

Common Interface for Every Kind of Astronomical Data Service

Yuji Shirasaki

National Astronomical Observatory of Japan

JVO

VOQL+DAL Joint session

JVO Skynode Implementation Experience

Yuji SHIRASAKI

National Astronomical Observatory of Japan/JVO

yuji.shirasaki @ jvo.nao.ac.jp

Under Preparation **JVO**

VOQL Joint session



Contents

- Objective of this talk
- Concept for adapting SkyNode Interface on DAL service
- JVO SkyNode experimental implementation
 - SkyNode Toolkit
 - Application to the science use case
- Required Specification
 - Data Service Classification and Table Data model
 - Definition of minimum subset of ADQL
 - Minor update on ADQL
 - Standard on usage of VOTable
 - Access to the Object data type
 - Metadata access interface
 - Standardization of error message

Contents

- Introduction of JVO SkyNode toolkit

- used free software:
- architecture of JVO skynode

- Implementation problems

- XML → Java deserialization problem in AXIS
- Namespace problem of ADQL → JVO (v0.8) vs NVO (v0.74)
- Namespace problem of VOTable → JVO (v1.1) vs NVO (v1.0 or no namespace)
- Usage of VOTable → id, name attributes ...
- Complexity of STC object
- ...

- Proposal

- Simplify the ADQL and STC → Define minimum subset of ADQL and STC and freeze them (never update, never change the namespace)
- VOTable transfer → attachment or URL
- Standardize the error message

Objective of this talk

- From a Simple Protocol to an Intelligent Protocol
 - SIAP and SSAP are simple parameter-based query protocol.
 - ADQL is a SQL-based query protocol and have a capability to describe more complicated query conditions.
- What is the merit ?
 - Multiple regions condition, directly or by VOTable
 - Homogeneous data access to both of the catalog and observational data
 - Query to the multiple data services (Portal)
- Effect on the current SIAP and SSAP spec.
 - Introducing the ADQL to the DAL service does not affect to the data model of a returned VOTable.
 - Just introduce an additional query interface to the DAL service.

Concept for adapting SQL to Image Database.

I have shown that image query can be described in SQL syntax by introducing virtual column concept.

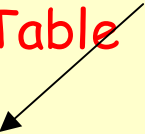
SIAP Query

POS [mandatory]
SIZE [mandatory]
FORMAT [mandatory]
INTERSECT [option]
...

```
http://jvo.nao.ac.jp/Image?  
Pos=34.3,-5.11&Size=0.01&Format=VOTable&  
INTERSECT=OVERLAPS
```

SIAP parameters and returned metadata are taken as columns of the virtual table.

Image Query on Virtual Table



Pos	Size	Format	ImageURL
(23,+30)	1.0	fits	http://jvo.nao.ac.jp/ Image?id=124214
(23,+23)	0.3	jpeg	http://jvo.nao.ac.jp/ Image?id=124215
...

```
Select imageURL  
From imageTable  
Where Pos = Point(23,+30)  
and Size = 1.0 and  
Format = 'fits'
```

This is a virtual table which has infinite number records.

Development of the JVO SkyNode Toolkit

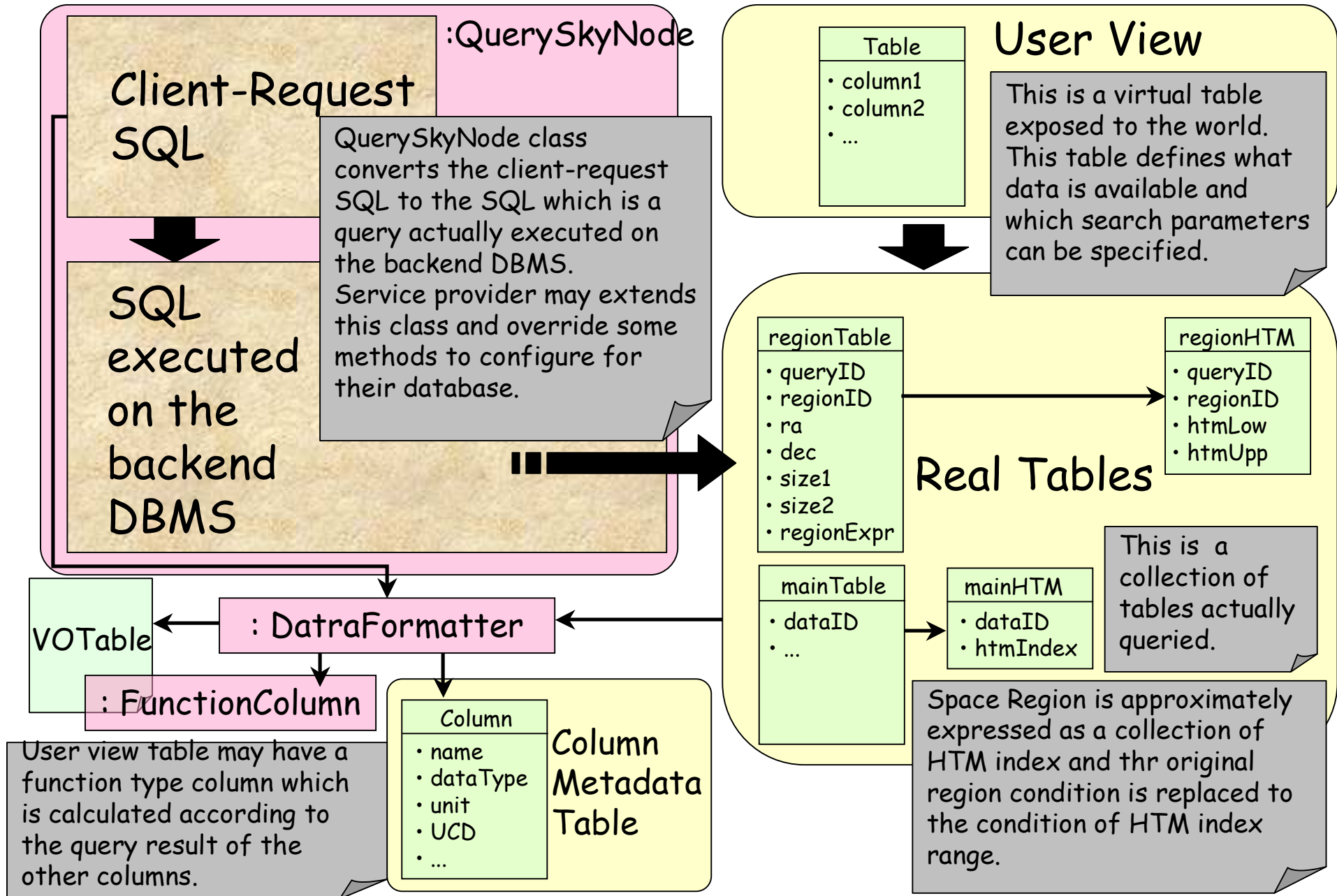
- **Primary aim:** to provide a reference implementation for a DAL service using SkyNode/ADQL/VOTable interface
- Independent from the type of backend DBMS (PostgreSQL, Oracle, MySQL, ...) JDBC driver is required.
- **Restrictions:**
 - Not all the ADQL syntax are supported.
 - String representation of ADQL is JVOQL.
- **Experimental Release:**
 - <http://jvo.nao.ac.jp/download/skynode-toolkit>

Software used

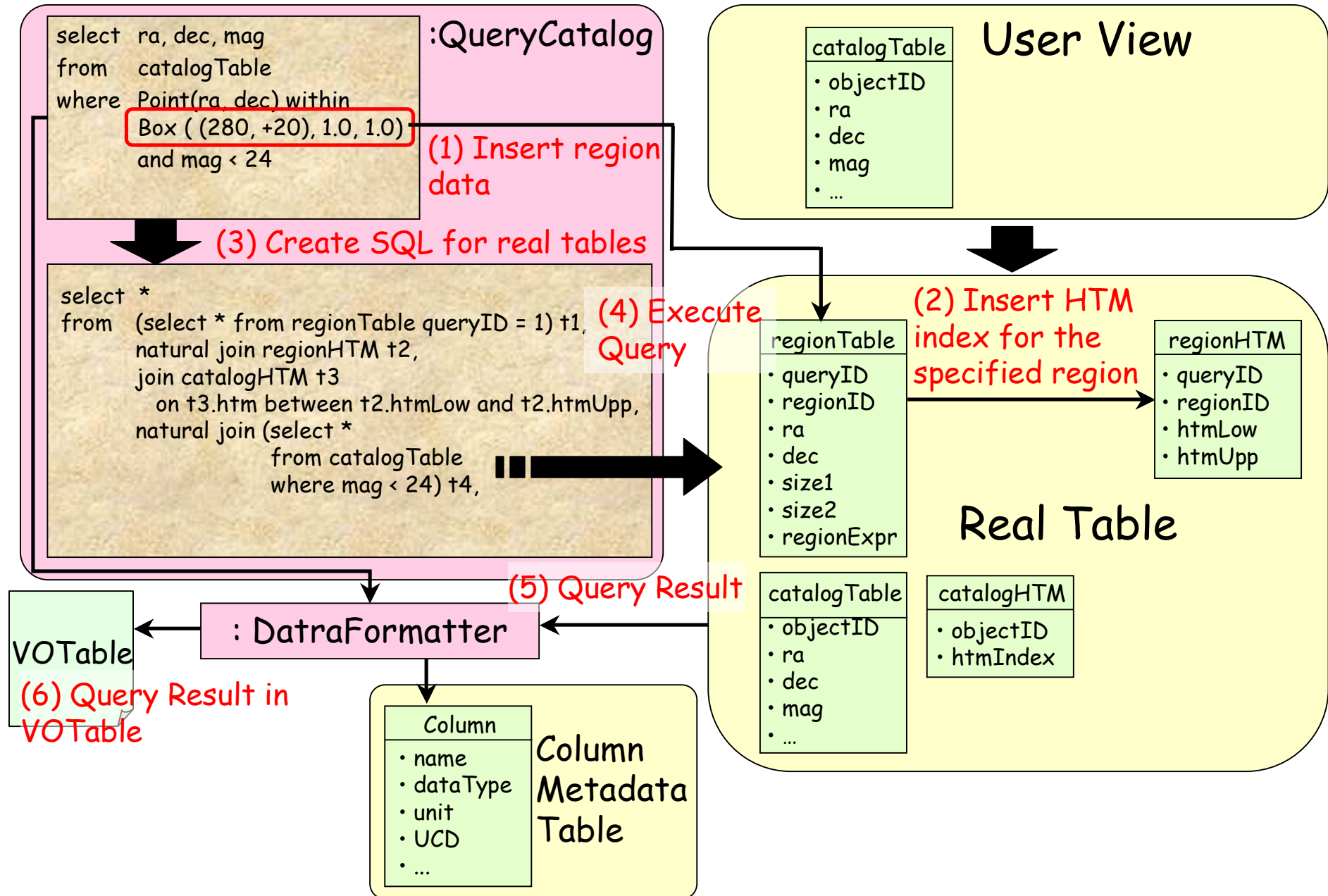
- Tomcat 4.1.31 ---- servlet container
- Axis 1.2RC1 ---- web service engine
- J2SDK 1.4.2 ---- Java compiler & library
- Ant 1.6.1 ---- Java-based build tool
- JavaCC 3.2 ---- parser generator for Java
- PostgreSQL 7.4.7 ---- DBMS
- Java HTM library (JHU) ---- spherical indexing
- Java FITS library (HEASARC) ---- FITS IO lib
- ...

Architecture

JVO SkyNode Toolkit Flow Chart



Catalog Data Query by ADQL



Catalog Data Xmatch Query with VOTable

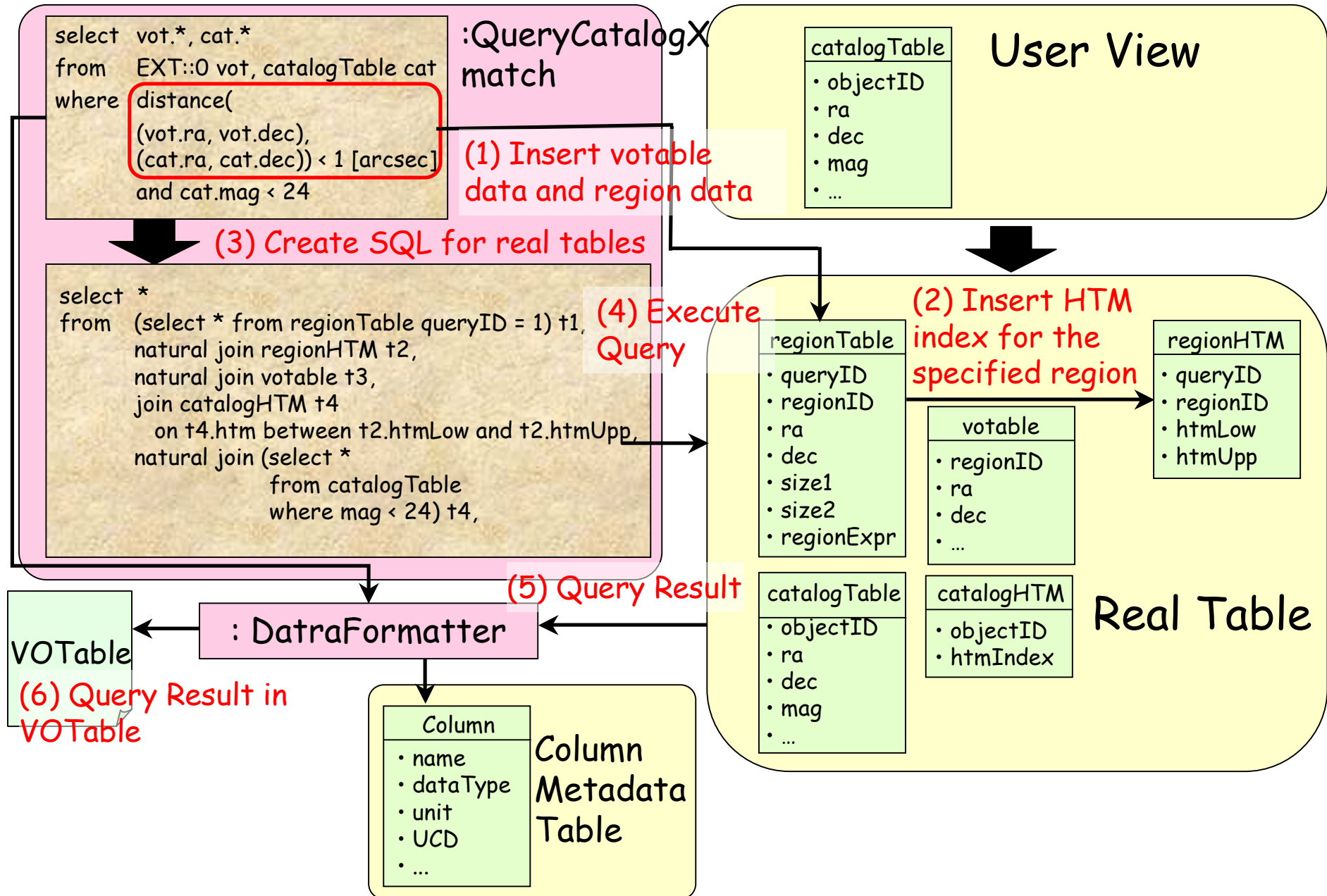


Image Data Query by ADQL

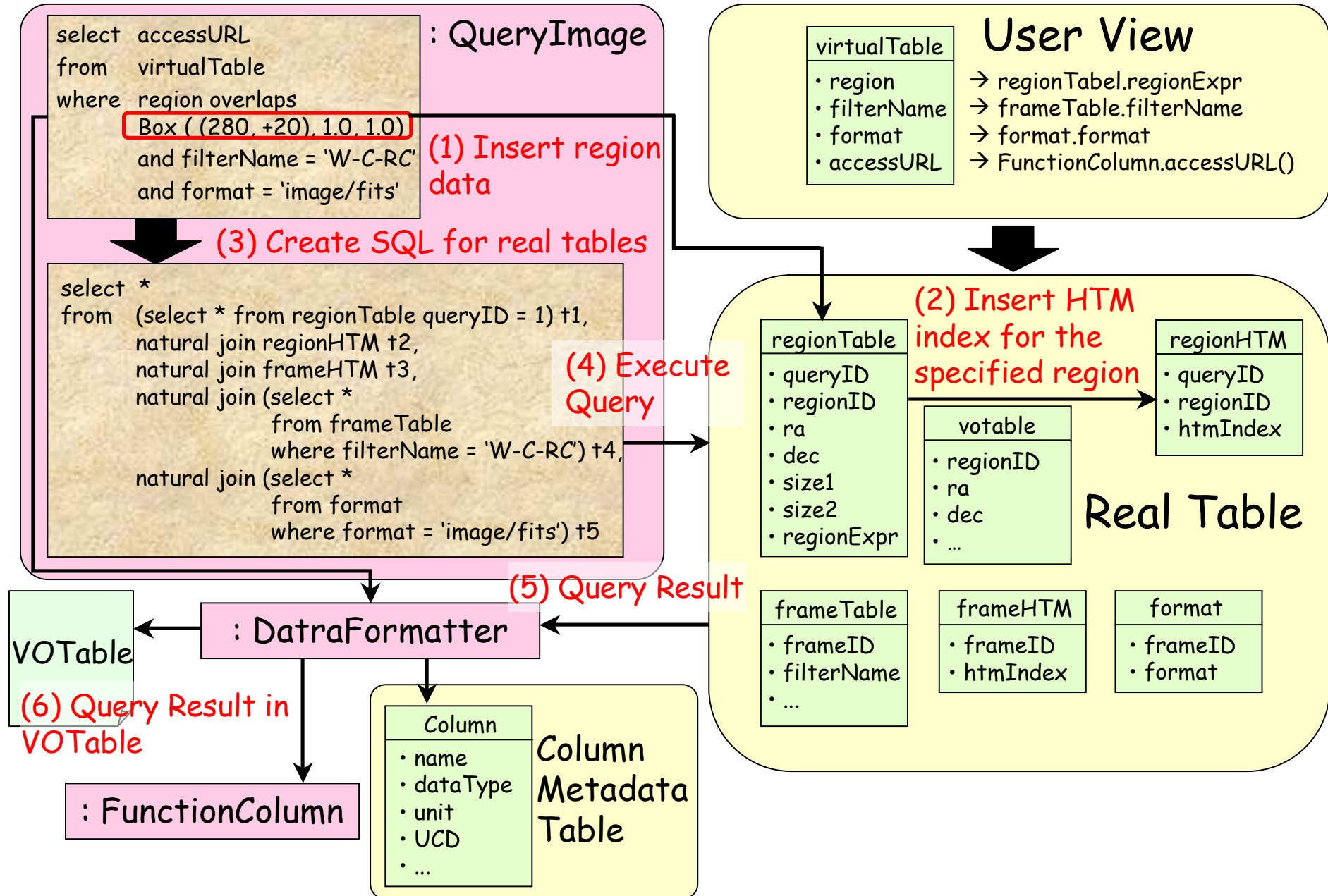
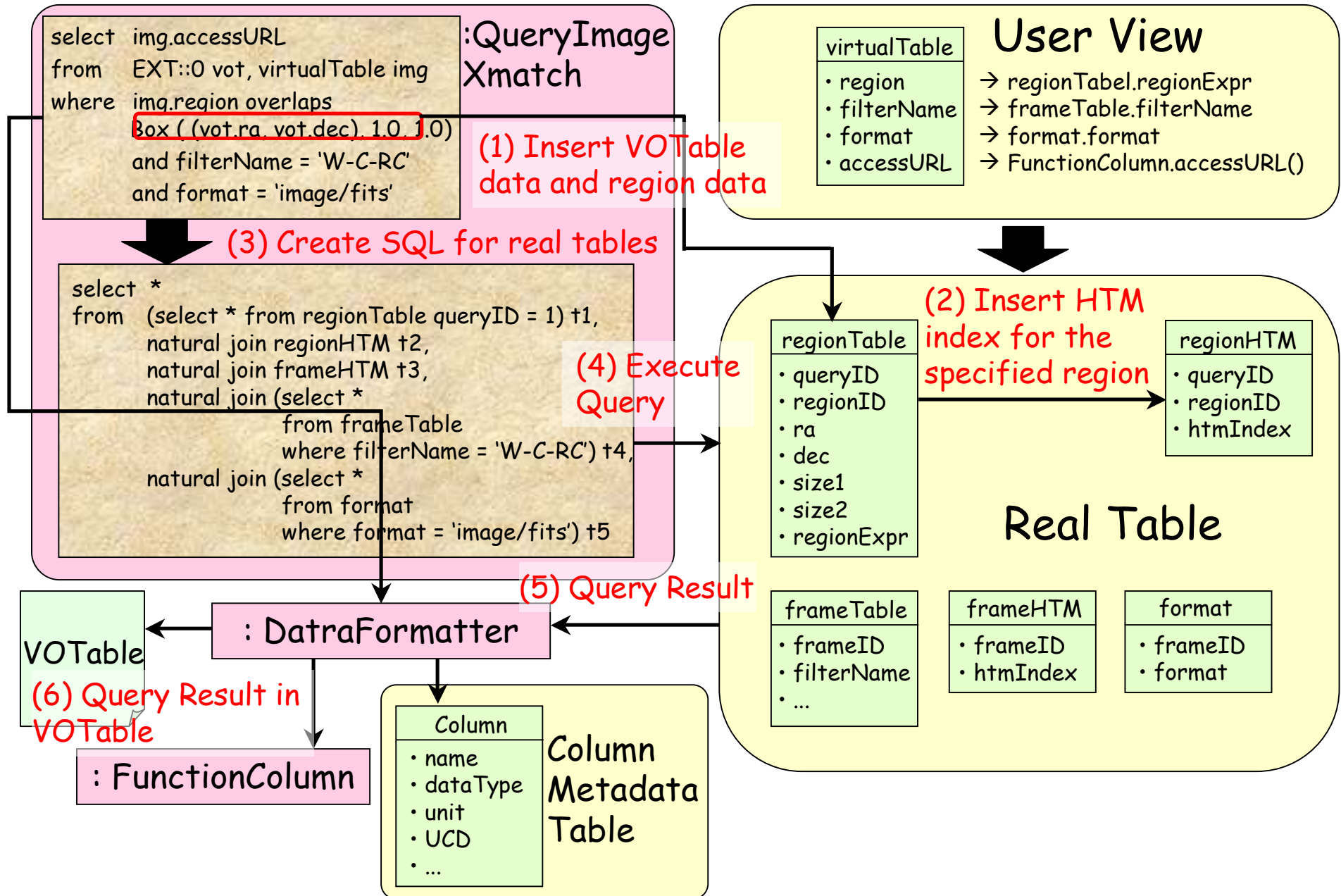
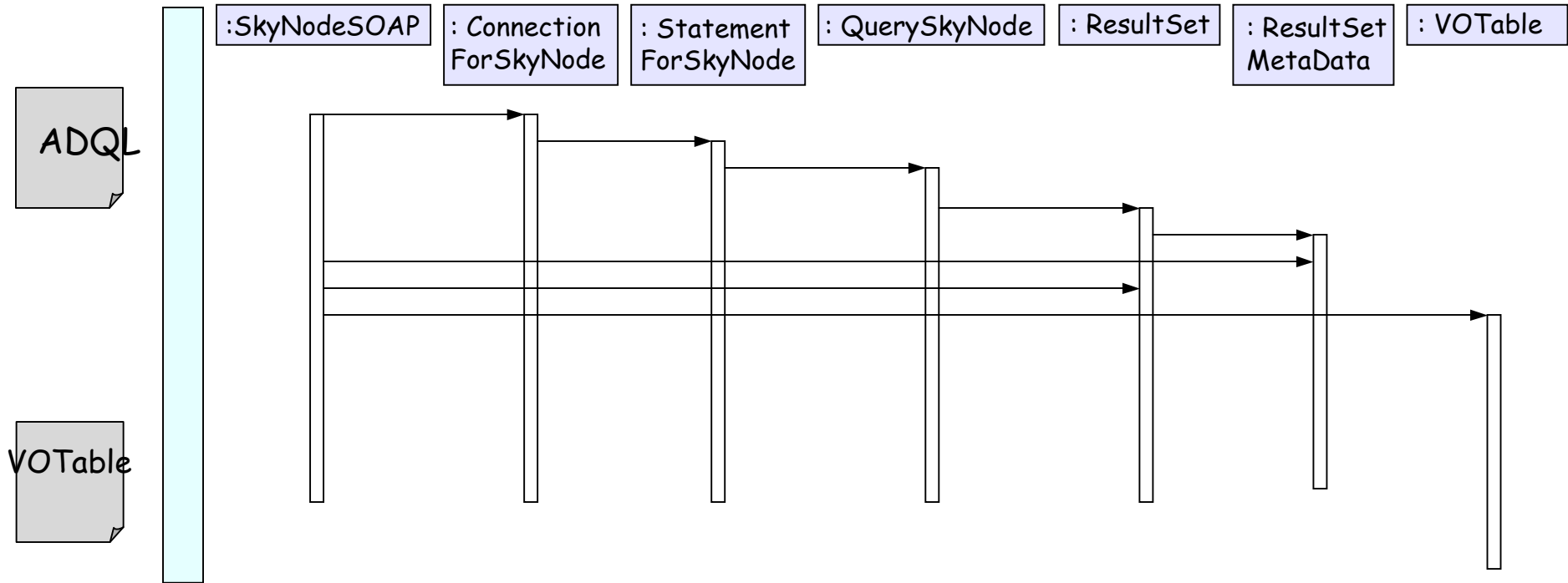


Image Data Xmatch Query with VOTable



Sequence Diagram

Axis engine



Science Use Case 2
QSO/Galaxy
Clustering Study

What we can learn from the QSO/Galaxy clustering ?

- Origin of the large scale structure

- QSO is a tracer of high density regions in the universe ← hierarchical clustering model
- Comparison between the observation and theoretical prediction is required.

- Origin of the QSO activity

- Why is the QSO so powerful.
- Test of the galaxy merger model

Work flow for studying the QSO/Galaxy clustering

1. Select QSO coordinates from the QSO catalog → Query to the Skynode Catalog Database.
2. Search deep imaging data which covers the QSO regions → Query to the Skynode of Subaru Image Database
3. Create catalog from the imaging data → Invoke the SExtractor Web service.
4. Estimate the distance to the objects around the QSO → Invoke the HyperZ Web service
5. Try Clustering Analysis → Invoke the clustering analysis web service.

```
SELECT cat.*, img.*
FROM jvo.misc.qso_veron_2003 cat,
     jvo.test.smoka.spcam_img img
WHERE POINT(cat.raj2000, cat.dej2000)
      WITHIN
      CIRCLE((189.20625, 62.216111), 0.10)
      AND cat.v_mag < 20
      AND img.format = 'image/fits'
      AND img.region OVERLAPS
      BOX((cat._raj2000, cat._dej2000), 0.02, 0.02)
```

```
SELECT cat.*
FROM jvo.misc.qso_veron_2003 cat
WHERE POINT(cat.raj2000, cat.dej2000)
      WITHIN
      CIRCLE((189.20625, 62.216111), 0.10)
      AND cat.v_mag < 20
```

SkyNode
QSO
Catalog
DB

SkyNode
Subaru
SuprimeCam
Image DB

JVO
Portal

(1)

(2)

(3)

(4)

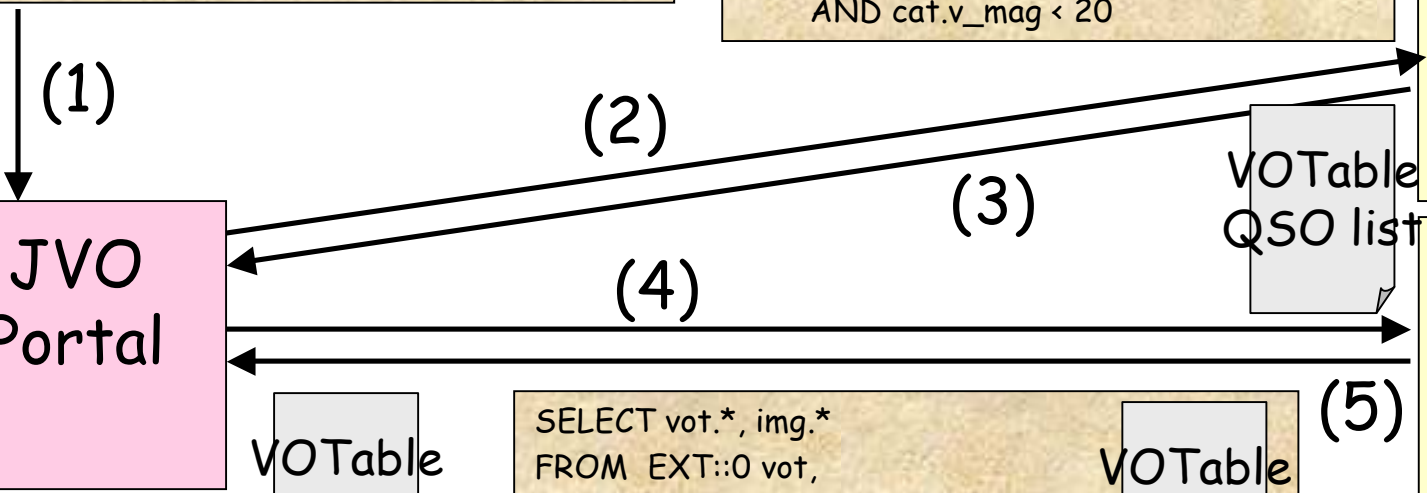
(5)

VOTable
QSO list

VOTable
QSO+Image

VOTable
QSO list

```
SELECT vot.*, img.*
FROM EXT:::0 vot,
     jvo.test.smoka.spcam_img img
WHERE img.format = 'image/fits'
      AND img.region OVERLAPS
      BOX((vot.raj2000, vot.dej2000), 0.02, 0.02)
```



QSO-Galaxies Search

[Status](#) | [Registry](#) | [Search](#) | [Result](#) | [Database](#) | [QSO Search](#) | [Image Viewer](#) | [Logout](#)

⇒ [Query](#) | [Catalog](#) | [PhotoZ](#)

Data Search

Search

ID for your query

Observation Name

Region

RA: [deg] Dec: [deg] Radius: [deg] Image Size: [arcmin]

Brightness

V_mag between [mag] and [mag]

Redshift

z between [mag] and [mag]

QSO Catalog Search

Image Data Service Search

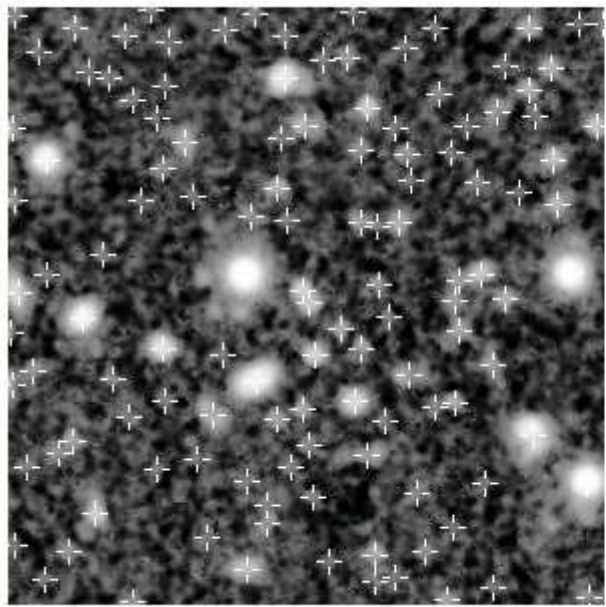
Search

User ID	User Name	Group	Last Login
yshirasa	Yuji Shirasaki	jvo	Sat Mar 12 21:36:49 JST 2005

Image Viewer

[Status](#) | [Registry](#) | [Search](#) | [Result](#) | [Database](#) | [QSO Search](#) | [Image Viewer](#) | [Logout](#)

Name	Origin	Scale	Contrast
image1	http://hete.mtk.nac	hist	min = 0.0 max = 65535.0 auto = true



Scale :

Contrast :

min =

max =

VOTable :

User ID	User Name	Group	Last Login
yshirasa	Yuji Shirasaki	jvo	Fri Mar 11 01:06:42 JST 2005

Total memory = 77684kB Used momory = 62095kB (79%)

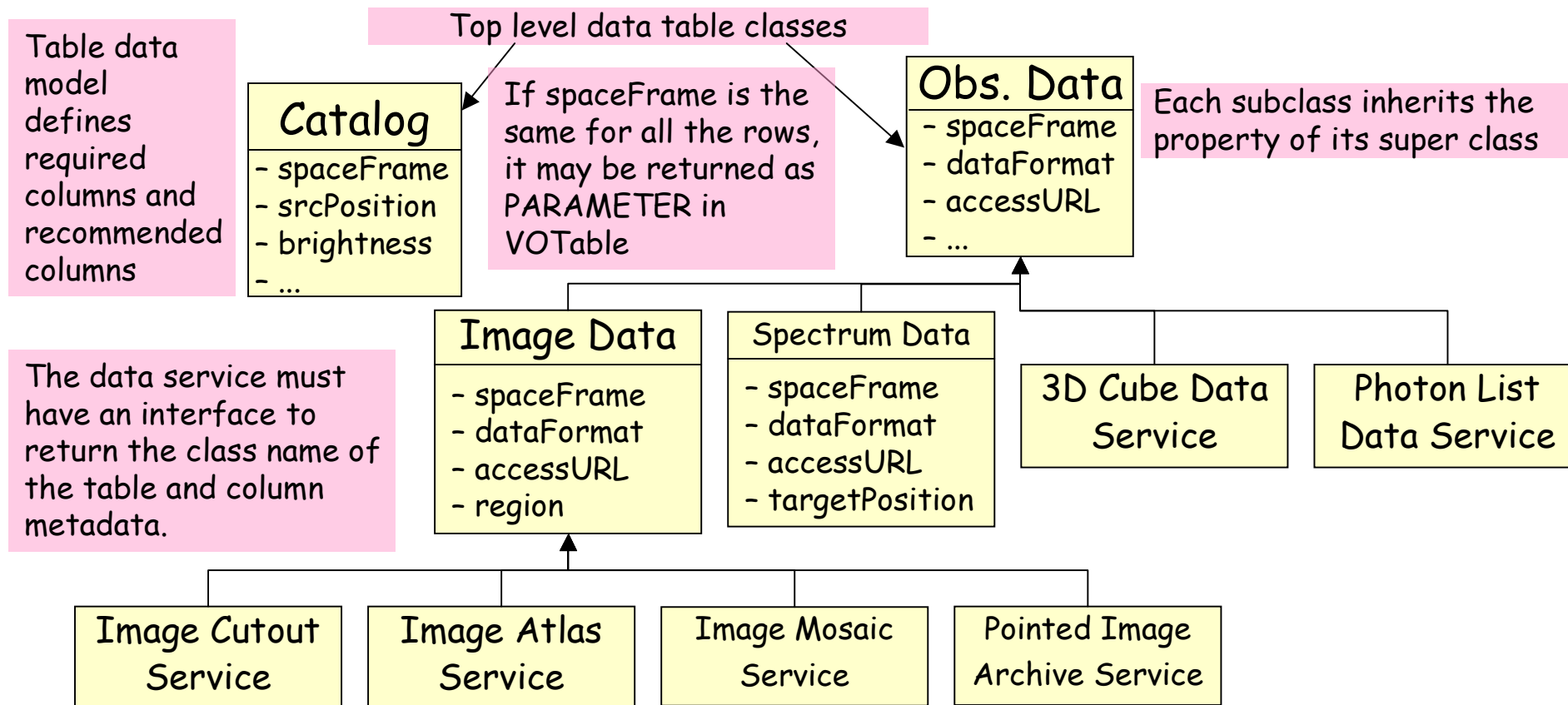
How to treat object data

- Most of the recent RDB support object data type (SQL99).
- relate one object type to one table.
- make it flat

Required Standard

Table Data Model for Astro-Data Service

- Query is described in SQL construct (ADQL).
- Query result is returned in a tabular form (VOTable).
- To write a SQL, table structure (column name) need to be known in advance.
- Table Data Model for each data service are required.



Column metadata

- All the column must have the following attributes:
 - **name**, **datatype**, **unit**, **arraysize**, **width**, **precision**, **ucd**, **utype**. These are used in the FIELD element attributes of VOTable. A coordinate column must have an **coordinateFrame** attribute.
 - column name can be specified by a data provider.
 - datatype must follows the VOTable specification.
 - expression of unit must follows the VOTable spec.
 - if "arraysize" is not defined "*" is assumed for "char" data type and "1" for the other data types.
 - utype is supplied by VO data model naming scheme. For columns not defined in the data model, utype is not required.
 - appropriate ucd is assigned by the data provider.

Data Service Type	Required Columns (utype) (TBR)	Query Spec. of "Where" part
Catalog	Coverage.Location.Sky.Value	<position> within 'STC Region'
Image	Coverage.Region.Sky Data.Format Data.AccessURL	<region> overlaps 'STC Region' and <format> = 'Image/FITS'
Spectrum	Coverage.Location.Sky.Value Data.Format Data.AccessURL	<position> within 'STC Region' and <format> = 'VOTable'
3D Cube (Space x Spectrum)	Coverage.Region.Sky Data.Projection Data.Format Data.AccessURL	<region> overlaps 'STC Region' and <projection> = 'Image' and format = 'Image/JPEG'
Photon list	Coverage.Region.Sky Data.Projection Data.Format Data.AccessURL	<region> overlaps 'STC Region' and <projection> = 'Spectrum' and format = 'Spectrum/FITS'

Minimum Requirement for Query Spec.

Current ADQL:

- 😊 describe most of the SQL92 syntax and some astro-extension
- 😞 many elements (33) & many data types (69), hard to build a data service. It is helpful to define minimum subset ADQL spec as a base line

- **SelectionItemType** (13 sub types in ADQL 0.8)
 - Only **AliasSelectionItem** and **ColumnReferenceType**
 - All the column must have **alias name**
- **FromTableType** (3 sub types in ADQL 0.8)
 - Only **TableType**
- **SearchType** (16 sub types in ADQL 0.8)
 - Only **IntersectionSearchType** and **ComparisonPredType**
 - Region Search Criterion can be specified by **ComparisonPredType**.

Minimum subset of ADQL for basic Skynode Service

Element: 33 (full) → 12 (basic)

Simple Type: 13 (f) → 4 (b)

Complex Type: 56 (f) → 12 (b)

Fundamental Type

xs:unsignedInt
 xs:string(*)
 xs:double(*)
 xs:long(*)

Simple Type

aggregateFunctionNameType
 allOrDistinctType
 binaryOperatorType
 comparisonType(*)
 jointTableQualifierType
 mathFunctionNameType
 orderDirectionType
 trigonometricFunctionNameType
 unaryOperatorType

Element

Allow	Restrict
Arg(*)	Select(*)
Column	SelectionList(*)
Condition(*)	Set
EndComment	Sigma
Expression(*)	StartComment
From(*)	Table(*)
GroupBy	TableName
Having	Tables
InTo	Unit(*)
Item(*)	Where(*)
Literal(*)	fromTableType
Name	selection
Nature	
Order	
OrderBy	
Params	
Pattern	
Qualifier	
Region(*)	

Complex Type

ArrayOfFromTableType	includeTableType	searchType
ConstantListSet	inclusionSetType	selectType
aggregateFunctionType	inclusiveSearchType	selectionItemType
aliasSelectionItemType(*)	integerType(*)	selectionLimitType
allSelectionItemType	intersectionSearchType(*)	selectionListType
archiveTableType	intoType	selectionOptionType
atomType(*)	inverseSearchType	stringType(*)
betweenPredType	jointTableType	subQuerySet
binaryExprType	likePredType	tableType(*)
closedExprType	literalType(*)	trigonometricFunctionType
closedSearchType	mathFunctionType	unaryExprType
columnReferenceType(*)	notBetweenPredType	unionSearchType
comparisonPredType(*)	notLikePredType	userDefinedFunctionType
dropTableType	numberType	whereType(*)
exclusiveSearchType	orderExpressionType	xMatchTableAliasType
fromTableType	orderOptionType	xMatchType
fromType(*)	orderType	
functionType	realType(*)	
groupByType	regionSearchType	
havingType	scalarExpressionType	

stc-v1.20.xsd (element name)

AZ_EL	GEOCENTER	MARS_C	Region	VENUS
AstroCoordArea	GEO_C	MARS_G	RegionFile	VENUS_C
AstroCoordSystem	GEO_D	MERCURY	SATURN	VENUS_G
BARYCENTER	GSE	MERCURY_C	SATURN_C_III	VelInterval
BODY	GSM	MERCURY_G	SATURN_G_III	velocityInterval
CARTESIAN	GenericCoordFrame	MOON	SM	velocitySphere
CatalogEntryLocation	HCD	NEPTUNE	SPHERICAL	Xaxis
Center	HCI	NEPTUNE_C_III	STCResourceProfile	
Coord2VecInterval	HEE	NEPTUNE_G_III	STCmetadata	
Coord3VecInterval	HEEQ	Name	SUPER_GALACTIC	
CoordArea	HELIOCENTER	ObsDataLocation	SearchLocation	
CoordFlavor	HGC	ObservationLocation	SpaceFrame	
CoordFrame	HiLimit	ObservatoryLocation	SpaceRefFrame	
CoordInterval	HiLimit2Vec	OffsetCenter	SpatialInterval	
CoordOrigin	HiLimit3Vec	PLUTO	SpectralFrame	
CoordScalarInterval	ICRS	PLUTO_C	SpectralInterval	
CoordSys	JUPITER	PLUTO_G	Sphere	
CustomRefPos	JUPITER_C_III	POLAR	StartTime	
CustomSpaceRefFrame	JUPITER_G_III	PixelCoordArea	StopTime	
DopplerDefinition	LOCAL_GROUP_CENTER	PixelCoordFrame	TOPOCENTER	
ECLIPTIC	LSR	PixelCoordSystem	TimeFrame	
EMBARYCENTER	LSRD	PixelSpace	TimeInterval	
Equinox	LSRK	PlanetaryEphem	TimeRefDirection	
FK4	LUNA_C	Pole_Zaxis	TimeScale	
FK5	LUNA_G	PositionInterval	UNITSPHERE	
File	LoLimit	RELOCATABLE	UNKNOWNFrame	
Frame	LoLimit2Vec	Radius	UNKNOWNRefPos	
GALACTIC_CENTER	LoLimit3Vec	RedshiftFrame	URANUS	
GALACTIC_I	MAG	RedshiftInterval	URANUS_C_III	
GALACTIC_II	MARS	ReferencePosition	URANUS_G_III	

stc-v1.20.xsd (complexType and simpleType name)

astroCoordAreaType	spatialIntervalType	coordEquinoxType
astroCoordSystemType	spectralFrameType	dopplerDefinitionType
catalogEntryLocationType	spectralIntervalType	planetaryEphemType
coord2VecIntervalType	sphereType	
coord3VecIntervalType	stcDescriptionType	
coordAreaType	stcMetadataType	
coordFlavorType	stcResourceProfileType	
coordFrameType	stdRefPosType	
coordIntervalType	timeFrameType	
coordScalarIntervalType	timeIntervalType	
coordSysType	velocityIntervalType	
customRefPosType	velocitySphereType	
customSpaceRefFrameType		
fkType		
icrsType		
obsDataLocationType		
observationLocationType		
observatoryLocationType		
pixelCoordAreaType		
pixelCoordSystemType		
pixelSpaceType		
positionIntervalType		
redshiftFrameType		
redshiftIntervalType		
referencePositionType		
regionFileType		
regionType		
searchLocationType		
spaceFrameType		
spaceRefFrameType		

region-v1.20.xsd

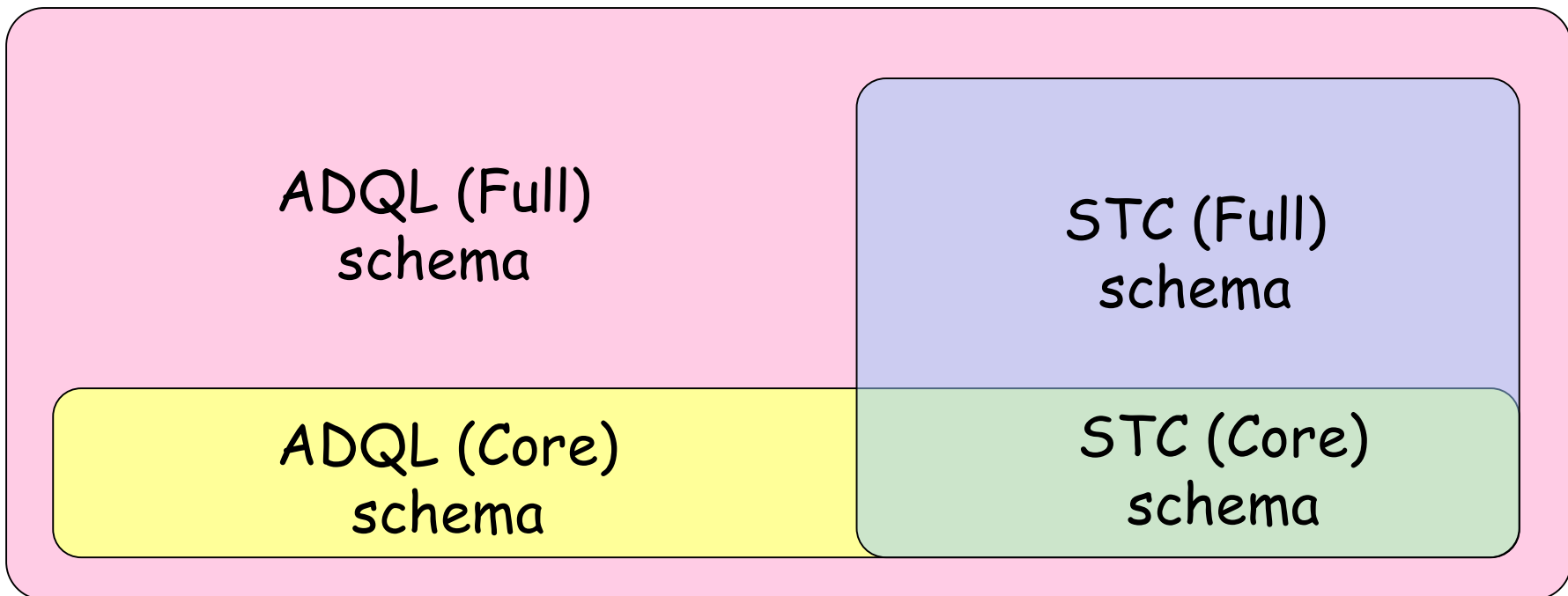
AllSky	boxType
Box	circleType*
Center*	constraintType
Circle*	convexHullType
Constraint	convexType
Convex	ellipseType
ConvexHull	intersectionType
Ellipse	negationType
Intersection	polygonType
MinorRadius	regionType*(abstract)
Negation	sectorType
Offset	shapeType*(abstract)
Point	skyIndexType
Pole	smallCircleType
Polygon	unionType
PosAngle	vertexType
PosAngle1	
PosAngle2	
Position	
Radius*	
Region	
Sector	
Size	
SkyIndex	
SmallCircle	
Union	
Vector	
Vertex	

coords-v1.20.xsd

AbsoluteTime	ErrorRef	Position1D	TimeOffset	angleUnitType
AstroCoords	FITSFile	Position2D	TimeOffsetRef	double2Type*
CError	FITSPosition	Position3D	TimeOrigin	double3Type
CError2	FITSRedshift	Redshift	Timescale	double4Type
CError3	FITSSpectral	RelativeTime	Value	double9Type
CPixSize	FITSTime	Resolution	Value2	doubleArrayType*
CPixSize2	FITSVelocity	Resolution2	Value2Ref	posAngleReferenceType
CPixSize3	ISOTime	Resolution2Matrix	Value3	posUnitType
CResolution	ISOTimeRef	Resolution2PA	Value3Ref	spectralUnitType
CResolution2	JDTime	Resolution2Ref	ValueRef	timeScaleType
CResolution3	JDTimeRef	Resolution3	velocity	timeUnitType
CSize	MJDTime	Resolution3Matrix	velocity1D	unitType
CSize2	MJDTimeRef	Resolution3PA	velocity2D	velTimeUnitType
CSize3	Name	Resolution3Ref	velocity3D	
CValue	PixSize	ResolutionRef		astroCoordsFileType
CValue2	PixSize2	ScalarCoordinate		astroCoordsType
CValue3	PixSize2Matrix	Size		astronTimeType
CoordFile	PixSize2PA	Size2		coordFITSColumnsType
CoordValue	PixSize2Ref	Size2Matrix		coordinateType
Coordinate	PixSize3	Size2PA		coordsType
Coords	PixSize3Matrix	Size2Ref		fitsType
Error	PixSize3PA	Size3		pixelCoordsType
Error2	PixSize3Ref	Size3Matrix		posAngleType
Error2Matrix	PixSizeRef	Size3PA		scalarCoordinateType
Error2PA	PixelCoordinates	Size3Ref		size2Type
Error2Ref	PixelCoords	SizeRef		size3Type
Error3	PosAngle	Spectral		timeCoordinateType
Error3Matrix	PosAngle1	StringCoordinate		vector2CoordinateType
Error3PA	PosAngle2	Time		vector3CoordinateType
Error3Ref	Position	TimeInstant		

Proposed ADQL Schema structure

- STC (Full) schema imports STC (Core) schema
- ADQL (Core) imports STC (Core)
- ADQL (Full) imports ADQL (Core) and STC (Full)
- ADQL (Core) and STC (Core) should not be changed. These are critical standard for interoperability.
- Each schema has a difference namespace.



Substitution of Column Name by UCD and UType

- Without knowing the column names, we cannot write an SQL...
- Possible Solution → use UCD or Utype for describing the query condition and selection list.
- Introduce "ucd" and "utype" attributes to the `columnReferenceType`

```
<Item xsi:type="columnReferenceType" Name="ra" Table="qso"/>
```

```
<Item xsi:type="columnReferenceType" ucd="pos.eq;src" Table="qso"/>
```

```
<Item xsi:type="columnReferenceType" utype="Target.pos" Table="qso"/>
```

Example Query

```
<Select xmlns:region="urn:region-core"
        xmlns="http://www.ivoa.net/xml/ADQL/core"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <SelectionList>
    <Item Table="t" Name="*" xsi:type="columnReferenceType"/>
  </SelectionList>
  <From>
    <Table Name="image" Alias="t" xsi:type="tableType"/>
  </From>
  <Where>
    <Condition comparison="overlaps" xsi:type="comparisonPredType">
      <Arg Table="t" Name="region" xsi:type="columnReferenceType"/>
      <Arg xsi:type="circleType" unit="deg" coord_system_id="FK5">
        <region:Center>20.0 30.0</region:Center>
        <region:Radius>1.0</region:Radius>
      </Arg>
    </Condition>
  </Where>
</Select>
```

Permanent namespace

Red characters
are user-supplied
parameters

Use "comparisonPred"
for describing the
search condition

Only "CircleType" and
"AllSky" are defined in
region-core schema

```
SELECT t.*
FROM image t
WHERE region OVERLAPS
      Circle((20.0,30.0,'FK5'), 1.0)
```

Define Standard Usage of STC Condition

- A basic SkyNode MUST support "comparisonPredType".
- A basic SkyNode MAY support the other searchType.
- A basic SkyNode MUST recognize an unsupported searchType as a "trueType".
- A basic SkyNode MUST support the following construct:
 - `<STC_columnName> <STC_operator> <STC_searchLocationType>`
 - `STC_operator ::= within, overlaps, outside`
 - point within 'Circle ICRS 200 -20 2.0' (Catalog Query)
 - region overlaps 'Circle ICRS 200 -20 2.0' (Image Query)
 - spectrum overlaps 'SpectralInterval A 4000 7000' and observationTime within 'TimeInterval 2004-05-01 2004-05-31' (Spectrum Query)

STC-type column must be compared with an object of STC searchLocationType

utype defines which column represents STC.

ADQL-x version of point within 'Circle ICRS 200 -20 2.0'

It might be convenient to define substitution types for frequently used frames such as `<SpaceFrame xsi:type="ICRSFrameType"/>`

```
<Condition comparison="within" xsi:type="comparisonPredType">
  <Arg Table="t" Name="point" xsi:type="columnReferenceType"/>
  <Arg xsi:type="searchLocationType">
    <AstroCoordSystem ID="ICRS">
      <SpaceFrame>
        <ICRS/><BARYCENTER/><SPHERICAL coord_naxes="2"/>
      </SpaceFrame>
    </AstroCoordSystem>
  <AstroCoordArea ID="SearchRegion" coord_system_id="ICRS">
    <Region>
      <Circle unit="deg"> <Center>200 -20</Center> <Radius>2.0</Radius> </Circle>
    </Region>
  </AstroCoordArea>
</Arg>
</Condition>
```

Which frame should be defined as a "must be supported" frame

Returned VOTable

- FIELD id → If alias name is specified "adql:aliasName", otherwise "adql:tableAliasName.columnName".
- FIELD name → free
- STC metadata for STC column
 - If each record in the STC column has a different STC metadata, STC metadata column must be included. Otherwise STC metadata must be included as PARAMETER.
- Some of the columns must always be included in the output VOTable. These columns are defined in SIAP, SSAP and SXAP document.

Problem encountered
in implementation

Problems encountered in implementation (1)

- Name space problem (as of 2005 Jan)
 - JVO → ADQL v0.8 + VOTable v1.1 + STC v1.1
 - NVO → ADQL v0.74 + VOTable <v1.0 + NVO-STC
 - External interface → ADQL v0.74, VOTable v1.0
 - Internal interface → ADQL v0.8, VOTable v1.1
 - Namespace exchanger
- Possible Solutions:
 - Define a core part of ADQL as a minimum subset and assign a permanent namespace. Never update, never change the namespace of the core part.

Problems encountered in implementation (2)

- XML → Java deserialization in AXIS
 - With the standard usage of AXIS, an XML document (ADQL, VOTable) is, as a default, deserialized to Java objects.
 - Server memory is easily exhausted.
 - Even several hundreds records of VOTable suffers out of memory error.
- Possible Solution :
 - Stop auto-deserialization of AXIS (learned from an AstroGrid person)
 - return VOTable as an attachment
 - return a reference URL to retrieve the VOTable

Problems encountered in implementation (3)

- Usage of VOTable
 - id and name attribute → what should be filled ?
 - Where column alias name should be put. This information is required for post-search processing on portal side.
 - Information on the origin of the column data should be kept anywhere in VOTable. Where ?

Problems encountered in implementation (4)

- Complexity of ADQL & STC object
- Number of defined element
 - ADQL v0.8 → 33
 - STC v1.20 → 250
- Number of defined ComplexType and SimpleType
 - ADQL v0.8 → 69
 - STC v1.20 → 88

Problem for interoperability

- Column name must be known in advance for writing ADQL.
 - We can get column names by "Columns" interface and write ADQL, but it requires human intervention.
 - A possible solution:
 - use UCD or Utype for specifying a column
 - Introduce "ucd" and "utype" attributes to the columnReferenceType
- ```
<Item xsi:type="columnReferenceType" Name="ra" Table="qso"/>
<Item xsi:type="columnReferenceType" ucd="pos.eq;src" Table="qso"/>
<Item xsi:type="columnReferenceType" utype="Target.pos" Table="qso"/>
```
- If the specified utype or ucd is not found in the queried table,
    - ignore the condition for that column
    - return PARAMETER of "NaN" for that column