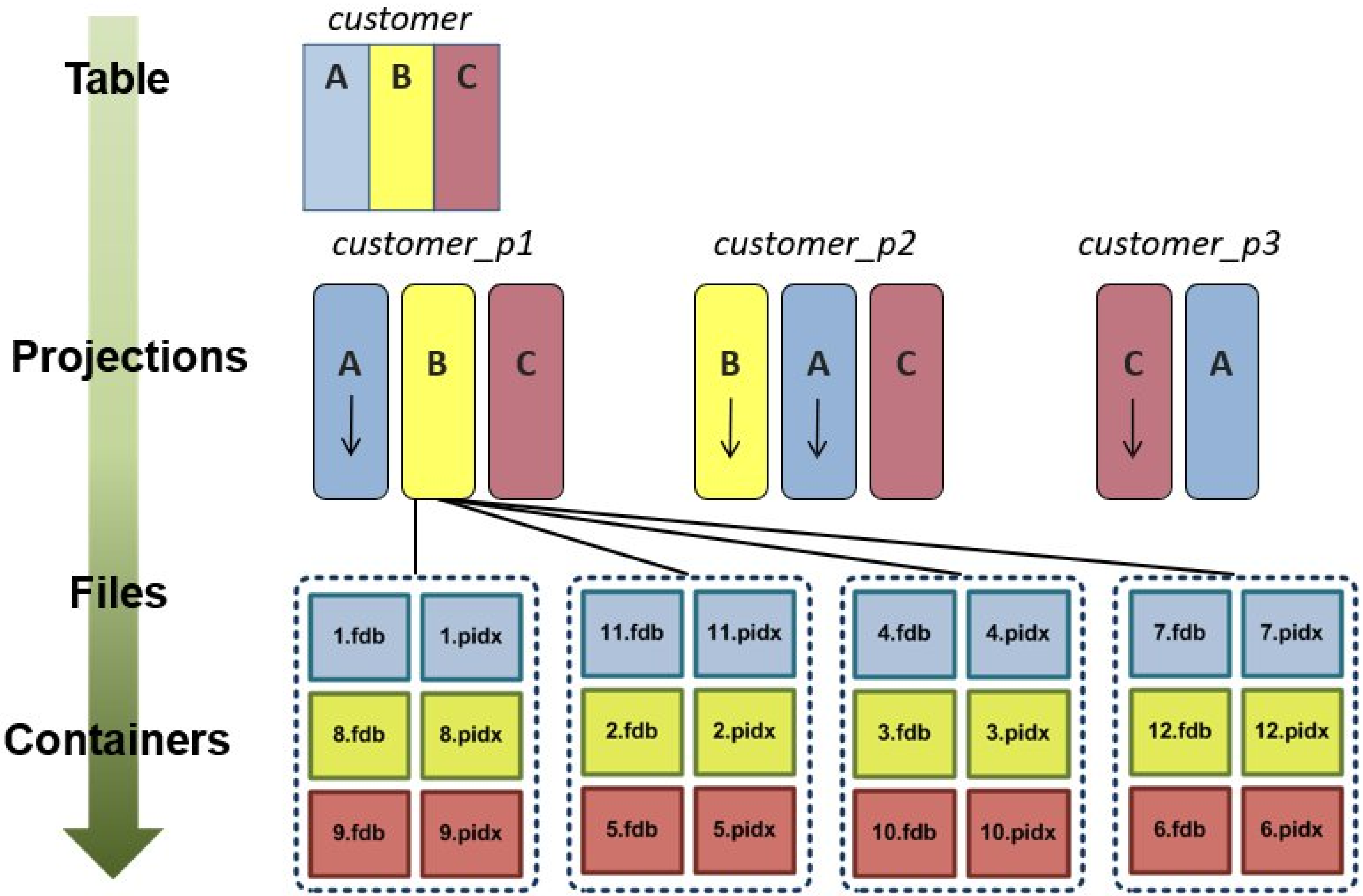
What is a Projection?

- Collection of table columns
- Stores data in a format to optimize query execution
- Similar in concept to materialized views

Projection Basics

- Anchor table is not stored
- Data is stored in a sorted and compressed format
- No need for indexing
- Transparent to the end user and applications
- Created by Database Designer (DBD)
  - Can be manually created
- Best projection for a query is chosen by the Optimizer at query execution time

Vertica Object Hierarchy



opentext™

Projection Types

- Superprojection
  - Projection that contains all columns of a logical table
- Query-specific projection
  - Optimized for a specific query or class of queries
  - Can include a subset of columns
- Buddy projection
  - Projection with the same columns and segmentation on different nodes to provide high availability (HA)

How to Create Projections

- Automatically on first data load into a table
  - Unoptimized superprojections
- Automatically using the Database Designer
  - Optimized superprojections
  - Possible query-specific projections
- Manually using CREATE PROJECTION syntax in a vsql statement

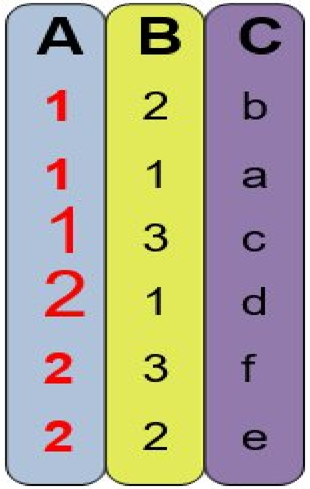
Projection Properties

- Encoding/Compression
- Sort Order
- Replication or Segmentation
- K-safety

Encoding/Compression and Sorting

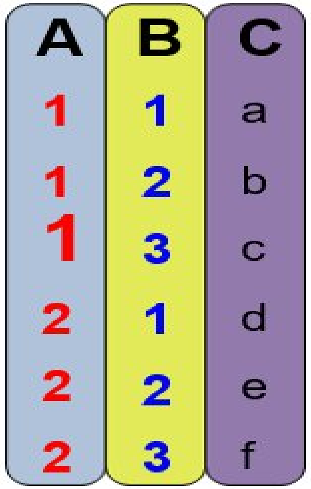
- Each column is always encoded, compressed, or both
- Vertica can work directly with encoded data; compressed data must first be uncompressed
- All projections contain at least one column in the ORDER BY statement

ORDER BY A



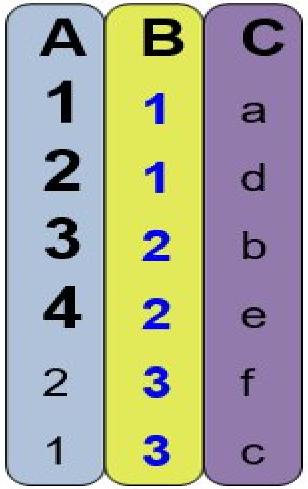
A	B	C
1	2	b
1	1	a
1	3	c
2	1	d
2	3	f
2	2	e

ORDER BY AB



A	B	C
1	1	a
1	2	b
1	3	c
2	1	d
2	2	e
2	3	f

ORDER BY B

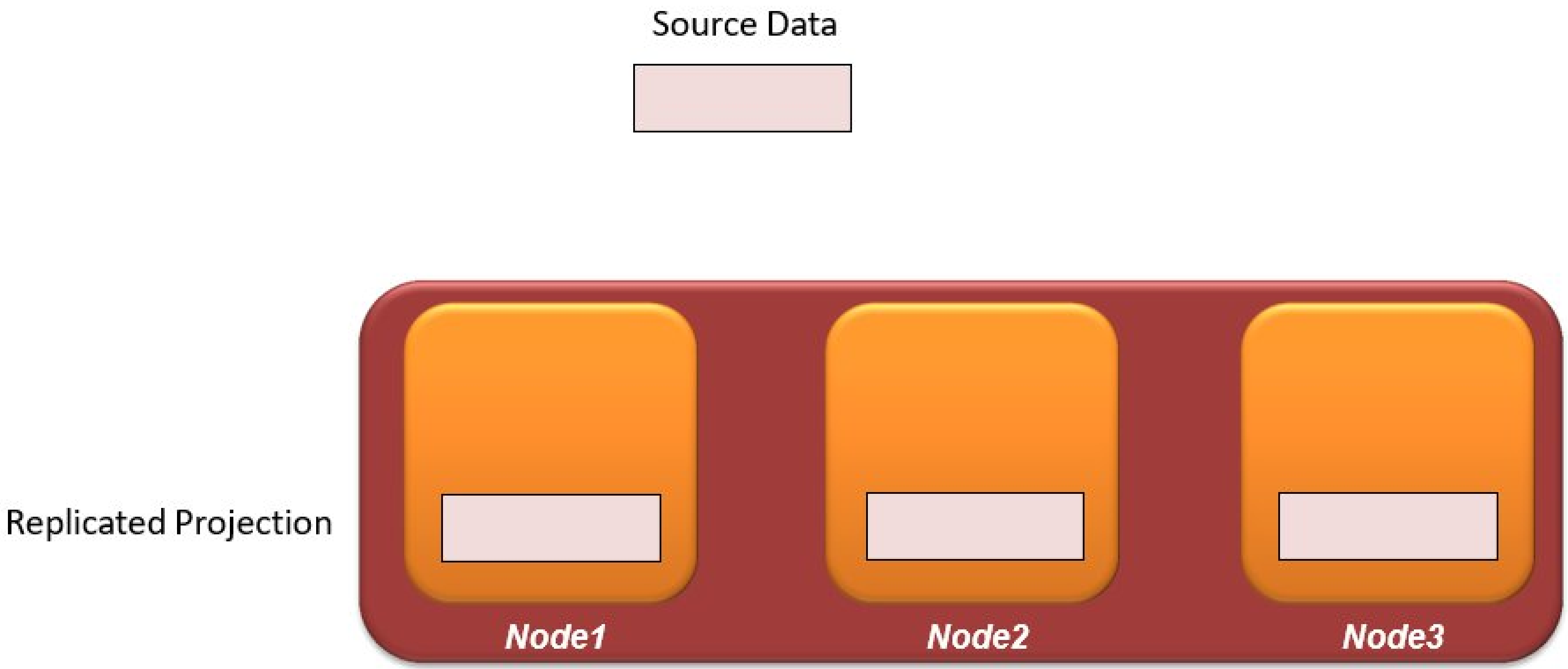


A	B	C
1	1	a
2	1	d
3	2	b
4	2	e
2	3	f
1	3	c

opentext™

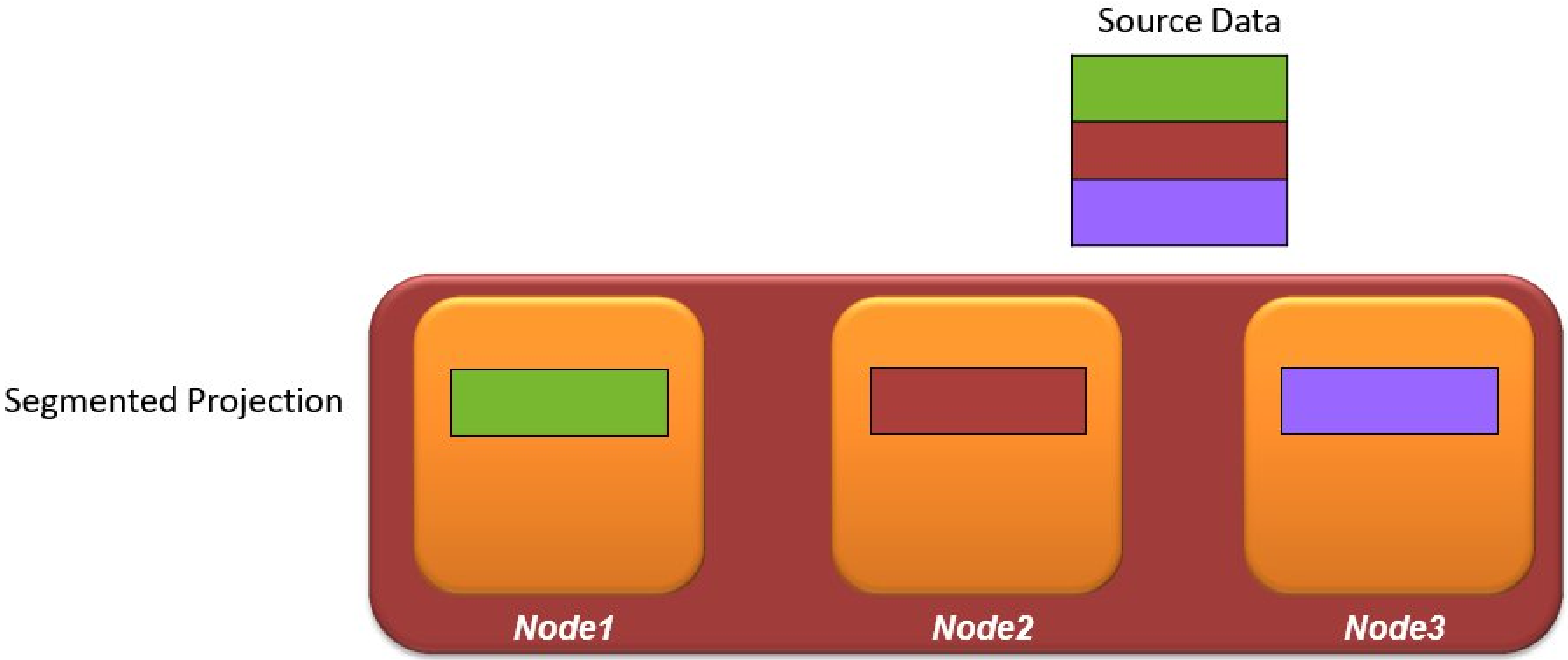
Replication

- For a small projection, copy the full projection to each node
  - Inherently provides high availability of the projection



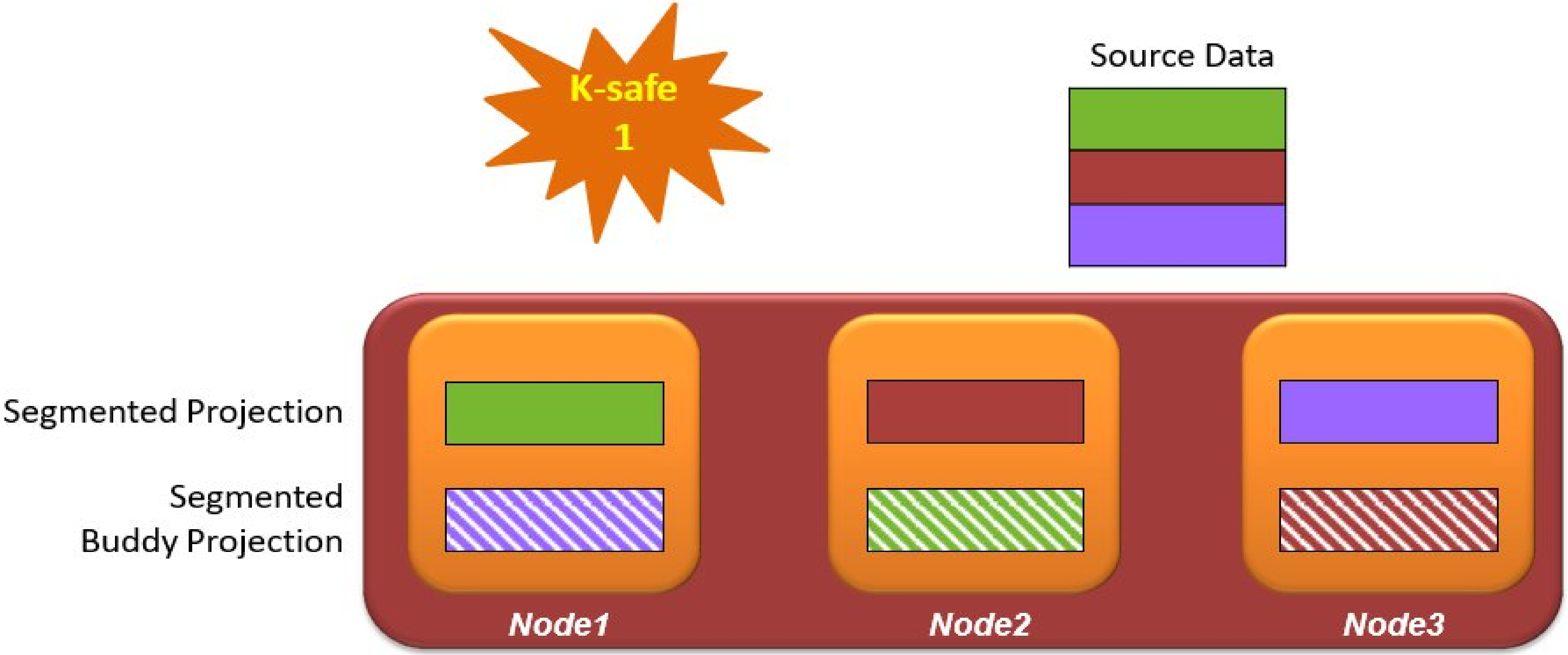
Segmentation

- For a large projection, distribute projection data across multiple nodes
  - Segmenting the projection splits the data into subsets of rows
  - Design for random data distribution



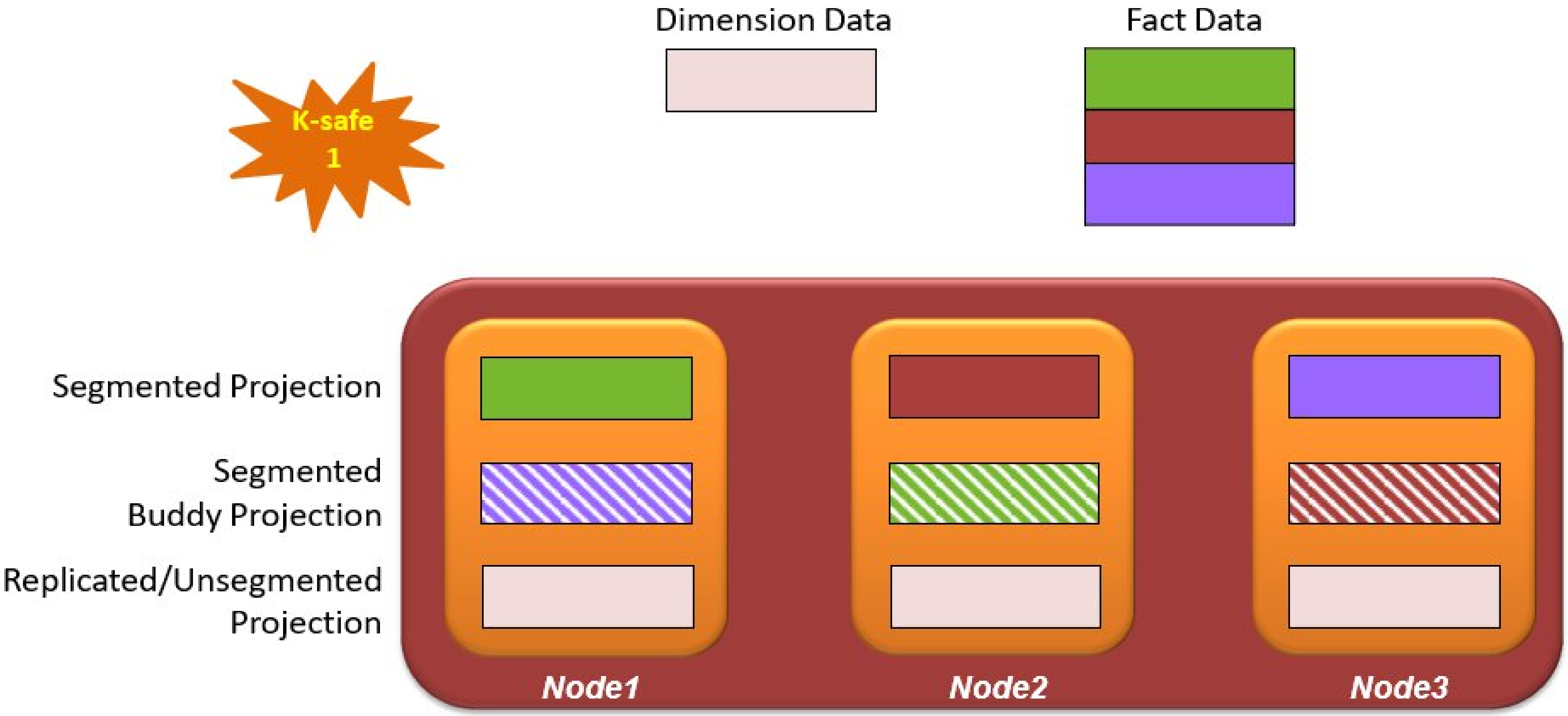
Segmentation and High Availability

- Buddy projections provide high availability for segmented projections
  - Buddy projection is a copy of a projection segment that is stored on a different node
  - K-safety is the number of replicas of data that exist on the cluster

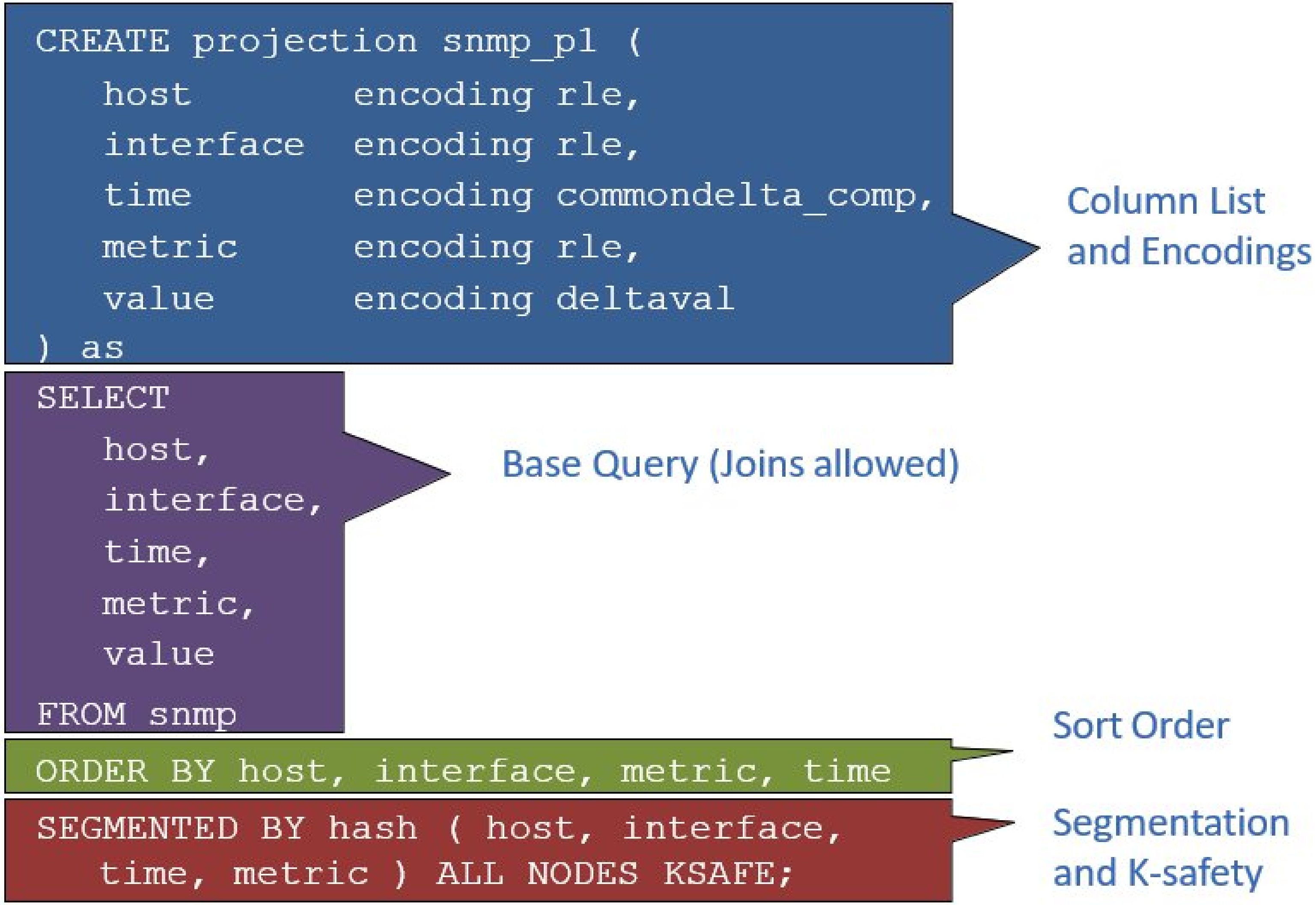


Segmentation, Replication, K-safety

- Segmentation with buddy projections
- Replication



Create Projection DDL



```
CREATE projection snmp_p1 (
   host        encoding rle,
   interface   encoding rle,
   time        encoding commondelta_comp,
   metric      encoding rle,
   value       encoding deltaval
) as
SELECT
   host,
   interface,
   time,
   metric,
   value
FROM snmp
ORDER BY host, interface, metric, time
SEGMENTED BY hash ( host, interface,
   time, metric ) ALL NODES KSAFE;
```

Column List
and Encodings

Base Query (Joins allowed)

Sort Order

Segmentation
and K-safety

Projection Basics: Maintenance

- Data is loaded directly into projections
- No need to rebuild or refresh projections after the initial refresh
- New projections can be created at any time, either manually or by running the Database Designer
- Old projections can be dropped

opentext™

Thank you