



第3章 数字矿山数据获取与管理

主讲人：毕林

2024年8月13日



目录

- 3.1 地理空间参考系统
- 3.2 数据来源与输入
- 3.3 数据规范化和标准化
- 3.4 数据处理
- 3.5 数据质量
- 3.6 数字矿山数据分类与管理



3.1

地理空间参考系统

矿山开采生产与经营管理新模式的变革背景





地理实体空间位置、分布、形态、空间关系（距离、方位、拓扑、相关场）等基本特征的精确描述依赖于空间参考系统，空间参考系统定义了地理空间三维表面的空间坐标系统及各坐标系统间的数学关系。所有的地理要素只有按经纬度或者特有的空间坐标系统进行严格的空定位，才能使具有时序性、多维性、区域性特征的空间要素进行复合和分解，将其中隐含的信息变为显示表达，形成空间和时间上连续分布的综合信息基础，支持空间问题的分析、处理与决策。地理空间参考系统是对地理空间的精准表达并成为数字模拟的基础。

地理空间坐标系统是空间位置的度量衡，是确定空间位置、空间距离、空间方位、空间关系等信息必需的工具，是空间数据分析的基础和前提。





地球椭球体

01

参考椭球体

为了用数学方法描述和表达地球表面，需要选择一个与地球形状、大小接近的球体来近似代替。我国1978年采用1975年国际大地测量和地球物理学联合会（IUGG）推荐的地球椭球体建立新的西安1980大地坐标系。

02

椭球定向和椭球定位

建立椭球体后，需要进行椭球定向。椭球定向是指确定椭球旋转轴的方向，即旋转椭球体需要套在地球的一个适当的位置上，这个位置就是大地原点，所有大地坐标均以大地原点作为坐标计算的起算点。椭球定位是指确定椭球中心的位置，可分为局部定位和地心定位两类。局部定位要求在一定范围内椭球面与大地水准面有最佳的吻合，对椭球的中心位置无特殊要求；地心定位要求在全球范围内椭球面与大地水准面有最佳的吻合，同时要求椭球中心与地球质心一致或最为接近。



坐标系统

01

天球坐标系

天球坐标系是惯性坐标系，其坐标原点及各坐标轴指向在空间保持不变，用于描述天体、卫星位置和状态。在天球坐标系中，天体 S 的空间位置可用天球空间直角坐标系或天球球面坐标系两种方式来描述。

02

地球坐标系

地球坐标系是为了描述地面点的位置建立的一个与地球相关联的坐标系。常用的地球坐标系有：地心坐标系、站心坐标系、参心坐标系、极移现象与协议地球坐标系。

03

常用椭球面上的坐标系统

包括：国家大地坐标系、高斯平面直角坐标系统与UTM坐标系、WGS-84世界大地坐标系、独立坐标系。





在GIS中，空间现象的表征通常以平面图形表示。受传统地图空间数学基础理论和方法的影响，目前几乎所有的GIS均沿用地图投影作为自己参考系的数学基础。

地图投影方法建立了地球椭球表面上的点与地图平面上点之间的一一对应关系，将地球表面的球面坐标转换为地图平面坐标，就可以方便地进行距离、方位、面积的量测。





共有
256种

投影类型

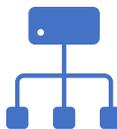
构成方式		方向	投影面与球面关系	变形性质
几何投影	方位投影	正轴投影	割投影	等角投影
	圆柱投影			
	圆锥投影	斜轴投影		等距投影
伪方位投影	横轴投影		切投影	
伪圆柱投影				
伪圆锥投影				
非几何投影	多圆锥投影			

按照投影面与地球相对位置，地图投影分为正轴投影（投影面的中心轴与地轴重合）、斜轴投影（投影面的中心轴与地轴斜交）和横轴投影（投影面的中心轴与地轴垂直），若投影面与地球表面相割称为割投影，相切则为切投影；根据投影可能引起的变形性质，可以分为等角投影、等积投影和任意投影三类；按照构成方法分为几何投影和非几何投影，后者不借助几何面，主要根据某些约束条件用数学解析法确定球面与平面之间点与点的函数关系；按照经纬线形状又分为伪方位投影、伪圆柱投影、伪圆锥投影和多圆锥投影。



地图投影选择的一般原则

- 1) GIS所采用的投影系统应与本国的基本地图系列所采用的投影系统一致;
- 2) 各比例尺GIS中的投影系统应与相应比例尺主要信息源地图的投影一致;
- 3) 各地区的GIS投影系统应与该地区所使用的投影系统一致;
- 4) 一般选择1~3种投影系统以保证地理定位框架的统一。



地图投影选择

我国GIS采用的地图投影

目前我国各种类型地理信息系统中均采用了与我国基本比例尺一致的投影系统，即大比例尺用高斯-克吕格投影，中小比例尺用Lambert投影。

我国GIS中配置高斯克吕格投影和Lambert投影的目的主要在于统一存储参照以及分析与运算的地理基础，并不意味着GIS仅处理这两种投影所提供的数据或仅以这两种投影作为输出模式。



3.2

数据来源与输入

由开采生产与经营管理新模式的变革背景





数据来源与输入

数字矿山数据是指用来表示矿山空间实体的位置、形状、大小及其分布特征诸多方面信息的空间数据，主要包括地表地形、探矿工程、矿体资源、井巷工程、围岩、断层、生产设备等方面的空间数据。

下面重点介绍下GNSS测量技术、倾斜摄影测量技术、三维激光扫描技术、图纸矢量化技术、以及基于文字报告的OCR技术获取来源数据的原理与方法。





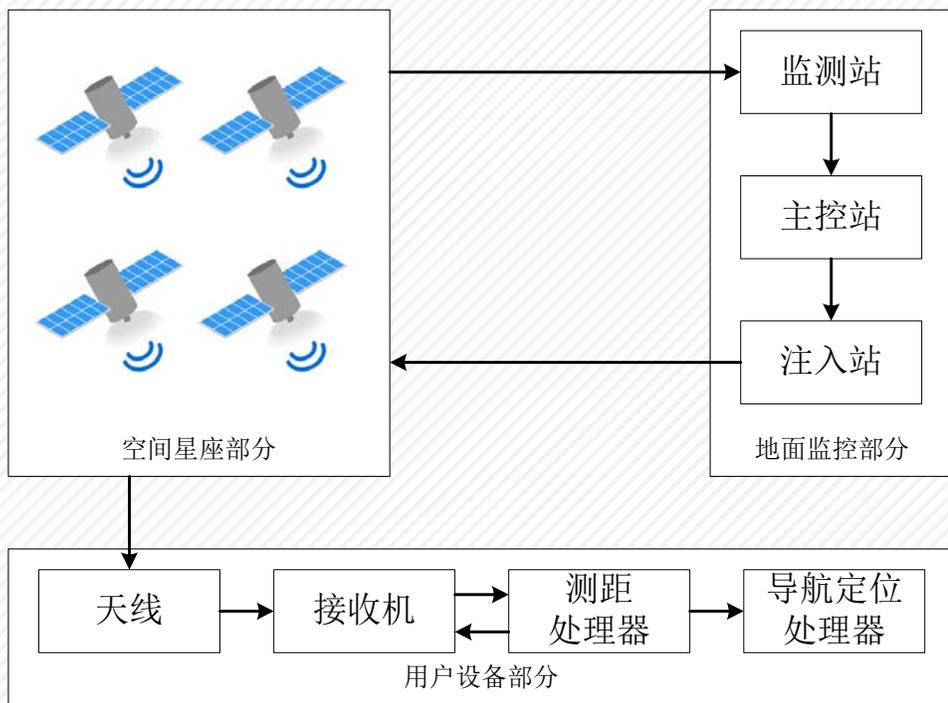
全球导航卫星系统（GNSS）是能在地球表面或近地空间的任何地点为用户提供全天候的三维坐标和速度以及时间信息的空基无线电导航定位系统。



全球包括4大卫星导航系统供应商，包括中国的北斗卫星导航系统（BDS）、美国的全球定位系统（GPS）、俄罗斯的格洛纳斯卫星导航系统（GLONASS）和欧盟的伽利略卫星导航系统（GALILEO）。



GPS是世界上第一个建立并用于导航定位的全球系统；BDS则是中国自主建设运行的全球卫星导航系统，为全球用户提供全天候、全天时、高精度的定位、导航和授时服务。



GNSS系统组成

空间星座部分

空间星座部分是GNSS系统的核心组成部分，负责接收并存储发自地面监控站的导航电文等信息，处理地面监控站的控制指令，向用户播发定位数据信息。

地面监控部分

地面监控部分是维持整个卫星导航系统平稳运行的关键，由主控站、监测站和注入站等设施组成，主控站是整个地面监控部分的核心，负责协调和控制地面设施的工作。

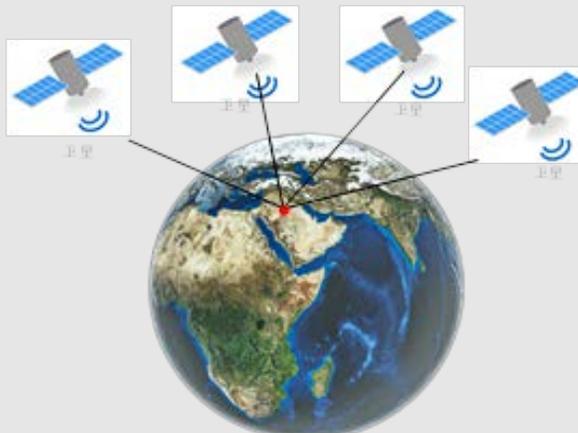
用户设备部分

GNSS用户设备主要指卫星接收机，主要作用是跟踪可见卫星，接收导航卫星发出的信号，根据收到的卫星星历、伪距观测数据以及载波观测测量，计算出载体在相应时间的位置和速度信息。



GNSS测量技术

GNSS测量技术利用高轨卫星作为已知点，通过轨道积分推算出卫星位置，接收机测量信号传输时间，计算距离。



测距原理

根据测距原理，GNSS定位方式分为伪距定位、载波相位测量定位、GNSS差分定位。



后方交会原理

当接收机同时接收4颗及以上卫星信号时，可以利用后方交会原理解算出地面测站点的坐标位置。





伪距定位

伪距定位可分为单点定位和多点定位，单点定位需锁定4颗以上卫星，用距离交会原理解算出天线所在点的三维坐标。



多点定位

多点定位中，多台GNSS接收机同时锁定相同卫星进行伪距测量，可消除大气电离层和对流层折射误差、星历误差影响。

单点定位

单点定位精度不高，C/A码定位精度为25m，P码定位精度为10m，未考虑大气电离层和对流层折射误差、星历误差影响。



载波相位定位

载波L1、L2的频率比测距码（C/A码和P码）频率高，波长比测距码短很多， $\lambda_1=19.03\text{cm}$ ， $\lambda_2=24.42\text{cm}$ 。

使用载波L1或L2作测距信号，将卫星传播到接收机天线的余弦载波信号，与接收机基准信号比相求出相位延迟计算伪距，可获得很高的测距精度。

如果测量L1载波相位移误差为1/100，伪距测量精度可达 $19.03\text{cm}/100=1.9\text{mm}$ 。





实时差分定位

安置GNSS接收机

在已知坐标点上安置一台GNSS接收机作为基准站，用已知坐标和卫星星历算出观测值的校正值，通过数据链发送给移动站。

实时动态差分法

实时动态差分法（RTK），也称为载波相位差分技术，可以实现厘米级的高精度定位。



移动站观测值改正

移动站用接收到的校正值对自身GNSS观测值进行改正，消除卫星钟差、接收机钟差、大气电离层和对流层折射误差。





01

差分改正值

基准站和流动站分别接收来自同一时间、相同GNSS卫星的信号，通过比较基准站的观测值和已知位置信息，得到GNSS差分改正值。

02

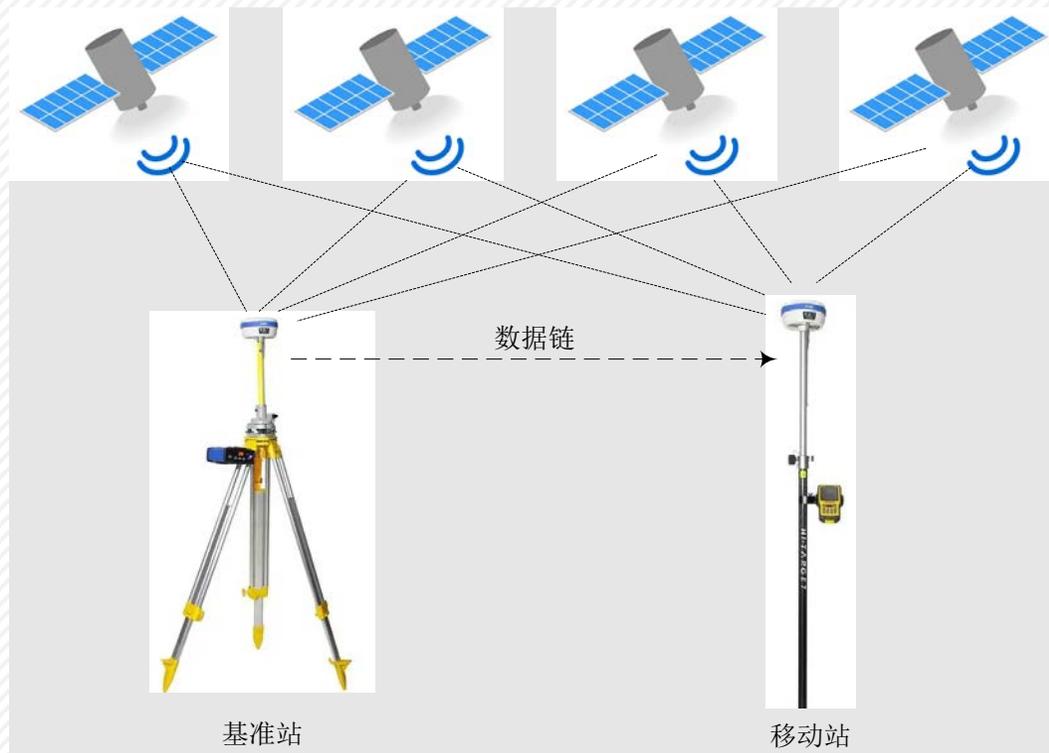
实时处理

改正值通过无线电数据链电台传递给移动站，移动站对接收到的数据实时处理，完成双差模糊度的求解、基线向量的解算、坐标的转换。

03

准确位置

最终得到移动站较准确的实时位置，RTK测量工作原理示意图如图所示，为GNSS测量领域带来更精确、更高效的结果。





观测时间短与高度自动化

采用GNSS布设一般等级的控制网时，在每个测站上的观测时间一般在1-2个小时左右。

灵活与成本低

选点灵活、不需要造标、费用低、GNSS测盘不要求测站间相互通视，不需要建造觇标。



全球覆盖

全天候作业，任何时间、任何气候条件下均可以进行GNSS观测，大大方便了测量作业。



高精度定位

GNSS观测的精度明显高于常规测量手段，基线向量相对精度一般在1mm~10mm之间。



信号干扰

在城市等拥挤的地区，由于信号干扰的存在，以及高大建筑物的多路径效应等因素，有时会导致接收器无法准确接收到卫星信号，从而影响测量的精度。



易受大气影响

由于GNSS技术需要依赖卫星发射的信号进行测量，而大气层对信号传播会产生一定的影响。例如，由于电离层等原因，在某些特殊的天气条件下，GNSS系统的测量精度会变得不稳定，需要采取相应的校正方法来消除大气影响。



数据安全性

由于GNSS技术本质上是无线传输技术，而且信号具有一定的强度，此特点使得其容易受到信号仿冒和干扰的攻击。为了保证测量数据的安全性，需要采取一系列的防护措施，比如加密传输和身份认证等。



由于GNSS测量具有高精度定位、选点灵活、测量成本低、高度自动化以及全球覆盖等优点，已被广泛应用于航空航天、交通运输、海洋渔业、军事领域、矿业领域、以及应急救援等诸多领域。其中矿业领域已在以下三个方面进行了应用。

1

矿山测量

2

露天矿边坡监测

3

尾款库坝体变形监测



矿山测量

GNSS技术广泛应用于矿区大比例尺地形图测量、矿区变形测量、露天矿区控制测量、爆破孔放样等方面。



自动化变形监测

基于GNSS自动化变形监测系统，可实现长期稳定监视监测点、无线传送高精度监测数据，全年连续观测。

1

矿山测量

2

露天矿边坡监测

3

尾款库坝体变形监测



露天矿边坡监测

利用GNSS技术建立自动化监测系统，实时采集边坡数据信号，通过传输模块传送至控制中心。

预警值设置

根据监测出的边坡变形量，如果达到了设置的预警值，就会发出预警信号，以便工作人员提前做出预防措施。

监控点三维坐标计算

在控制中心，对每个监控点的三维坐标进行计算，并与初始三维坐标进行比较，得出监控点的变化。

边坡变形量与速率

通过计算监控点的变化，得出边坡变形量、变形速率、变形加速度等参数，以便及时发现边坡变形情况。

1

矿山测量

2

露天矿边坡监测

3

尾款库坝体变形监测



01

尾矿库坝体变形监测

利用GNSS技术建立尾矿库坝体位移监测系统，实现全天候、高精度、高效率、实时动态地监测坝体变形状况。

02

有效获得变形过程

通过监测系统，能够有效地获得尾矿库坝体变形的动态演化过程，加强尾矿库的安全监管，掌握尾矿库的安全现状。

03

具有重要意义

对减少尾矿库事故的发生具有重要意义，有助于提高尾矿库的安全性和稳定性，确保尾矿库的安全运行。

1

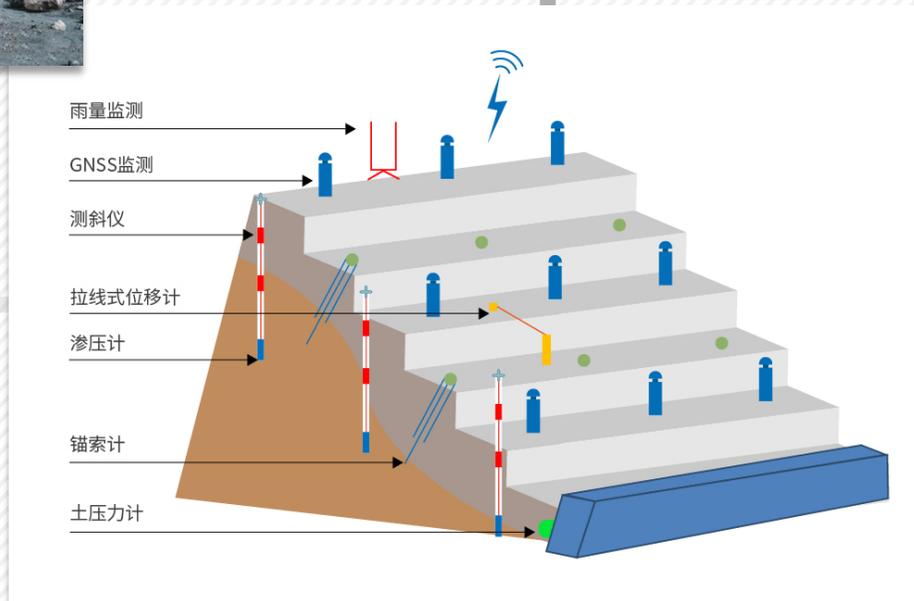
矿山测量

2

露天矿边坡监测

3

尾矿库坝体变形监测



应用案例

以某露天矿土质边坡表面水平位移监测为例，通过采用多台高精度型GPS接收机及其配套设施（GPS天线、软件等，来采集观测点坐标数据，通过多点GPS高精度解算技术来解算GPS观测点的坐标，从而达到实时监测表面位移（如位移方向、位移速率、累计位移等）的目的。如左图所示。



倾斜摄影测量技术 是一种新兴的摄影测量技术，作为传统摄影测量技术的一种新发展，它基于同一飞行平台上搭载的多台传感器，同时从垂直、侧向和前后等角度采集图像，能够比较完整的获取丰富的侧面纹理信息。

结合现有的具备协同并行处理能力的倾斜影像数据处理软件，可快速实现大范围的三维建模，这很大程度上提高了三维模型的生产效率。

倾斜摄影测量技术凭借其多视角、高真实性、全要素等优点，在矿业领域已得到了广泛应用。





倾斜摄影测量技术

倾斜摄影测量是在无人机飞行平台上使用多镜头相机或可改变拍摄视角的单镜头相机，从不同角度对兴趣物进行拍摄，以获取地物信息。



五镜头相机系统

倾斜摄影测量最常用的影像采集系统是五镜头相机，该系统具有前、后、中、左、右五个镜头对地物进行影像采集，提高外业工作效率。



全球卫星导航系统

倾斜摄影测量融合全球卫星导航系统 (GNSS) 和惯性导航单元 (IMU) 等技术，可以记录多角度影像数据在拍照曝光瞬间的位置和姿态信息。



无人机倾斜摄影测量

无人机倾斜摄影测量是倾斜摄影测量的实际应用，通过在无人机飞行平台上搭载多镜头相机或单镜头相机，从不同角度对地面目标进行拍摄。



01

低成本、高效率、高真实性

倾斜摄影测量所建立的模型的几何精度可以达到厘米级，纹理信息来自于外业航空影像的自动批量映射，高精度的位置信息和真实的纹理信息能够逼真的反映地物的真实情况，很好的弥补了传统三维建模真实性不足的缺点。

02

数据成果多样

通过倾斜摄影测量技术得到的成果多样化。不仅可以进行高精度的实景三维场景重建，而且可以根据需要生成不同的数据成果，如DLG、DSM、DOM、DEM等成果，飞行一次获得的数据即可获得多套测绘成果，从而满足不同的数据需求。

03

信息共享

传统人工构建的三维模型数据量往往较大，在进行网络发布和共享之前需要对模型进行轻量化，而这必然会对模型的精细程度造成影响，通过倾斜摄影测量技术三维重建的模型数据量小，形成的数据成果格式便于进行网络发布和信息共享。

04

单张影像的可量测性

通过无人机航飞获得的影像数据经过处理后，可以在成果影像上进行地物的长度、高度、角度、面积等基础信息的量测，量测的结果精度较高而且可靠，弥补了传统正直摄影测量在影像量测应用中的不足。



1、分辨率不一致

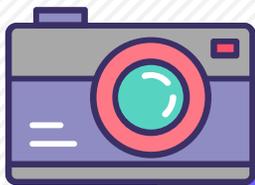
倾斜摄影测量技术由于相机倾角的存在，会使获得的影像的分辨率不一致，这就导致了同一地物在不同视角的影像中会有明显的几何变形，给影像匹配带来一定困难

2、摄影盲区影响

倾斜摄影测量技术由于飞行高度的限制，仍有部分地物遮挡导致影像信息获取不到，出现摄影盲区，从而给后期的数据处理造成影响。

3、影响实展效果

生成的三维模型在存在遮挡的区域会存在模型缺失、纹理拉花的现象，严重影响实景三维模型的展示效果。



01

摄影测量应用广泛

摄影测量技术在问世后不久就被广泛地运用于各个领域，如今已被大量地运用于测制地图、工程质量管理、建筑物监测、气象监测、环境保护以及自然灾害防治等方面。

02

矿山地质考察应用

在矿山地质考察方面，通过多种测量手段，可以进行矿区地表形态的测量，水文地质、工程地质的测量和矿房施工断面的测量与设计等，特别是在岩体结构面分析研究方面的运用获得了极大的成功。

03

露天矿开采现状测量应用

在露天矿开采现状测量方面也得到广泛成熟的应用，通过无人机摄影测量技术进行外业航拍、内业数据处理，可得构建非常逼真的露天矿开采现状模型，并可进行坐标信息查询与方量计算。



摄影测量技术优势

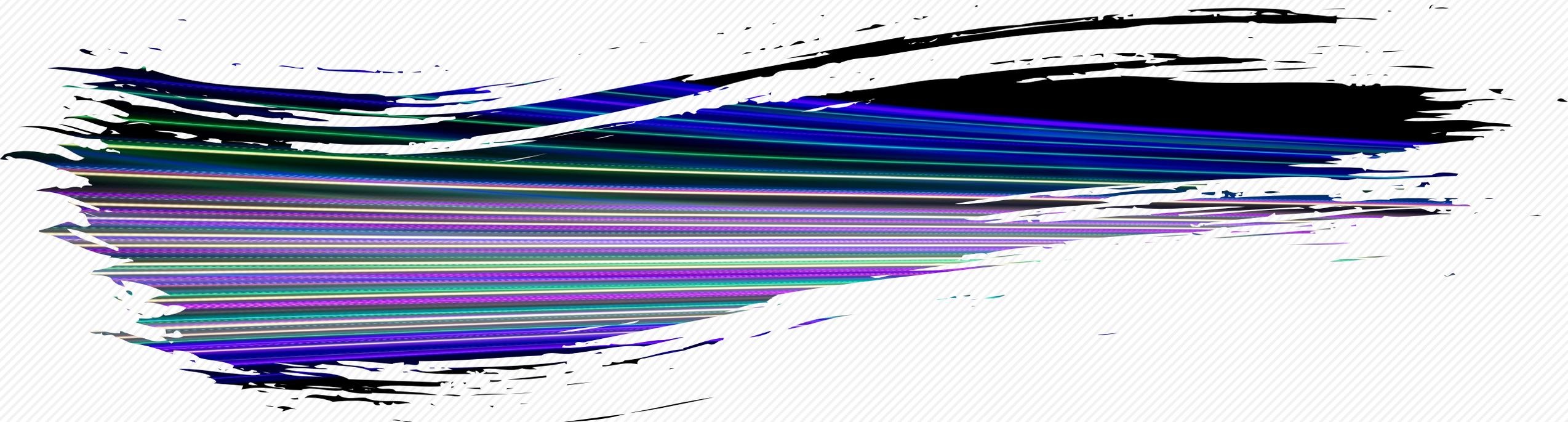
摄影测量方法采集物体空间位置信息的技术由于其非接触性、可同时获得大量标志点信息而得到了广泛的应用，目前，摄影测量技术是矿山测量领域最先进方法之一。

摄影测量技术作用

摄影测量技术可以创建一个实时的信息交流和反馈环境，提高信息采集的效率，降低不完整信息和信息丢失的可能性，大大提高作业效率与测量结果的准确性与真实性。



三维激光扫描技术是集光、机、电和计算机于一体的非接触测量技术，具有测量速度快、自动化程度高、分辨率高、可靠性高和相对精度高的特点，其扫描结果直接显示为点云，利用点云数据，可快速建立结构复杂、不规则的场景的三维可视化模型，既省时又省力，其三维激光扫描头如左图所示。



该方法与传统测量方式相比具有很大的优越性，显著地提高了生产效率和质量。它可以对复杂的环境及空间进行扫描操作，并直接将各种大型的、复杂的、不规则、标准或非标准的实体或实景的三维数据完整的采集到计算机中，进而重构出目标的三维模型以及点、线、面、体、空间等各种制图数据。



按测量方式划分

脉冲式和相位差是两种常见的测量方式，脉冲式激光测距是利用发射和接收激光脉冲信号的时间差来实现对被测目标的距离测量。

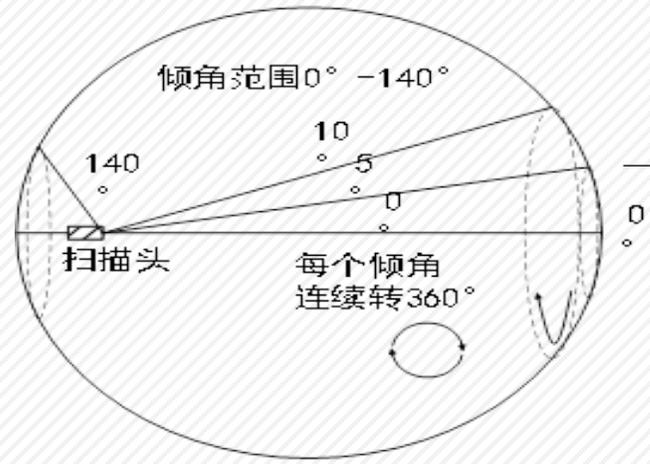
按照载体划分

机载、车载、地面和手持型是四种常见的载体方式，机载型主要用于测量大型的、复杂的、不规则的实体或实景的三维数据。

按用途划分

室内型和室外型是两种常见的划分方式，室内型主要用于测量室内的三维数据，室外型主要用于测量室外的三维数据。





基于空区监测系统 (CMS) 的三维激光扫描技术.

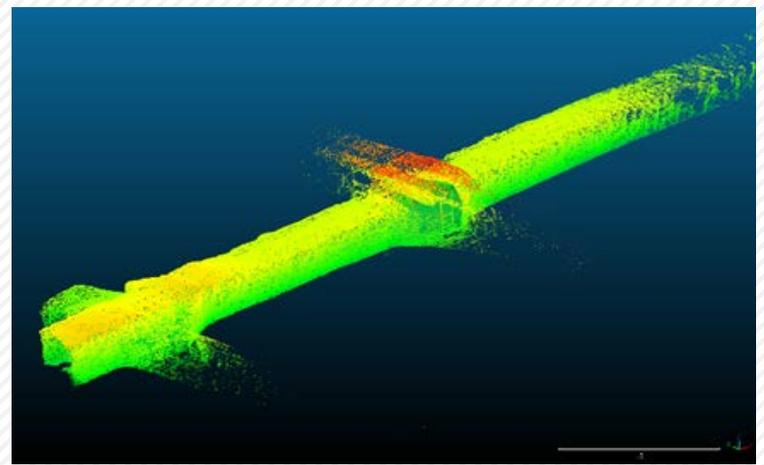
01

扫描仪内置反光镜可将激光束水平偏转，以实现激光水平方向的扫描功能。每当水平扫描一个周期后，步进一次，以进行第二次水平扫描，如此同步下去，最终实现对整个空间的扫描过程。

02

基于SLAM的三维激光扫描技术

激光雷达通过旋转或移动的方式对周围环境进行全方位的扫描，收集各个方向的数据点；SLAM算法会根据数据点之间的关系来估计设备的位姿（位置和方向），并通过位姿的变化来构建地图；在这个过程中，算法还会对数据进行滤波和优化，以消除误差和噪声，提高地图的精度；经过SLAM算法处理后，可以得到设备在环境中的准确位置和姿态，以及由数据点构建出的三维地图。





01 高效性

激光扫描技术可以快速记录被测物体的大面积空间信息，提高测量效率。

02 非接触性

激光扫描技术利用激光反射计算距离和角度，无需接触测量目标，实现快速重构。

03 数字化

三维激光仪器扫描、影像获取、数据存储均自动化完成，数据实时传输至计算机。

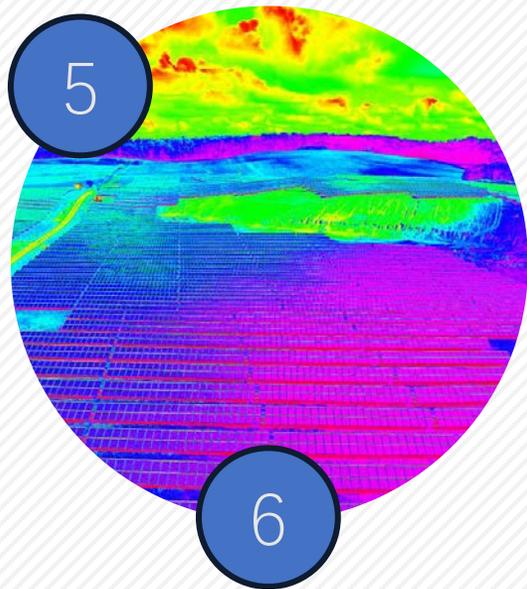
04 环境适应性强

扫描可以不受光照的限制，有无光照都可以进行作业。此外三维激光扫描系统有较强的抗辐射性和抗耐潮性，增强了三维激光扫描仪的生命力。



精度高

三维激光扫描系统测量数据精度能够达到亚毫米级，满足建筑工程、矿业工程、地质工程的测量要求。



穿透性强

通过设定激光频率，可以过滤掉特殊物质，如透明水、玻璃材质物体和稀疏植被。

数据获取全

三维激光扫描仪内置相机具有自动对焦功能，根据被测物体不同位置改变相机视距。保证不同测距物体散距效应真实性，实现扫描点云与被测物逼近。



数据兼容性强

扫描系统具备及时采集储存功能，这些点云数据通过三维激光扫描仪配套的软件可导出多种文件格式。原点云数据通过格式转换输出其他兼容性文件格式，兼容性软件可以利用这些数据进行多角度分析点云数据，进而达到工程对数据的需求。



适用场景

作为新的高科技产品，三维激光扫描仪已经成功的在文物保护、城市建筑测量、地形测绘、变形监测、工厂、大型结构、管道设计、飞机船舶制造、公路铁路建设、隧道工程、桥梁改建等领域里应用。

在采矿业方面

用三维激光扫描仪进入到一些人员不方便到达或有危险的区域进行三维扫描，可有效解决露天及地下矿山生产作业中遇到的台阶变形监测和空区塌陷体积计算等问题。



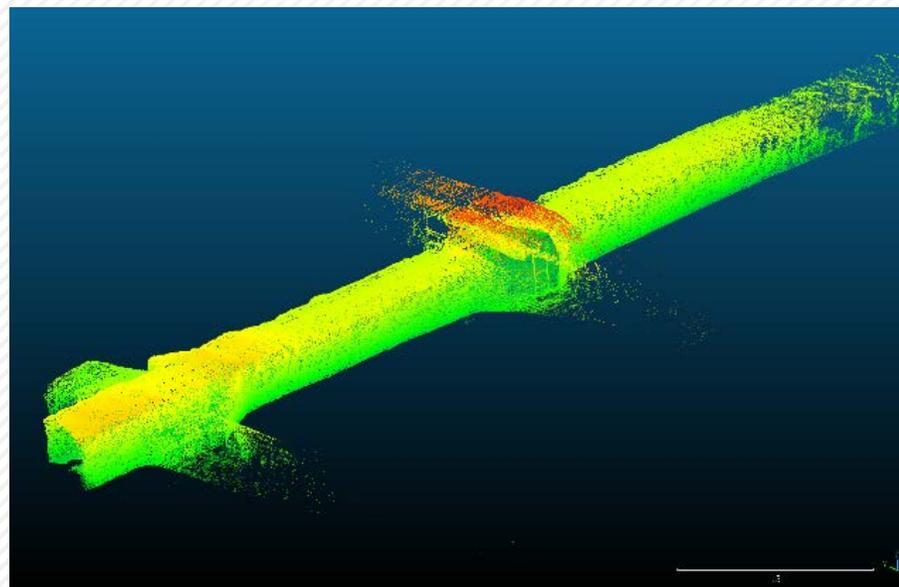


基于SLAM的三维激光扫描技术

基于SLAM的三维激光扫描技术则具有数据采集量大、精度高、速度快、自动化程度高、操作灵活、作业时间短等特点，能有效解决传统测量不方便、作业时间长、工作效率低等难题。

地下矿巷道测量应用

以某地下矿巷道测量为例，利用SLAM三维激光扫描技术对该矿的某条巷道的扫描测量，取得了很好的应用效果。





矿山地图数据承载了地质勘探、资源勘探、生产勘探、开采过程以及闭坑复垦等矿山全生命周期的各类信息，其呈现方式主要有数字化图形与纸质图形。



数字化图形

是矿山数字化应用过程中的基础，具有存储空间小、易于编辑等优点。但一些矿山早期缺乏数字化观念和意识，生产资料数字化程度不高，部分矿山仍存在一些手绘图纸，要高效应用此类图件，需先通过扫描形成jpg、pdf等格式，然后对图纸进行矢量化处理。



纸质图形矢量化处理

是对各种类型的数字工作底图如纸质地图、黑图或聚酯薄膜图，使用扫描仪及相关扫描图像处理软件，把底图转化为光栅图像，对光栅图像进行诸如点处理、区处理、帧处理、几何处理等，在此基础上，对光栅图像进行矢量化处理和编辑。



01

基于细化的矢量化方法

细化就是通过对图像外边界进行逐层腐蚀操作，直至图像宽度减少为单像素的宽度，然后提取能够表征其特征的骨架点。基于细化的方法就是通过搜索跟踪这些特征点来进行矢量化操作。最初的矢量化方法大多采用基于细化的方法。

优点：算法简单、对硬件要求不高、比较容易实现。

缺点：容易受噪声影响，细化操作一般会造成节点畸变、直线抖动、产生间断等，检测速度慢、计算量大。

02

基于轮廓匹配的矢量化方法

先对图像进行轮廓预处理，然后对图像的轮廓进行跟踪，提取图像的轮廓，然后进行轮廓匹配，通过计算轮廓对的中轴点拟合出矢量图元的中心线，最后对所有中心线进行搜索、拼接操作得到完整的图形对象。

优点：能够避免遍历全部像素点，保留线宽信息，能够保留原图像的拓扑结构。

缺点：无法再相交处进行正确的轮廓匹配，有时还会造成中断，准确率不高。

03

基于邻接域的矢量化方法

记录相同像素值的起始位置与像素值，然后基于游程的宽度与一致性生成具有相似性质的像素群集合，即邻接域，并获得邻接域之间的拓扑关系。利用邻接域的形状，结合邻接域的拓扑信息，拟合出完整的图形。

优点：不需要经过细化处理，抗噪性能明显加强，邻接域结构有效减少了存储空间。

缺点：对于复杂图像处理效果不理想。



04

像素追踪方法

像素跟踪方法从任意中点处按照一个预先设定的方向进行跟踪。每一步形成一个中点。因为它的跟踪方法只有两个方向（水平和竖直），对于斜线，它采用ZigZig跟踪。当整个跟踪完成时，由中点序列拟合出矢量图元。

优点：缩短了处理时间，在一定程度上解决交叉问题，而且还能保留线宽信息。

缺点：对于交叉区域大于跟踪步长或图像本身质量差，该方法不能一次性地完成识别，跟踪不能停止。

05

基于整体的方法

第一阶段为构造种子段：根据直线的几何特性，通过一些约束条件获得种子段，并得到直线的斜率以及线宽信息。第二阶段为识别阶段：此阶段以种子段为基础进行跟踪延长拓展，在扩展的过程中，根据探测点和实际位置进行比较，进行动态校正，以提高识别率。

优点：准确率高、检测时间短、交接点和端点定位准确。

缺点：对长度较短直线和曲线等复杂情况的识别率有很大的局限性。



文字报告

矿山业务的重要资料，其内容完整性和可靠性程度对矿山管理至关重要。矿山的文字报告主要包括地质勘查报告、基建探矿报告、生产探矿报告、项目建议书、初步可行性研究报告等。



数字矿山数据获取

文字报告也是数字矿山数据获取的重要来源之一，尤其对于数字化程度较低的矿山企业，大多以纸质文档的形式进行存档。因此，如何将纸质的文字报告转换为计算机可以度量的数据，一直是难题。



OCR技术

OCR技术是一种将图像形状转变为文本字符的技术，能很好地实现将纸质文档转变为计算机可识别的数据。



OCR 技术的原理



图像采集

是OCR技术实现的第一步。图像采集可以采用扫描仪、相机、或者其他数字设备进行采集，采集过程需要注意图像的分辨率、清晰度等。

图片的预处理是为了更好的进行文字识别，其主要目的是去除图像中的噪声、增强图像的对比度以及亮度等。图片预处理包括光影预处理、倾斜预处理、扭曲预处理等。

预处理图片



是OCR技术的核心步骤，是对图像中的每个字符进行分析，提取出其特征。特征可以是字符的形状、大小、颜色等。特征提取的方法包括模板匹配、神经网络等。



特征提取

在特征提取之后，计算机会将提取出的特征与已知的字符进行比对，以确定图像中的字符。字符识别的方法包括模板匹配、神经网络、支持向量机等。

识别字符



后处理的目的是去除识别错误、修正识别结果等。后处理的方法包括语言模型、规则检测等。

后处理



OCR技术会将识别结果输出为可编辑文本。输出结果可以是文本文件、电子表格、数据库等。

结果输出





在矿山的应用

OCR技术在矿山的应用主要包括勘探、测量、采矿研究、设计报告与方案等纸质文档转换为可编辑的数字文档；矿山企业工作过程中形成的扫描与拍照文档转化为可编辑文档。





3.3

数据规范化和标准化

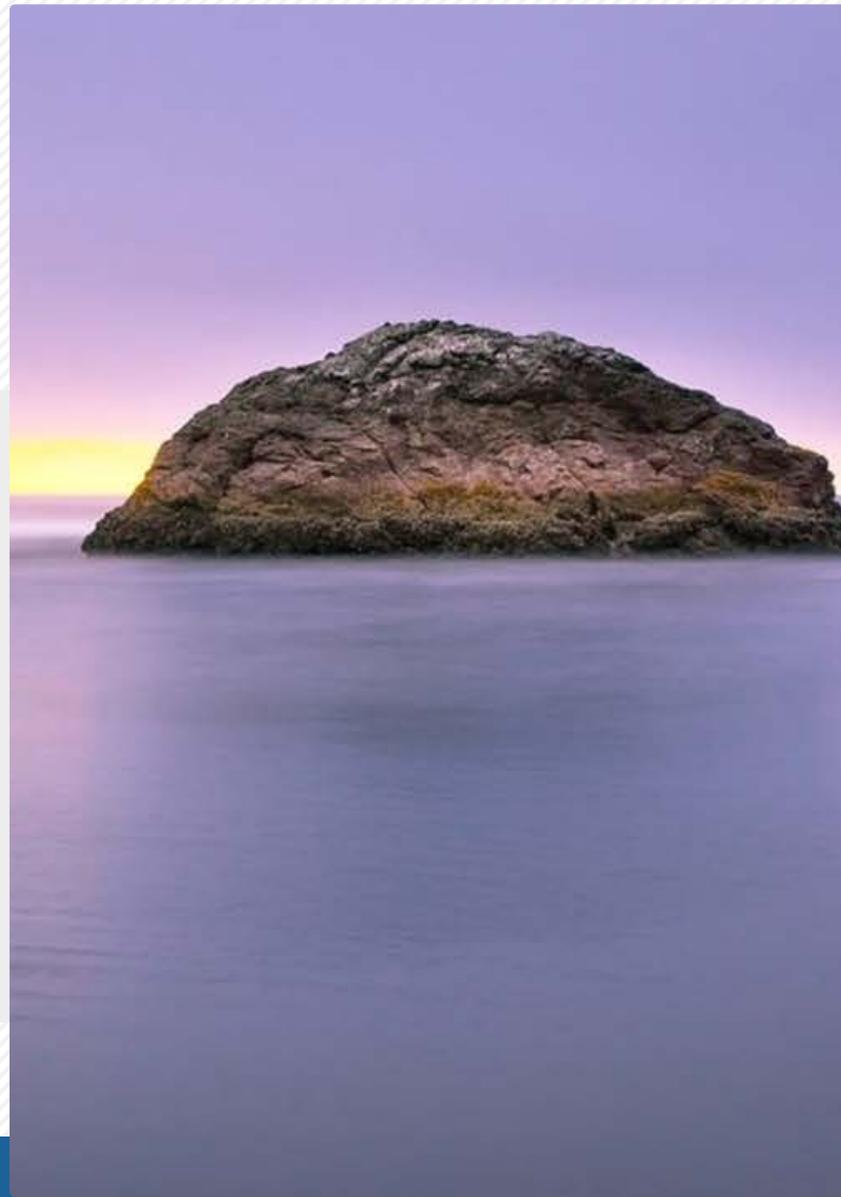
—— 开采生产与经营管理新模式的变革背景





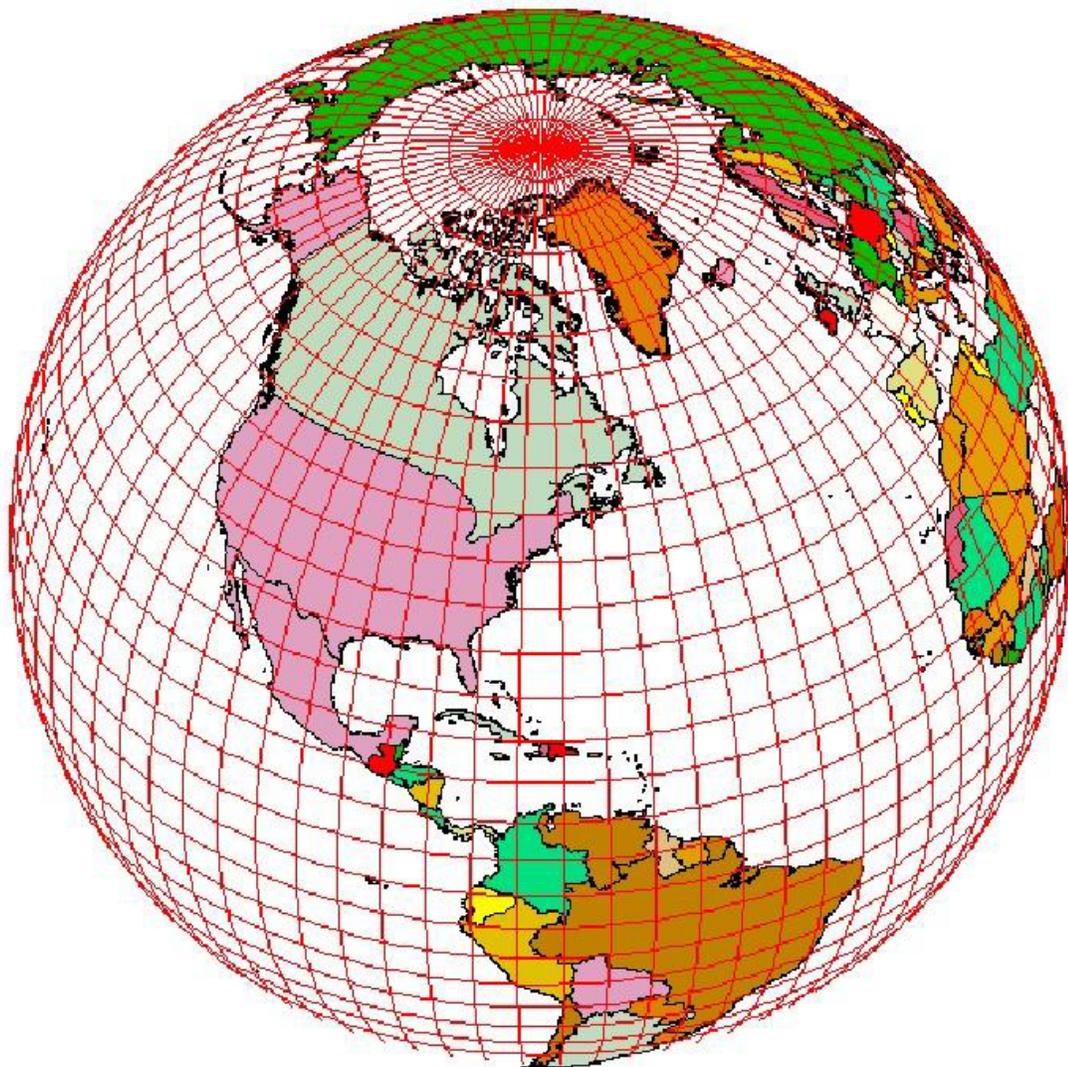
规范化与标准化

数据规范化与标准化是数据共享的基础，其直接影响着数字矿山与智能矿山建设的成效。若不进行数据的规范化与标准化操作，将出现“信息孤岛”现象，数据将无法在矿山全生命周期各业务之间互联互通，进而严重阻碍了我国数字矿山与智能矿山建设的进程。因此，为实现数据的共享与互联互通，数据规范化与标准化是一项十分紧迫的任务。





地理基础是数字矿山信息数据表达格式与规范的重要组成部分。统一的地理基础主要包括统一的地图投影系统、统一的地理坐标系统以及统一的地理编码系统。因此，各种来源的矿山数据与信息，通过投影坐标、地理坐标、网格坐标对数据进行定位处理，然后在共同的地理基础上反映出它们的地理位置和地理关系特征。





数据的分类与编码

01

数据的分类与编码是指对数据进行分类、排列，并用易于计算机和人识别的符号体系表示的过程。

建立分类体系

05

针对智能化矿山数据，应建立分类体系、编码体系和代码元素集合，这是智能化矿山数据融合共享规范体系的基础。

数据标准化的基础

02

数据的分类与编码是数据标准化的基础，也是数据存储、组织管理、检索以及交换的基础。

统一分类与编码

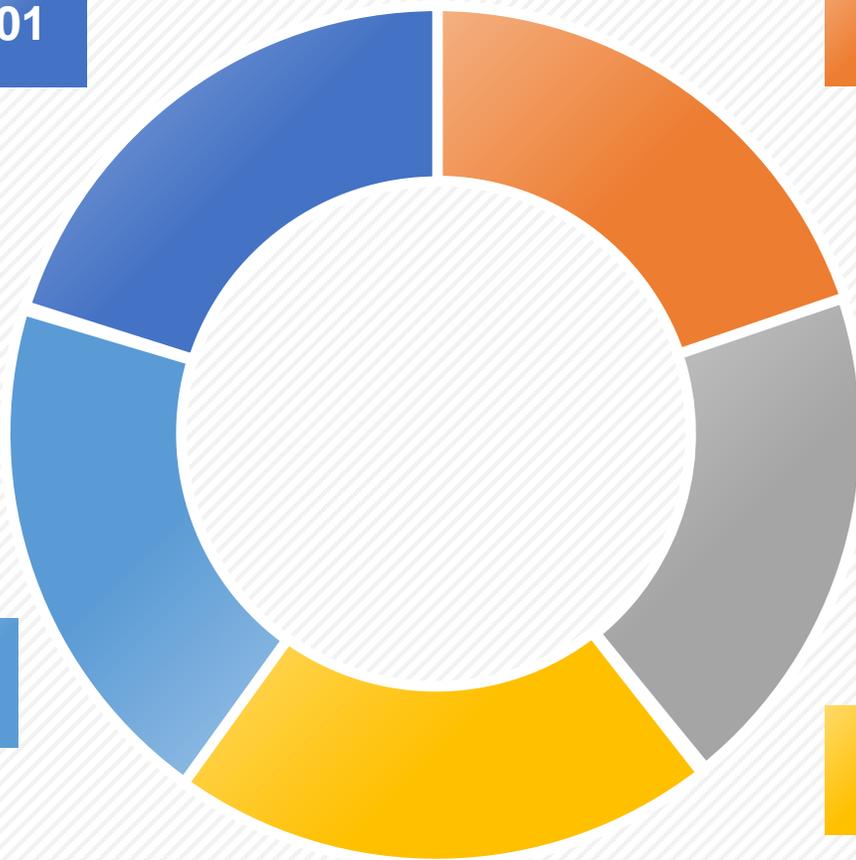
03

通过统一地分类与编码，形成数据标准，确保数据在矿山全生命周期内的互通与共享。

数字矿山数据的分类

04

数字矿山数据的分类与编码应遵循科学性、系统性、实用性、统一性、完整性、稳定性等原则。





数据交换格式标准 是规定数据交换时采用的数据记录格式，主要用于不同系统之间数据交换。

由于数字矿山与智能采矿相关的软硬件产品来自于不同的生产厂商，各生产厂商的软硬件产品又都有其特有的数据格式，而各生产厂商的软硬件产品之间进行数据交换又不可避免，但目前矿业领域暂无相应的数据交换格式标准。

因此，亟需制定矿业领域的数据交换格式国家级或行业级标准，并向整个矿业领域推广应用。





格式要求

第一原则

制定的数据交换格式应尽量简单实用，能独立于数据提供者和用户的数据格式、数据结构和硬软件环境。

第二原则

数据格式应便于修改扩充和维护，便于同国内外重要的数字矿山与智能采矿硬件系统的数据格式进行交换，保证较强的通用性。



数据采集

- 数据采集标准化是数据加工、交换、传输、存储与应用等的前提与基础。数据采集涉及到矿山全生命周期的各个业务阶段及其各个业务，要想确保采集到的数据在矿山全生命周期互联互通以及充分利用，必须从数据采集这一源头实施标准化与规范化。
- 数据采集标准化能有效地解决数据采集的重复性与不完备性问题。因为矿山全生命周期各个业务阶段及其各个业务对数据的需求不一致，关注点不一致，则采集的数据不一致，缺少从矿山全生命周期的角度整体考虑，会导致数据采集重复、不完备以及噪声大等问题。2023年6月26日正式发布的《智能化矿山数据融合共享规范》，填补了长久以来矿山数字化与智能化数据采集规范的空白，为数据采集标准化提供了规范依据。



3.4

数据处理

山开采生产与经营管理新模式的变革背景





1

BJ-54坐标系与WGS-84坐标系转换方法

- WGS-84坐标与BJ-54坐标的转换，可用下列步骤实现：
- ①将两个坐标系的坐标都转为直角坐标；
- ②按所采用的转换方法(三参数或七参数)求解出转换参数；
- ③根据所求参数进行坐标转换；
- ④根据需要，将直角坐标再转为大地坐标。

2

BJ-54坐标系与1980西安坐标系转换方法

- ①利用1954年北京坐标系与1980西安坐标系之间的二维向量差值 Δx 、 Δy 与平面坐标 (x, y) 的关系，采用回归分析方法建立新旧坐标转换数学模型。
- ②利用最小二乘原理得到未知参数的估计量；
- ③利用未知参数向量的估计值，分别确定平面坐标 (x, y) 分量的回归方程。即可利用1954年北京坐标系下任意点的平面坐标 (x_{54}, y_{54}) ，得到1980西安坐标系下的二维坐标 (x_{80}, y_{80}) ，并进行精度评定。

3

地方独立坐标系与国家坐标系之间的转换方法

- ①进行两坐标系转换的最直接办法是求算地方坐标系相对于国家坐标系的旋转角度和平移量。
- ②间接变换法的出发点是把地方坐标系的建立与国家高斯-克吕格直角坐标等同起来。坐标系变换的实质就成为投影带的变换，可以由地方直角坐标反解大地坐标，再根据大地坐标正解国家高斯直角坐标。



数据的预处理是对采集的各种数据，按照不同的方式方法对数据进行编辑运算，清除数据冗余，弥补数据缺失，形成符合工程要求的数据文件格式。处理内容主要包括：数据编辑、数据压缩、数据变换、数据格式转换、空间数据内插、边沿匹配、数据提取等。数据处理对于空间数据有序化、检验数据质量、实现数据共享、提高资源利用效果都具有重要意义。本节简要的介绍GPS数据、三维激光扫描数据以及雷达遥感数据的空间数据预处理。

GPS数据处理

从基本流程上分析，可将GPS网的数据处理流程划分为数据预处理、格式转换、基线解算、无约束平差以及约束平差等五个阶段

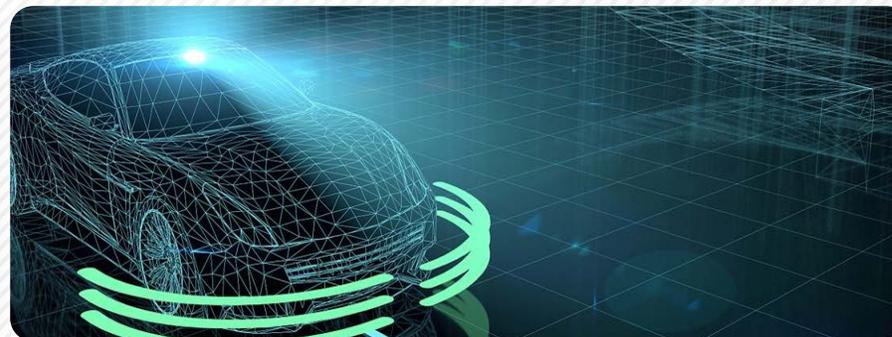


三维激光扫描数据处理

对空间三维物体特征点快速扫描，精确获取目标的空间三维信息，便于结构复杂、非接触式场景的三维可视化建模

雷达遥感数据处理

雷达遥感数据的处理包括辐射校正和几何纠正、图像整饰、投影变换、镶嵌、特征提取、分类等内容





01

① 仪器误差

在仪器结构、制造方面，每一种仪器具有一定的精确度，因而使观测结果的精确度受到一定限制。

02

② 观察误差

观测者感官鉴别能力有一定的局限性。观测者的习惯因素、工作态度、技术熟练程度等也会给观测者成果带来不同程度的影响。

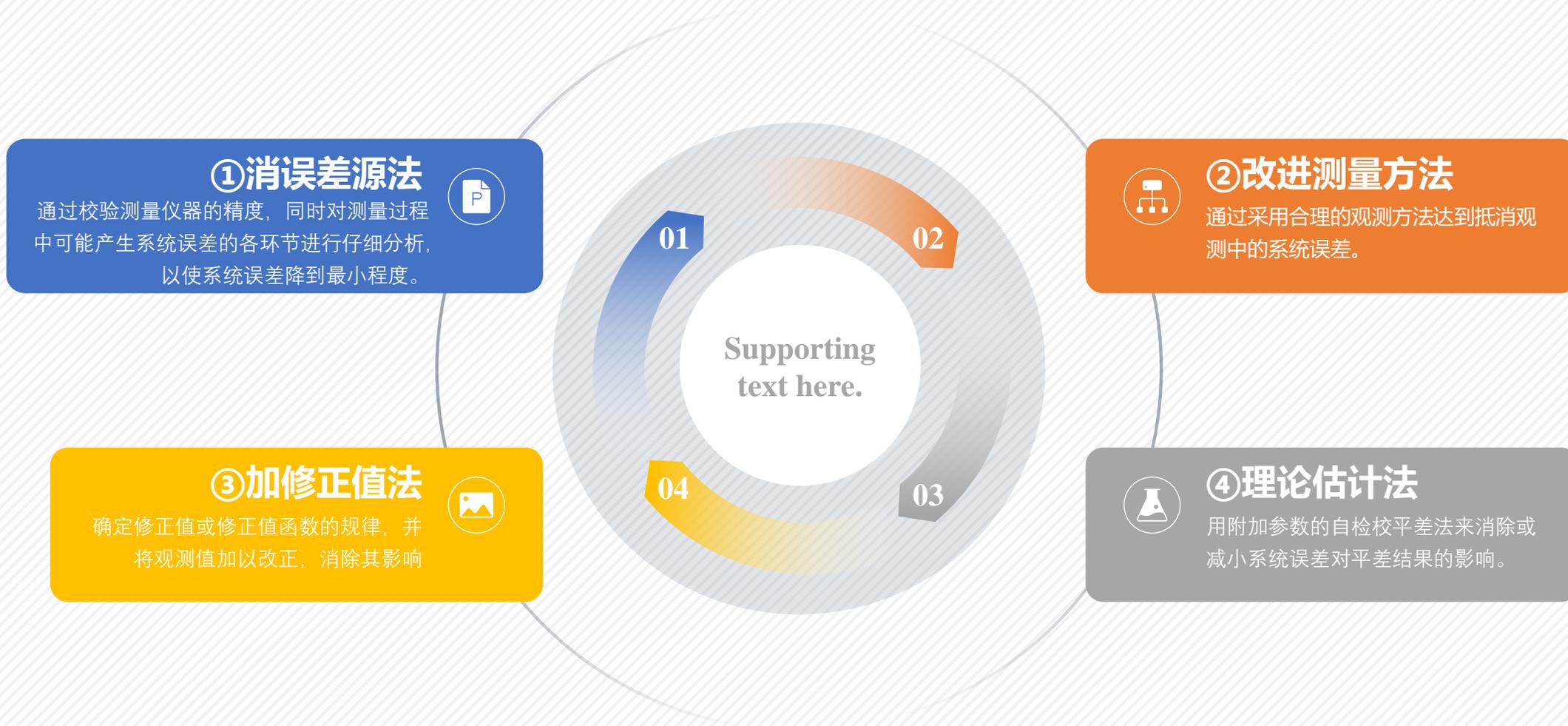
03

③ 外界条件误差

外界环境如温度、湿度、风力、大气折光等因素的变化，均会使观测结果产生误差。如温度变化使钢尺产生伸缩阳光曝晒使水准气泡偏移，大气折光使望远镜的瞄准产生偏差，风力过大使仪器安置不稳定等。



系统误差的处理





偶然误差的处理

由于引起偶然误差的因素是随机的，所以产生的偶然误差是不可修正的，所以，要想减小偶然误差的产生，应该在测量初期适当提高测量仪器的等级，并进行多次观测求取其平均值作为测量的结果。实验统计表明，对同一量值重复多次测量后所得数据符合统计规律。当测量次数无限次增多时，偶然误差与偶然误差出现的概率密度的关系符合正态分布规律。偶然误差有如下特性：

01

对称性

绝对值相等的正、负偶然误差出现的概率相等。

单峰性

绝对值小的偶然误差比绝对值大的偶然误差出现的机会多。

02

03

有界性

在一定测量条件下，偶然误差绝对值不会超过一定范围。

04

抵偿性

当测量次数 n 无限增多时，偶然误差的算术平均值趋向于零。



3.5

数据质量

开采生产与经营管理新模式的变革背景





/ 01

数据的质量是指数据的可靠性与精度，通常用数据误差来度量。数字矿山的基础与本质特征是矿山开采环境、对象及过程信息的数字化，而数字化的过程中数据的可靠性与精度问题总是存在的。

/ 02

从数据的采集、加工处理、存储、输出以及使用的整个过程，均会产生不同的数据质量问题。因此，从某种程度上讲，某些数据质量问题是无法避免的，但可以通过分析数据质量问题的产生原因，降低数据在质量上承担的风险。下面将从定位精度、属性精度、时间性、完整性、逻辑一致性以及数据档案等六个方面进行阐述。





定位精度是指矿山空间坐标数据与其真实的地理位置之间的误差。这种误差主要有两种：第一种是偏差；第二种是偏移的分布。

定位精度

属性精度是指矿山相关数据库与文件中点、线、面的属性数据正确与否。属性定义往往也会有误差，除人为因素外，还有技术因素。

属性精度

矿山数据采集和输入是一个相当长的过程，是随着工程的不断推进而不断变化，从而造成数据时间性的差异。

时间性

数据完整性包括数据层的完整性、分类的完整性和检验完整性。

完整性

逻辑一致性是指数据之间要维护良好的逻辑关系，如多边形的闭合、节点匹配、拓扑关系的正确性或一致性等。

逻辑一致性

数据档案主要是指数据集合生产历史，原始数据以及处理这些数据所使用的处理步骤等。

数据档案



空间数据的误差检查是指发现数据错误，探测数据精度与准确性。数据质量检查的方法主要有直接评价法、间接评价法和非定量描述法等。

直接评价法

01

包括用计算机程序自动检查和随机抽样检查。

计算机程序自动检查是指某些类型的错误可以用计算机软件自动发现；随机抽样检查是指随机抽取一部分数据，检查其质量指标。

间接评价法

02

是指通过外部知识或信息进行推理来确定空间数据质量的方法，用于推理的外部知识或信息如用途、数据历史记录、数据源的质量、数据生产的方法、误差传递模型等。

非定量描述法

03

是指通过对数据质量的各组成部分的评价结果进行综合分析来确定数据的总体质量的方法。



误差分析方法

敏感度分析法

为了从理论上了解输出结果如何随输入数据误差的变化而变化，可以人为地在输入数据中加上扰动值来检验输出结果对这些扰动值的敏感程度。然后根据适合度分析，由置信域来衡量由输入数据误差引起的输出数据的变化。

尺度不变空间分析法

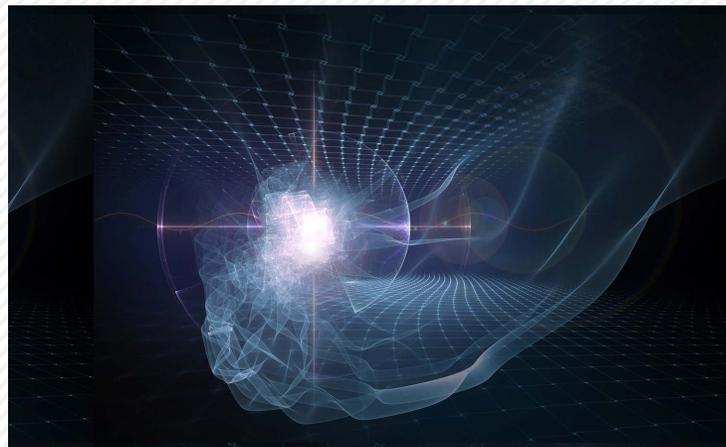
矿山数据的分析结果应与采用的空间坐标系无关，即尺度不变空间分析，包括比例不变和平移不变。尺度不变是数理统计中常用的一个准则：一方面能保证用不同方法得到的结果一致；另一方面又可在同一尺度下合理地衡量估值的精度。

蒙特卡洛实验仿真

蒙特卡洛实验仿真首先依据经验对数据误差的种类和分布模式进行假设，然后利用计算机进行模拟实验，将所得结果与实际结果进行比较，找出与实际结果最接近的模型。

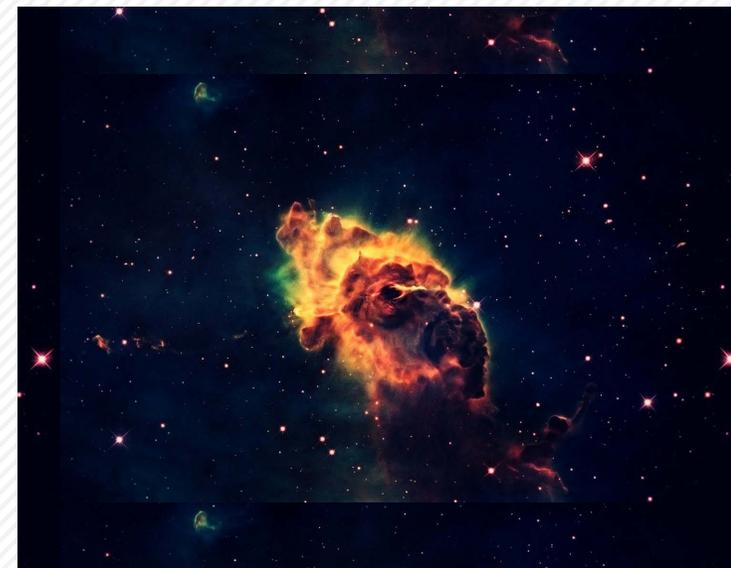
空间滤波

空间滤波分高通滤波和低通滤波。前者指从含有噪声的数据中分离提取噪声信息的过程；而后者指从数据中提取信号的过程。经高通滤波后可得到一个点(或线、面)的随机噪声场，然后按随机过程理论或方差-协方差分量估计理论求得数据采集误差。



由数据的误差来源, 可知数据从采集、加工、存储、输出以及使用的各个阶段都可能引入误差。而正因为误差的存在, 导致矿山各要素之间不能套合以及不同时间的数字化成果不能精确联结等。因此, 空间数据必须进行误差校正, 使之满足实际使用要求。

一般情况下, 数据编辑处理只能消除或减少在数字化过程中因操作产生的局部误差或明显误差, 但图纸变形和数字化过程产生的随机误差, 必须经过几何校正才能消除。由于造成数据变形的因素很多, 对于不同因素引起的误差, 其校正方法也不同, 具体采用何种方法应根据实际情况而定。





从理论上讲，根据图形的变形情况，计算出其校正系数，然后根据校正系数，即可对图形进行校正变换。常用的校正方法有一次变换、二次变换与高次变换、最小二乘法线性校正以及分块校正。

一次变换

同素变换和仿射变换均为一次变换。

二次变换与高次变换

制图资料上的直线经变换后，可能为二次曲线或高次曲线，它适用于原图有非线性变形的情况。

最小二乘法线性校正

如果所给已知点数多于方程所要求的个数，为了使其尽量满足各个已知点，可运用最小二乘法求解其校正系数。

分块校正

将已知点形成一个个小区域，使该区域内的点仅受相应区域上的已知点控制，来校正整幅图形，使其满足精度要求。



3.6

数字矿山数据分类与管理

— 开采生产与经营管理新模式的变革背景



数据分层 为了便于管理和更新，将数据逻辑上划分为基础层、专业层、感知层和管理层。



基础层

基础层为基础地理数据，重点是以基础地理框架为基底的正射遥感影像数据，以及地名、行政境界等基础地理信息。



专业层

专业层是地质、测量、采矿、生产计划等专业产生的数据，主要由点、线、面、体要素组成的空间数据图层。



感知层

感知层是矿山生产过程中实时采集的数据，一般通过传感器、采集终端、通讯网络，感知、传输而获得的数据。



管理层

管理层是矿产资源等管理过程产生的数据，随管理业务实时更新，主要是由坐标串构成的空间数据及统计表格组成的属性数据。



各数据层之间的关系



基础层

作为所有数据的基础，基础层为各类数据提供了统一的空间参考。



专业层

反映矿山开采全生命周期的专业数据，是管理层之本底。



感知层

矿山开采过程的现实数据，体现过程与现状，是管理层的支持层。



管理层

矿山开采管理过程及行为的记录，是管理过程及结果“沉淀”在专业层上的信息。





从时效上将空间数据分为时空数据和版本化数据，时空数据具有较强的时效性。

01

版本化数据：要素根据需要发生更改，如矿体形态变化、设计更改，版本化数据是要素发生更改的依据。

02

数据
存在形式



地质数据：原始的地质数据。包括各类地质勘探、生产勘探数据。注意，没有基于原始地质数据建立的三维模型，如矿体模型、岩层模型等，而是将这些数据定义为设计数据（加工数据）。以属性表（table）的形式存在。

测量数据：原始的测量数据。包括大地测量、矿井测量数据。注意，没有基于测量数据建立三维模型，如地形模型、巷道模型等，而是将这些数据定义为设计数据（加工数据）。以属性表（table）的形式存在。

设计数据：包括基于测量数据建立的地形模型、巷道模型等；包括基于地质数据建立的矿体模型、岩层模型等；包括各种开采设计：开拓、采准、回采等。数据形式有中心线、三维模型。以要素层（layer）的形式存在。

开采计划数据：开采计划编制过程中，很少产生基础数据，更多地是应用基础数据，所以可以考虑将其定义为管理层。

感知数据：以实时采集数据为主，通过采集服务对数据进行抽取、转换（通过服务完成）后进入关系型数据库。

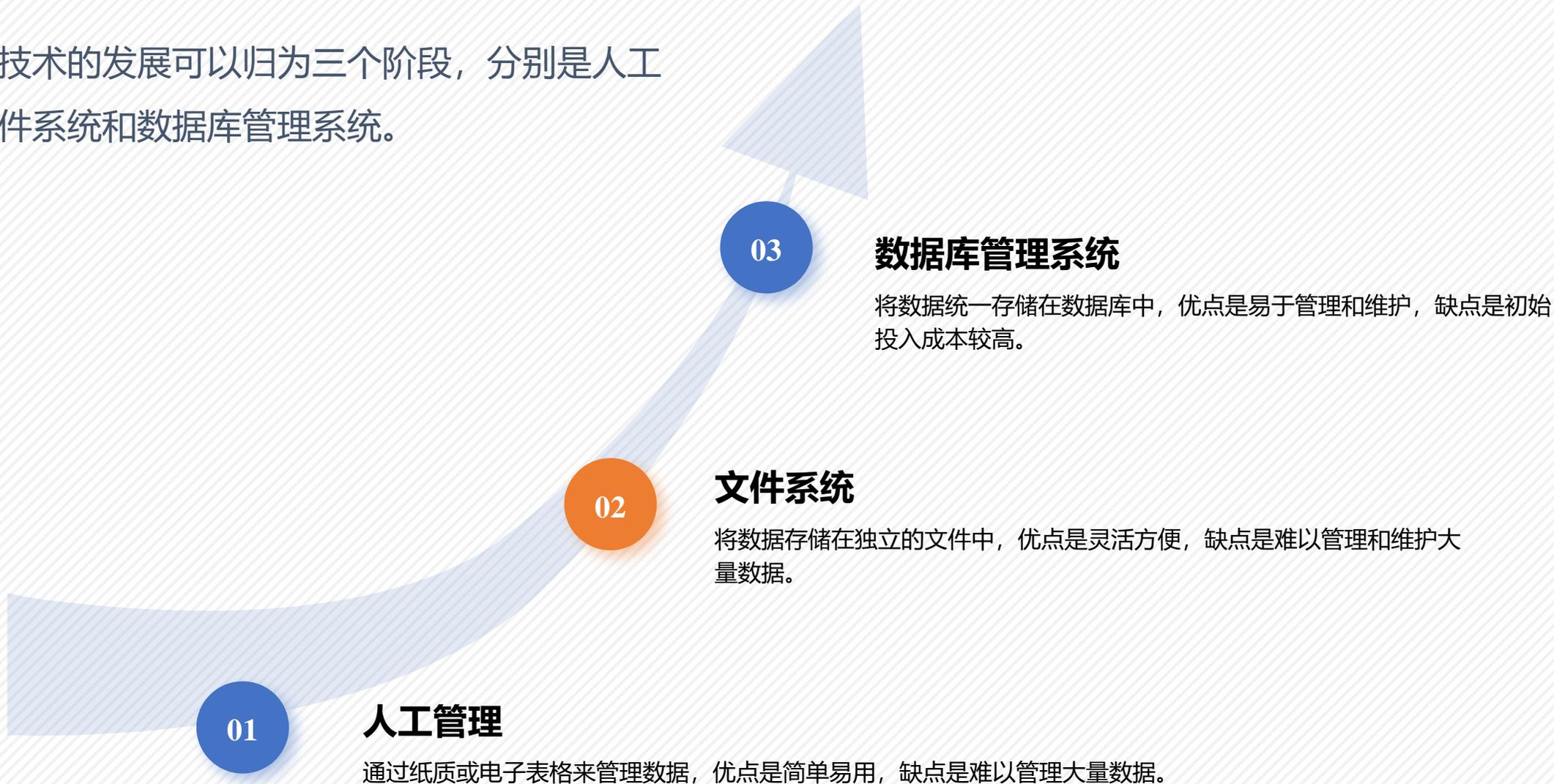
管理数据：在矿山企业管理过程中产生的数据，通常为关系型数据。

各类图件：包括地质图件、设计图件等。





数据管理技术的发展可以归为三个阶段，分别是人工管理、文件系统和数据库管理系统。



人工管理阶段（20世纪50年代中期以前），计算机主要用于科学计算。外部存储器只有磁带、卡片和纸带等还没有磁盘等直接存储设备。软件只有汇编语言，尚无数据管理方面的软件。数据处理方式基本是批处理。这个阶段有如下几个特点：

01

计算机系统不提供对用户数据的管理功能

程序和数据是一个不可分割的整体。数据脱离了程序就无任何存在的价值，数据无独立性。

02

数据不能共享

不同的程序均有各自的数据，这些数据对不同的程序通常是不相同的，不可共享；即使不同的程序使用了相同的一组数据，这些数据也不能共享。

03

不单独保存数据

基于数据与程序是一个整体，数据只为本程序所使用，数据只有与相应的程序一起保存才有价值，否则就毫无用处。所以，所有程序的数据均不单独保存。



文件系统阶段（20世纪50年代后期至60年代中期）计算机不仅用于科学计算，还利用在信息管理方面。随着数据量的增加，数据的存储、检索和维护问题成为紧迫的需要，数据结构和数据管理技术迅速发展起来。

此时，外部存储器已有磁盘、磁鼓等直接存取的存储设备。软件领域出现了操作系统和高级软件。操作系统中的文件系统是专门管理外存的数据管理软件，文件是操作系统管理的重要资源之一。数据处理方式有批处理，也有联机实时处理。

这个阶段有如下几个特点：

- 01.** 数据以“文件”形式可长期保存在外部存储器的磁盘上
计算机的应用转向信息管理，因此对文件要进行大量的查询、修改和插入等操作。
- 02.** 数据的逻辑结构与物理结构有了区别
由操作系统的文件系统提供存取方法（读 / 写）。
- 03.** 文件组织已多样化
有索引文件、链接文件和直接存取文件等。
- 04.** 数据面向应用
数据不再属于某个特定的程序，可以重复使用。
- 05.** 对数据的操作以记录为单位
文件的建立、存取、查询、插入、删除、修改等所有操作，都要用程序来实现。



随着数据管理规模的扩大，数据量急剧增加，文件系统显露出一些缺陷如下：

1

数据冗余

由于文件之间缺乏联系，造成每个应用程序都有对应的文件，有可能同样的数据在多个文件中重复存储。

2

不一致性

在进行更新操作时，稍不谨慎，就可能使同样的数据在不同的文件中不一样。

3

数据联系弱

由于文件之间相互独立，缺乏联系造成的。



随着数据管理规模的扩大，数据量急剧增加，文件系统显露出一些缺陷如下：

文件系统阶段是数据管理技术发展中的一个重要阶段。在这一阶段中，得到充分发展的数据结构和算法丰富了计算机科学，为数据管理技术的进一步发展打下了基础，现在仍是计算机软件科学的重要基础。

1

数据冗余

由于文件之间缺乏联系，造成每个应用程序都有对应的文件，有可能同样的数据在多个文件中重复存储。

2

不一致性

在进行更新操作时，稍不谨慎，就可能使同样的数据在不同的文件中不一样。

3

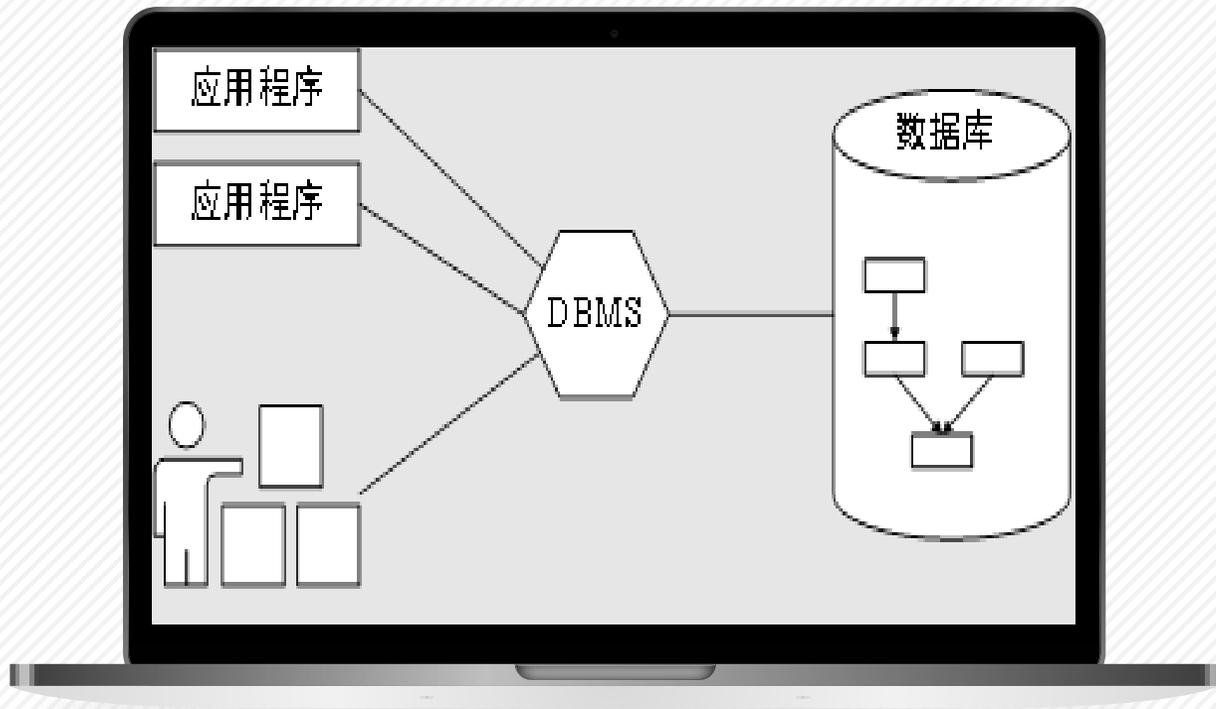
数据联系弱

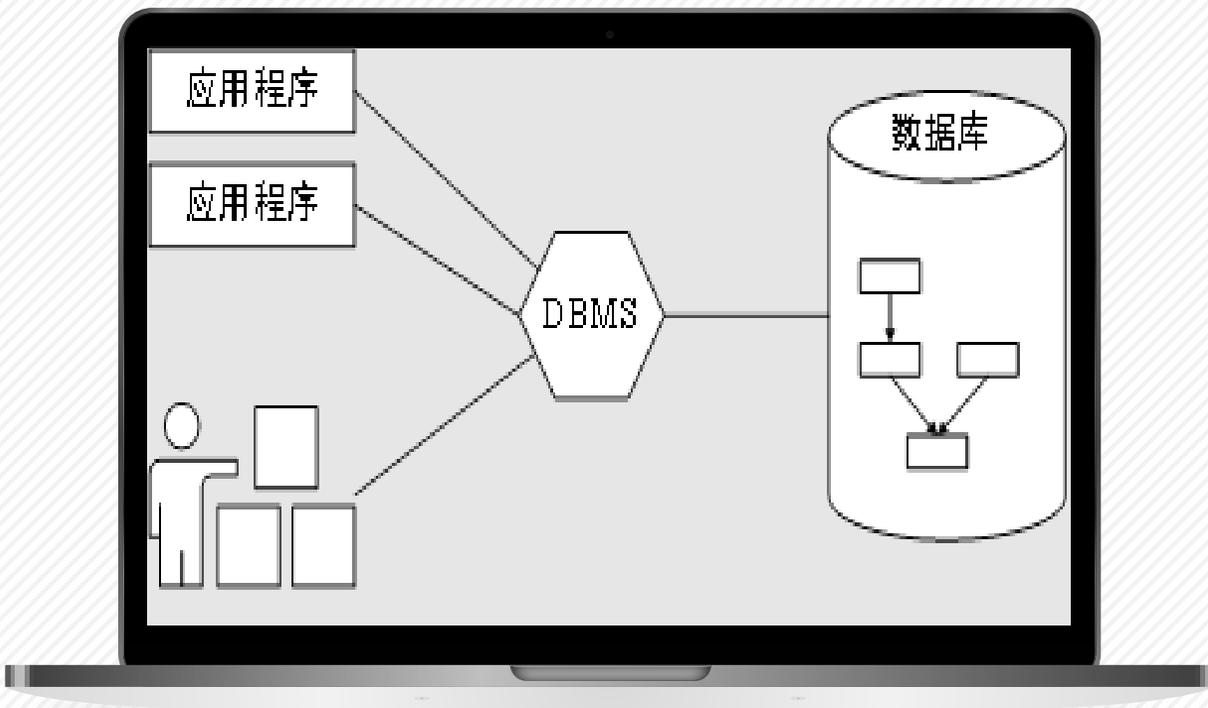
由于文件之间相互独立，缺乏联系造成的。



数据库管理系统

这一阶段（60年代后期），数据管理技术进入数据库系统阶段。数据库系统克服了文件系统的缺陷，提供了对数据更高级、更有效的管理。这个阶段的程序和数据的联系通过数据库管理系统来实现（DBMS），见右图





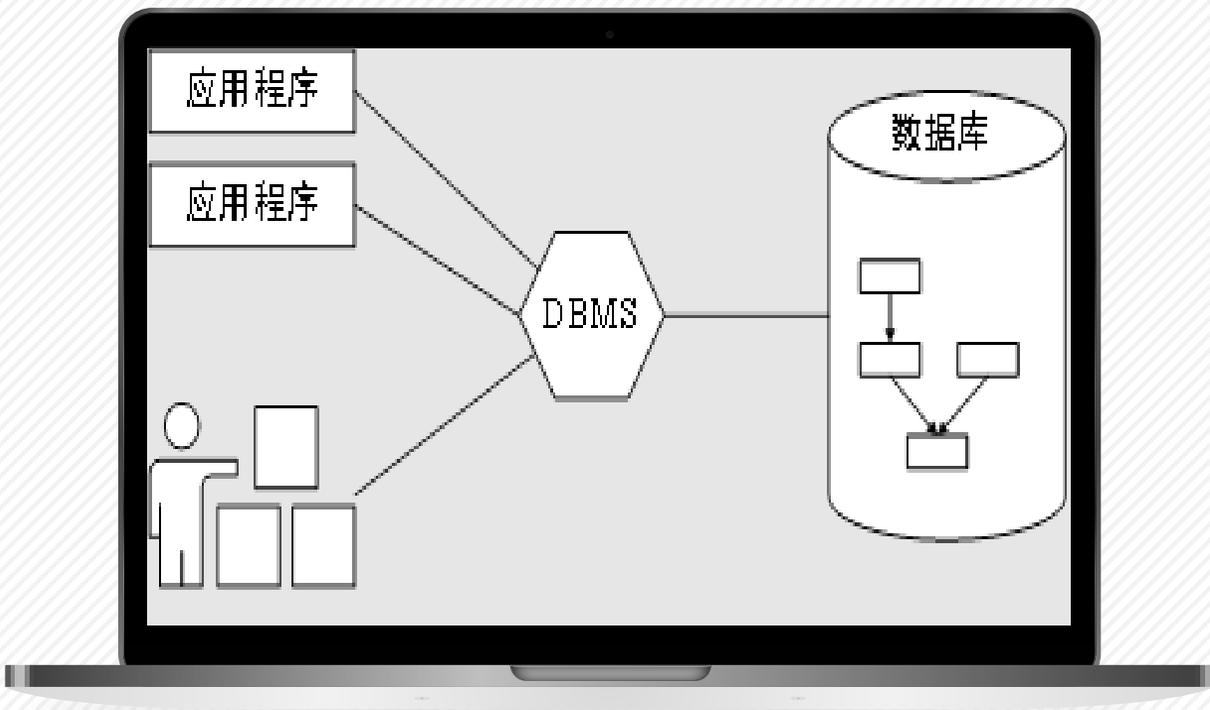
05 增加了系统的灵活性

04 提供了数据控制功能

03 为用户提供了方便的用户接口

02 有较高的数据独立性

01 采用数据模型表示复杂的数据结构



数据库管理系统 应用最为成熟的是关系型数据库管理系统，但近年来随着地理信息技术不断发展，空间数据的大量应用，空间数据库管理系统也发展迅速，但大都是基于传统关系型数据库扩展而来的，比如 Oracle Spatial、SQL Server Spatial等。



数据库系统

数据库系统是管理大量、持久、可靠、共享的数据的工具，由数据库、计算机设备和数据库管理系统组成。



数据库

数据库是存储数据的空间，可以看作是存储数据的容器，实际上是由多个文件组成的。



数据库系统中的数据库

在数据库系统中，通常包含多个数据库，每个数据库中又包含了一定数量的以一定格式存储的数据集合。



数据库的特点

概括起来，数据库的特点是持久化、共享化、规范化、安全性和完整性。



01

数据库定义

数据库管理系统必须首先能充分定义并管理各种类型的数据项。例如，关系型数据库管理系统必须建立数据库和数据表，定义字段的数据类型、限制以及数据之间的关联等。

02

数据库处理

数据库管理系统必须能为用户提供对数据库存取的能力，这些能力包括增加、删除、修改和查询等。有时候并不是所有的要求都可以由数据库管理系统提供，因此需要编制相应的应用程序来满足特殊的需求。

03

数据字典

数据字典（Data Dictionary, DD）中存放着对实际数据库各级模式所做的定义，即对数据库结构的描述。对数据库的使用和操作都要通过查阅数据字典来进行。

04

数据库控制

数据库管理的核心工作是对数据库的运行进行管理，包括：



04

数据库控制

数据库管理的核心工作是对数据库的运行进行管理，包括：

01. 数据库安全性控制功能

应该具备创建用户帐号、相应的口令以及设置权限等功能。

03. 并发控制功能

应能对多个用户的并发操作加以控制、协调。

02. 数据库完整性控制功能

完整性是数据的准确性和一致性的测度。

04. 数据库恢复功能

数据库管理系统需要为用户提供准确、方便的备份功能。



结构化查询语言

结构化查询语言 (Structured Query Language) 简称SQL, 是一种特殊目的的编程语言, 是一种数据库查询和程序设计语言。



功能

SQL能完成提取查询数据、插入修改删除数据、生成修改和删除数据库对象、数据库安全控制、数据库完整性及数据保护控制。



底层结构

结构化查询语言是高级的非过程化编程语言, 允许用户在高层数据结构上工作, 不要求用户指定对数据的存放方法。



语句嵌套

结构化查询语言语句可以嵌套, 具有极大的灵活性和强大的功能, 使它具有极大的灵活性和强大的功能。



功能分类

SQL按其功能分为三大部分, 包括数据定义语言、数据操作语言、数据控制语言, 为数据库系统提供稳定可靠的数据管理。



01



数据库定义语言

CREATE: 用于创建数据库对象

DECLARE: 创建临时表

DROP: 删除任何用CREATE:
创建对象

ALTER: 修改某些数据库对象
的信息

02



数据库操纵语言

SELECT: 从表中查询符合数
据

DELETE: 删除已有表的数据

UPDATE: 更新已有表的数据

INSERT: 向已有表中插入数
据

03



控制语言

GRANT: 授予用户权限

REVOKE: 撤消用户权限

COMMIT: 提交事务, 可以
使数据库的修改永久化

ROLLBACK: 回滚事务



空间数据库管理系统介绍

空间数据库 (Spatial DBMS) 是一种能够高效存储、操作和查询空间数据的数据库管理系统。

空间数据类型

空间数据类型用于指定空间对象，包括点、线、面等，以满足地图制图和遥感图像处理的需求。

三大要素

空间数据库系统的三大要素为空间数据类型、空间索引和空间分析函数。







空间数据库通常提供专用数据类型来存储空间数据，并提供空间索引来优化对空间数据集的访问。例如，空间索引允许有效地检索与其他对象一定距离内的点。此外，空间DBMS提供了对对象执行操作或操作对象的功能。例如计算距离、合并或交叉对象以及计算对象的属性，例如多边形的面积。空间数据库业界代表包括Oracle Spatial、IBM的DB2 Spatial Extender、微软的SQL Server Spatial、开源的PostGIS、GeoMesa等。



非关系型数据库 (NoSQL)

非关系型数据库存储数据的方式不同于传统的表格形式，它们可以使用键值对 (Key-Value)、文档 (Document)、列族 (Column Family) 或图 (Graph) 等形式来组织数据。非关系型数据库具有高可伸缩性和高性能的特点，常见的有MongoDB、Cassandra和Redis等。优点是能够快速存储和检索大量非结构化数据，但对数据一致性和事务处理的支持相对较弱。

内存数据库 (In-Memory Database)

内存数据库将数据存储在内存中，以提供高速的数据访问和处理能力。相比于传统的磁盘存储方式，内存数据库具有更低的延迟和更高的并发能力，常用于对实时性要求较高的应用场景，如金融交易和实时数据分析等。Redis和Memcached是常见的内存数据库。

分布式文件系统 (Distributed File System)

分布式文件系统将数据分散存储在多个服务器上，以提供可靠性和高可扩展性。分布式文件系统可以自动处理故障恢复和负载均衡等问题，并支持分布式计算和存储。Hadoop的HDFS和Google的GFS是常见的分布式文件系统。



Q&A